



# ASA 2022

## Data-Driven Decision Making

BOOK OF SHORT PAPERS

edited by

Enrico di Bella

Luigi Fabbris

Corrado Lagazio

  
FIRENZE  
UNIVERSITY  
PRESS

 GENOVA  
UNIVERSITY  
PRESS

PROCEEDINGS E REPORT

ISSN 2704-601X (PRINT) - ISSN 2704-5846 (ONLINE)

### ***Scientific Program Committee***

Enrico di Bella (University of Genoa, co-chair)  
Luigi Fabbris (ASA, Tolomeo Studi e Ricerche, Padua, co-chair)  
Corrado Lagazio (University of Genoa, co-chair)  
Fabrizio Antolini (University of Teramo)  
Rossella Berni (University of Florence)  
Bruno Bertaccini (University of Florence)  
Silvia Maria Biasotti (CNR)  
Gian Lorenzo Boracchia (Liguria Region)  
Eugenio Brentari (University of Brescia)  
Maurizio Carpita (University of Brescia)  
Giulia Cavrini (University of Bolzano-Bozen)  
Alessandro Celegato (AICQ-AISS, PSV Project Service and Value)  
Giuliana Coccia (ASVIS)  
Cristina Davino (Federico II University of Naples)  
Giulia De Candia (ISTAT)  
Loretta Degan (Galgano Group)  
Tonio Di Battista (G. D'Annunzio University of Chieti and Pescara)  
Angela Maria Digrandi (CNR-National Council for Research)  
Simone Di Zio (G. D'Annunzio University of Chieti and Pescara)  
Patrizia Elli (Assirm)  
Benito Vittorio Frosini (Sacred Heart Catholic University of Milan)  
Antonio Giusti (University of Florence)  
M. Gabriella Grassia (Federico II University of Naples)  
Salvo Ingrassia (University of Catania)  
Enrico Ivaldi (University of Genoa)  
Paolo Mariani (Bicocca University of Milan)  
Stefania Mignani (University of Bologna)  
Marta Nai Ruscone (University of Genoa)  
Chiara Parretti (University Guglielmo Marconi of Rome)  
Francesco Palumbo (Federico II University of Naples)  
Francesco Porro (University of Genoa)  
Fabio Rapallo (University of Genoa)  
Alessandra Petrucci (University of Florence)  
Alfonso Piscitelli (Federico II University of Naples)  
Alessandra Righi (Istat)  
Fabrizio Ruggeri (CNR-National Research Council)  
Giorgio Tassinari (University of Bologna)  
Nicola Tedesco (University of Cagliari)  
Venera Tomaselli (University of Catania)  
Domenico Vistocco (Federico II University of Naples)  
Mariangela Zenga (Bicocca University of Milan)

### ***Local Program Committee***

Enrico di Bella (University of Genoa, co-chair)  
Corrado Lagazio (University of Genoa, co-chair)  
Gian Lorenzo Boracchia (Regione Liguria)  
Fabrizio Culotta (University of Genoa)  
Giulia De Candia (ISTAT)  
Enrico Ivaldi (University of Genoa)  
Marta Nai Ruscone (University of Genoa)  
Fabio Rapallo (University of Genoa)  
Eva Riccomagno (University of Genoa)  
Francesco Porro (University of Genoa)  
Sara Preti (University of Genoa)

ASA 2022  
Data-Driven Decision Making

BOOK OF SHORT PAPERS

edited by  
Enrico di Bella  
Luigi Fabbris  
Corrado Lagazio

ASA 2022 Data-Driven Decision Making : book of short papers / edited by Enrico di Bella, Luigi Fabbris, Corrado Lagazio. – Firenze : Firenze University Press ; Genova : Genova University Press, 2023. (Proceedings e report ; 134)

<https://books.fupress.com/isbn/9791221501063>

ISSN 2704-601X (print)  
ISSN 2704-5846 (online)  
ISBN 979-12-215-0106-3 (PDF)  
ISBN 979-12-215-0107-0 (XML)  
DOI 10.36253/979-12-215-0106-3

Cover graphic design: Lettera Meccanica SRLs

Front cover image: Orizzonte del centro di Genova, Italia © sepavo|123rf.com

Volume published with the support of Università di Genova, Scuola di Scienze Sociali - Dipartimento di Scienze Politiche e Internazionali.



Università  
di Genova

DISPI DIPARTIMENTO  
DI SCIENZE POLITICHE  
E INTERNAZIONALI

DIEC DIPARTIMENTO  
DI ECONOMIA

#### Partners



#### Peer Review Policy

Peer-review is the cornerstone of the scientific evaluation of a book. All FUP publications undergo a peer-review process by external experts under the responsibility of the Editorial Board and the Scientific Boards of each series (DOI 10.36253/fup\_best\_practice.3).

#### Referee List

In order to strengthen the network of researchers supporting FUP evaluation process, and to recognise the valuable contribution of referees, a Referee List is published and constantly updated on FUP - Genova University Press's website (DOI 10.36253/fup\_referee\_list).

#### Firenze University Press Editorial Board

M. Garzaniti (Editor-in-Chief), M.E. Alberti, F. Vittorio Arrigoni, E. Castellani, F. Ciampi, D. D'Andrea, A. Dolfi, R. Ferrise, A. Lambertini, R. Lanfredini, D. Lippi, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, A. Orlandi, I. Palchetti, A. Perulli, G. Pratesi, S. Scaramuzzi, I. Stolzi.

The online digital edition is published in Open Access on [www.fupress.com](http://www.fupress.com).

Content license: except where otherwise noted, the present work is released under Creative Commons Attribution 4.0 International license (CC BY 4.0: <http://creativecommons.org/licenses/by/4.0/legalcode>). This license allows you to share any part of the work by any means and format, modify it for any purpose, including commercial, as long as appropriate credit is given to the author, any changes made to the work are indicated and a URL link is provided to the license.

Metadata license: all the metadata are released under the Public Domain Dedication license (CC0 1.0 Universal: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>).

© 2023 Author(s)

Published by Firenze University Press and Genova University Press

Firenze University Press  
Università degli Studi di Firenze  
via Cittadella, 7, 50144 Firenze, Italy  
[www.fupress.com](http://www.fupress.com)

*This book is printed on acid-free paper  
Printed in Italy*

## Table of contents

Assessing the predictive capability of Invalsi tests on high school final mark <i>Silvia Bacci, Bruno Bertaccini, Alessandra Petrucci, Valentina Tocchioni</i>	11
Profiling students' satisfaction towards university courses with a latent class approach <i>G. Damiana Costanzo, Michelangelo Misuraca, Angela Cosca</i>	17
The relation between students' educational performances and their access test results: a focus on an Italian case <i>Matteo Corsi, Luca Persico, Sara Preti, Agnese Sechi</i>	23
Structure and dynamics of immigration in the municipalities of northwestern Italy <i>Simona Ballabio, Arianna Carra, Flavio Verrecchia, Alberto Vitalini</i>	29
Are Italian youngsters adequately equipped for an after-pandemic upswing? <i>Luigi Bollani, Simone Di Zio, Luigi Fabbris</i>	35
How Italians coped with COVID-19 lockdown: evidence from a survey promoted through social networks <i>Margherita Silan, Riccardo Bellide</i>	41
Official statistics for measuring the sustainability of tourism: the UNWTO initiative <i>Emanuela Recchini</i>	47
Misinformation and Disinformation in Statistical Methodology for Social Sciences: causes, consequences, and remedies <i>Giulio Giacomo Cantone, Venera Tomaselli</i>	53
The impact of economic insecurity on life satisfaction among German citizens <i>Demetrio Panarello, Gennaro Punzo</i>	59
Cultural and sensorial correlates of Trebbiano wine consumption <i>Luigi Fabbris, Alfonso Piscitelli</i>	65
Tourism and territorial economy: beyond satellite accounting <i>Fabrizio Antolini, Antonio Giusti, Francesca Petrei</i>	71
Short-term forecasts on time series for tourism in Lombardy <i>Andrea Marletta, Roberta Rossi, Elena Diceglie</i>	77
Interventions for non-self-sufficiency – Focus on care and social policies in South Tyrol <i>Giulia Cavrini, Nadia Paone, Evan Tedeschi</i>	83
The territorialisation of the 2030 Agenda: a multilevel approach <i>Raffaele Attanasio, Manlio Calzaroni, Alessandro Ciancio, Federico Olivieri, Giovanni Siciliano</i>	89
Sustainable development goals: classifying European countries through self-organizing maps <i>Cristina Davino, Nicola D'Alesio</i>	95

Individual and social aspects of after-Covid-19 pandemic depression <i>Pasquale Anselmi, Daiana Colledani, Simone Di Zio, Luigi Fabbris, Egidio Robusto</i>	101
Spread of Covid-19 epidemic in Italy between March 2020 and February 2021: empirical evidence at provincial level <i>Fabrizio Antolini, Samuele Cesarini, Francesco Giovanni Truglia</i>	107
New perspectives for the quality of sub-municipal data with the Italian permanent population and housing census <i>Giancarlo Carbonetti, Stefano Daddi, Giampaolo De Matteis, Marco Di Zio, Davide Fardelli, Raffaele Ferrara, Fabio Lipizzi, Enrico Orsini</i>	113
The Land Cover/Use Code of the new Istat Census cartography <sup>1</sup> <i>Stefano Mugnoli, Alberto Sabbi, Fabio Lipizzi</i>	119
Trusted smart statistics: new statistics for decision makers. Istat's experience <i>Massimo De Cubellis, Gerarda Grippo</i>	125
Quality of life in Health Care: focus on patients <i>Laura Benedan, Angela Digrandi, Paolo Mariani, Cinzia Pilo, Mariangela Zenga</i>	129
Access to emergency care services and inequalities in living standards: some evidence from two Italian northern regions <i>Andrea Marino, Marco Pesce, Raffaella Succi</i>	135
Population ageing and sustainability in South Tyrol: measuring the economic implications of an ageing society <i>Giulia Cavrini, Elisa Cisotto, Alex Weissensteiner</i>	141
Job loss and financial struggle among the older age groups in 2021: Lessons from the European Union <i>Demetrio Panarello, Giorgio Tassinari</i>	145
On the use of auxiliary information in spatial sampling <i>Chiara Bocci, Emilia Rocco</i>	151
Measures of interrater agreement when each target is evaluated by a different group of raters <i>Giuseppe Bove</i>	157
A Natural Language Processing approach to measuring expertise in the Delphi-based scenarios <i>Yuri Calleo, Simone Di Zio, Francesco Pilla</i>	163
Exploring Globalization with Cosmopolitics <i>Maria Serena Causo, Erika Cerasti, Fabrizio De Fausti, Monica Scannapieco</i>	169
Professional choices and personal values: Similarities and differences between Schein's career anchors and Schwartz basic values <i>Maria Cristiana Martini, Aldo Arra</i>	175
Factors affecting tertiary education decisions of immigrants in Italy <i>Michele Lalla, Patrizio Frederic</i>	181
Internet use, feeling of unacceptance and Loneliness: immigrants of first and second generation in Italy <i>Giovanni Busetta, Maria Gabriella Campolo, Antonia Cava</i>	187

A composite indicator to measure regional investment policies on R&D and innovation <i>Sergio Salamone, Alessandro Faramondi, Stefania Della Queva</i>	193
Assessing intimate partner violence in African countries through a model-based composite indicator <i>Anna Maria Parroco, Micaela Arcaio</i>	197
Students' feedback on the digital ecosystem: a structural topic modeling approach <i>Annalina Sarra, Adelia Evangelista, Tonio Di Battista</i>	203
The digitization of the private sector. A non-aggregative method to monitor the NRRP agenda at macro-area level <i>Susanna Traversa, Enrico Ivaldi</i>	209
Digital.VET: an innovative approach for teaching and training <i>Teresa Maltese, Maria Santarcangelo, Vito Santarcangelo, Diego Sinitò, Aneta Poniszewska-Marañda, Jure Šuligoj, Alcidio Jesus, Elisardo Sanchis</i>	215
The joint estimation of accuracy and speed: An application to the INVALSI data <i>Luca Bungaro, Marta Desimoni, Mariagiulia Matteucci, Stefania Mignani</i>	221
Ammonia emissions and fine particulate matter: some evidence in Lombardy <i>Alessandro Fusta Moro, Matteo Salis, Andrea Zucchi, Michela Cameletti, Natalia Golini, Rosaria Ignaccolo</i>	227
On the utility of treating a vineyard against <i>Plasmopara viticola</i> : a Bayesian analysis <i>Lorenzo Valleggi, Federico Mattia Stefanini</i>	233
Trust and security in Italy <i>Silvia Golia</i>	239
Topic modeling for analysing the Russian propaganda in the conflict with Ukraine <i>Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Michelangelo Misuraca, Agostino Stavolo</i>	245
The relationship between religiosity, religious coping, and anxieties about the future: a multidimensional analysis on the Evangelical churches of Naples <i>Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Agostino Stavolo</i>	251
An application of the Agency for Digital Italy guidelines and CSA Star self-assessment: A Docustar case study <i>Pierluigi Calabrese, Paola Lunalbi, Vincenzo Ribaudò, Saverio Crisafulli, Antonio Ruoto, Vito Santarcangelo, Diego Sinitò, Carlo Bonelli, Giuseppe Stella</i>	257
Remote working in Italy: Just a pandemic accident or a lesson for the future? <i>Luigi Bollani, Simone Di Zio, Luigi Fabbri</i>	263
Repression of the future-oriented disposition of Italians by a never-ending pandemic <i>Simone Di Zio, Luigi Fabbri</i>	269
Monitoring and evaluation of gender equality policies <i>Giuliana Coccia, Emanuela Scavalli</i>	275
An experimental annotation task to investigate annotators' subjectivity in a misogyny dataset <i>Alice Tontodimamma, Stefano Anzani, Marco Antonio Stranisci, Valerio Basile, Elisa Ignazzi, Lara Fontanella</i>	281



Potential risk of gambling products and online gambling among European adolescents <i>Elisa Benedetti, Gabriele Lombardi, Rodolfo Cotichini, Sonia Cerrai, Marco Scalese, Sabrina Molinaro</i>	287
An Open Data platform for decision making in local public administration <i>Giuseppe Sindoni, Matteo Massenzio</i>	293
Educational mismatch and productivity: evidence from LEED data on Italian firms <i>Laura Bisio, Matteo Lucchese</i>	299
A paradata-driven statistical approach to improve fieldwork monitoring: the case of the Non-Profit Institutions census <i>Gabriella Fazzi, Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti, Valentino Parisi, Roberta Piergiovanni, Luigi Arlotta, Maura Giacummo</i>	305
Gender INequality Indicator for Academia (GINIA) <i>Margherita Silan, Giovanna Boccuzzo</i>	311
Given $N$ Forecasting Models, What To Do? <i>Fabrizio Culotta</i>	317

# Preface

The Association for Applied Statistics (ASA), the Department of Political and International Sciences and the Department of Economics and Business Studies of the University of Genoa, jointly with the partners AICQ (Italian Association for Quality Culture), AICQ-CN (Italian Association for Quality Culture North and Centre of Italy), AISS (Italian Academy for Six Sigma), ASSIRM (Italian Association for Marketing, Social and Opinion Research), Istat, SIS (the Italian Statistical Society) and with the support of the School of Social Sciences of the University of Genoa, the InLiguria tourism agency of Regione Liguria and the Tourism Office of the Municipality of Genoa, have organised a scientific conference titled “*Data-Driven Decision Making*”. The conference discussed how applied statistics can support public and private decision-makers. Over 130 participants attended the conference in presence, and more than 60 were online. Eighty-four spontaneous communications were presented in twenty-eight parallel sessions, enriched by four plenary sessions organised by the ASA's partners. The opening of the conference was preceded by an event organised by Istat on '*Italian Statistics for Local Decisions*' introduced and chaired by the Istat President, Prof. Gian Carlo Blangiardo.

This book includes 53 peer-reviewed short papers submitted to the scientific conference. The works published in this book follow the order of the conference programme.

On behalf of the Scientific Program Committee, we would like to thank the authors for submitting and presenting their interesting and inspiring works in the context of the evaluation of policies, the partners, the chairs, the discussants, and the Local Organizing Committee.

Finally, we are thankful to the members of the Scientific Committee for helping with the peer-reviewing process.

Genoa (Italy), February 2023

Enrico di Bella, Luigi Fabbris, Corrado Lagazio



# Assessing the predictive capability of Invalsi tests on high school final mark

Silvia Bacci, Bruno Bertaccini, Alessandra Petrucci, Valentina Tocchioni

## 1. Introduction

Educational achievement can be considered a multifaceted issue, which takes into account many domains of learning at different levels of the educational path. In Italy, during the secondary school years, such achievements are measured through the administration of the INVALSI tests, which are standardized tests on a national scale that students carry out at different stages of their career, to identify their level of competence in subjects like literacy, numeracy, and English reading and listening proficiencies. They are applied each year to trace a history of students' skills and knowledge, but also to assess the correspondence between skills and competences acquired with respect to ministerial educational programs. Moreover, the high school final mark may be considered an overall result of performance at the end of secondary school, a sort of synthesis of several achievements and marks in different subjects.

The aim of the present work is to discover if and how the INVALSI scores and the high school final marks are related. More specifically, we intend to verify how the INVALSI scores are associated with students' high school final mark, taking into account students' characteristics as well as school observed (mainly, type of high school) and unobservable characteristics.

The present contribution represents a preparatory work to analyse the predictive capability of INVALSI scores and/or high school final marks on university students' careers. For this reason, the analysis is carried on the INVALSI dataset related with students enrolled in an Italian university.

In the next section, we describe data and statistical methods used in the study. Then, we illustrate the main results. A preliminary discussion of results and some final remarks about future research conclude the work.

## 2. Data and methods

To analyse university students' career in light of their performances during high school we use MobySU.it, a database that integrates multiple data sources, such as the Anagrafe Nazionale Studenti (ANS) data file, the INVALSI data file and the High School database. ANS is a government administrative database on the population of students enrolled in an Italian university between 2010 and 2020. The ANS data contain information on university students' career, individual characteristics, and high school background. The INVALSI data collect information on high school students' performances who obtained the high school diploma in 2019 and 2020. For each student, the following information are available: Economic and Social Status indicator (ESCS), students' INVALSI test scores in English (reading and listening), Italian, and Maths for grades 10<sup>th</sup> and 13<sup>th</sup> (i.e., high school second and fifth year), parents' education and type of employment, as well as other information about school, class and the student him/herself. These two sources of information at the student-level are merged using exact matching. Finally, the High School database includes aggregate data on all Italian high schools between 2015 and 2020, providing information on school characteristics (e.g., geographical area in which it is located, type of released degrees, and so on) and the number of students (grouped by gender) admitted to the final exam and of those who got the diploma.

We select 194,778 students who obtained the high school diploma in Italy in 2019 and enrolled

Silvia Bacci, University of Florence, Italy, [silvia.bacci@unifi.it](mailto:silvia.bacci@unifi.it), 0000-0001-8097-3870

Bruno Bertaccini, University of Florence, Italy, [bruno.bertaccini@unifi.it](mailto:bruno.bertaccini@unifi.it), 0000-0002-5816-2964

Alessandra Petrucci, University of Florence, Italy, [alessandra.petrucci@unifi.it](mailto:alessandra.petrucci@unifi.it), 0000-0001-9952-0396

Valentina Tocchioni, University of Florence, Italy, [valentina.tocchioni@unifi.it](mailto:valentina.tocchioni@unifi.it), 0000-0002-0793-6122

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Silvia Bacci, Bruno Bertaccini, Alessandra Petrucci, Valentina Tocchioni, *Assessing the predictive capability of Invalsi tests on high school final mark*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.03, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 11-16, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

in an Italian university in the academic year 2019/2020. To verify if and how INVALSI scores are associated with students' high school final mark, we estimate a random intercept proportional odds model (Goldstein, 2010; Liu and Agresti, 2005; Snijders and Bosker, 2012) with students as lower-level units and high schools as upper-level units, formulated as follows:

$$\text{logit} [P(Y_{ij} > y_c | X_{ij})] = \beta X_{ij} + \gamma Z_j + u_j - \alpha_c$$

with  $i$  the generic student ( $i = 1, \dots, 194778$ ),  $j$  the high school ( $j = 1, \dots, 5203$ ), and  $c = 1, 2, 3, 4$  the four thresholds corresponding to the five categories in which the students' high school final mark was classified. As currently the high school final mark in Italy ranges from 60 to 100 cum laude, the response variable of the model was constructed by defining five ordinal categories: categories 1 to 4 represent 10 points of the high school mark range (i.e., 60-69, 70-79, 80-89, 90-99) and category 5 collects together 100 and 100 cum laude. Moreover,  $\beta$  and  $\gamma$  denote the vectors of regression coefficients for individual and school-level covariates,  $X_{ij}$  and  $Z_j$  respectively;  $u_j$  is the random intercept capturing the unobserved heterogeneity due to unobservable differences among schools, and  $\alpha_c$  is a response category-specific threshold parameter. The random effects  $u_j$  are assumed to be normally distributed, with mean 0 and constant variance.

The explanatory variables of primary interest are the four students' INVALSI scores on English (reading and listening), Italian, and Maths on grade 13<sup>th</sup>, and are included as standardized, continuous variables. The effect of INVALSI scores is controlled for both students' and schools' characteristics: at student-level we consider the student's gender, citizenship (Italian or not) and the student's macroarea of residence (North, Centre, and South/Islands); at school-level we consider the type of high school management (public vs. private), the type of high school attended (classified in seven categories: see Table 2 below), the percentage of high school graduates older than expected age at graduation, and the average ESCS of the school.

### 3. Results

Table 1 shows the median and mean values obtained in INVALSI scores by female students and male students, respectively, and predicted probabilities of obtaining one of the five mark categories for a median, female student and a median, male student<sup>1</sup>. The median female student has the highest probability (nearly 4 out of 10 students) of obtaining a high school final mark between 70 and 79 points, whereas the median male student has the highest probability (more than 1 out of 2 students) of obtaining a high school final mark between 60 and 69 points, namely the lowest category. Despite both groups have a low probability of obtaining a score equal to 90 or above, female median students seem to obtain higher scores than their male counterparts.

Table 2 shows predicted probabilities of a high school final mark between 60 and 69 points and between 100 and 100 cum laude, for a female/male student that obtained extreme scores to the INVALSI test, namely equal to the 10<sup>th</sup> percentile and to the 90<sup>th</sup> percentile in all four INVALSI scores (other control variables were set at the reference value), by type of high school. On one hand, predicted probabilities of a low final mark (60-69) are very high for those students who obtained an INVALSI score at the 10<sup>th</sup> percentile. This result is confirmed throughout the different types of schools and for both genders, confirming how low scores on INVALSI tests are associated with low high school final marks. Nevertheless, students from vocational institutes, especially female students, report lower predicted probabilities, thus suggesting that these types of school may tend to give higher final scores than other schools, on average. Moreover, predicted probabilities of a low final mark are always higher for male students than for female students across all schools, thus suggesting that female students outperform male students.

---

<sup>1</sup> A median student is an Italian student that lives in the North of Italy, obtained a median score in the four INVALSI tests, and attended a scientific high school with a median percentage of high school graduates older than expected and a median ESCS at the school level.

**Table 1:** Median and mean values of INVALSI scores and predicted probabilities of obtaining a high school final mark for the median profile by gender.

	Female student	Male student
<b>Median value (mean value)</b>		
INVALSI score on Italian	216.8 (216.5)	218.2 (216.9)
INVALSI score on Maths	207.9 (209.0)	229.4 (228.4)
INVALSI score on English reading	220.4 (217.5)	222.6 (219.0)
INVALSI score on English listening	214.8 (214.0)	217.1 (216.2)
<b>Predicted probability</b>		
Pr(60-69 score)	0.359	0.509
Pr(70-79 score)	0.388	0.338
Pr(80-89 score)	0.159	0.101
Pr(90-99 score)	0.071	0.039
Pr(100-100L score)	0.024	0.012

**Table 2:** Predicted probabilities of high school final mark categories, by gender and type of high school. Extreme profiles (10<sup>th</sup>/90<sup>th</sup> percentile of INVALSI scores)

	Pr(60-69 score)	Pr(100-100L score)	Pr(60-69 score)	Pr(100-100L score)	
		<i>Scientific high school</i>		<i>Classical high school</i>	
10percentile F	0.852	0.002	0.713	0.005	
10percentile M	0.917	0.001	0.823	0.002	
90percentile F	0.054	0.203	0.023	0.368	
90percentile M	0.099	0.118	0.044	0.238	
		<i>Applied sciences high school</i>		<i>Foreign language high school</i>	
10percentile F	0.861	0.002	0.723	0.004	
10percentile M	0.922	0.001	0.831	0.002	
90percentile F	0.058	0.191	0.024	0.357	
90percentile M	0.106	0.110	0.046	0.229	
		<i>Technical institute</i>		<i>Vocational institute</i>	
10percentile F	0.628	0.007	0.399	0.020	
10percentile M	0.759	0.004	0.551	0.010	
90percentile F	0.015	0.460	0.005	0.685	
90percentile M	0.029	0.315	0.011	0.540	
		<i>Other high school</i>			
10percentile F	0.619	0.007			
10percentile M	0.751	0.004			
90percentile F	0.014	0.470			
90percentile M	0.028	0.324			
<b>Sample size</b>	<b>194,778</b>				

Note: other covariates are set at the reference value/mean value.

On the other hand, INVALSI scores at the 90<sup>th</sup> percentile tend to be associated with high final marks (100 and 100 cum laude), with differences varying according to the type of school. More precisely, students who attended scientific high schools and applied sciences high schools report predicted probabilities lower than 0.25, whereas students who attended technical institutes and vocational institutes show probabilities of high final marks definitely higher. This result outlines the presence of a significant interaction effect between type of high school and INVALSI score on the high school final mark. Coherently with low final marks, predicted probabilities of a high final mark are always higher for female students than for male students across all schools, thus suggesting

again that female students outperform male students.

Finally, coherently with a positive association between INVALSI scores and high school final mark, predicted probabilities of having a high final mark are very unlikely for those students who obtained an INVALSI score at the 10th percentile, as well as predicted probabilities of having a low final mark are unlikely for those students who obtained an INVALSI score at the 90th percentile.

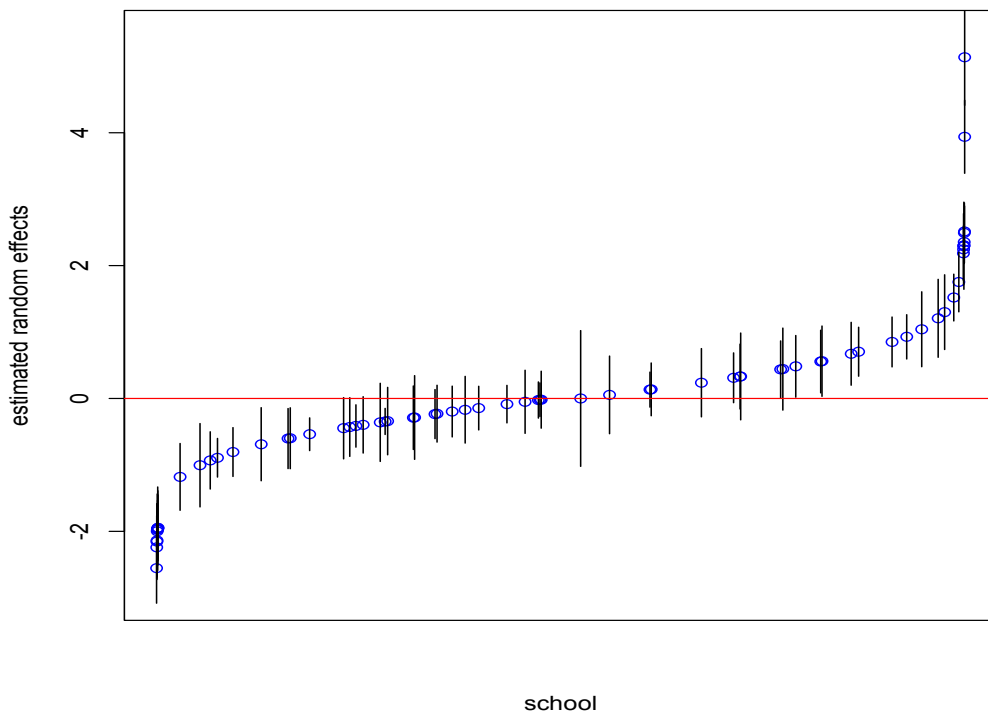
Lastly, Table 3 shows the estimated coefficients for all covariates included in the models. To sum up the effect of variables on high school final mark, all INVALSI scores are positively associated with the school final mark, as well as female students (with respect to male students), residing in the Centre and South of Italy (instead of residing in the North), attending a private school (in comparison with a public school) have all a positive effect on the likelihood of a high final mark. Conversely, being a foreign student has a negative effect on a high final mark. As for the type of high school, all schools have a positive effect on the likelihood of a high final mark with respect to students attending a scientific high school, except students attending an applied science high school, whose coefficient is negative (but only slightly significant). Finally, the two second-level covariates appear to be significant: indeed, both the high school ESCS and the percentage of graduates over 19 in the high school have a negative association with the high school final mark.

**Table 3:** Model coefficients for the multilevel proportional odds model on high school final mark categories (Sample size: 194,778).

	<b>Coeff.</b>	<b>SE</b>	<b>P-value</b>
INVALSI score on Italian	0.559	0.007	0.000
INVALSI score on Maths	0.856	0.007	0.000
INVALSI score on English reading	0.243	0.007	0.000
INVALSI score on English listening	0.296	0.007	0.000
Gender (ref. Male)			
Female	0.689	0.010	0.000
Citizenship (ref. Italian)			
Foreign	-0.343	0.024	0.000
Macroarea of residence (ref. North)			
Centre	1.039	0.032	0.000
South	1.974	0.029	0.000
School property (ref. Public)			
Private	0.455	0.053	0.000
Type of high school (ref. Scientific high school)			
Classical high school	0.913	0.044	0.000
Applied sciences high school	-0.081	0.042	0.053
Foreign language high school	0.857	0.042	0.000
Other high school	1.384	0.042	0.000
Technical institute	1.339	0.043	0.000
Vocational institute	2.385	0.073	0.000
High school ESCS	-0.424	0.039	0.000
% graduates over 19 in high school	-0.004	0.001	0.001
Thresholds			
First: 60-69 score	-0.065	0.037	
Second: 70-79 score	1.786	0.037	
Third: 80-89 score	3.049	0.037	
Fourth: 90-99 score	4.556	0.038	
Random part			
Variance at the high school level	0.542	0.014	

Finally, from Table 3 we observe that the school-level variance is statistically significant and represents the 14% (intraclass correlation coefficient) of the total variance of the response variable explained by the hierarchical structure of data. In more detail, the estimated school-level random effects are displayed in Figure 1 together with the related 95% confidence intervals. For ease of readability, the caterpillar plot reports only a sub-sample of schools: the ten schools with the lowest random effects (on the left side of the plot), the ten schools with the highest random effects (on the right side of the plot), and other fifty randomly selected schools (in the centre of the plot). It is worth to outline how schools at the extremes of the plot significantly differ from the other schools. Moreover, in the two extremes we found different schools (i.e. technical institutes such as classical and scientific high schools), as well as divers geographical location (i.e., Sicily, Tuscany, or Emilia-Romagna) without showing a precise pattern (for example, high schools with a positive influence are located both in South and in the Centre of Italy). At first sight, we could not find any systematic difference between high schools that may have a positive or negative influence on INVALSI scores, but a deeper interpretation is needed to check if potential differences exist.

**Figure 1:** Caterpillar plot: school-level estimated random effects with 95% confidence intervals for a sub-sample of schools.



#### 4. Preliminary conclusions and future research

Our preliminary analyses show that the INVALSI scores are positively associated with the high school final mark, which may be considered an overall performance outcome at the end of the high school career, with higher INVALSI scores corresponding also to higher high school final marks. Despite it, some highlights are worth to be stressed. First, female students achieve high school final marks higher than male students, keeping constant the INVALSI scores and other characteristics. Second, differences by type of high schools are visible too, being constant the INVALSI scores and



other characteristics. Third, the association between INVALSI scores and high school final marks seems to be stronger for lower scores/marks. These issues rise some doubts. On one side, they question about the real capability of INVALSI tests to predict the performance at the high school final examination; on the other side, the high school final evaluation is not exempt from disparities according to gender and type of school, irrespective the INVALSI scores.

Given these preliminary results, we will proceed with a deeper analysis of our results in the light of eventual differences on individual characteristics – such as student’s geographical area of residence – and on school-level characteristics – such as high school quality (for example, in terms of percentage of graduates over 19). Moreover, in light of the discrepancies between INVALSI scores and high school final marks above outlined, both these types of information will be object of interest in a next step concerning the academic career of students in terms of credits earned at the first year of university. In particular, it will be of primary interest to investigate the predictive capability of INVALSI scores and the high school final mark, and the differences between them, also taking into account the high school of origin and the gender. More precisely, to analyse the predictive capability of the INVALSI scores and the high school final mark on the academic students’ career (evaluated in terms of credits earned in the first year), we will estimate a multilevel model, to take into account that students are nested within athenaeums. Then, the functional form of the model will be chosen in accordance with the distribution of the number of credits earned in the first year, which, at first sight, does not seem distributed as a normal variable and shows one or two peaks around zero and/or sixty credits in most athenaeums. We will interpret our results in the light of assessing potential divergences in students’ performances during the transition from high school to university.

## References

- Goldstein, H. (2010). *Multilevel Statistical Models*. 4th Edition, John Wiley & Sons, Ltd
- Liu, I., Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, **14**(1), pp. 1-73.
- Snijders, T. A.B., Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London, Sage Publishers.

# Profiling students' satisfaction towards university courses with a latent class approach

G. Damiana Costanzo, Michelangelo Misuraca, Angela Coscarelli

## 1. Introduction

Collecting and analysing students' opinions on their learning experiences during enrollment in an academic program is widely recognised as a key strategy for evaluating tertiary education quality. Academic institutions require students to participate every year in specific surveys, aiming to gather their viewpoints about the organisation of the single courses and their feelings about the teaching activity's traits and effectiveness. The *Standards and guidelines for quality assurance in the European Higher Education Area* (ESG, 2015), for example, underline the relevance of students' voices in the assessment processes. Likewise, students' and graduates' opinions constitute essential information for the quality assurance of the Italian university system. The National Agency for the Evaluation of Universities and Research Institutes (ANVUR), to harmonise the data collection in all the universities, provides guidelines to build the questionnaire administered in the surveys on students' opinions. Initially released in 2013, these guidelines were updated in 2017 and are currently in use. The survey, under article 1 of Law 370/1999, is mandatory and autonomously carried out every academic year by the different institutions, representing one of the fundamental sources for the so-called AVA system<sup>1</sup> (i.e., self-assessment, periodic assessment, accreditation), introduced in Italy by Law 240/2010 and Legislative Decree 19/2012. This system aims to improve teaching and research quality in universities, applying a model based on internal planning and management procedures.

This study investigates students' satisfaction towards courses delivered at the University of Calabria, focusing on the academic year 2020–2021. In this period, almost corresponding to the occurrence of the second and third waves of COVID-19 outbreak in Italy, traditional teaching methods were substantially disrupted by the social distancing actions pursued by the Italian government, enhancing the use of blended and hybrid learning (e.g., Aboagye et al., 2021; Chaturvedi et al., 2021). Here we considered the first-year courses since students enrolled in 2020–2021 programs experienced the completion of their previous educational degrees in the first wave of COVID-19. We carried out a latent class analytical strategy to profile students' satisfaction at a course level, taking into account their interest in each course and their perceptions about the course organisation and the instructor's behaviour. Since the items listed in the survey are expressed as 4-point balanced scales, we used the so-called Latent Profile Analysis (LPA) to identify unobserved course profiles, starting from students' responses to the continuous indicators concerning course satisfaction.

## 2. Theoretical framework and data structure

LPA is a statistical approach belonging to finite mixture models. It can be seen as a variant of latent class analysis (LCA, Oberski, 2016) aiming at identifying a set of discrete, exhaustive, and non-overlapping groups of subjects characterised by different patterns of responses on indicator items, typically represented by ordinal or continuous manifest variables. Each subject

---

<sup>1</sup><https://www.anvur.it/attivita/ava>

is assigned to the most likely latent group, i.e. an unobserved profile that generates patterns of responses on the indicators. LPA may be considered, therefore, a case-centred analytic tool focusing on similarities and differences among subjects rather than relations among variables (Bergman and Magnusson, 1997).

Assuming that the continuous (or ordinal) variables are normally distributed within each latent profile, a model of  $G$  components aims at representing the distributions of the observed subjects' scores on a set of indicator items  $\mathbf{x}_i$  ( $i = 1, \dots, n$ ), given the latent categorical variable  $\Theta$ , as a function of the probability of the subjects to be typed into a profile and the profile-specific normal density:

$$f(\mathbf{x}_i|\Theta) = \sum_{k=1}^G \pi_k f_k(\mathbf{x}_i|\theta_k) \quad (1)$$

where  $\pi_k$  and  $\theta_k$  represent the probability of belonging to the  $k$ -th latent profile (with  $\pi_k$  summing to 1 across the different profiles) and the estimation of the mean and the set of variances/covariances for  $k$ , respectively (Tein et al., 2013).

LPA has been recently used in the educational domain, for example, to identify students' time use profiles (Fosnacht et al., 2018), to explore motivation patterns in learning environments (Hodis and Hodis, 2020), to define types of social support for student resilience during the COVID-19 pandemic (Mai et al., 2021). In the following, LPA is used to identify academic course profiles, considering students' opinions about the courses they attended.

The yearly survey about students' satisfaction carried out at the University of Calabria (Italy) was used to build a dataset of course response patterns. The questionnaire administered in the survey is built following the ANVUR guidelines to harmonise the data collection in all the different Universities.

We focused, in particular, on the academic year 2020–2021. During this year, because of the COVID-19 pandemic, courses have been delivered in presence, in distance (via online platforms), or by mixing the two types. The total number of the collected questionnaires was 77,049 for the courses of 73 academic programs included in the entire university catalogue. After filtering only the questionnaires completed by first-year students that attended at least half of the lectures for each first-year course (24,064), we selected 10 different items concerning for each course the *Interest* of the students, the *Instructor* behaviour and the *Teaching* characteristics. Table 1 lists the three dimensions and the corresponding items.

Table 1: Dimensions and indicator items.

Dimension	Item	Short label
Interest (INTE)	<i>Interest towards the course</i>	INT_1
	<i>Interest stimulated by the instructor</i>	INT_2
Instructor (INST)	<i>Instructor clear in the explanations</i>	INS_1
	<i>Instructor available for tutoring</i>	INS_2
	<i>Instructor on time at the lectures</i>	INS_3
Teaching (TEAC)	<i>Course coherent with the syllabus</i>	TEA_1
	<i>Examination rules clearly defined</i>	TEA_2
	<i>Adequate prior knowledge for the course</i>	TEA_3
	<i>Adequate learning material</i>	TEA_4
	<i>Workload proportional to ECTS</i>	TEA_5

Since the items listed in the survey are expressed as 4-point balanced scales (*definitively no, more no than yes, more yes than no, definitively yes*), we converted the ranks in scores from 1 to

4. Moreover, since the items INT\_1 and INS\_3 showed a certain number of missing data (10.8% and 6.7%, respectively), we performed a multiple imputation procedure in order to save all the selected cases in the dataset. Finally, the cases were collapsed at a course level by averaging the individual response patterns, and hence the course response patterns were standardised. The resulting  $657 \times 10$  matrix was used in the analysis. The imputation of missing data was performed by using the R library `mice`, whereas LPA was performed by using the library `mclust` (van Buuren and Groothuis-Oudshoorn, 2011; Scrucca et al., 2016).

### 3. Model selection and main findings

A key question in finite mixture modelling is how many latent classes should be included. The selection of the best model was carried out by jointly evaluating the Bayesian Information Criterion (BIC) and the Integrated Complete-data Likelihood (ICL) criterion proposed by Biernacki et al. (2000). ICL appears more robust than BIC, adding a penalty on solutions with greater entropy or classification uncertainty.

In addition to the number of profiles, the model can be specified in terms of whether and how the variable variances and covariances are estimated. Geometric features (shape, volume, orientation) of the clusters are determined by the covariances. We estimated two kinds of models, considering in both cases profiles with equal volume and shape. In the *EEI* model, the indicator variables are set to have zero covariances within and across profiles. Indicator-variable variances are allowed to vary within profiles but are constrained to be equal between them. In the *EEE* model, the complete variable (co)variance matrix is estimated, with variances and covariances constrained to be the same across the profiles.

Table 2 shows the fit indices of the two models, including the log-likelihood  $\ell$  (with the corresponding degrees of freedom), the BIC and the ICL.

Table 2: Data fit of EEI and EEE models.

N. of profiles	EEI				EEE			
	$\ell$	$df$	BIC	ILC	$\ell$	$df$	BIC	ILC
1	-9317.31	20	-18764.37	-14304.68	-6941.49	65	-18764.37	-14304.68
2	-8081.98	31	-16365.07	-16404.11	-6871.95	76	-14236.97	-14293.81
3	-7553.06	42	-15378.61	-15444.33	-6765.99	87	-14096.41	-14160.06
4	-7304.36	53	-14952.57	-15051.02	-6628.85	98	-13893.49	-13912.09
5	-7188.40	64	-14792.00	-14927.39	-6640.24	109	-13987.64	-14472.57

The log-likelihood had lower values for each model that increased of one latent profile, with EEE models showing lower values than EEI models. Regarding the information criteria, we observed that the two indices confirmed that EEE models are better than EEI models. Jointly considering the results of the three selection criteria listed above, we selected the EEE model with 4 latent clusters. To validate our choice, we performed a bootstrap likelihood ratio test (BLRT) to verify the null hypothesis that a  $(k + 1)$ -profile model is equal to or better than a  $k$ -profile model, i.e. that an increase in the number of profiles increases fit (Nylund et al., 2007). Table 3 shows the results of the test and the corresponding  $p$ -values, suggesting that a 4-profile solution is optimal.

The four profiles allowed to identify different satisfaction patterns for the first-years courses under investigation. Table 4 reports the mean values of the different items per profile, looking at each as a factor score with a mean equal to 0 (due to standardisation).

Profile 1 (90.26% of the courses) showed for each item a score above 0, identifying courses with an average level of student satisfaction. Profile 2 (5.48% of courses) showed for the *Interest*

Table 3: BLRT for the EEE models.

	1 vs 2	2 vs 3	3 vs 4	4 vs 5
LRTs	139.072	211.926	274.282	-22.787
p-value	0.001*	0.001*	0.001*	0.217

Table 4: Mean values of the ten items per profile.

	Item	Profile 1	Profile 2	Profile 3	Profile 4
INTE	<i>Interest towards the course</i>	0.01	0.14	-0.08	-0.68
	<i>Interest stimulated by the instructor</i>	0.05	0.16	-1.21	-1.18
INST	<i>Instructor clear in the explanations</i>	0.07	-0.07	-1.10	-1.53
	<i>Instructor available for tutoring</i>	0.07	0.14	-0.58	-3.77
	<i>Instructor on time at the lectures</i>	0.12	0.14	-3.60	-1.17
TEAC	<i>Course coherent with the syllabus</i>	0.04	0.12	-0.94	-1.05
	<i>Examination rules clearly defined</i>	0.03	0.03	-0.77	-0.61
	<i>Adequate prior knowledge for the course</i>	0.02	-0.16	0.37	-0.98
	<i>Adequate learning material</i>	0.06	-0.56	-0.59	-0.60
	<i>Workload proportional to ECTS</i>	0.14	-2.13	0.17	-0.38
	% membership	90.26	5.48	2.74	1.52

and *Instructor* dimensions – as well as the items of the *Teaching* dimension related to course organisation – scores above 0, with values greater than the corresponding values of Profile 1. The only negative value was observed for the clarity of instructors, with a value just below 0. Nevertheless, the courses likely belonging to this profile showed negative scores for the adequacy of the prior knowledge required to attend the lectures and the learning materials, with a peak for the course workload that is not perceived as proportional to the ECTS. Profile 3 (2.74% of the courses) is characterised by a fair share of students’ dissatisfaction, particularly concerning the instructors’ activities. Courses likely belonging to this profile showed very negative scores for the interest stimulated by the instructor, the clarity of this latter in the explanations, and a peak for the punctuality of the instructors. On the other hand, the profile had positive scores for the adequacy of the prior knowledge and the workload. Profile 4 (1.52% of the courses), finally, was even more characterised by dissatisfaction, with scores significantly above 0 for all the items. In particular, the *Instructor* dimension showed very negative scores, together with a low score for the interest stimulated by the instructors themselves. At the same time, we observed negative scores for the items of the *Teaching* dimension, with an unfavourable evaluation of the coherence of the courses with respect to their syllabus and an inadequacy of prior knowledge and learning materials. A joint lecture on the profile membership and some covariates can offer valuable insights into the less satisfying courses.

To characterise the profiles, we focused at this stage of the research only on the nature of academic programs in which courses are embodied, taking into account if the courses belong to undergraduate and single-cycle programs (namely, *Laurea* and *Laurea Magistrale a Ciclo Unico*) or to master programs (namely, *Laurea Magistrale*). Focusing on the profiles that showed a certain degree of dissatisfaction, we observed that 83.3% of courses in Profile 2 are embodied in master programs. At the same time, the 66.7% and the 50.0% of courses in Profile 3 and Profile 4, respectively, are embodied in master programs. This aspect can help characterise the source of the satisfaction (or dissatisfaction) perceived by students, helping academic institutions perform targeted interventions on the courses showing specific shortcomings. Con-

sidering, for example, the courses belonging to Profile 2, preparing more effective learning materials and re-designing the course programs may improve students' satisfaction, taking into account the higher expectations of master students. Nevertheless, the data used in this study referred to the academic year 2020–2021, which occurred during the second and third waves of the COVID-19 pandemic in Italy. This means that a comparison with other academic years is necessary to detect potential structural weaknesses that deserve greater attention.

#### **4. Final remarks and future research**

This study analysed students' satisfaction towards courses attended during the first year of the academic programs delivered at the University of Calabria. The different course types were depicted using a latent class approach known as LPA, a finite mixture model able to measure the impact of an unobserved categorical variable defining different latent profiles on a set of continuous variables. The survey structure did not allow us to evaluate satisfaction at a student-level since each questionnaire is registered with a different ID due to the privacy policies implemented at the University of Calabria. For this reason, we evaluated satisfaction at a course level, trying to identify different course types and analyse their characteristics. By contrast, the data averaging did not allow considering the variability of students' opinions for each course, posing a possible limitation of the present study.

The response patterns expressing students' satisfaction/dissatisfaction levels to the different aspects concerning teaching and the organisation of learning activities characterised a 4-profile model, revealing for each one the most critical aspects. Remarkably, together with a profile encompassing the majority of courses and revealing a general degree of satisfaction, we identified three course profiles expressing a different degree of dissatisfaction instead. A focus on the academic programs the latter courses belong – considering if they were included in first-cycle, single-cycle or second-cycle programs – showed that graduate students have potentially higher expectations than undergraduate students, evaluating in a more critical way the course organisation and the workload required by each course. A noteworthy aspect is that these students experienced the rapid change of teaching induced by COVID-19 during the last year of their first-cycle degree, starting a new cycle of study in the uncertainty caused by the ongoing social distancing and the limitations established by the Italian government.

The analytical strategy employed in the study can be easily implemented as a visual tool helping academic institutions at a department level (e.g., by the Joint Teaching-Student Committees) or at a university level (e.g., by the Independent Evaluation Units) in the quality assurance systems, giving a hint of which courses have to be carefully monitored and at the same time of which aspect are perceived as more critical by the students. Currently, a different version of the survey has been tested in some universities by the National Agency for the Evaluation, and it is ready to be released in the short term a new version of the guidelines.

Several future developments of the study can be considered. First, the effect of some meta-data (as covariates) may help better define and characterise the profiles, taking into account the cycle dimension and the attendance rate, and the domain of the courses (e.g., courses based on quantitative or qualitative methods). Second, a longitudinal analysis may help in evaluating how a shock like the COVID-19 pandemic or a change in the academic programs' regulation influences the perception of courses' quality, estimating the transition of a course (in a probabilistic fashion) from a latent profile to another one in different periods (Collins and Lanza, 2013). This latter variant of latent models can enrich the analytical strategy's informative power, allowing for evaluating the quality across time.

## References

- Aboagye, E., Yawson, J.A., Appiah, K.N. (2021). COVID-19 and e-learning: The challenges of students in tertiary institutions. *Social Education Research*, **2**(1), pp. 1–8.
- Bergman, L.R., Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, **9**(2), pp. 291–319.
- Biernacki, C., Celeux, G., Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(7), pp. 719–725.
- Chaturvedi, S., Purohit, S., Verma, M. (2021). Effective Teaching Practices for Success During COVID 19 Pandemic: Towards Phygital Learning. *Frontiers in Education*, **6**(646557)
- Collins, L.M., Lanza, S.T. (2013). *Latent class and Latent Transition Analysis: with applications in the social, behavioral, and health sciences*. Wiley, New York, (NY).
- ENQA, ESU, EUA, EURASHE (2015). *Standards and Guidelines for Quality Assurance in the European Higher Education Area*. Brussels, Belgium.
- van Buuren, S., Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of statistical software*, **45**(3), pp. 1–67.
- Fosnacht, K., McCormick, A.C., Lerma, R. (2018). First-year students' time use in college: A latent profile analysis. *Research in Higher Education*, **59**(7), pp. 958–978.
- Hodis, F.A., Hodis, G.M. (2020). Patterns of motivation and communication in learning environments: a latent profile analysis. *Social psychology of Education*, **23**, pp. 1659–1685.
- Mai, Y., Wu, Y.J., Huang, Y. (2021). What type of social support is important for student resilience during COVID-19? A Latent Profile Analysis. *Frontiers in Psychology*, **12**(646145)
- Nylund, K.L., Asparouhov, T., Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: a multidisciplinary journal*, **14**(4), pp. 535–569.
- Oberski, D. (2016). Mixture Models: Latent Profile and Latent Class Analysis, in *Modern statistical methods for HCI*, eds. J. Robertson, M. Kaptein, Springer, Cham, pp. 275–287.
- Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, **8**(1), pp. 205–233.
- Tein, J.-Y., Coxse, S., Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural equation modeling: a multidisciplinary journal*, **20**(4), pp. 640–657.

# The relation between students' educational performances and their access test results: a focus on an Italian case

Matteo Corsi, Luca Persico, Sara Preti, Agnese Sechi

## 1. Introduction, Data and Descriptive analysis

This paper aims at analyzing the relationship between the university performances of freshman students, measured by the University Credits (CUs)<sup>1</sup> gathered during the first semester, the results achieved in T.E.L.E.MA.CO. (*TEst di Logica E MAtematica e COmprensione verbale*) test and their social-demographic characteristics. Starting from the Bologna Declaration of 1999 (ministerial decree of November 3, 1999, no. 509), the Italian university system has seen important changes at the organizational, educational, and financial levels. The training credit model was introduced for harmonizing national and international university systems. Another change of major importance in the reform consisted of the reorganization of degree courses into homogeneous classes. The reform established a three-cycle higher education system comprising undergraduate (3-years bachelor's degrees), master's or specialist degrees (2-years master equivalent degrees), and doctoral studies. The education system also provides for the possibility of attending other courses such as first and second-level masters. Furthermore, in 2004 non-selective admission tests were introduced for all bachelor's degrees.

The Department of Economics and Business Studies (DIEC) of the University of Genoa (Italy), which has open-enrolment courses, adopted T.E.L.E.MA.CO. test, a very important tool for verifying initial knowledge considered functional to the effective participation of a university course. It consists of two sections: a common core for all degree programs, aimed at proving the basic skills of comprehension of Italian texts (literacy), and logical reasoning skills (numeracy), and a differentiated section according to the chosen program<sup>2</sup>. Additional mandatory tasks will be assigned to students who gain a score lower than the established thresholds.

Data are collected by the DIEC. The main dataset derives from three different sources: the first one contains information related to sociodemographic characteristics and students' educational backgrounds; the second one is about information relating to the university career; the last one concerns the results of the T.E.L.E.MA.CO admission test. The main dataset records information on 488 students enrolled in the Department of Economics of the University of Genoa; they are all pure freshmen (first matriculation in the university) and not exempted from the obligation to take the test<sup>3</sup>. The considered attributes are age, gender, high school, diploma grade, course of study, results of T.E.L.E.MA.CO. test, and average number of CUs.

Once the main dataset has been assembled, we performed a descriptive analysis of the students' characteristics. The average age of the students is 19 years, the females represent 31% of the sample. 55% of students are enrolled in Business Administration, 27% in Economics of Maritime Business, Logistics and Transport, and 18% in Economics. The average high school final grade is 74.78, and 25% of the students have a grade higher than 81. Women in the sample

---

<sup>1</sup>CUs represent indicators that measure the workload required to attend the lessons and prepare for the specific exam.

<sup>2</sup>Students pass the common core if and only if obtain a score equal to or higher than 12 out of 20. Then, those who have passed the common core and who have achieved a score equal to or greater than 6 in the individual sections (literacy and numeracy) can access the T (text) and M (mathematical) extensions respectively.

<sup>3</sup>Students who are exempted are students who have achieved a high school final score equal to or greater than 90/100 or in other peculiar situations listed at the following link: <https://unige.it/studenti/telemaco#cosaTELEMACO>.

Matteo Corsi, University of Genoa, Italy, [matteo.corsi@edu.unige.it](mailto:matteo.corsi@edu.unige.it), 0000-0001-7545-3600

Luca Persico, University of Genoa, Italy, [luca.persico@unige.it](mailto:luca.persico@unige.it), 0000-0002-5436-2627

Sara Preti, University of Genoa, Italy, [sara.preti@edu.unige.it](mailto:sara.preti@edu.unige.it), 0000-0002-7424-9998

Agnese Sechi, University of Genoa, Italy, [agnese.sechi@edu.unige.it](mailto:agnese.sechi@edu.unige.it)

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Matteo Corsi, Luca Persico, Sara Preti, Agnese Sechi, *The relation between students' educational performances and their access test results: a focus on an Italian case*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.05, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 23-28, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



are on average better than males in terms of high school final grades: female students have a mean equal to 76.75, while men’s one is equal to 73.88. A t-test confirmed a significant difference on average between the two groups.

To have a whole picture of the scenario, it is interesting to deepen into how the unlike performances are related to the different types of high schools. Table 1 shows the frequency distribution of students’ high school and university performances by school of origin. High school performances are measured by the average grade of diploma, while university performances by the number of CUs gained during the first semester and by the average score of the Common Core Score (named *CC Score* in Table 1) in the TE.L.E.MA.CO. test. It is worth noting that about 40% of students enrolled in Economics in the year 2021 come from the scientific high school, followed by the technical institute with 30% and the vocational institute and linguistic high school with 9%. Regarding the TE.L.E.MA.CO. test results, 346 students out of 488 students have been successful: 65% of the total girls and 74% of the total males who do the test, pass it<sup>4</sup>. Focusing on the sample distribution of the scores gained by students grouped by gender in the common core of the TE.L.E.MA.CO. test, there are no gender gaps in the scores obtained in the literacy section; on the other hand, differences emerge in the scores in the numeracy section. If there is a gap in favor of females relating to high school performances, the scenario tips up and male students perform better than females in the numeracy section, a result that has been confirmed with a t-test<sup>5</sup>. These two results may be consistent. Indeed, we do not know if the differences in STEM<sup>6</sup> subjects performances (which occur in our sample for the numeracy section) in favor of males also exist in the grades of the high school STEM tests or not. On average, we know that females get higher graduation marks, but we do not know what their performance in STEM subjects is. It should be considered that our sample examines students who must necessarily take the test (therefore not the best in terms of school performance) and that the male-female ratio which comes from scientific high schools (students with a stronger propensity in scientific subjects) is very high, compared to other institutes. There is therefore certainly a problem with sample selection and balancing, which does not allow us to interpret the problem of the gender gap exhaustively and completely.

**Table 1:** Distribution of students’ high school and university performances by school of origin

<i>School Type</i>	<i>Number of students</i>	<i>Female</i>	<i>Male</i>	<i>Average grade</i>	<i>CC Score</i>	<i>CUs</i>
Other types	13	4	9	73.78	11.92	10.38
Vocational Institute	43	19	24	75.21	11.02	5.44
Technical Institute	145	44	101	76.23	12.48	11.98
Classical High School	24	6	18	77.08	14.17	13.88
High School for Human Sciences	23	12	11	75.96	12.70	7.04
Linguistic High School	43	25	18	75.79	13.93	12.14
Scientific High School	197	43	154	73.04	14.20	16.99

Focusing on the students’ background, Figure 1 shows that on average people who come from scientific and classical high schools perform better than the others in all the sections, and on average students who attended vocational or other types of high school (such as music or artistic high school) do not pass the common core of the TE.L.E.MA.CO. test. Students who come from the scientific school perform much better than others, even compared to the students of the classic school, with regard to the extension of mathematics. Finally, we examine the performance of students during the first semester by looking at the number of CUs (which ranges from 0 to 27); 33% of students do not pass any exams (0 CUs), while 28% reach 27 CUs

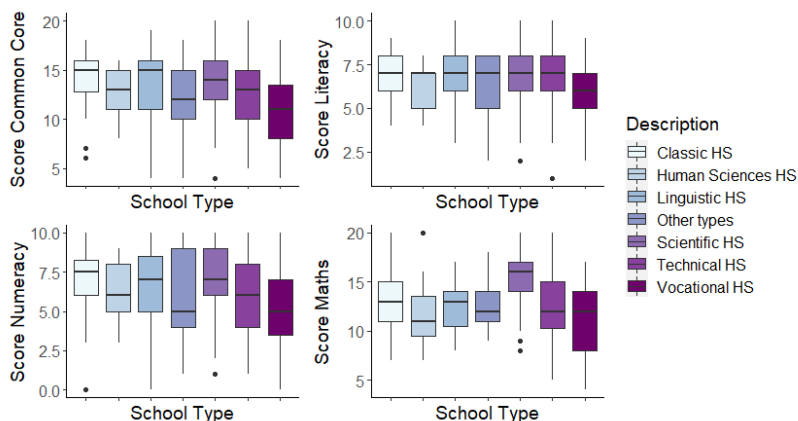
<sup>4</sup>Moreover, 253 students pass the mathematical extension: 43% of the total girls and 55% of the total males who do it, pass it. No one is allowed to do the text extension.

<sup>5</sup>This result is consistent with the literature about the gender gap in STEM courses (Priulla et al., 2021).

<sup>6</sup>STEM is an acronym for the fields of science, technology, engineering, and math.

threshold. The number of CUs earned at the end of the first semester follows the same trend for both male and female students. Looking at the backgrounds, students from vocational institutes, human sciences, or other types of high schools perform worse, while people with scientific and classical backgrounds earn a greater number of CUs.

**Figure 1:** TE.L.E.MA.CO. test scores' distribution per school type



Source: Computed on the basis of data from DIEC, 2021

## 2. Empirical Model

In this section, we perform two different models, a logistic and an ordered logistic, to study the probability of acquiring CUs. These approaches are useful to understand when and how timely policies and programs can be implemented to avoid losing students, a frequent trend, especially in the first semester of the first year. Specifically, the main goal of the logit model is to represent the probability of getting at least 18 CUs<sup>7</sup> during the first semester, with respect to students' characteristics and their TE.L.E.MA.CO. test results. This model and the idea of expressing the dependent variable as a dummy depend on the fact that, after only a few months from the start of a university career, a student has necessarily given few, if any, exams. This implies the existence of a minimum number (0) and a maximum number (27) of credits which prompted us to consider the exceeding or not of the threshold as a proxy of academic performance. The binary dependent variable is equal to 1 if students gain at least 18 CUs (2 exams) at the end of the first semester, and 0 otherwise. The independent variables included in the model are the following: gender (dummy variable); age at enrolment<sup>8</sup>; high school final mark (which are normalized from 60 to 100); type of school; university courses; two variables that capture the literacy and numeracy scores<sup>9</sup>; a variable that measures the distance in km between home (we use the high school address as a proxy) and the university; and a variable which represents the average income in the municipality where they reside, as a proxy of the students' parents income. We suppose that both variables have an important, even if indirect, impact on students' performances. The idea that commuting or changing the habit and home (especially at

<sup>7</sup>We have chosen this threshold because it represents 2 out of 3 exams since in the first semester there are only 9-credit exams by default.

<sup>8</sup>The variable is dichotomous in  $\leq 19$  and  $> 19$ ; the dummy assumes the value 0 if the student has a regular or early schooling path, otherwise it takes the value 1.

<sup>9</sup>We do not consider the mathematical extension score because this variable hides the effects of other covariates, even though only a part of the sample accesses the test.

the early beginning) may negatively affect performances is widespread in the literature (Tigre et al., 2016). Also, socio-economic situations can influence school achievement. The left side of Table 2 reports the main results of the logit model (odds ratios, estimated coefficients, standard errors, and p-value significance).

**Table 2:** Logit and Ordered Logit estimates

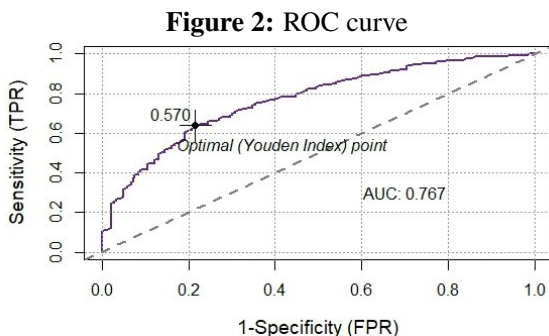
	<i>Logit model</i>				<i>Ord. Logit model</i>			
	<i>OR</i>	<i>Coef.</i>	<i>SE</i>	<i>Signif.</i>	<i>OR</i>	<i>Coef.</i>	<i>SE</i>	<i>Signif.</i>
Inactive — 1 exam					0.138	-1.982	0.223	***
1 exam — 2 exams					0.311	-1.169	0.214	***
2 exams — 3 exams					1.058	0.056	0.209	
Intercept (Logit)	3.125	1.139	0.259	***				
Gender (M)	0.758	-0.277	0.242		0.777	-0.252	0.198	
Age at enrollment	0.861	-0.150	0.274		0.880	-0.127	0.234	
Technical HS	0.455	-0.788	0.272	**	0.394	-0.933	0.214	***
Vocational HS	0.095	-2.354	0.477	***	0.102	-2.283	0.382	***
Classic HS	0.438	-0.826	0.482	.	0.391	-0.940	0.377	**
Linguistic HS	0.259	-1.353	0.381	***	0.291	-1.233	0.315	***
Humans Sciences HS	0.160	-1.833	0.528	***	0.121	-2.114	0.451	***
Other types	0.345	-1.063	0.673		0.326	-1.121	0.176	***
HS grade	1.081	0.078	0.014	***	1.075	0.072	0.011	***
Score Literacy	1.081	0.078	0.070		1.075	0.072	0.060	
Score Numeracy	1.149	0.138	0.050	**	1.166	0.154	0.042	***
Economics of MB	0.995	-0.005	0.249		0.818	-0.201	0.210	
Economics	0.909	-0.095	0.291		0.929	-0.074	0.238	
Distance from home	0.998	-0.002	0.000	*	0.999	-0.001	0.001	
Average Income	1.000	-0.000	0.000		1.000	-0.000	0.000	
Signif. codes	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ''			1

The baseline student has the following profile: female, who comes from the scientific high school, with an age of 19 years at most (therefore regular from the academic point of view), with a final grade equal to 74.78 (average diploma grade of the sample) and who has reached the average sample results in both literacy and numeracy sections. In addition, this student attends Business Administration, has an income equal to the average of the sample, and has a zero distance from the university.

Proceeding with the analysis of the results obtained from the logit regression, the intercept shows that for the baseline student the probability to gain at least 18 CUs is 76% and the odds ratio is 3.125 with a significance of (with  $p < 0.01$ ). Regarding the school types, we can see that students attending different high schools to the scientific one are less likely to obtain the credit threshold with a high significance. The Other types high school category, on the other hand, is not significant. Another relevant variable is the High School final grade; for a unit increase in the final grade, the log odds of CUs increases by 1.081 (with  $p < 0.01$ ). About the admission test, we can see that the score achieved in the numeracy section is the only significant: with a probability of 53% students who have a score higher than the mean, perform better. Distance also has a significant impact on students' performance: the further away a student is from the university, the less likely it is to take two out of three exams. In literature, the role of commuting as a penalty in student performance has already been addressed, although not extensively: the waste of time associated with the hours of travel, the physical and mental stress of being far away, and also the greater difficulty in creating work and friendship groups are certainly some of the main components.

To assess the performance of the logit model we use the area under the receiver operating characteristic (ROC) curve (AUC). The AUC value of the logit model is equal to 0.767; since

the larger the AUC, the more accurate will be the prediction model, the logit model can be considered as sufficiently accurate. Another way to assess the model performances is to examine the agreement between actual observations and predictions, through a contingency table. In order to transform the student's predicted probability (probability of obtaining at least 18 CUs) into a predicted class (if the student has obtained at least 18 CUs) is sufficient to define a specified cut-off probability value. This value is computed using the *Youden's index*<sup>10</sup> (Youden, 1950), and it is equal to 0.570, as shown in Figure 2. Finally, we consider the actual and predicted classification to measure the goodness of the logit model: the percentage of correctly classified is 70%.



*Source:* Computed on the basis of the logit model's output

Since in the first semester, students have done only 0, 9, 18, or 27 CUs, and every exam has the same number of CUs (9) and so the same difficulty, we have decided to perform an ordered logistic model trying to capture more information. Also in this model, the dependent variable is the students' performances in terms of CUs, but this time it is measured on an ordinal scale in 4 categories: 0 exams (inactive) corresponding to 0 credits, 1 exam to 9, 2 to 18, and 3 to 27. The right side of Table 2 shows the main results of the ordered logistic model. As we can see, there are three estimates of the intercept because, being four the variables, three are the cutoffs from one category to another. About the last cutoff, it is worth noting that the third and fourth categories (2 exams and 3 exams respectively) are not significantly different, therefore they could be aggregated without consequences. Also in this case it is more interesting to comment on the coefficients, which confirm the results of the logistic model, even if some differences emerge: the variable *Other Types* becomes significant, and the influence on the dependent variable of other covariates (Technical, Classic, Score Numeracy) increases. However, the Distance from home loses its significance. Compared to the baseline, set as previously, males rather than females, students of other schools than the scientific, and with a lower than average diploma and numeracy grades are more likely to obtain fewer CUs. We also perform a Brant test to check the hypothesis of parallelism and the test suggests that ordered logit's regression assumptions are met. In addition to the results of ordered logit coefficients, marginal effects are used to predict the effect and the magnitude of change. Concerning the high school type, we can see that students who came from a high school other than the scientific (model baseline), have a lower probability to reach two or three exams; in particular, the probability is much lower for the vocational and the human science high schools (in these cases also the likelihood of students to get one exam is lower). Furthermore, students who attended classic and technical high schools have a higher probability to take at least one exam: for example, a student from a classical

<sup>10</sup>The Youden's index, also called Youden's J statistic, was developed in 1950 by W.J. Youden and represents a single statistic that captures the performance of a dichotomous test. The index considers both the true positive rate (Sensitivity), and the true negative rate (Specificity), and it is given by Sensitivity+Specificity-1.

high school has a probability of 0.327 of getting two exams higher than a student from human sciences. Moreover, if the student's high school grade or the score in the numeracy section increases by one point, then the likelihood of taking zero exams decreases by 1.26% and 2.69% respectively.

### 3. Conclusions

The objective of this work was to analyze the relationship between students' university performances, measured by the University Credits (CUs) gathered during the first semester, and the results achieved in T.E.L.E.MA.CO. test, a useful tool for orientation and access to university studies based on solid scientific methodologies, and their social-demographic characteristics. A logit and an ordered logit model are used to compute the probabilities to reach at least 18 CUs (logit) or to obtain 0, 9, 18, and 27 CUs (ordered logit). What emerges from the models is that various factors are determinants. About the students' background, the graduation grade and the type of school predict the success at exams (especially in a negative way for vocational, linguistic, and human sciences high schools). As for the test, the evaluation of the numeracy section is the main determinant of success in performance. Based on a consistent statistical approach, our result seems to confirm the ability of the admission test to predict academic success in the first year (Bestetti et al., 2020; Migliaretti et al., 2017; Carrieri et al., 2013; CISIA, 2020). Furthermore, given the fact that students we consider obtain a diploma grade lower than 90, the admission test is also significant in the presence of the high school grade, providing additional information that the latter element fails to provide. Also for this reason the test can be a powerful tool and a good alternative to the high school final mark as a university admission indicator, often the only information used. It would be interesting as future work to understand if additional and perhaps differentiated approaches are necessary according to the background of each student, especially at the beginning of their university careers. In addition, hybrid solutions for distance and face-to-face teaching could be implemented to facilitate off-site students.

### References

- Bestetti, R.B., Couto, L.B., Roncato-Paiva, P., Romão, G.S., Faria-Jr, M., Furlan-Daniel, R.A., Geleilate, T., Jorge-Neto, S., Mendonça, F., Garcia, M., Durand, M.T. (2020). University Admission Test Associates with Academic Performance at the End of Medical Course in a PBL Medical Hybrid Curriculum. *Advances in Medical Education and Practice*, 11, 579.
- Carrieri, V., D'Amato M., Zotti R. (2013). *Selective Admission Tests and Students' Performances. Evidence from a Natural Experiment in a Large Italian University*. (No. 0/00). CELPE-Centre of Labour Economics and Economic Policy, University of Salerno, Italy.
- Cisia (2020). *Il Test D'Ingresso. Valenza Predittiva e Rilevanza Nei Percorsi di Orientamento Universitario*. Edizioni CISIA, Pisa, pp. 1-68.
- Migliaretti, G., Bozzaro S., Siliquini R., Stura I., Costa G., Cavallo F. (2017). Is the admission test for a course in medicine a good predictor of academic performance? A case-control experience at the school of medicine of Turin. *BMJ Open*, 7(11), pp. 1-7.
- Priulla, A., D'Angelo, N. & Attanasio, M. (2021). An analysis of Italian university students' performance through segmented regression models: gender differences in STEM courses. *Genus*, 77, 11.
- Tigre, R., Sampaio, B., Menezes, T. (2016). The impact of commuting time on youth's school performance. *Journal of Regional Science*, 57(1), pp. 28-47.
- Youden, W.J., (1950). Index for rating diagnostic tests. *Cancer*, 3(1), pp. 32-35.

# Structure and dynamics of immigration in the municipalities of northwestern Italy

Simona Ballabio, Arianna Carra, Flavio Verrecchia, Alberto Vitalini

## 1. Introduction

In less than a century, Italy has been characterized by profound changes in migration phenomena. From a country of origin to a country of destination of international migration flows, it has seen a strong and rapid intensification of incoming migration, and then reached a phase of stabilization. In the first phase, immigration mainly affects metropolitan areas and industrial zones. While in the second phase, the presence of foreigners becomes a structural phenomenon, characterized by a prolonged presence. In particular, there is a trend toward the territorial spread of the phenomenon and the increasing peripheral configuration of areas with a high concentration of foreigners (Costarelli and Mugnano, 2017; Bergamaschi et al. 2021), a consequence of a suburbanization process, that is, the progressive displacement of the foreign population from the center to the more peripheral areas of cities (Avallone and Torre, 2016) and in suburban municipalities around major cities and metropolitan areas (Borruso and Murgante, 2013). Although migration flows develop without planning and control, settlement regularities are observed, knowledge of which is crucial for the implementation of effective policies at the local level. Regularities that seem to depend mainly on the fact that specific patterns of residential settlement related to each ethnic group emerge, often shaped by vocational occupation (Costarelli and Mugnano, 2017). Specifically, three prevalent settlement patterns stand out: a metropolitan pattern, attributable to communities with a strong imbalance in gender structure, employed mostly in family services or commercial activities, such as the Filipino community, which has a substantial presence in the Milanese context; a diffuse pattern in the face of a greater range of employment opportunities, as in the case of three of the most widespread ethnic groups: Romanians, Albanians and Moroccans; and a border pattern, of communities coming from countries bordering Italy (Istat, 2022e). The aim is to identify the spatial pattern of the presence of foreigners in the Northwest, one of the Italian areas that most attracts migration flows. With this in mind, in the next section we introduce spatial autocorrelation data and techniques, while in the following paragraphs we focus on analyzing the share of foreigners at the municipal level both to observe current spatial concentrations and to outline the evolution of spatial clusters in recent decades. Concluding remarks close the paper.

## 2. Data and methods

The Istat census data dissemination system was used both for the latest available data (Istat, 2022a, 2022b, 2022c, 2022d) and for the years 2001 and 2011 (Istat, 2015). For the construction of the indicators, a spatial reconstruction was necessary, which in the first instance involved new municipalities established through mergers<sup>1</sup>. Foreign population shares were used to study the

---

<sup>1</sup> Val di Chy, Valchiusa, Alto Sermenza, Cellio con Breia, Gattico-Veruno, Cassano Spinola, Alluvioni Piovera, Lu e Cuccaro Monferrato, Montalto Carpiasio, Maccagno con Pino e Veddasca, Cadrezzate con Osmate, Bellagio, Colverde, Tremezzina, Alta Valle Intelvi, Centro Valle Intelvi, Solbiate con Cagno, Vermezzo con Zelo, Torre de' Busi, Sant'Omobono Terme, Val Brembilla, Cornale e Bastida, Corteolona e Genzone, Colli Verdi, Piadena Drizzona, Borgo Virgilio, Borgo Mantovano, Borgocarbonara, Lessona, Campiglia Cervo, Quaregna Cerreto, Valdilana, Verderio, La Valletta Brianza, Valvarrone, Castelgerundo, Borgomezzavalle, Valle Cannobina.

phenomenon. Measurement of the dynamics at ten-year intervals was made possible by using the differences, in terms of percentage points, of the raw shares of foreigners in municipal territories.

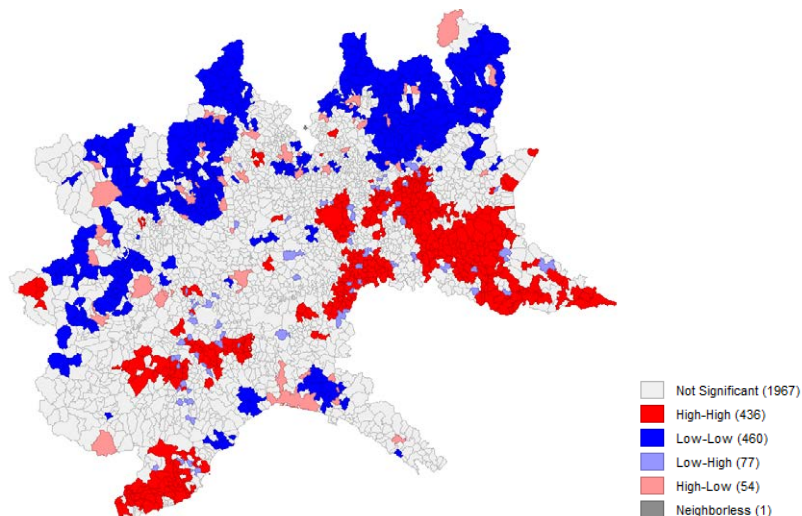
The identification of local clusters is crucial for the study and understanding of the spatial variability of the share of foreigners. A fundamental part of the clustering process is the measurement of spatial auto-correlation between the units studied, i.e. the degree to which the values of a variable are clustered or dispersed in space. Here we use LISA - Local Indicator of Spatial Association (Anselin, 1995) which is, in discursive terms, a measure of the similarity between the value of a variable measured in an areal unit of analysis (e.g. municipality) and the values of the same variable in neighbouring units, as defined by a spatial weighting matrix. A LISA value can be calculated for each spatial unit of analysis. Since the population varies across the areas under consideration, the precision of each share will also vary. For areas with small populations, the value of the share will be less reliable and, vice versa, the larger the population, the greater the reliability. Therefore, to avoid the risk of a false representation of the spatial distribution of the underlying phenomenon, the share of foreign population were corrected for this inherent instability by using Empirical Bayes (EB) technique (Anselin, Lozano-Gracia, and Koschinky, 2006): a kind of smoothing approaches which improve on the precision of the raw rate by *borrowing strength* from the other observations. The EB technique consists of computing a weighted average between the raw share of foreign population for each municipality and the Northwest regional average, with weights proportional to the resident population in a municipality. Simply put, municipalities with a small population will tend to have their shares adjusted substantially, whereas for larger municipalities the share will hardly change. In the end, given the EB share values and the generic spatial weighting matrix element, for each municipality, a positive value of LISA indicates a high value surrounded by high values (high-high) or a low value surrounded by low values (low-low), while a negative value indicates a high value surrounded by low values (high-low) or a low value surrounded by high values (low-high).

The LISA Cluster Map is the most intuitive way to graphically represent the information provided by LISA values and to visualise local clusters and local spatial outliers. The Cluster Map is, in fact, a thematic map showing only those municipalities with statistically significant LISA values, classified according to five categories: i. Not Significant (areas that are not significant at the 0.05 level); ii. High-High (High indicator value and neighbouring municipalities with high indicator values); iii. Low-Low (Low indicator value and neighbouring municipalities with low indicator values); iv. Low-High (Low indicator value and neighbouring municipalities with high indicator values); v. High-Low (High indicator value and neighbouring municipalities with low indicator values). In addition, we will apply a LISA-based analysis technique called LISA Cluster Transitions Analysis, which studies the dynamics of the spatial distribution of the share of foreigners in Lombardy municipalities, grouping municipalities according to their changes or transitions of LISA values from one period to another (Anselin, 2018; Martin et al., 2016, Brooks, 2019). Simplifying, LISA Cluster Transitions Analysis, consists of classifying the different types of transitions present in a transition matrix between two states. For example, a municipality, which was High-High in both 1990 (high value surrounded by high values in the same period) and 2011, has the value 11 (alternately HH, HH); a municipality, which is Low-Low in both periods, is 22 (LL, LL); and a municipality, which has transitioned from Not Significant to High-High, is 01 (NS, HH). In this paper considering 2001 and 2011, twenty-five transitions between the LISA categories are possible, most of them with little substantive significance; as the literature suggests (Martin et al., 2016; Brooks, 2019), the focus must be on the ability of the transitions to show where the share is persistent over time and where it is changing, therefore the following transitions will be analysed: High-high in both periods; Low-low in both periods; from Non-significant to High-High; from Non-significant to Low-low; from High-High to Non-significant; Low-low to Non-significant. A thematic map of municipalities will be used in the presentation of results, associating different colors with different types of transitions.

### 3. Foreign presence in northwestern Italy

In 2020, the foreign population in northwestern Italy amounted to 1,766,425 residents: in Lombardy 1,190,889 people (with an average regional share of 11.9%), in Piedmont 417,279 people (9.8%), in Liguria 149,862 people (9.9%) and in Aosta Valley 8,395 people (6.8%). The picture of the resident foreign presence in the Northwest shows rather different concentrations. Starting with Liguria, it is evident how, at the end of the observed 20-year period, there is a high concentration of foreigners in the province of Imperia, the western area of the province of Savona, the regional capital. Although in Genoa the share does not exceed 10%, higher concentrations are observed in the province. In the eastern part of the region, only the provincial capital of La Spezia has a high share of foreign population (12.7%). As far as Piedmont is concerned, a greater concentration of foreign population can be observed in the provinces of Cuneo, Alessandria and Asti, in some localities in the eastern part of the province of Turin and in the capital itself (14.1%), as well as in Novara. Conversely, apart from a few exceptions, the foreign presence is more contained in the province of Verbano-Cusio-Ossola, in the upper Vercellese, in the Biella area and in municipalities along the western borders of the provinces of Cuneo and Turin, where the most notable singularity is the territory between Bardonecchia, Salza di Pinerolo and Claviere. In Aosta Valley, only three municipalities have a foreign population share above 10 percent (Challand-Saint-Anselme, Valtournenche and Verrès). In Lombardy, foreigners tend to be concentrated in the Milan area (18.2% in the capital municipality), in the southern area of the region - that is, in the provinces of Pavia, Lodi, Cremona and Mantua - and in the southernmost areas of the provinces of Bergamo and Brescia. A conspicuous presence is also recorded in Como and Lecco (in the two provincial capitals, the share is 14.4% and 10.7%, respectively). In contrast, the phenomenon appears less widespread in the northernmost parts of Lombardy, particularly in the province of Sondrio and northern areas of the provinces of Brescia and Bergamo, i.e. in the Alpine areas. The map of LISA clusters, highlights local clusters with significant information for both areas with the highest concentration and areas where the phenomenon is of low intensity (Figure 1).

Figure 1 - LISA representation of the EB share of foreigners, Northwest, 2020



### 4. Ethnic differences in migration movements in northwestern Italy

According to EESC (EESC, 2018), the absence of migrants in European countries would have negative consequences, especially in relation to population aging. "Population to grow in some MS,



*to shrink in others... but to age in all*" reads the presentation accompanying the European Commission's Ageing Report 2021 (EC, 2021). Economically and socially, in the countries of Southern Europe, migrants contribute to the functioning of health care systems and assistance in personal services. Also in the Northwest, immigration contributes to the labour force in agriculture and construction, helps counter depopulation in some territories and plays a positive role in the balance of pension systems. At the same time, as migratory pressure increases, so does the need to invest locally in integration, to avoid conflicts between host communities and migrant due to socio-cultural differences, including through implementation of policies aimed at countering risks related to the spread of undeclared work, territorial segregation, and discrimination. The migration flows of the past two decades have resulted in differentiated ethnic concentrations. Taking two provinces in the northwestern perimeter (Mantua and Imperia) as examples, significant differences emerge. In Mantua, a province still highly specialised in industrial production, the Asian component is notable (37.6%) with large shares of Indians (17.3%), Chinese (8.8%) Bangladeshi (4.1%) and Pakistani (4.0%). A completely different story is observed, however, in the province of Imperia, which as a strong tourist vocation, where foreigners are predominantly European. In addition to Romanians and Albanians, whose diffusion in fact covers the entire national territory, French and Germans, although they have seen their relative weight decrease (at the beginning of the century they represented a fifth of the foreign population overall) continue to have a significant presence.

## 5. The evolution of immigration in the last two decades in northwestern Italy

The evolution of the migration phenomenon over the past two decades in the Northwest is characterised by two phases that differ both quantitatively and qualitatively. In the first decade (2001-11), the population of foreigners increased significantly (about 1,000,000 more), tripling the total amount. The largest relative increase in the presence of foreigners is concentrated in southern Piedmont and south-central Lombardy. The areas of greatest attraction are the large urban centers (e.g. Milan) and more traditional industrialised areas and industrial districts characterised by the presence of small and medium-sized enterprises. In the second decade (2011-20), the increase is significantly smaller and is around 24% (in absolute terms it increases by 340 thousand foreign residents). At this stage, the expansion of immigration is spread over almost the entire Northwest, with a particularly large area of expansion in the Milanese hinterland, an outcome, in line with the literature on the suburbanization process. On the other hand, there is a marked slowdown in terms of the change in share in the eastern area of Lombardy (Brescia, Bergamo and Cremona).

Figure 2 - Dynamics of foreigners, Northwest, 2001-11 (change in percentage points of raw share)

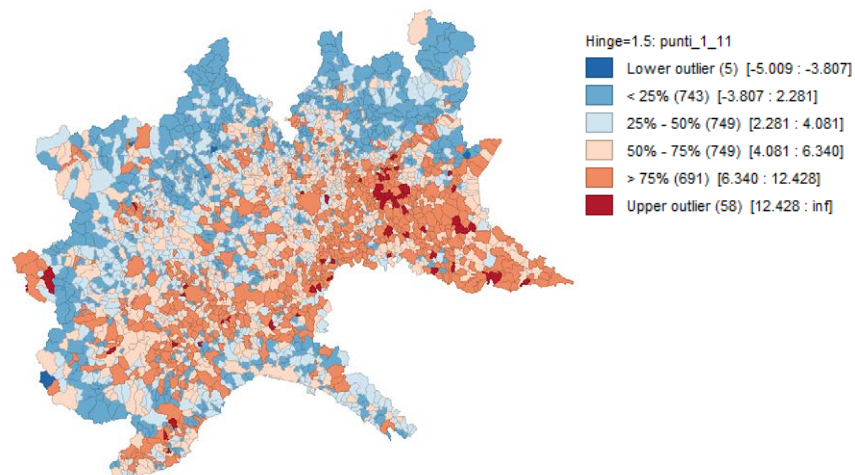
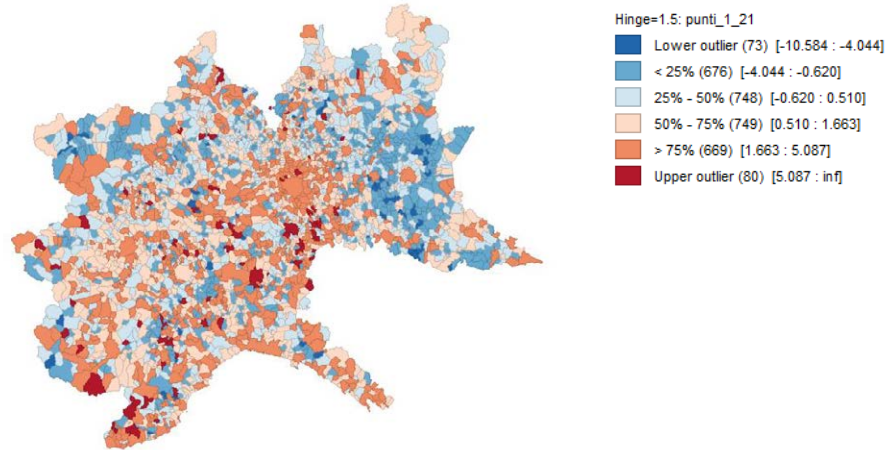


Figure 3 - Dynamics of foreigners, Northwest, 2011-20 (change in percentage points of raw share)

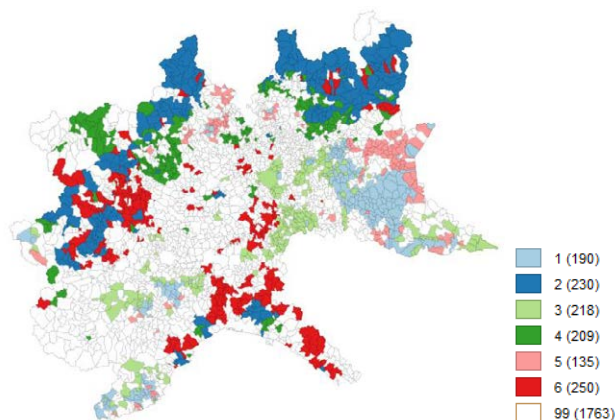


## 6. The spatial dynamics of clusters

Between 2001 and 2020 we can observe that clusters tend to strengthen and expand while at the same time, new places of concentration of foreigners appear (Figure 4). In particular there are:

- Areas of persistence with high share (Figure 4, 1. light blue). Extensive clusters in the border area between the provinces of Bergamo, Brescia, Cremona and Mantua; less extensive clusters in the Imperia province; clusters of interest also in the area between Asti and Cuneo;
- Expanding areas: i. non-significant to high share (Figure 4, 3. light green). This includes areas of new clusters (in the Milan province where the high share was initially limited to the provincial capital; municipalities in the Pavia and Lodi areas) and areas of expansion of already existing high share clusters (in the Mantua and Bergamo areas, which is an extension of the Brescia-Bergamo cluster; the area on the border between Imperia and Savona; the Asti-Alessandria area); ii. low share to not significant (Figure 4, 6. red). These are areas where the share increases and approaches the average value: discontinuous areas in the mountains of both Turin and Valtellina; Pavia area; several clusters in Liguria;
- Areas of persistence with low share (Figure 4, 2. blue). Cluster in the Genoa province; Alpine arc: mountain areas in Lombardy, Piedmont (especially Biella and Verbania) and Aosta Valley;

Figure 4 - Foreign location change, Northwest, 2001-20 (types based on LISA EB share of foreigners)



- Shrinking areas: i. non-significant to low share (Figure 4, 4. green): extensive cluster in Aosta Valley and in the mountainous border area between the provinces of Biella and Vercelli); ii. From high share to non-significant (Figure 4, 5. light pink): extended cluster in the province Brescia - North Mantua; cluster in the Varese area.

## 7. Concluding remarks

The study examined immigration in municipalities in northwestern Italy. Data from official statistics were considered in the analysis. In particular, foreign population shares based on ISTAT data for the past decades were used. The results, determined by the complementarity of different methods of spatial analysis, made it possible to identify clusters of municipalities and to understand both differences and migration dynamics. Areas of persistence of high share of foreigners, areas of expansion and areas of contractions emerged. The proposed analysis can be considered a useful reference for public policy development at the local level.

## References

- Anselin L. (1995). Local Indicators of Spatial Association – *LISA*, *Geographical Analysis*, **27**, pp. 93-115.
- Anselin, L. (2018). Global spatial autocorrelation. *GeoDa website*, accessed in March 2019. [https://geodacenter.github.io/workbook/5a\\_global\\_auto/lab5a.html](https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html).
- Anselin, L., Lozano-Gracia, N., and Koschinsky, J. (2006). Rate Transformations and Smoothing in *Technical Report. Urbana, IL: Spatial Analysis Laboratory, Department of Geography, University of Illinois*.
- Avallone, G., Torre, S. (2016). Dalla città ostile alla città bene comune. I migranti di fronte alla crisi dell’abitare in Italia. *Archivio di Studi Urbani e Regionali*, **115**, pp. 51-74.
- Bergamaschi, M., Daconto, L., Montesano, M.G. (2021). Diffusione periferica e concentrazioni. La dinamica insediativa dei residenti stranieri a Bologna e Milano. *Mondi Migranti*, **2**, pp 151-168.
- Borruso, G., Murgante, B. (2013). Analisi dei fenomeni immigratori e tecniche di autocorrelazione spaziale. *Territorio Italia*, **1**, pp. 27-38.
- Brooks, M.M. (2019). The Advantages of Comparative LISA Techniques in Spatial Inequality Research: Evidence from Poverty Change in the United States, *Spatial Demography*, **7**, pp. 167–193.
- Costarelli, I., Mugnano, S., (2017). Un’analisi della condizione abitativa degli stranieri a Roma e Milano: micro-segregazione e periferizzazione. *Fuori luogo. Rivista Sociologia del Territorio, Turismo, Tecnologia*, **2**, pp. 39-47.
- EESC (2018). Without migrants, Europe's economic and social model is in danger, says the EESC. *European Economic and Social Committee website*, accessed in November 2022. <https://www.eesc.europa.eu/en/news-media/press-releases/without-migrants-europes-economic-and-social-model-danger-says-eesc>.
- EC (2021). 2021 Ageing Report: Ageing populations and fiscal sustainability. Presentation in European Commission (2021). *European Commission website*, accessed in August 2022. [https://economy-finance.ec.europa.eu/system/files/2022-05/2021\\_ageing\\_report\\_0.pdf](https://economy-finance.ec.europa.eu/system/files/2022-05/2021_ageing_report_0.pdf).
- Istat (2015). La ricostruzione dei valori censuari dei comuni con variazioni territoriali nel periodo 1991-2011. *Istat website*, accessed in June 2022. <https://ottomilacensus.istat.it/>.
- Istat (2022a). *Il Censimento permanente della popolazione in Liguria*. Anno 2020. Istat.
- Istat (2022b). *Il Censimento permanente della popolazione in Lombardia*. Anno 2020. Istat.
- Istat (2022c). *Il Censimento permanente della popolazione in Piemonte*. Anno 2020. Istat.
- Istat (2022d). *Il Censimento permanente della popolazione in Valle d’Aosta*. Anno 2020. Istat.
- Istat (2022e). *Rapporto annuale 2022*. Istat, pp. 186-190.
- Martin, M. J. R., Matthews, S. A., Lee, B. A. (2016). The spatial diffusion of racial and ethnic diversity across U.S. counties. *Spatial Demography*, **5**(3), pp. 1–25.
- Perna, R. (2015). L’immigrazione in Italia. Dinamiche e trasformazioni in tempo di crisi. *Politiche sociali*, **1**, pp. 89-116.

# Are Italian youngsters adequately equipped for an after-pandemic upswing?

Luigi Bollani, Simone Di Zio, Luigi Fabbris

## 1. Introduction

Since the onset of the severe acute respiratory syndrome (SARS) epidemic caused by coronavirus, many studies (e.g. Xiang et al., 2014; Mental Health Commission of Canada, 2020) demonstrate that such a large scale and long-lasting infectious disease—being a traumatic social shock—has a grave impact on public mental health, causing strong negative emotions and psychological and mental disorders, such as depression and anxiety.

In this research, we analyse the results of a survey on the effects of the COVID-19 pandemic on Italian youth, focusing on the disrupting effects of the coronavirus outbreak on people's perception of their possible future. We hypothesise that as the pandemic is approaching its end, an imminent radical change in lifestyles can occur that would not only recover the pre-pandemic normality but also frame social behaviours in a more sustainable way than before. For youth, the future contains a projection of the strategic roles one is prepared to play in the natural process of replacing the older generation.

This paper describes how Italian youth experienced the COVID-19 pandemic and how they tend to face their future. The research questions are as follows:

H1: Did COVID-19 infection cause youth depression and, consequently, affect their perception of the future?

H2: Is youth depression blurring their vision of their future social role?

H3: Do proactivity and self-efficacy counterbalance depressive symptoms and malaise, creating a positive vision of the future among youth?

H4: Which characteristics make youth more prone to having a blurred vision in the after-pandemic upswing?

The rest of the paper is organised as follows: Section 2 describes the researched sample and introduces the model and the methodological aspects for the data analysis; Section 3 presents the main results of the statistical analysis of the collected data; and finally, Section 4 interprets the data with reference to the mainstream literature and concludes the work.

## 2. Data and methods

### 2.1. The data

A sample of Italian adults was surveyed from June to November 2021 using the computer-assisted web-based interviewing (CAWI) technique. A total of 817 respondents collaborated with the survey, filling in an electronic questionnaire, of which 428 respondents, aged between 18 and 34, were chosen as the sample for analysis. The sample is moderately unbalanced toward central and northern Italy, being 74% of respondents against 64% of Italians aged 18-34.

Below is a set of descriptors of youth mental states and their possible predictors. The variables used in the relational model are as follows:

**Y:** The respondent has a clear vision of what they will do after the pandemic. Although this

Luigi Bollani, University of Turin, Italy, [luigi.bollani@unito.it](mailto:luigi.bollani@unito.it), 0000-0002-2488-3659

Simone Di Zio, University of Chieti-Pescara G. D'Annunzio, Italy, [s.dizio@unich.it](mailto:s.dizio@unich.it), 0000-0002-9139-1451

Luigi Fabbris, Tolomeo studi e ricerche, Padua, Italy, [fabbris@stat.unipd.it](mailto:fabbris@stat.unipd.it), 0000-0001-8657-8361

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Luigi Bollani, Simone Di Zio, Luigi Fabbris, *Are Italian youngsters adequately equipped for an after-pandemic upswing*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.07, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 35-40, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

question was posed in a dichotomous way, it was the last in the series of questions on the attitudes to the pandemic experience, so the responses to the question can be considered informed and will be referenced to study the youth's ability to determine their future.

$X_1$ : Full-blown depression (dichotomous; computed using the nine-item PHQ – Patient Health Questionnaire proposed by Spitzer et al. (1999) and translated into Italian by Mazzotti et al. (2003); the value of cumulative responses  $\geq 10$  identifies the diagnosis of major depression).

$X_2$ : Passive attitude (dichotomous; obtained by a factor analysis of a set of eight items related to pessimism-proactivity, keeping the standardised scores below -0.25; the 8 items were selected from the 20 items proposed by Beck et al. (1974) to construct the Beck Hopelessness Scale).

$X_3$ : Proactive attitude (dichotomous; obtained by a factor analysis of the set of the eight items related to pessimism-proactivity mentioned above, keeping the standardised scores above 0.40).

$X_4$ : Self-efficacy score (continuous; obtained by a factor analysis of a set of nine items related to individual self-effectiveness and resilience: the items were selected from the 25-item Connor and Davidson [2003] resilience scale and translated in Italian by authors).

$X_5 \div X_{24}$ : see the description in Table 1.

## 2.2 The analytical model

The model for the data analysis includes having clear ideas about the after-pandemic future as a dependent variable  $Y$  and two sets of possible regressors, from  $X_1$  to  $X_4$  and from  $X_5$  to  $X_{24}$ . This relationship can be expressed as:

$$Y=f(X_1 \div X_4, X_5 \div X_{24}).$$

After a bivariate or trivariate correlation analysis between  $Y$  and the first set of possible regressors, a multiple logistic regression model was fitted. The logistic regression can be expressed as follows (Hosmer and Lemeshow, 2000):

$$\text{logit} [p(Y = 1)] = \beta_0 + \beta_1 X_1 + \dots + \beta_{24} X_{24},$$

where  $\text{logit}(p) = \ln[p/(1-p)]$ , and  $\beta_i$  measures the relation between  $Y$  and  $X_i$  when all other variables in the model remain fixed. A regressor enters the model only if it is statistically significant.

The statistical analyses were carried out in the  $R$  environment (R Core Team, 2022); a logistic regression model for a binary response variable was performed with the  $glm$  function from the MASS package. Moreover, the  $stepAIC$  function was utilised to perform stepwise model selection with the  $AIC$  criterion.

## 3. Results

As shown in Table 1, 62.6% of the Italian youth have a clear view of their after-pandemic role, whereas 37.4% are unable to imagine how their future life could develop, which can stem from the individual pandemic experience, mental health status and character traits.

The diffusion of mental health problems, measured with a depression diagnosis, concerns almost one out of two young people: an estimate of 44.4% depressed signifies that youngsters are a population layer that has suffered mentally due to the pandemic more than others. This depression rate is significantly higher than that of older adults (33.7%).

Bearing in mind that young Italians undergoing a therapy for a mental disease comprised 4.4% before the survey, that mental illness is difficult to assess in young people and the quota remains concealed, and that our data were collected when the health crisis was still ongoing, it can be stated that the pandemic has caused a flood of psychic disturbances, including eating disorders (34.1%), sadness/desperation (31.8%) and self-harming disposition (6.1%).

It can also be noted that the number of young individuals consuming wine or beer at meals and consuming spirits has increased by 9.1% and 9.3%, respectively (minor consumption is 19.4% and 24.1%, respectively), compared to the numbers before the pandemic. Thus, social isolation has not curtailed drinking habits.

**Table 1.** Mean of the variables used in the statistical analysis of the youth in Italy, 2021

<i>Variable</i>	<i>mean</i>	<i>Variable</i>	<i>mean</i>
Y: Clear vision of the future	0.626	X <sub>12</sub> : Suffered psychologic damages	0.451
X <sub>1</sub> : Full-blown depression	0.444	X <sub>13</sub> : Suffered physical damages	0.154
X <sub>2</sub> : Passive attitude	0.416	X <sub>14</sub> : Had controls through swabs	0.666
X <sub>3</sub> : Proactive attitude	0.248	X <sub>15</sub> : Had a psychic disease	0.044
X <sub>4</sub> : Self-efficacy score	0.000	X <sub>16</sub> : Male (gender)	0.357
X <sub>5</sub> : Trusted scientists	0.787	X <sub>17</sub> : Age 25-34 (vs. 18-24)	0.236
X <sub>6</sub> : Family doctor available during pandemic	0.402	X <sub>18</sub> : Higher education degree	0.348
X <sub>7</sub> : Hospitals were a source of contagion	0.103	X <sub>19</sub> : Worker (vs. else)	0.154
X <sub>8</sub> : Vaccinated: Yes	0.638	X <sub>20</sub> : Living in a couple family	0.320
: Not yet	0.271	X <sub>21</sub> : Remote learning/working	0.930
: Never	0.091	X <sub>22</sub> : Eating disorders	0.341
X <sub>9</sub> : Doubting about vaccine efficacy	0.098	X <sub>23</sub> : Sadness, desperation	0.318
X <sub>10</sub> : Infection: personal	0.143	X <sub>24</sub> : Self-harming	0.061
X <sub>11</sub> : Infection: parents	0.222		

The analysis of correlations (Table 2) reveals that depression causes the difficulty in perceiving one's role in the future ( $r = -0.340$ ,  $p < 0.001$ ) and passive attitudes ( $r = -0.357$ ,  $p < 0.001$ ) and that individual clarity about the future correlates with one's proactivity in facing life problems ( $r = 0.354$ ,  $p < 0.001$ ) and self-efficacy attitude ( $r = 0.303$ ,  $p < 0.001$ ).

**Table 2.** Correlation coefficients between the variables used in the statistical analysis of the youth in Italy, 2021 (used to test H1, H2 and H3)

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>
Y: future clearness	-0.340	-0.357	0.354	0.303	0.080	0.052	-0.251	-0.058
X <sub>1</sub> : depression		0.257	-0.197	-0.355	0.012	-0.013	0.334	0.165
X <sub>2</sub> : passivity			-0.484	-0.278	-0.032	-0.006	0.112	0.007
X <sub>3</sub> : proactivity				0.268	0.014	0.032	-0.107	-0.035
X <sub>4</sub> : self-efficacy					0.075	0.097	-0.116	0.043

It should be highlighted that no viral infection (either in respondents or parents) is statistically correlated with any of the psychological variables  $Y$  and  $X_1$  through  $X_4$  (columns  $X_{10}$  and  $X_{11}$  in Table 2). In a similar vein, the physical consequences of the disease co-vary with depression ( $r = 0.165$ ;  $p = 0.001$ ), although without any other psychological status. Instead, the psychological consequences of the pandemic highly correlate with both the difficulty in forming one's own outlook ( $r = 0.251$ ,  $p = 0.001$ ) and one's depressive status ( $r = 0.334$ ,  $p = 0.001$ ). The youth are at risk of psychological distress not because of the contagion itself but because of the contextual conditions of the pandemic. Instead, reduced physical contact with peers, the manner in which incumbent health risks were communicated as well as procrastinated closing of the emergency are likely to be at the root of such a diffused malaise.

The analysis (Table 3) proves that youth perceive their future more clearly if the pessimistic views due to the pandemic are limited and if they possess proactive and other positive attitudes. The only physical variable that improves the individual outlook on the future was involvement in distance learning and remote working that the youth practiced during lockdowns and occasionally after that. As a whole, 93% of youth were involved in activities from remote. We can conjecture that keeping youth busy and favouring their participation in the management of

the pandemic could have enhanced their disposition to the future instead of fostering inactivity and removal of responsibilities, which have opened the Pandora's box of youth mental problems.

Another relevant result is the absence of gender's role as a predictor, although being a female correlated with both difficulties in the outlook on the future and depression, which means that the variables in the model explain the gender differential.

**Table 3.** Beta estimates of the regression model with clear vision of future as the criterion variable (forward stepwise selection of regressors, n=428; Nagelkerke pseudo-R<sup>2</sup>=38.1%; AIC criterion=441.6)

<i>Regressor</i>	$\hat{\beta}$	<i>se</i> ( $\hat{\beta}$ )	<i>Signific.</i>
Intercept	0.469	0.459	NS
X <sub>3</sub> : Proactive attitude	1.766	0.429	***
X <sub>1</sub> : Full-blown depression	-0.801	0.255	**
X <sub>2</sub> : Passive attitude	-0.713	0.255	**
X <sub>12</sub> : Suffered psychological damages	-0.783	0.247	**
X <sub>4</sub> : Self-efficacy score	0.200	0.067	**
X <sub>15</sub> : Had a psychic disease	-1.518	0.685	*
X <sub>21</sub> : Remote learning/working	1.034	0.451	*

\*\*\* < 0.001; \*\* < 0.01; \* < 0.05; NS= Not significant

#### 4. Discussion and conclusion

This work aimed to highlight that the worries that the pandemic caused among Italian youth can threaten their future. The research reveals that 45% of young Italians felt depressed and 38% were unable to imagine their future after the pandemic. This worrying outcome recurs in many studies (e.g. Ettman et al., 2020; Eurofound, 2021; Renaud-Charest et al., 2021).

In general, young people have been lightly affected by the disease, showing a lower risk of contagion and even lower consequences. As Commodari and La Rosa (2020) suggest, young individuals perceived the disease as less damaging. Moreover, the threat of susceptibility to and the severity of a potential infection with the virus has notably decreased during the pandemic, particularly following the discovery of the vaccine (Rupprecht et al., 2022).

Imposed confinement did not increase anxiety-depression symptoms; in fact, these symptoms decreased during lockdowns (Muzi et al., 2021). Youngsters used various means of communication to stay connected with their schoolmates and friends at any time of the pandemic, more or less in the same manner as they used to do before. They did not suffer from a lack of communication; on the contrary, it was the frequent use of social media as a potential source of health news regarding COVID-19 that may have caused psychological disorders, further disposing youngsters to panic, distress and anxious-depressive symptoms (Higuchi et al., 2020). Moreover, the endless prolongation of the emergency, complete change of all structured occupations (school, work and training) and economic and occupational concerns may have contributed to the overall anticipation of an insecure and worrying future, causing psychological distress and depressive ailment, worsening pre-existing vulnerabilities and repressing proactive attitudes (Power et al., 2020; Steele, 2020; Esposito et al., 2021; Muzi et al., 2021; Rania and Coppola, 2022; Chadi et al., 2022; Rupprecht et al., 2022) so much so that the mental problems left on the ground by the pandemic have far surpassed the less frequent and harming effects of the virus contagion (Shuster et al., 2021).

Individuals who have experienced such a traumatic event not only have difficulties in finding their own strategy to cope with the trauma and its sequelae but are also conditioned to trying to define a strategy for their future (Liang et al., 2020). The changes brought about by

the pandemic have been so pervasive and increased people's insecurity so much that it has become a common assumption that the changes do not end with the health emergency—rather, pessimism has become a generalised feeling (Barrafrem et al., 2020).

Youth, by nature, develop imagining their future day by day and looking for the means to construct it. The power of choosing, changing, creating and even fighting to impose their will is intrinsic to youth development. Therefore, if the future is perceived as too worrying or insecure, youth can lose the sense of time continuity, which can transform their lives into a series of empty times. The pandemic outbreak diffused apathy and pessimism, slowed down social growth and instilled discontent and depression in youth minds. Moreover, the perception of insurmountable and prolonged social and economic difficulties caused by the possible lack of resources added pressure in youth minds. Even during the decade before the pandemic, young individuals showed high levels of mental disorder with a feeling of helplessness, depression and thoughts of suicide. The pandemic exacerbated the situation for many and helped just those who could spend more time within their families. Scholars argue that young individuals have the necessity to take an active part in societal activities in order to gain confidence for the future. Hence, resuming offline school activities as much as possible could have helped students because schools, being inclusive and safe, provide them with opportunities to engage with their communities and be mentored by supportive adults.

We are not able to forecast how long this dramatic situation could last, in particular for marginalised groups. Economists conjecture that social booms and busts are temporary phenomena. Though, studies (Power et al., 2020) show that the effects of social shocks persist for long, in particular for those who enter the job market during a recession. And, even worst, for those who are not able to enter the job market. What is going to happen to the youth who are going to enter their productive life after such a lengthy pandemic?

While it is not possible to forecast how long this deplorable situation will last, especially for marginalised groups, studies (e.g. Power et al., 2020) show that the effects of social shocks persist for long, in particular for those who enter the job market during a recession, and even more so for those who are not able to enter the job market. This can be especially detrimental for the future of the youth who are yet to enter their productive lives after a lengthy pandemic.

## References

- Barrafrem, K., Västfjäll, D., Tinghög, G. (2020). Financial well-being, COVID-19, and the financial better-than-average-effect. *Journal of Behavioral and Experimental Finance*, 28: 10041.
- Beck, A. T., Weissman, A., Lester, D., Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, 42(6): 861–865.
- Chadi, N., Castellanos Ryan, N., Geofroy, M.-C. (2022). COVID-19 and the impacts on youth mental health: Emerging evidence from longitudinal studies. *Canadian Journal of Public Health*, 113: 44–52.
- Commodari, E., La Rosa, V.L. (2020). Adolescents in quarantine during COVID-19 pandemic in Italy: Perceived health risk, beliefs, psychological experiences and expectations for the future. *Frontiers in Psychology*, 11: 10.3389/fpsyg.2020.559951.
- Connor, K.M. Davidson, R.T. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC), *Depression and Anxiety*, 18: 76–82.
- Esposito, S., Giannitto, N., Squarcia, A., Neglia, C., Argentiero, A., Minichetti, P., Cotugno, N., Principi, N. (2021). Development of psychological problems among adolescents during school closures because of the COVID-19 lockdown phase in Italy: A cross-sectional survey. *Frontiers in Pediatrics*, 8: 628072.
- Ettman, C.K., Abdalla, S.M., Cohen, G.H., Sampson, L., Vivier, P.M., Galea, S. (2020). Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. *Journal of the American Medical Association Network Open*, 3(9): e2019686.



- doi:10.1001/jamanetworkopen.2020.19686.
- Eurofound (2021). *Living, Working and COVID-19 (Update April 2021): Mental Health and Trust Decline across EU as Pandemic Enters Another Year*. Luxembourg: Publications Office of the European Union.
- Higuchi, S., Mihara, S., Kitayuguchi, T., Miyakoshi, H., Ooi, M., Maezono, M., Nishimura, K., Matsuzaki, T. (2020). Prolonged use of Internet and gaming among treatment seekers arising out of social restrictions related to COVID-19 pandemic. *Psychiatry and Clinical Neurosciences*, 74: 602-631.
- Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression (2nd ed.)*. New York: John Wiley & Sons.
- Liang, L., Ren, H., Cao, R., Hu, Y., Qin, Z., Li, C., Me, S. (2020). The effect of COVID-19 on youth mental health. *Psychiatric Quarterly*, 91: 841–852.
- Mazzotti, E., Fassone, G., Picardi, A., Sagoni, E., Ramieri, L., Lega, I., Camaioni, D., Abeni, D., Pasquini, P. (2003). Il Patient Health Questionnaire (PHQ) per lo screening dei disturbi psichiatrici: Uno studio di validazione nei confronti della Intervista Clinica Strutturata per il DSM-IV asse I (SCID-I). *Italian Journal of Psychopathology*, 9(3): 235-242.
- Mental Health Commission of Canada (2020). *LOCKDOWN LIFE: Mental Health Impacts of COVID-19 on Youth in Canada*. Ottawa, Canada.
- Muzi, S., Sansò, A., Pace, C.S. (2021). What's happened to Italian adolescents during the COVID-19 pandemic? A preliminary study on symptoms, problematic social media usage, and attachment: Relationships and differences with pre-pandemic peers. *Frontiers of Psychiatry*, 12: 590543. doi: 10.3389/fpsy.2021.590543.
- Power, E., Hughes, S., Cotter, D., Cannon, M. (2020). Youth mental health in the time of COVID-19. *Irish Journal of Psychological Medicine*, 37: 301–305.
- Rania, N. Coppola, I. (2022). The fear of contagion and the attitude toward the restrictive measures imposed to face COVID-19 in Italy: The psychological consequences caused by the pandemic one year after it began. *Frontiers of Psychology*, 13: 805706. doi: 10.3389/fpsyg.2022.805706.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, AT (<http://www.R-project.org>, accessed 26 June 2022).
- Renaud-Charest, O., Lui, L.M.W., Eskander, S., Ceban, F., Ho, R., Di Vincenzo, J.D., Rosenblat, J.D., Lee, Y., Subramaniapillai, M., McIntyre, R.S. (2021). Onset and frequency of depression in post-COVID-19 syndrome: A systematic review. *Journal of Psychiatric Research*, 144: 129-137.
- Rupprecht, F.S., Martin, K., Kamin, S.T., Lang, F.R. (2022). COVID-19 and perceiving finitude: Associations with future time perspective, death anxiety, and ideal life expectancy. *Psychology and Aging*, 37(2), 260–271.
- Shuster, A., O'Brien, M., Luo, Y, Berner, L.A., Perl, O., Heflin, M., Kulkarni, K., Chung, D., Na, S., Fiore, V.G., Gu, X. (2021). Emotional adaptation during a crisis: Decline in anxiety and depression after the initial weeks of COVID-19 in the United States. *Translational Psychiatry*, 11: 435; <https://doi.org/10.1038/s41398-021-01552-y>.
- Spitzer, R.L., Kroenke, K., Williams, J.B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ Primary Care study. *Journal of the American Medical Association*, 282(18): 1737-1744.
- Steele, H., (2020). COVID-19, fear and the future: An attachment perspective. *Clinical Neuropsychiatry*, 17(2): 97-99.
- Xiang, Y.T., Yu, X., Ungvari, G.S., Correl, C.U., Chiu, H.F. (2014). Outcomes of SARS survivors in China: not only physical and psychiatric co-morbidities. *East Asian Archives of Psychiatry*, 24: 37–38.

# How Italians coped with COVID-19 lockdown: evidence from a survey promoted through social networks

Margherita Silan, Riccardo Bellide

## 1. Introduction

In December 2019, a coronavirus, SARS-COV2 (severe acute respiratory syndrome coronavirus), responsible for a respiratory syndrome with severe complications, appeared in Wuhan, China. The worldwide spread of the virus was rapid and on 11<sup>th</sup> March 2020, the World Health Organization (WHO) declared a global *pandemic* status. The lack of scientific information, effective drugs, the absence of vaccines, and the state of panic caused by the contagiousness of the coronavirus, as well as the awareness of being unprepared in the face of a totally unforeseen situation, brought the world to adopt habits and impose restrictions unthinkable until then. The first measures consisted of the adoption of non-pharmaceutical practices such as the use of masks, disinfectants, social distancing, and travel bans.

During the first wave of the pandemic, several studies were implemented to investigate the social economic and health-related consequences of the COVID-19 pandemic. Among them, an international project, the SEBCOV study, was born in five countries: Italy, Slovenia, Malaysia, Thailand, and the United Kingdom (Osterrieder et al., 2021). Its objective consisted of evaluating the social, ethical, and behavioral aspects of the COVID-19 pandemic through an online survey. In this work, we focus on the analysis of Italian data from this survey promoted through social networks and carried out through two different sampling designs.

## 2. Questionnaire and survey

The SEBCOV survey questionnaire consisted of 36 questions concerning social, ethical and behavioural aspects of the COVID-19 pandemic. In particular, the questionnaire is composed of five sections: (1) socio-demographic information; (2) income, occupational status and economic impact of COVID-19; (3) preferences and perceptions regarding COVID-19-related communication and occurrence of fake news; (4) perceived level of knowledge about COVID-19, the use of non-pharmaceutical interventions and behavioural changes; (5) concerns and coping strategies related to restrictions.

It was spread in two different stages among people between 18 and 75 years old, residing in Italy. In fact, the study participants were adults who gave their informed consent, and were able to use a computer or smartphone. In the first stage, which took place from April 21 to May 4, 2020, the questionnaire was filled out via social networks such as Facebook and Instagram following a quota sampling. During this stage, 1002 responses were collected. The second stage started on 1 May and ended on 30 June 2020 and was promoted on the Facebook page dedicated to the SEBCOV study. During these two months, the questionnaire was advertised in all countries involved in the study. The responses received in this phase were 712. The two samples differ because of the timing of data collection and the sampling technique.

Margherita Silan, University of Padua, Italy, silan@stat.unipd.it, 0000-0001-5541-0603

Riccardo Bellide, University of Padua, Italy, rbellide@gmail.com

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Margherita Silan, Riccardo Bellide, *How Italians coped with COVID-19 lockdown: evidence from a survey promoted through social networks*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.08, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 41-45, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

Table 1: Sample composition according to gender, age, education, geographical location and number of household components in the first and second stages, and in the Italian population (ISTAT).

		First stage	Second Stage	Italian population
Gender	Female	55.1%	68.9%	50.2%
	Male	44.9%	31.1%	49.8%
Education	Low	56.3%	30.5%	82.6%
	High	43.7%	69.5%	17.4%
Geographical Location	Northwest	26.2%	16.3%	26.3%
	Northeast	19.5%	50.6%	19.8%
	Center	28.1%	18.9%	19.7%
	South/Islands	26.3%	14.2%	34.2%
Age	18-34	22.2%	38.3%	25.1%
	35-54	39.0%	35.2%	40.0%
	55+	38.9%	26.6%	34.9%
Number of Households Components	1	17.1%	14.8%	26.7%
	2	30.8%	32.4%	25.2%
	3	26.2%	24.6%	22.7%
	4	18.4%	20.8%	18.1%
	5+	7.6%	7.5%	7.3%

### 3. Sample weights

The samples collected in the two stages present different socio-demographic characteristics, from each other and from the whole Italian population. Thus, producing distorted results without the use of weights (Table 1) (Mercer et al., 2017). We considered two weighting methods: *post-stratification*, which considers the frequencies resulting from mutually exclusive intersections between the modalities of selected variables; and *raking*, which considers only the marginal frequencies of selected variables, neglecting intersections between them (Battaglia et al., 2004).

Variables we selected to compute post-stratification weights (B in Figure 1) are gender, education, geographic location and age, according to data availability on the ISTAT website. We calculate the raking weights with the same subset of variables (A in Figure 1) and including an additional variable (C in Figure 1): the number of household members. This last variable is extremely relevant in the analysis of the reaction of individuals to COVID-19 lockdown.

In order to compare results, we considered as a benchmark the percentage of individuals who adopted smart-working among those who continued to work during the pandemic. According to an ISTAT survey about daily activities at the time of the coronavirus, this percentage is equal to 44% between 5 and 21 April 2020. The weighting method that produces less biased results with respect to the benchmark question is the raking with 5 variables (thus including also the number of household members). Using this set of weights, the percentages of individuals that adopted smart-working among those who continued to work during the pandemic is 45.3% in the first stage and 50.48% in the second. These two percentages may be different mainly for two reasons: the sampling technique or the timing of the data collection.

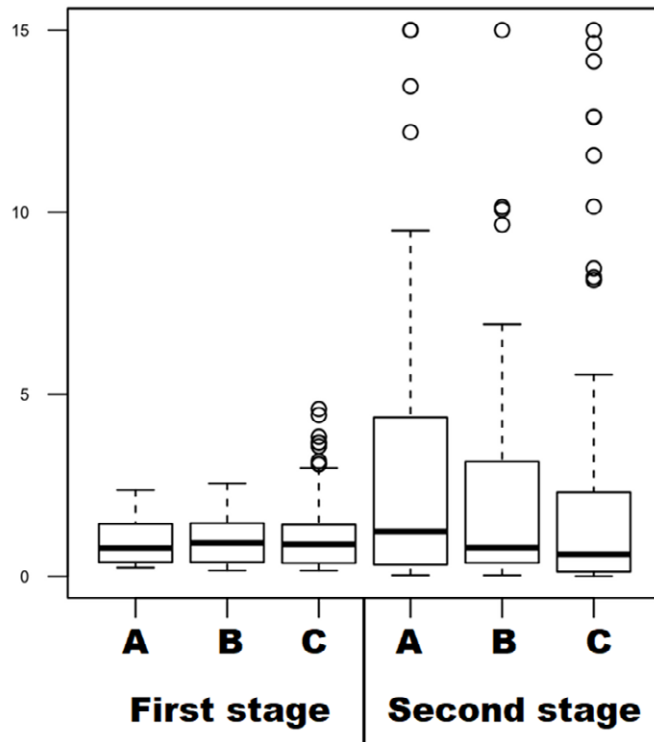


Figure 1: Weights calculated with different combinations of techniques and variables.

#### 4. Samples comparison

Having two samples that differ both in time and data collection method, we use a Chi-Squared test and a Chow test to check for structural differences in the two datasets due to time reasons, assuming that trimmed weights computed with raking and 5 variables helped us reduce the self-selection bias.

The Chow test is an econometric test that consists of verifying structural differences in two datasets using regression models (Wooldridge, 2015).

Table 2: P-values of the Chi-square test and the Chow test for some variables contained in the questionnaire.

Variable	$\chi^2$ TEST	CHOW TEST
Reduction of working hours	< 0,01***	< 0,01***
Smartworking	< 0,01***	< 0,01***
Changed social behaviour before restrictions	0,73	0,72
Physical concerns related to lockdown	0,27	0,07*
Concerns about mental well-being during lockdown	0,08*	0,05*

The two testing methods mentioned above lead to almost concordant results. The first two rows of Table 2 refer to work-related questions; in this case, the tests underline a significant structural difference in the answer between the two shots, probably due to the different timing of the surveys. In fact, after the 4<sup>th</sup> May 2020, Italians experienced an ease of mobility restrictions

and the resumption of work activities.

On the other hand, the following three rows in Table 2 relate answers to questions such as *"Did you change your social behaviour before the implementation of government restrictions?"* where a precise reference period is specified in the question, whether it is *"before the implementation of government restrictions"* or *"during the lockdown"*. In these cases, tests did not find significant structural differences between the two stages of the survey after being weighted.

Thus, we may conclude that an important element that prevents us from pooling together the data coming from the two stages is the fact that they represent situation completely different: the first stage when Italians were locked inside their houses without the possibility of going out, sometimes not even going to work; and the second one when the restrictions were already eased. However, the dissimilarities we observed in our samples may also be due to different sampling strategies that affect other factors not properly taken into account by the post-stratification techniques used in the analysis.

## 5. The impact of COVID-19 lockdown

In this section we show some results regarding data collected during the first shot that better represent the lockdown period.

One of the most sensitive topics during the lockdown period was the impossibility of going to work. In this regard, 64% of workers experienced a decrease in income, as a consequence of the forced reduction of work activity. This figure is particularly dramatic for workers with a low-medium level of education (high school diploma or lower). In addition, there has been a real work suspension for 27% of workers before the COVID-19 lockdown period, it was temporary in some cases, but also permanent. In contrast, some categories have been subjected to greater work pressure. Indeed, one must remember the contribution of health care personnel subjected to gruelling shifts to cope with the emergency, 56% of whom suffered a greater workload compared to 15% of workers in other sectors.

Many respondents expressed concern about a possible deterioration of their financial situation if they were unable to leave home except for essential needs. This concern was mainly expressed by the less educated respondents (45% vs. 55%) and among people under 35 years of age and between 35 and 54 years of age (respectively 62% and 57% vs. 40% of those over 55 years of age).

Restrictions on movement and social interaction imposed for longer or shorter periods can produce health consequences and induce states of anxiety among the population. Even before the government restrictions, many respondents (48%) had adopted new behaviors: 77% of them did so by trying to avoid contact with elderly people or those with pre-existing medical conditions, and 10% by moving from their usual home to parents or relatives. In fact, during the period of the pandemic, there was a phenomenon by which, in anticipation of the restrictions, a part of the population decided to change its home, mainly for space or social reasons. This behaviour was recorded mainly in large families consisting of 4, 5 or more individuals (18-19%), while in small families with less than 3 components it is less evident (no more than 8%). Indeed, one of the most worrying aspects was the limitation of social interactions together with mental health, with some differences between age classes and gender.

One of the sections of the questionnaire was about concerns when unable to leave home except for essential or work-related needs. Respondents are particularly concerned about their mental and physical state and maintaining mental well-being (65% of respondents), especially if they are unable to leave home for really long periods of time. Another major concern relates to being unable to see and hang out with relatives or friends and thus the risk of social isolation (68%). This state was more prevalent among 18-34 year olds (82% versus 62% for adults

and 64% for elderly). The percentage of concerns about social interactions is higher among men, 71%, compared to 65% among women. Worries about care-giving responsibilities (that refer to both caring for children and elderly) vary, of course, with age: higher percentages of respondents worried by care-giving are between 35-44, 45-54, and 55-64 years old (respectively, 56%, 62%, and 62%); while lower percentages are in younger age classes between 18 and 24 and between 25 and 34 (respectively, 36%, and 44%). Furthermore, 60% of the respondents said that during the lockdown that they had tried to improve their health status by implementing exercise or introducing food deemed healthier into their diet.

Almost all respondents (96%) declared that they had spent the lockdown period connecting with other people through the social network that had a fundamental role in this challenging period, for all age classes. Indeed, 49% used alternative methods (via the web) to carry out their activities (educational or work). Obviously, for the latter, the difference is considerable depending on the level of education (67% among those with a bachelor's degree compared to 45% among those with a lower level of education). The strong influence of the Internet on everyday life during this period helped keep people close but also encouraged the spreading of fake news; indeed, almost all respondents received fake news on several topics.

## 6. Conclusions

It can be concluded that the use of non-probabilistic surveys, particularly those taken through social networks such as the SEBCOV survey, can be a powerful tool in health emergency situations (Grow et al., 2020). In these circumstances, as demonstrated in this work, the health condition and people's perceptions of it change rapidly. As shown in our analysis, the timing of surveys is a very important aspect and the spreading of the questionnaire should be well advertised especially among quotas that are more difficult to reach in order to fill quotas quickly and reduce the duration of the survey. Using a quota sampling allows smaller weights and thus a lower variance of estimates even with small samples.

In this sense, it seems that the SEBCOV survey allowed for an accurate snapshot of the effects of lockdown on the lives of Italians.

## References

- Battaglia, M. P., Izrael, D., Hoaglin, D. C., Frankel, M. R. (2004). Tips and tricks for raking survey data (aka sample balancing). *Abt Associates*, **1**, 4740-4744.
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., Zagheni, E. (2020). Addressing public health emergencies via Facebook surveys: advantages, challenges, and practical considerations. *Journal of medical Internet research*, **22**(12), e20653.
- Mercer, A. W., Kreuter, F., Keeter, S., Stuart, E. A. (2017). Theory and Practice in Nonprobability Surveys: Parallels between Causal Inference and Survey Inference. *Public Opinion Quarterly*, **81**(S1), 250-271.
- Osterrieder, A., Cuman, G., Pan-Ngum, W., Cheah, P. K., Cheah, P., and Peerawaranun, P., Silan, M., Orazem, M., Perkovic, K., Groselj, U., Schneiders, M. L., Poomchaichote, T., Waithira, N., Asarath, S., Naemiratch, B., Ruangakajorn, S., Skof, L., Kulpijit, N., Mackworth-Young, C., Ongkili, D., Chanviriyavuth, R., Mukaka, M., Cheah, P. Y. (2021). Economic and social impacts of COVID-19 and public health measures: results from an anonymous online survey in Thailand, Malaysia, the UK, Italy and Slovenia. *BMJ Open*, **11**(7), e046863.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.



# Official statistics for measuring the sustainability of tourism: the UNWTO initiative

Emanuela Recchini

## 1. Introduction

The worldwide ongoing digital transformation is facilitating the availability of an ever increasing amount of data. The demand for data-driven decision-making is stimulated more and more by the increase itself of information. On the other side, even more attention is to be paid to make sure that the available data is accurate and, given the importance now attached to sustainability, the accuracy at issue concerns distinctly economic, social and environmental statistics and their integration.

Since tourism has been one of the fastest growing sector in the recent past before the appearance of the COVID-19 pandemic, in the last few decades this sector has been increasingly drawing the attention of agencies and stakeholders, focused on how tourism might deter or even support efforts towards sustainable development, especially in the face of challenges like climate change or poverty alleviation.

In order to make the tourism sector more responsible and its development more sustainable, the availability of data that are relevant, integrated and timely and the establishment of a statistical system devoted to sustainable tourism that is worldwide trusted are more important than ever. Data from official statistics, characterized by the highest quality possible inasmuch as they are produced in compliance with the United Nations Fundamental Principles of Official Statistics and the European Statistics Code of Practice, are best suited to meet this need.

The United Nations World Tourism Organization (UNWTO) is the agency with the UN mandate to promote tourism as a driver of economic growth, inclusive development and environmental sustainability. Along these lines, UNWTO is involved in a range of projects to support the sustainability of tourism. An initiative known as Measuring the Sustainability of Tourism (MST), launched by UNWTO in late 2015 in partnership with the United Nations Statistics Division (UNSD), is particularly relevant from a statistical standpoint.

As a long-term purpose which is particularly close to decision makers' needs, the MST initiative intends to propose an international statistical standard that not only can provide methodological guidance for statistics on the sustainability of tourism but can support measurement of progress towards the UN Sustainable Development Goals (SDGs), part of the 2030 Development Agenda, on the basis of indicators that are relevant as far as those targets directly related to tourism are concerned<sup>1</sup>.

On a methodological ground, the main effort of MST is to establish a Statistical Framework for measuring the role of tourism in sustainable development (SF-MST). Official statistics are at the core of said framework, since this is supposed to provide crucial guidance for countries to produce statistical data that is credible, comparable, integrated and enriched by harmonised metadata.

In the present paper, after an overview of the concept of sustainable tourism and the UNWTO

---

<sup>1</sup> Relevant targets within SDGs are the following: target 8.9 (“devise and implement policies to promote sustainable tourism that creates jobs and promotes local culture and products”); target 12.b (“develop and implement tools to monitor sustainable development impacts for sustainable tourism that creates jobs and promotes local culture and products”); target 14.7 (“increase the economic benefits to small island developing States and least developed countries”) (UN General Assembly, 2015).



MST initiative, a specific focus is on SF-MST. Concluding remarks and annotations on future developments complete the paper.

## **2. The sustainability of tourism and the UNWTO initiative**

Tourism is a multidimensional phenomenon relying on and having impacts on economy, environment and society. Its role in supporting or deterring efforts towards sustainable development (e.g. by creating jobs, on the one hand, or by contributing to pollution, on the other hand) is now universally recognized. This awareness has been consolidating in the framework of the debate on sustainability started since the early nineties following the appearance in 1987 of “Our Common Future”, the Brundtland Commission report on sustainable development (World Commission on Environment and Development, 1987), and the subsequent Rio Earth Summit of 1992<sup>2</sup>.

According to UNWTO, “sustainable tourism” is a “tourism that takes full account of its current and future economic, social and environmental impacts, addressing the needs of visitors, the industry, the environment and host communities” (UNEP, UNWTO, 2005). More specifically, according to the community of experts working on UNWTO projects, sustainable tourism is one that, in addition to making optimal use of environmental resources, should respect the socio-cultural authenticity of host communities, conserve their built and living cultural heritage and traditional values, and contribute to inter-cultural understanding and tolerance; furthermore, in addition to ensuring viable, long-term economic operations, sustainable tourism should provide socio-economic benefits to all stakeholders that are fairly distributed, including stable employment and income-earning opportunities and social services to host communities, and contributing to poverty alleviation.

UNWTO found that for the collection of statistical information suitable to describe sustainability aspects of tourism no standardized basis was available for the time being. This gap was deemed worth to be filled in order to make possible a proper support to decision makers involved in advancing sustainable tourism.

With a view to achieve this, the UNWTO Committee on Statistics has set up a multidisciplinary Working Group of Experts on Measuring the Sustainability of Tourism (WGE-MST), by engaging experts from national statistical offices, tourism administrations and observatories, international and regional organizations, academia and the private sector. The major task of this group of experts not only is to lead the needed technical development but also to support engagement among stakeholders. Given the relevance of the environmental-economic dimension of sustainability, this leading role is played by WGE-MST in coordination with the United Nations Committee of Experts on Environmental-Economic Accounting (UNCEEA).

A very important objective already realized is the almost finalized drafting of the above mentioned statistical framework, in a sense the core element of MST since SF-MST is envisaged to be adopted as the much needed standardized statistical basis.

A number of pilot studies, including examples of policy applications, have been carried out according to the conceptual structure of SF-MST. Italy, with Istat, is among the first countries that have realized pilot studies for the purposes of MST (Istat, 2019; Tudini, Ardi, Recchini, 2018).

The current draft of SF-MST is the result of several rounds of consultations among the members of UNWTO Committee on Statistics and of WGE-MST (UNWTO, 2018a). In addition to that, global consultations have been carried out, obtaining comments from about 20 countries, including Italy with Istat and Ministry of the Environment, as well as from international agencies and academic institutions.

It’s worth noting that on the occasion of the International Year of Sustainable Tourism for Development 2017, the first draft of SF-MST was a core component of the conference programme of the “Sixth UNWTO International Conference on Tourism Statistics: Measuring

---

<sup>2</sup> <https://www.un.org/en/conferences/environment/rio1992>

sustainable tourism” held in Manila in the same year<sup>3</sup>.

As highlighted within the tourism statistician community, this Conference is considered a historical milestone for tourism statistics. It was the first time ever that a UNWTO event united ministers, statistical chiefs, policy experts, statisticians, private sector and academics dedicated to the measurement of sustainable development and tourism. Not only all parties fully supported the SF-MST, but the Conference concluded with the adoption of the Manila Call for Action on Measuring Sustainable Tourism, which represents a global commitment to create a consistent statistical approach to measuring the full impact of tourism. It emphasizes that effective sustainable tourism policies require integrated, coherent, comparable and robust data.

Along with the development of overall methodological work, a number of specific though somewhat cross-cutting conceptual research areas have been addressed by WGE-MST, namely the following: social sustainability of tourism, employment in tourism industries, defining spatial areas, implementation strategy, communication strategy, tourism SDG indicators. For the different research areas, ad-hoc sub-groups have been established, each led by an expert from a different country/agency. An expert from Istat – the Institute representing Italy in the WGE-MST – leads the sub-group on social sustainability of tourism (Recchini, 2018; Recchini, Costantino, 2019).

### **3. Measuring the full impact of tourism based on official statistics data: the statistical framework**

As anticipated, the ambition of the MST initiative is to provide a standardized statistical structure allowing to measure and monitor the full impact of tourism and supporting decision-making towards any preventive and/or corrective action/policy/measure.

SF-MST plays a fundamental role in providing a common and harmonized set of relevant concepts, definitions, classifications and measurement scopes, thus developing a standardized and comparable language in the field of quantitative measurement.

Particularly important in achieving this goal is the crucial role of official statistics in SF-MST. As a matter of fact, it is within official statistics that common understanding on concepts, definitions and related terminology for measurement purposes is ensured and proper support to the measurement of changes over time and of differences between locations can be provided. Official statistics, by their nature, provide reliable, impartial, transparent, accessible and relevant information produced according to the highest possible quality criteria and strict conditions concerning processes and conceptual methods. In official statistics metadata is no less important than the data itself.

In principle, SF-MST builds upon existing internationally agreed statistical standards and guidance developed for the three dimensions of sustainable tourism, economic, environmental, social. By integrating these different domains, SF-MST intends to overcome the fragmentation due to no underlying alignment between the corresponding statistics.

Among international statistical standard, the International Recommendations for Tourism Statistics 2008 (IRTS 2008) (UN, UNWTO, 2010) together with the Framework for the Development of Environment Statistics 2013 (FDES 2013) (UN, 2017) are two essential references for the definition and collection of internationally comparable tourism statistics that take into account also the environmental dimension of sustainability.

SF-MST is to a great extent inspired by accounting concepts. In this perspective, the System of National Accounts 2008 (SNA 2008), with its comprehensive, consistent and flexible set of macroeconomic accounts provides the globally accepted accounting framework supporting decision-making, analysis and research work in the economic field (UN et al., 2009), thus being the basic statistical standard for addressing the economic dimension of sustainability. Of course,

---

<sup>3</sup> <https://www.unwto.org/archive/asia/event/6th-international-conference-tourism-statistics-measuring-sustainable-tourism>

in addition to SNA 2008, the Tourism Satellite Accounts: Recommended Methodological Framework 2008 (TSA: RMF 2008) (UN, UNWTO, Eurostat, OECD, 2010) is the international statistical standard specific for describing the economic aspects linked to tourism. Furthermore, for the environmental aspects concerning tourism, in addition to the above mentioned IRTS 2008, the System of Environmental-Economic Accounting 2012 – Central Framework (SEEA-CF 2012) (UN et al., 2014) is a very key international statistical standard which SF-MST builds upon.

The scope of existing international statistical standards that are actually used for measuring tourism is largely economic for the time being. Systems of tourism statistics in line with the international statistical standards specifically focused on tourism mentioned above have been developed by many countries, but the growing need of decision makers and stakeholders for an overall system covering the three dimensions of sustainability has led SF-MST to acknowledge the multifaceted nature of sustainable tourism, without trying to provide a univocal operational definition of this concept. In practice, SF-MST is meant to provide a single reference point for extending the current range of tourism statistics to include the three dimensions of sustainable tourism at relevant spatial scales. The integration of economic, environmental and social statistics on tourism at appropriate spatial scales represents the key aspect of SF-MST.

The linking of TSA: RMF 2008 and SEEA-CF 2012, both aligned with SNA principles and structure, is a central feature of SF-MST: the former provides guidance for measuring the direct economic impact of tourism, the latter for the measurement of the relationships between tourism as an economic activity and the natural environment.

A specific output of MST has been the development of a Technical Note linking SEEA and TSA, which has been prepared under the joint auspices of the UNWTO Committee on Statistics and the UNCEEA. This SEEA-TSA Technical Note describes a core part of the overall SF-MST by providing a framework to link the economic and environmental dimensions of sustainable tourism. It is structured to provide a starting point for compilers of tourism and environmental-economic accounts to consider ways in which their accounts can be adapted and extended to organize information for assessing sustainable tourism (UNWTO, 2018b).

Based on an accounting approach, SF-MST points to sustainability assessment by measuring a broad set of capitals (produced, natural, human and social capital) and the flows of related incomes and benefits.

As regards the social dimension of the sustainability of tourism, further effort is needed, however, because social statistics are particularly complex and in general they are relatively less mature, compared e.g. to the economic data (Recchini, Costantino, 2019).

The social dimension, in fact, is the weakest pillar of the measurement of sustainable development, due to different theoretical and analytical bases still under debate. Nevertheless, the concept of social capital, despite the current unavailability of a standard accounting system due to its intangible and multi-dimensional nature not allowing its direct measurement, is deemed to be appropriate for integrating the social dimension of the sustainability of tourism into the multiple capitals-based approach (Recchini, 2018).

Turning to implementation aspects concerning SF-MST, it is expected that application work would be flexible and modular, allowing countries to take into consideration only those aspects and those spatial levels considered most relevant, also on the basis of available resources.

#### **4. Concluding remarks and future developments**

An increasingly globalized and interconnected world, which is also better aware about sustainable development, enhances the need for accurate information to better target decision-making. Official statistics, based on the UN Fundamental Principles of Official Statistics and the European Statistics Code of Practice, are best suited to meet this need.

Regarding tourism – given its impacts on economy, environment and society – we are moving towards the production of data reflecting a sustainability perspective. SF-MST, the main effort of UNWTO in terms of methodological development for the purposes of MST, addresses decision

makers' demand for integrated statistics on tourism reflecting the three dimensions of sustainability and is proposed as a standardized basis for the collection of relevant information. This is supposed to integrate statistics on different domains in order to measure the role of tourism in sustainable development at appropriate spatial scales.

The finalization of SF-MST – after an active process of research, discussion and worldwide consultation across multiple experts, sectors and stakeholders – is currently at a quite advanced stage. The United Nations Statistical Commission (UNSC), noting the strong interest from countries in this work, has encouraged the finalization of SF-MST (UNSC, 2022). The final version of the document is expected to be submitted to the UNSC at its next session, for approval as an international statistical standard.

SF-MST, involving a wide range of agencies and stakeholders, plays a key role in providing an integrated information basis for development of data and metadata and derivation of indicators supporting more effective decision-making towards sustainable outcomes.

## References

- Istat (2019). *Integrated economic and environmental accounts for tourism: pressures of tourism industries on the natural environment*. Experimental statistic, 14 march 2019. <https://www.istat.it/en/archivio/228239>
- Recchini E. (2018). *Note on measuring the social dimension of the sustainability of tourism*. Methodological contribution prepared for the UNWTO Second meeting of the Working Group of Experts on Measuring the Sustainability of Tourism, Madrid, Spain, 24-25 October 2018. [https://webunwto.s3-eu-west-1.amazonaws.com/imported\\_images/50458/italy\\_mst\\_discussion\\_note\\_social\\_issues.pdf](https://webunwto.s3-eu-west-1.amazonaws.com/imported_images/50458/italy_mst_discussion_note_social_issues.pdf)
- Recchini E., Costantino C. (2019). *Measuring the social dimension of sustainable tourism*. Technical report for the UNWTO Committee on Statistics, Nineteenth meeting, Madrid, Spain 26-27 February 2019. [https://webunwto.s3-eu-west-1.amazonaws.com/imported\\_images/51342/19th\\_statcom\\_item\\_5.1\\_doc.pdf](https://webunwto.s3-eu-west-1.amazonaws.com/imported_images/51342/19th_statcom_item_5.1_doc.pdf)
- Tudini A., Ardi C., Recchini E. (2018), *Italy's MST Pilot Study*, contribution for the UNWTO Committee on Statistics, Eighteenth meeting, Madrid, Spain 27-28 February 2018. [https://webunwto.s3-eu-west-1.amazonaws.com/s3fs-public/2020-09/18thmeeting\\_item\\_4.3.1.pdf](https://webunwto.s3-eu-west-1.amazonaws.com/s3fs-public/2020-09/18thmeeting_item_4.3.1.pdf)
- UN (2017). *Framework for the Development of Environment Statistics (FDES 2013)*. New York. <https://unstats.un.org/unsd/environment/fdes/FDES-2015-supporting-tools/FDES.pdf>
- UN, EC, FAO, IMF, OECD, World Bank (2014). *System of Environmental Economic Accounting 2012 — Central Framework*. New York. [https://unstats.un.org/unsd/envaccounting/seearev/seea\\_cf\\_final\\_en.pdf](https://unstats.un.org/unsd/envaccounting/seearev/seea_cf_final_en.pdf)
- UN, EC, OECD, IMF, World Bank (2009). *System of National Accounts 2008*. New York. <https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf>
- UNEP, UNWTO (2005). *Making Tourism More Sustainable: A Guide for Policy Makers*. Paris, France; Madrid, Spain.
- UN General Assembly (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. 21 October 2015, A/RES/70/1. <https://www.refworld.org/docid/57b6e3e44.html>
- UNSC (2022). *Report on the fifty-third session (28 February–2 March and 4 March 2022)*. New York. <https://unstats.un.org/unsd/statcom/53rd-session/documents/2022-41-FinalReport-E.pdf>
- UN, UNWTO (2010). *International Recommendations for Tourism Statistics 2008*. New York. [https://unstats.un.org/unsd/publication/seriesm/seriesm\\_83rev1e.pdf](https://unstats.un.org/unsd/publication/seriesm/seriesm_83rev1e.pdf)
- UN, UNWTO, Eurostat, OECD (2010). *Tourism Satellite Account: Recommended*

- Methodological Framework 2008*, United Nations publication.  
[https://unstats.un.org/unsd/publication/seriesf/seriesf\\_80rev1e.pdf](https://unstats.un.org/unsd/publication/seriesf/seriesf_80rev1e.pdf)
- UNWTO (2018a). *Statistical Framework for Measuring Sustainable Tourism – Consultation Draft*. Draft prepared for discussion with the Working Group of Experts on Measuring the Sustainability of Tourism, October 2018. [https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2020-09/wge\\_mst\\_2nd\\_item\\_2.1\\_doc\\_0.pdf](https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2020-09/wge_mst_2nd_item_2.1_doc_0.pdf)
- UNWTO (2018b). *Linking the TSA and the SEEA: A Technical Note*. Madrid, Spain.  
<https://webunwto.s3-eu-west-1.amazonaws.com/2019-08/tsaseeatechnote.pdf>
- World Commission on Environment and Development (1987), *Our common future*, Oxford University Press.

# Misinformation and Disinformation in Statistical Methodology for Social Sciences: causes, consequences, and remedies

Giulio Giacomo Cantone, Venera Tomaselli

## 1 Introduction: the replicability of the Social Sciences

This paper concerns the prevalence and the causes of low replication rates in Social Sciences. The aim is to frame unintentional errors as scientific misinformation, and questionable research practices as disinformation. In Section 3 is presented Multiverse Analysis, which helps the assessment of the uncertainty about scientific claims and reduces false discoveries.

In order to introduce the topic of replication rate in Science, it is important to clarify the epistemological conditions to claim a scientific result to be replicated:

1. A scientific result consists of a claim  $A$  that is deduced through a procedure that can be reproduced by a third party (Goodman et al., 2016). A proper scientific result should be reported in an authoritative scientific venue, usually a peer-reviewed journal.
2. Others try to refute  $A$  by reproducing the same procedure on a different sample or adopting advanced but theoretically coherent alternative procedures on the original sample.
3. But these attempts fail: new results are not incompatible with  $A$ .

A replicated scientific theory is a collection of connected claims that are, for most, individually replicated (Lakatos, 1976; Schmidt, 2009). A replication rate is the rate of replicated results given a grouping variable: an author, an institution, or a scientific field. High replication rates are observed in exact sciences. Often, these replications are implicit: after a few successful experiments, a scientific theory is applied to more complex theories or technologies. The application of a theory is an implicit process of scientific replication (Feigenbaum and Levy, 1996). Methods of Social Sciences are not exact but probabilistic, harder to reproduce (e.g. due to changes in society), and applications into social policies are more nuanced than the vertical integration of natural sciences into technology.

Often claimed causal effects in Social Sciences are just statistical artifacts. Even meta-analyses are biased by so-called ‘publication bias’ (Nissen et al., 2016). It has been empirically demonstrated, indeed, that not significant estimates are less likely to be published in scientific venues (van Zwet and Cator, 2021). Prof. Breznau’s research group provided the same dataset to 73 independent teams of quantitative social scientists, for a total of 161 people. He asked them to estimate the effect of immigration rates on public support for welfare-oriented political agenda. A sample of  $n > 1,200$  estimate values for the effect has been drawn through this survey. Of the estimates, 25% were significantly negative, 17% significantly positive, and 57.7% of the times the specified model failed to reject the null hypothesis (Breznau et al., 2022). Impressively, based on this result, not only it is almost impossible to claim that a general effect exists, but even to fully deny it, because it is always possible to assert that an effect holds under specific conditions.

The U.S. Agency for Defense Advanced Research Projects (DARPA) understood the problem of traditional approaches for Meta-Analysis and Causal Inference and launched the Systematizing Confidence in Open Research and Evidence (SCORE) Project to understand how to

predict if a study is deemed to fail to replicate. Preliminary findings have not been rosy: with exception of Economics, social scientists believe that their own fields produce more not replicable claims than replicable ones, i.e. there are more false discoveries than not. Economics seems to suffer of overconfidence in itself (Gordon et al., 2020). These results came after a large study led by Brian Nosek that attempted to replicate 100 claims in Psychology journals: less than half passed a replication attempt (OPEN SCIENCE COLLABORATION, 2015). Journals with high bibliometric scores do not perform better than other sources: evidence is in the direction of zero or negative correlation between bibliometric performances (e.g. journal impact factor) and replication rates (Szucs and Ioannidis, 2017; Brembs, 2018; Camerer et al., 2018).

## 2 Misinformation and disinformation

Ioannidis (2005) summarised predictors of low replication rates: small sample sizes, small effect sizes, and more than one hypothesis being tested on the same sample. On top of this, he stresses the incentives to look for novel findings instead of replication studies, too. He claims that papers on new theories are always more cited than their replication attempts, even when replication is not attained! This is a case of misinformation: inaccurate claims spread more than their corrections. Disinformation is a distinct phenomenon, where false claims are justified through a process of fabrication (West and Bergstrom, 2021). It is not necessary to report *fake data* to fabricate a fake result. The insidious alternative is to *omit* observed results. This behaviour is called “hacking the science” in the scientific community, by analogy with the method of *bruteforcing* many random combinations of inputs until a singular desired outcome is achieved by chance, e.g. hacking a password (Imbens, 2021).

### 2.1 Misinformation: is Dunning-Kruger effect a statistical artifact?

It is commonly observed that the correlation between performance and self-assessment of performance is significantly negative. Since performance depends on skill, the theory of Dunning-Kruger Effect (Kruger and Dunning, 1999) or DK, explains this correlation through the claim that unskilled people have a tendency to overestimate their own skills. The original study, with more than 8,000 citations, is foundational for modern Pedagogy. A concurrent to DK is the “better than average” theory (Krueger and Mueller, 2002), or BTA. It claims that all people have a tendency to self-assess their skills above the average, independently of their skill. These two theories can coexist but if BTA is true, then the DK effect is overestimated.

Consider the conservative case of two actors: one with a true skill score  $x_1 = 40$  and the other with a true skill score  $x_2 = 60$ . Their average is  $\bar{x} = 50$ . Assume the claim of BTA: actor 1 and actor 2 have exactly the same model of assessment of self-score: they adopt the average plus an expected positive error  $\epsilon^+$ . In this case, it holds

$$|x_1 - (\bar{x} + \epsilon^+)| > |x_2 - (\bar{x} + \epsilon^+)|, \forall \epsilon^+ \quad (1)$$

where  $|x - (\bar{x} + \epsilon^+)|$  is the absolute error between true skill and self-assessed skill. It follows that: even with absolutely no cognitive differences between classes of actors (i.e.  $\epsilon^+$  is unique across actors), the less skilled actor has a larger absolute deviation. In this case, even if DK is not true, then the parameter  $\epsilon^+$  would induce a negative correlation. With few generalisations it is shown that any model that parameterises the self-assessed score to  $\mu_X + \epsilon^+$ ;  $\forall X : \{x_1, x_2, x_3, \dots, x_n\}$  would lead into an artificial DK effect, even when DK is not true. The effect would hold even for normally distributed positive  $\epsilon^+_{actor}$ .

A meta-analytical study that adopted advanced statistical techniques found that, given the observed scores in the literature, DK is likely to be a statistical artifice due to BTA (Gignac and

Zajenkowski, 2020). Another study reports only partial support for a true DK effect while confirming BTA (Jansen et al., 2021). Here no information has been concealed or fabricated. The authors did not adopt any questionable research practices. They lacked the correct specification of their null model.

## 2.2 Disinformation: six degrees of separation and even more

The expression “small world” refers to a network where a part of the connections happens with a uniform probability, and another part happens with a higher probability to form triadic closures (fully connected triangles of nodes). As emergent propriety, small world networks have a “characteristic average path length”  $L$ : for any given node in the network, any other node can be reached only by crossing paths with an expected length equal to  $L$ , independently by the number of nodes in the network.

Formation and structure of small-world networks have been described in the Watts-Strogatz model (Watts and Strogatz, 1998), but the description of this network goes back to Milgram (1967). Indeed, the implicit claim of Milgram is that in modern societies (pre-Internet) there is a characteristic path length  $L$  between human connections and that  $L$  is relatively short. Curiously, the paper with the experiment that originated the catchphrase “six degrees of separation” (Travers and Milgram, 1969) has been published only 2 years after a theoretical paper (Milgram, 1967) claiming the emergency of  $L$  in human societies. Together, the two papers collected more than 13.000 citations and, a rare case for a social science theory, they inspired new ideas not only in business (marketing, etc.) but also in engineering (transports, etc.).

It was a surprise for Judith Kleinfield (2002) to discover that the paper presenting the actual report of the *in vivo* experiment of the theory (Travers and Milgram, 1969) is actually poor in terms of statistical results. 296 participants have been recruited for the study. Their task was to send a document to one of their pre-existing social ties with the final aim that this document could reach a specific male broker in Boston. These 296 participants have been sampled across three populations: not brokers in Nebraska, brokers in Nebraska, and brokers in Boston.

This stratification would have been helpful if just enough documents reached their final destination: only 214 original participants sent the document and only 64 documents reached Boston’s broker, after  $s$  stages. Among these 64, the observed average path length  $l = 5.2$ . The territorial variable was the only statistically significant. The number 6 (degrees of separation) is never explicitly mentioned, however, in footnote 4 the authors mention that they adjusted  $l$  through a not better specified marginal distribution of probabilities of reaching the final node at  $s + 1$  stage (see paramter  $Q_i$ ). In footnote 4, they claim a confidence interval for  $L$  between 5 and 7. Is there sufficient evidence for claiming that  $L$  exists? From the sample of not brokers from Nebraska, only 18 documents reached the destination, with  $l = 5.7$ . This result could be generalised to the U.S. population but the sample size would be small.

Kleinfield (2002) investigated Milgram’s archives, looking for more. She only found concerning details:

- Milgram (1967) mentions a pilot study where a document has been received by a woman in only four days. Kleinfield found the pilot’s report and concluded that Milgram picked an interesting anecdote but he never published more details about the pilot because it was a failure. Attrition in the pilot was so high to make meaningless the observed statistics.  $Q_i$  is never mentioned in the pilot.
- Travers and Milgram (1969) tried to alter the attrition rate in two ways: avoiding to recruit social outcasts and modifying the document from a single piece of paper to a “passport” in bright colours.
- She found an anonymous manuscript about a third attempt with inconsistent results.



### 2.3 *p*-hacking

The first case study falls under the category of ‘misinformation within science’ because it regards how the reputation of theories spreads within science even when a new model has been proven more consistent. The second case study is different: researchers concealed results from their own research because these were inconclusive toward their hypothesis. This is relatable to the case of so-called *p*-hacking of the level of significance  $\alpha$  for rejection of the null hypothesis in statistical testing. *p*-hacking is a fraud because it omits to report the number of tests attempted before reaching a statistically significant result in data analysis (Simmons et al., 2011; Head et al., 2015). *p*-hacking is typically done in two ways:

1. Parallel *p*-hacking: many tests are arranged on different samples of the same population. Each sample has a minimal size but it is large enough to be deemed credible by the typical reader. Once a positive outcome is seen, no further test is necessary. In the reported result of the study, the number of tested samples is omitted and only the one associated with  $p < \alpha$  is reported. As a reference: if the parameter of the effect size is equal to 0 and the null hypothesis of the test is true; with  $\alpha = .05$ , after 14 tests (Bernoulli trials of parameter  $\alpha$ ), the probability to see a  $p < \alpha$  in at least a test is

$$\sum_{k=1}^{14} \alpha \cdot (1 - \alpha)^{k-1} > .51 \quad (2)$$

following the geometric distribution of the Bernoulli trials<sup>1</sup>.

2. Sequential *p*-hacking: a multivariate dataset is collected and a hypothesis is formalised with a simple model. If the statistics of the model are not significant, then the specification of the model is trivially adjusted (e.g., control variables are added to the model, outliers are removed, data is pre-processed differently, etc.) until a random  $p < \alpha$  is achieved. All of these operations are not reported. This is a fraudulent type of Hypothesising After Results are Known, or HARKing (Rubin, 2017).

## 3 Remedies: pre-registration and Multiverse Analysis

A possible remedy for *science hacking* is pre-registration, that is to record in a dedicated electronic archive an anonymous manuscript that details all the research questions and the methods of incoming research. This happens before the data collection, so in a peer-review authors can certify that their analysis is coherent with the original research design and that hypotheses are not drawn after knowing the sample statistics (Nosek et al., 2018). Pre-registration has two problems: (i) nothing prevents *p*-hacking a result, pre-registering its specification, then submitting the complete manuscript for peer-review (Yamada, 2018); (ii) it does not allow serendipitous discoveries incoherent with what is pre-registered (Simmons et al., 2021).

Looking back at the crowd-sourced estimation in Breznau et al. (2022), this approach is kindred to a meta-analytical paradigm called Multiverse Analysis: Gelman and Loken (2014) popularised the assumption that the robustness of a scientific model can be estimated through trivially altering its specification. They call “degrees of freedom of the researcher” the analytical choices in data analysis, e.g. the choice of a link function in binomial regression between *logit* and *probit*. Steegen et al. (2016) introduced the concept of the “multiverse” of a scientific claim. These degrees of freedom are the source of errors in estimation.

In particular, claims are formalised into models. Assuming that a *true parameter*  $\theta$  of the model exists, given a dataset, exists a set  $\Theta_j = \{\hat{\theta}_j\}$  of estimates from different *j*-specifications

---

<sup>1</sup>The equivalent command in R language is `pgeom(13, .05)`.

of the model such that each estimate  $\hat{\theta}_j$  sufficiently close to  $\theta$  and  $\mathbf{E}(\hat{\theta}_j) = \theta$  holds. How to draw a sample that is representative of  $\Theta_j$  in order to ascertain the uncertainty associated with the error of misspecification (model error)? Crowd-sourced estimation (Brenzau et al., 2022) draws a random sample of specifications and estimates just by surveying experts. Instead, Multiverse Analysis draws a systemic (not random) sample  $\hat{J}$  of specifications through mapping all the degrees of freedom of the researcher, e.g. inclusion/exclusion of control variables, operations in data pre-processing, modelling choices for overdispersion, etc. and combining them into  $\hat{J}$ , that is the multiversal sample of specifications or just the “multiverse”.

Multiverse Analysis assumes that measures of variability in the observed multiversal estimates  $\hat{\theta}_{j \in \hat{J}}$  are as much if not more informative than parametric or bootstrapped standard error or confidence intervals about the uncertainty involved in the estimation of  $\theta$  (Young and Holsteen, 2017; Simonsohn et al., 2020). An interesting application of Multiverse Analysis is for checking the Janus effect (Patel et al., 2015), which is when in the same multiverse co-exist statistically significant  $\hat{\theta}_j$ , but with different signs. Janus Effect is a red flag in the sample of so-called parametric type S error (Gelman and Tuerlinckx, 2000).

## References

- Brembs, B. (2018). Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers in Human Neuroscience*, 12.
- Brenzau, N. et al. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences*, 119(44):e2203150119.
- Camerer, C. F. et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9):637–644.
- Feigenbaum, S. and Levy, D. M. (1996). The technological obsolescence of scientific fraud. *Rationality and Society*, 8(3):261–276.
- Gelman, A. and Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102(6):460–466.
- Gelman, A. and Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics*, 15(3):373–390.
- Gignac, G. E. and Zajenkowski, M. (2020). The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. *Intelligence*, 80:101449.
- Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341):341ps12–341ps12.
- Gordon, M. et al. (2020). Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme. *Royal Society Open Science*, 7(7):200566.
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS biology*, 13(3):e1002106.
- Imbens, G. W. (2021). Statistical Significance, p-Values, and the Reporting of Uncertainty. *Journal of Economic Perspectives*, 35(3):157–174.
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8):e124.
- Jansen, R. A., Rafferty, A. N., and Griffiths, T. L. (2021). A rational model of the Dunning-Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour*, 5(6):756–763.
- Kleinfeld, J. S. (2002). The small world problem. *Society*, 39(2):61–66.
- Krueger, J. and Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average

- heuristic and statistical regression predict errors in estimates of own performance. *Journal of personality and social psychology*, 82(2):180.
- Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121.
- Lakatos, I. (1976). Falsification and the Methodology of Scientific Research Programmes. In Harding, S. G., editor, *Can Theories be Refuted? Essays on the Duhem-Quine Thesis*, Synthese Library, pages 205–259. Springer Netherlands, Dordrecht.
- Milgram, S. (1967). The small world problem. *Psychology today*, 2(1):60–67.
- Nissen, S. B., Magidson, T., Gross, K., and Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *eLife*, 5:e21451. Publisher: eLife Sciences Publications, Ltd.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11):2600–2606.
- OPEN SCIENCE COLLABORATION (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716.
- Patel, C. J., Burford, B., and Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, 68(9):1046–1058.
- Rubin, M. (2017). When Does HARKing Hurt? Identifying When Different Types of Undisclosed Post Hoc Hypothesizing Harm Scientific Progress. *Review of General Psychology*, 21(4):308–320.
- Schmidt, S. (2009). Shall we Really do it Again? The Powerful Concept of Replication is Neglected in the Social Sciences. *Review of General Psychology*, 13(2):90–100.
- Simmons, J. P., Nelson, D., and Simonsohn, U. (2021). Pre-registration: Why and How. *Journal of Consumer Psychology*, 31(1):151–162.
- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11):1359–1366.
- Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11):1208–1214. Number: 11 Publisher: Nature Publishing Group.
- Steen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci*, 11(5):702–712.
- Szucs, D. and Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, 15(3):e2000797.
- Travers, J. and Milgram, S. (1969). An Experimental Study of the Small World Problem. *Sociometry*, 32(4):425–443.
- van Zwet, E. W. and Cator, E. A. (2021). The significance filter, the winner's curse and the need to shrink. *Statistica Neerlandica*, 75(4):437–452. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/stan.12241>.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *nature*, 393(6684):440–442.
- West, J. D. and Bergstrom, C. T. (2021). Misinformation in and about science. *Proceedings of the National Academy of Sciences*, 118(15):e1912444117.
- Yamada, Y. (2018). How to Crack Pre-registration: Toward Transparent and Open Science. *Frontiers in Psychology*, 9.
- Young, C. and Holsteen, K. (2017). Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis. *Sociological Methods & Research*, 46(1):3–40.

# The impact of economic insecurity on life satisfaction among German citizens

Demetrio Panarello, Gennaro Punzo

## 1. Introduction

The concept of life satisfaction dates back to the Age of Enlightenment and became popular in the Nineteenth century as a synonym for ‘good life’. Understood as an overall assessment of the life a person leads (Veenhoven, 2017), since the 1960s there have been attempts to go beyond traditional economic criteria by broadening the definition and measurement of both the concept of well-being and life satisfaction on the basis of a wide range of indicators (Hasan, 2019; Hall et al., 2010). Although ‘money cannot buy happiness’, the economic dimension remains a crucial element in the assessment of many issues such as poverty, inequality, and deprivation (D’Ambrosio and Rohde, 2014). In particular, economic insecurity may also play a central role in assessing the well-being and life satisfaction of individuals and, by extension, of their family members, with inevitable repercussions for future generations (Linz and Semykina, 2010).

Economic insecurity has attracted the attention of researchers as a key aspect of socio-economic behaviour. It originates from unexpected economic loss (Giambona et al., 2022) due to feelings of failure and inability to recover and can be broadly defined as the sense of stress associated with an uncertain financial future (Panarello, 2021). Among other things, researchers observed associations between economic insecurity and political support (e.g. Colantone and Stanig, 2018; Guiso et al., 2017), body weight (Smith et al., 2013), mental health (Rohde et al., 2016), and environmental concern (Panarello, 2021). Therefore, there is reason to believe that economic insecurity can greatly affect individuals’ behaviour, as well as their well-being and satisfaction with life.

Based on the above, this paper analyses the impact of economic insecurity on workers’ life satisfaction over time in Germany. In particular, economic insecurity is investigated for its impact on the trajectories of life satisfaction over a time span of 29 years among working-age German citizens, taking into account their age and sector of economic activity.

The present article is structured as follows. Section 2 introduces the economic insecurity index and the growth models, which represent the key methodological ingredients of the study. Then, Section 3 illustrates the data, while Section 4 presents the main findings and closes with a brief summing-up.

## 2. Method

### 2.1 Economic insecurity index

Economic insecurity depends on the current level of income that each individual earns and its past changes, considering both the reserve role it may play in the event of future adverse events and the subjective prediction of how well the individual will handle possible future losses (D’Ambrosio and Rohde, 2014). In our analysis, we use the Bossert and D’Ambrosio’s (2016) economic insecurity index. This is an individual-level objective measure that considers income fluctuations between various consecutive years. Income gains and losses are assigned different weights, as well as more recent periods compared to those farther in the past, assuming that losses are more relevant than gains for the development of insecurity and that closer periods are more important than the remoter ones. The index can be defined as:

Demetrio Panarello, University of Bologna, Italy, demetrio.panarello@unibo.it, 0000-0003-1667-1936

Gennaro Punzo, University of Naples Parthenope, Italy, gennaro.punzo@uniparthenope.it, 0000-0001-6861-9553

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Demetrio Panarello, Gennaro Punzo, *The impact of economic insecurity on life satisfaction among German citizens*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.11, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 59-64, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

$$I^T(x) = l_0 \sum_{\substack{t \in \{1, \dots, T\}: \\ x_t > x_{t-1}}} \delta^{t-1}(x_t - x_{t-1}) + g_0 \sum_{\substack{t \in \{1, \dots, T\}: \\ x_t < x_{t-1}}} \delta^{t-1}(x_t - x_{t-1})$$

where  $x = (x_T, \dots, x_0)$  is an individual's yearly income;  $t$  is the distance from the current period, so that 0 refers to the current year and 2, for instance, refers to two years before;  $l_0$  and  $g_0$  are the weights assigned to past income losses and gains, respectively, and  $\delta$  is the weight based on the distance from the current period. We use  $l_0 = 1$ ,  $g_0 = 0.9375$  and  $\delta = 0.9$  for five years of income as in Bossert et al. (2019). Then, to run the models, the index is standardised with mean of zero and standard deviation of one.

## 2.2 Growth models

Latent Growth Curve Models (LGCs) were fitted to analyse over-time changes in workers' life satisfaction in relation to their economic insecurity. LGCs involve fitting a trajectory through each individual's repeated measures of life satisfaction to summarise its changes over the period 1989-2017 ( $T=29$ ).

To consider variation between individuals in the rate of change in life satisfaction (outcome variable) and its level at any time, a random slope growth model was fitted:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{ij}$$

where  $y_{ij}$  is the outcome variable at time  $i$  ( $i = 1, \dots, T$ ) for individual  $j$  ( $j = 1, \dots, n$ );  $x_{ij}$  is the economic insecurity evaluated at time  $i$  on individual  $j$ ;  $\beta_0$  is the intercept;  $\beta_1$  is the overall average slope;  $u_{0j}$  and  $u_{1j}$  are two individual-level random effects; and  $e_{ij}$  is an occasion-specific residual, detecting the effects on  $y$  of unobserved time-varying characteristics.

The growth rate for individual  $j$  is given by the sum of the overall average slope  $\beta_1$ , which is common to all individuals, and a random amount  $u_{1j}$  specific to individual  $j$ . It is assumed that  $u_{0j}$  and  $u_{1j}$  follow a bivariate normal distribution with zero mean:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(0, \Omega_u) \quad \text{where } \Omega_u = \begin{pmatrix} \sigma_{u_0}^2 & \\ & \sigma_{u_1}^2 \\ \sigma_{u_{01}} & & \sigma_{u_1}^2 \end{pmatrix}$$

$\sigma_{u_0}^2$  is the between-individual variance in the intercept;  $\sigma_{u_1}^2$  is the between-individual variance in the slope of  $x_{ij}$ ;  $\sigma_{u_{01}}$  is the covariance between individuals' intercepts and slopes.

The random slope growth model captures the within-individual correlation structure, relaxing the assumption of equal covariance between any pair of measurement occasions of the random intercept model. The correlation between responses is assumed to depend on the timing of each response and is expected to decrease as the time lag between observations increases. The random slope model allows the decomposition of the impact of economic insecurity on life satisfaction into a fixed component (the same for all individuals) and an individual-specific random component.

## 3. Data

LGCs were estimated on longitudinal data from the German Socio-Economic Panel (SOEP). Established in 1984, the SOEP has been running for almost 40 years. About 15,000 households and 30,000 individuals are currently part of the SOEP survey. The SOEP collects information from a representative sample of the German residential population aged 17 years and older, by means of questions of both objective (socio-demographic) and subjective (satisfaction, perceptions, attitudes, concerns) nature.

In this analysis, we consider a panel dataset of individuals aged from 16 to 64. The outcome variable is the current level of satisfaction with life, self-reported by respondents every year on a Likert scale going from 1 (low) to 10 (high).

All available waves until 2017 were used to build the dataset. We dropped the initial sample, interviewed in 1984. Then, as the economic insecurity index is computed over a five-year time span, we calculated the first value of the index for 1989, based on data from 1985-1989. Therefore, we finally consider complete data for twenty-nine SOEP waves (1989-2017).

Considering the observations with available data on economic insecurity and life satisfaction, we perform our analysis on a dataset of 195,004 observations from a sample of 31,496 individuals over a time span of 29 years.

Life satisfaction for the full estimation sample is shown in Table 1. Among the observations collected over time, about two thirds fall between the sixth and the eighth level on the life satisfaction scale, while the rest is equally distributed between levels 1 to 5 (16%) and 9 to 10 (16%).

**Table 1 Distribution of life satisfaction over the considered sample (195,004 observations, 31,496 individuals, 29 years)**

Current Life Satisfaction	Frequency	Percent	Cumulative
Level 1	502	0.26	0.26
Level 2	1,650	0.85	1.10
Level 3	3,951	2.03	3.13
Level 4	5,784	2.97	6.10
Level 5	18,708	9.59	15.69
Level 6	21,125	10.83	26.52
Level 7	47,158	24.18	50.71
Level 8	64,884	33.27	83.98
Level 9	24,211	12.42	96.39
Level 10	7,031	3.61	100.00
Total	195,004	100.00	

#### 4. Results and conclusions

Economic insecurity was investigated to assess its impact on life satisfaction trajectories over a 29-year time span among working-age German citizens, grouped by activity sector (secondary vs. tertiary) and age (16-29, 30-39, 40-49, 50-64).

**Table 2 Random slope model estimates – Secondary sector by age**

Parameter	16-29	30-39	40-49	50-64
	Estimate (Std. Err.)	Estimate (Std. Err.)	Estimate (Std. Err.)	Estimate (Std. Err.)
Intercept ( $b_0$ )	7.3477*** (.0204)	7.2048*** (.0200)	7.0737*** (.0197)	7.0148*** (.0218)
Average slope ( $b_1$ )	-0.0633*** (.0196)	-0.0815*** (.0191)	-0.1310*** (.0192)	-0.0898*** (.0118)
Between-individual intercept variance ( $\sigma_{u0}^2$ )	1.0142 (.0379)	1.2111 (.0366)	1.4221 (.0386)	1.5161 (.0438)
Between-individual slope variance ( $\sigma_{u1}^2$ )	0.0407 (.0183)	0.0496 (.0196)	0.0812 (.0243)	0.0047 (.0041)
Between-individual intercept-slope covariance ( $\sigma_{u01}$ )	-0.0092 (.0232)	0.0842 (.0232)	1.1015 (.0252)	0.0601 (.0159)
Within-individual variance ( $\sigma_e^2$ )	1.3864 (.0229)	1.1127 (.0153)	1.1438 (.0142)	1.2200 (.0146)
<i>Observations</i>	11921	15761	19367	18329
<i>Groups</i>	4024	4196	4850	4129
<i>Log-likelihood</i>	-21002.92	-26328.45	-32781.11	-31277.19

Note: \*\*\* stands for  $p$ -value < 0.01.

**Table 3 Random slope model estimates – Tertiary sector by age**

Parameter	Age 16-29	Age 30-39	Age 40-49	Age 50-64
	Estimate (Std. Err.)	Estimate (Std. Err.)	Estimate (Std. Err.)	Estimate (Std. Err.)
Intercept ( $b_0$ )	7.3259*** (.0153)	7.2362*** (.0144)	7.1194*** (.0144)	7.0847*** (.0155)
Average slope ( $b_1$ )	-0.0313*** (.0095)	-0.0796*** (.0110)	-0.0956*** (.0090)	-0.0607*** (.0080)
Between-individual intercept variance ( $\sigma_{u_0}^2$ )	1.0760 (.0287)	1.1828 (.0266)	1.5228 (.0287)	1.6021 (.0312)
Between-individual slope variance ( $\sigma_{u_1}^2$ )	0.0004 (.0001)	0.0226 (.0076)	0.0148 (.0062)	0.0171 (.0055)
Between-individual intercept-slope covariance ( $\sigma_{u_01}$ )	0.0204 (.0011)	0.0812 (.0141)	0.0864 (.0127)	0.0630 (.0114)
Within-individual variance ( $\sigma_e^2$ )	1.3316 (.0155)	1.1621 (.0115)	1.1330 (.0097)	1.1501 (.0097)
<i>Observations</i>	22202	29375	38517	39532
<i>Groups</i>	7327	7965	9509	8476
<i>Log-likelihood</i>	-38752.09	-49540.71	-65084.15	-66590.55

Note: \*\*\* stands for  $p$ -value < 0.01.

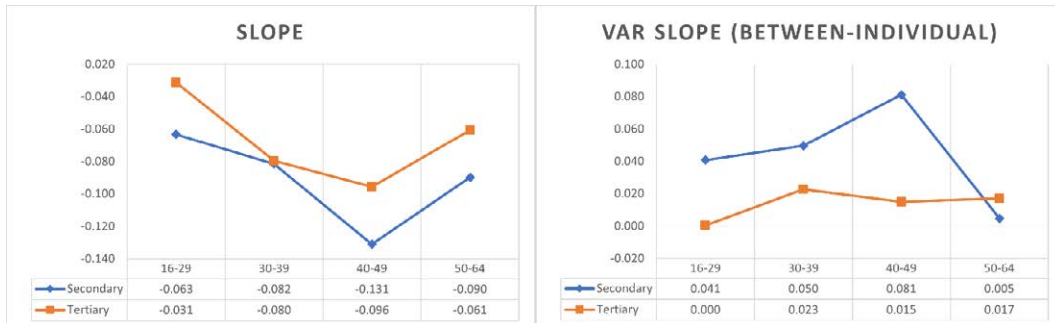
The results of our models are presented in Table 2 (for the secondary sector) and Table 3 (for the tertiary sector).

Random slope growth models allow us to adequately capture individual variation in over-time trajectories. As, in our case, the average slopes are negative ( $\beta_1 < 0$ ), the positive intercept-slope covariance ( $\sigma_{u_01} > 0$ ) shows that individuals with above-average intercepts ( $u_{0j} > 0$ ) tend to have flatter-than-average slopes ( $u_{1j} < 0$ ). Similarly, individuals with below-average intercepts ( $u_{0j} < 0$ ) tend to have steeper-than-average slopes ( $u_{1j} > 0$ ).

We describe the four main components of the random slope models graphically in Figs. 1-4. Each graph shows, separately for the two activity sectors, the values for the four age groups. The blue line depicts the secondary sector, while the orange line represents the tertiary sector; the four points on the x-axis represent the age groups (16-29; 30-39; 40-49; 50-64).

Figure 1 (left side) shows the average slopes. For each group of workers, there is a significant negative relationship between economic insecurity and life satisfaction; that is, a higher level of economic insecurity leads people to a lower level of life satisfaction, regardless of age and activity sector. In particular, for workers in both activity sectors, the negative impact of economic insecurity on life satisfaction is stronger for mid-career workers (40-49 age group) and less relevant for younger workers (16-29). The negative impact of economic insecurity on life satisfaction is consistently stronger for workers in the secondary sector than for those in the tertiary sector, except for workers in the 30-39 age group, for whom this impact is not significantly different between the two sectors.

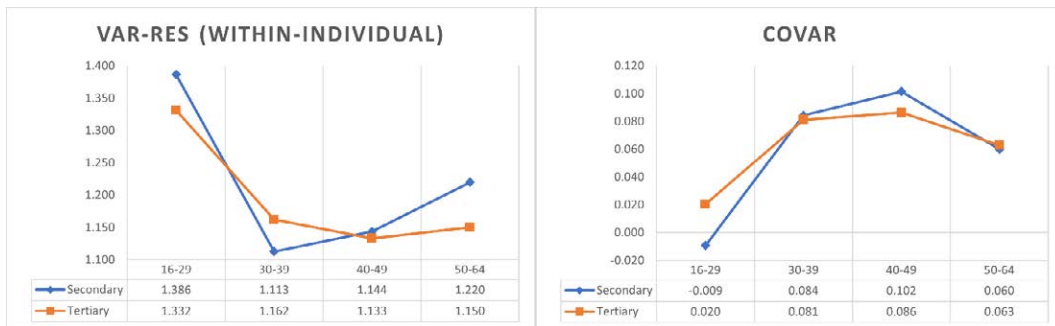
Figure 1 (right side) shows the between-individual slope variance, estimated individually for each worker in the sample, which illustrates the variability of the random component of the growth rate. The between-individual slope variance is higher in the secondary sector for the first three age groups, while it is higher in the tertiary sector for workers aged 50 and over. With reference to the 40-49 age group, the differences between workers in the two sectors, which already appeared quite large when considering the fixed component of the model, appear even larger when also considering the random component.



**Figure 1 Average slope (left) and between-individual slope variance (right), by activity sector and age group**

Figure 2 (left side) shows that the within-individual variance does not show large differences between workers in the two sectors. Greater variability is observed for the youngest class of workers (16-29). Therefore, within this age group, the impact of economic insecurity on life satisfaction has a greater variability over time; that is, economic insecurity affects young workers' life satisfaction in a more volatile way, meaning that the perception on satisfaction with life is less stable over time at young ages.

The between-individual intercept-slope covariance (Figure 2, right side) is generally positive and increasing up to the 40-49 age group. This means that workers who show an above-average level of life satisfaction at baseline also tend to show a below-average decline in their level of life satisfaction over time. By contrast, workers with a below-average level of life satisfaction at baseline tend to show an above-average decline in their level of life satisfaction. This trend is particularly relevant for mid-career workers (age group 40-49), especially for those employed in the secondary sector. For the age groups 30-39 and 50-64, there are no significant differences between workers in the two activity sectors.



**Figure 2 Within-individual variance (left) and between-individual intercept-slope covariance (right), by activity sector and age group**

In brief, the results show that economic insecurity has a consistent negative impact on life satisfaction, especially for mid-career people and for employees in the secondary sector. The higher within-individual over-time variability shows that economic insecurity affects life satisfaction more unpredictably for the youngest workers. These and other relevant differences between the considered groups give room for the implementation of policy measures aimed at reducing economic insecurity with a view to enhance individuals' satisfaction with life, specifically targeted on the different life stages and activity sectors.



## References

- Bossert, W., Clark, A., D'Ambrosio, C., Lepinteur, A. (2019). Economic Insecurity and the Rise of the Right. *CEP Discussion Papers* **1659**.
- Bossert, W., D'Ambrosio, C. (2016). Economic insecurity and variations in resources. *ECINEQ Working Paper* **422**.
- Colantone, I., Stanig, P. (2018). The trade origins of economic nationalism: Import competition and voting behavior in Western Europe. *American Journal of Political Science*, **62**(4), pp. 936-953.
- D'Ambrosio, C., Rohde, N. (2014). The Distribution of Economic Insecurity: Italy and the U.S. over the Great Recession. *Review of Income and Wealth*, **60**, pp. S33-S52.
- Giambona, F., Grassini, L., Vignoli, D. (2022). Detecting economic insecurity in Italy: a latent transition modelling approach. *Statistical Methods & Applications*, **31**, pp. 815-846.
- Guiso, L., Herrera, H., Morelli, M., Sonno, T. (2017). *Demand and Supply of Populism*. London, UK: Centre for Economic Policy Research.
- Hall, J., Giovannini, E., Morrone, A., Ranuzzi, G. (2010). A Framework to Measure the Progress of Societies. *OECD Statistics Working Papers* **2010/05**.
- Hasan, H. (2019). Confidence in Subjective Evaluation of Human Well-Being in Sen's Capabilities Perspective. *Journal of Happiness Studies*, **20**, pp. 1-17.
- Linz, S.J., Semykina, A. (2010). Perceptions of economic insecurity: Evidence from Russia. *Economic Systems*, **34**(4), pp. 357-385.
- Panarello, D. (2021). Economic insecurity, conservatism, and the crisis of environmentalism: 30 years of evidence. *Socio-Economic Planning Sciences*, **73**: 100925.
- Rohde, N., Tang, K.K., Osberg, L., Rao, P. (2016). The effect of economic insecurity on mental health: Recent evidence from Australian panel data. *Social Science & Medicine*, **151**, pp. 250-258.
- Smith, T.G., Stillman, S., Craig, S. (2013). The U.S. Obesity Epidemic: New Evidence from the Economic Security Index. *Annual Meeting of the Agricultural and Applied Economics Association Paper* **151419**.
- Veenhoven, R. (2017). Greater happiness for a greater number: Did the promise of enlightenment come true?. *Social Indicators Research*, **130**(1), pp. 9-25.

# Cultural and sensorial correlates of Trebbiano wine consumption

Luigi Fabbris, Alfonso Piscitelli

## 1. Introduction

The Trebbiano from Abruzzo is a variety of white grapevine cultivated in the Abruzzo region, Italy. The Trebbiano grapevine is also called ‘white Bombino’ or ‘Tuscan Trebbiano’ and is cultivated all over central and southern Italy, in the Emilia Romagna region and here and there in other northern Italy regions. The variety cultivated in Abruzzo and respecting the production protocol can be named Trebbiano from Abruzzo – quality assurance label (*Trebbiano d’Abruzzo DOC*).

Varieties of Trebbiano grapevine are cultivated in Italy since at least two millennia. At the time of ancient Romans, Pliny the Elder in his first century *Naturalis Historia* mentions a ‘vinum trebulanum’ whose name was associated to the word ‘trebula’, say farmhouse. This may highlight the large diffusion of this vine because its primary use was, particularly at those times, for home consumption of farmers. The semantics of its name, as some scholars object (Bacci, 1596, quoted in Labra et al., 2001; Hohnerlein-Buchinger, 1996) could differ. Also, the varieties of current Trebbiano wine do not show a common ancestor (see the DNA analysis of various Trebbiano-like strains in Labra et al., 2001). As a matter of fact, this grapevine has found on the Abruzzo hills such an ideal soil and climate to gain a foothold that the Abruzzo Trebbiano vine (from now on, Trebbiano) could now be considered an autochthon variety.

In this paper, we analyse the preference for Trebbiano wine by means of a sample of Italian consumers involved in an experimental wine tasting experience. Due to the small sample at hand, we keep our analytical model simple and assume cultural and sensorial characteristics of Italian consumers as possible predictors of preference of Trebbiano to other wines. In this way, we highlight the characteristics of the social groups who are favourable to its consumption and, on the contrary, those who dislike it, so to be able to campaign for a larger consumption of Trebbiano from Abruzzo.

The rest of the paper is organised as follows: Section 2 introduces the available data, the wine tasting experience that led to the data collection and the model for data analysis. Then, Section 3 presents the main results of the statistical analysis of the collected data. Finally, Section 4 discusses the results with reference to the mainstream literature on wine preference analysis.

## 2. Data and methods

### 2.1. The tasting experience

In September 2018, a sensory evaluation experiment was conducted on 12 white wines originating from four grape varieties (*Trebbiano d’Abruzzo*, *Pecorino d’Abruzzo*, *Passerina d’Abruzzo*, *Verdicchio dei Castelli di Iesi*). In the sensory experiment the *Trebbiano d’Abruzzo* was also blindly served as it were two different grape varieties (*Vino bianco DOC*; *Vino bianco da pasto*). Overall, in the experiment there were six grape varieties, and two different cellars were included for each grape. The tasting experiment consisted in evaluating the preferences for aspects of a set of four wines administered to the assessors according to a randomised fractional factorial design; only the name of the wines to evaluate was made explicit to the

Luigi Fabbris, Tolomeo studi e ricerche, Padua and Treviso, Italy, fabbris@stat.unipd.it, 0000-0001-8657-8361  
Alfonso Piscitelli, University of Naples Federico II, Italy, alfonso.piscitelli@unina.it, 0000-0001-6638-2759

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Luigi Fabbris, Alfonso Piscitelli, *Cultural and sensorial correlates of Trebbiano wine consumption*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.12, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 65-70, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

assessors. Though, in order to avoid complaisance towards the Trebbiano, also either of two anonymous wines (*Vino bianco DOC*; *Vino bianco da pasto*) –which were actually Trebbiano– was juxtaposed at each tasting trial to Trebbiano.

The pool of tasters included 48 individuals, of whom 30 typically consumed mild amounts of wine (mild consumers), and 18 were professional sommeliers belonging to the AIS-Abruzzo association. Both mild consumers and sommeliers were selected on the basis of their consensus to the experiment as well as their experience in wine consumption.

The wine characteristics considered in this experiment were selected through an anonymous paper questionnaire. This questionnaire asked participants to make judgements on 11 intrinsic attributes of appearance and an overall judgement of each tasted wine. Attributes were rated on a 10-point Likert scale from ‘Min Preference’ (1) to ‘Max Preference’ (10). The questionnaire also gathered data regarding the tasters’ background characteristics, their drinking habits, and the relevance of wine in their diet and social life.

Since it was deemed practical to serve only four out of the six possible varieties to each taster, the actual subset of wine varieties to be administered to each assessor was defined according to a fractional design with main factor grape-variety.

Therefore, four glasses were served in randomised order to each taster, and for each of the proposed varieties one of the two potential cellars was randomly selected. The wines were poured in a flight, and taster were supplied with a glass of water too. In the tasting session, the judges received six centilitres of each of the four randomly selected wine varieties, which were served at the same cold temperature. The protocol envisaged that tasters could taste and re-taste before concluding preferential judgements, and they would evaluate the intrinsic attributes of each tasted wine.

## 2.2 The analytical model

The model for data analysis of responses collected about Trebbiano includes the frequency of consumption of Trebbiano as a criterion variable,  $Y$ , a first regressor,  $X_1$ , describing the role of Trebbiano wine in an everyday outdoor dinner, and a selection of other  $J-1$  significant regressors, so that  $\mathbf{X} \equiv (X_1, X_2, \dots, X_J)$ . The relationship may be written as

$$Y = f(X_1, X_2, \dots, X_J).$$

The  $Y$  variable, measured on an ordinal *scale*, was dichotomised as follows:  $Y = 1$  if the respondent used to drink Trebbiano wine often or occasionally, and  $Y = 0$  if the respondent consumed it rarely or never.

The logistic regression model is written as follows (Hosmer and Lemeshow, 2000):

$$\text{logit} [p(Y=1)] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_J,$$

where  $\text{logit}(p) = \ln[p/(1-p)]$ , and  $\beta_i$  ( $i=0, 1, \dots, J$ ) measures the relation between  $Y$  and  $X_i$  when all other variables in the model remain fixed. Regressor  $X_1$  was forced into the model, while  $X_i$  ( $i = 2, \dots, J$ ) was selected in a stepwise fashion, block after block, according to its significance ( $< 0.10$ ). The goodness of fit is measured through the Nagelkerke pseudo- $R^2$  index.

The possible regressors were examined in blocks: firstly, the selection concerned the descriptors of consumers’ wine expertise and Trebbiano wine evaluation and, finally, the set of variables describing the personal and social aspects that, either in a positive or negative direction, may influence wine consumption. The characteristics of the assessed Trebbiano wines enter this analysis as distributional parameters (mean and absolute deviation) of the scores which single assessors assigned to the tasted wines.

Such a model identifies a relational scheme *à la Ajzen* (Fishbein and Ajzen, 1975; Ajzen, 1991), in which blocks of predictors positively and negatively correlated to the response both concur to the statistical fit of the propensity to consume the topical wine and then to its consumption in reality. Statistical analysis was performed using the SPSS package (IBM, 2020).

### 3. Results

From responses to the questionnaire, it resulted that Trebbiano wine was regularly consumed by the majority of the involved assessors: 29.2% consumed it often and 37.5% occasionally in a regular meal, while another 18.7% declared drinking it rarely, and only 14.6% never. Overall, our sample included a group of experts and a group of nonexperts. Of the 48 assessors, five (10.4%) considered themselves to be wine experts, and eight (16.7%) stated that they were able to recognise some wines but did not consider themselves to be wine experts. The majority of the 12 participating sommeliers classified themselves in the latter category. A large share of assessors (47.9%) indicated that they possessed sufficient knowledge of wine to adequately understand its quality. Finally, 25% of the assessors admitted that they knew little or very little about wine. Regarding wine practice, about 56% of assessors had been consuming wine for decades, usually with dinner. Several assessors (54.2%) had attended a wine-tasting session coordinated by a sommelier.

Both experts and nonexperts perceive that commonly people associate the label Trebbiano with low quality wines: 47.8% of assessors perceive that a general consumer evaluates it as a mediocre wine, 34.8% as just fair and only 17.4% as a fine quality wine. Instead, the evaluation of Trebbiano at the tasting experiment was generally positive and, in any case, better rated than the other labelled wines: indeed, the mean of the two tastes of Trebbiano obtained a mean evaluation of 6.48 (out of 10) and that of the tastes in which the Trebbiano label was evident of 6.91, against an overall mean of the four tasted wines of 6.72.

It is evident that the label of the tasted wine somewhat influenced the assessors: if the label of the Trebbiano was in fact ‘white wine’ (n = 32), the mean evaluation was 6.34 and in case the label was ‘white wine suited for meals’ (n = 32) it was 6.19. The difference between the Trebbiano-labelled wine and the generally ‘white wine’ labelled one was 0.65 (out of 10), which is statistically significant at 10% level only. This induces to conjecture the presence of a mild complaisance effect among the tasters who better scored the wines labelled Trebbiano than those which, actually being Trebbiano, were labelled in a more general, less inviting way.

Table 1 summarises the results of two applications of the regression analysis: Model 1 (**M1**, referred by columns 2 and 3) concern the analysis in which  $X_1$  was forced as a regressor and Model 2 (**M2**, referred by columns 4 and 5) without any forced variable. The fairly significant statistical fit of the Trebbiano consumption propensity (pseudo- $R^2 = 37.3\%$  for Model 1 and 30.1% for Model 2) supports the following claims:

**Table 1.** Parameter estimates  $\hat{\beta}$  of two regression models with Trebbiano wine consumption as criterion variable (forward stepwise selection of regressors, n = 48; \*\* < 0.01; \* < 0.05; ° < 0.10; NS= Not significant).

<i>Regressor</i>	$\hat{\beta}$ (M1)	Significance (M1)	$\hat{\beta}$ (M2)	Significance (M2)
Intercept	-6.936	*	-5.205	*
Role of Trebbiano wine in an outdoor dinner	0.452	°	=	=
Self-perceived expertise	(a)		(a)	
Average evaluation of two Trebbiano tastes	0.515	NS	0.684	°
Absolute deviation between Trebbiano tastes	1.055	°	1.127	°
Drinking wine regularly	1.590	*	1.234	°
<i>Nagelkerke Pseudo R<sup>2</sup></i>	0.373		0.301	

(a) Variable initially selected and then ejected because of its correlation with other significant predictors.

- The applied models show a fair fit of the criterion variable: the models involving only the significant regressors returned pseudo- $R^2$  values equal to 37.3% (**M1**) and 30.1% (**M2**). These quotas are encouraging (Smith at al., 2021), although there may be other socio-economic and contextual variables that, in combination with those considered in this application, could improve the fit.

- A model *à la Ajzen* —implying that the variable anticipating the behaviour, which in our case is the perception that a Trebbiano wine label may influence the wine choice in an outdoor dinner, the behaviour being the consumption of that wine— is supported by the data at hand. In other words, it may be conjectured a causal chain starting from a bipartite set of regressors (one part being the personal and social resources and the other being the personal and social problems related to the consumption of that typical wine) and crossing the competence of assessors as wine consumers which may influence the assessors' disposition towards that wine, and, finally, its regular consumption.
- Model 2 includes three regressors: 1) the mean evaluation of the two tasted Trebbiano wines, 2) a measure of evaluation variability between the two tasted wines, and 3) the habit of drinking wine at meals. Model 1 involves again three regressor, but the mean evaluation of the two tasted Trebbiano wines is substituted with  $R_0$ , the perception that Trebbiano can be considered a good wine for an outdoor meal. The measure of variability between the two tastes of Trebbiano —a low or null variability measuring the ability to similarly evaluate two wines from the same bottle but served as different— can be considered an indirect measure of tasting ability of the involved assessors, in the sense that the smaller the variability, the higher the ability of an assessor to discern wine quality. The relationship with Trebbiano consumption being positive, the presence of this predictor in both models highlights that Trebbiano, *ceteris paribus*, that in this case means for a given evaluation score, was more consumed by qualified tasters.
- Also, it may be noticed that positive evaluations of the tasted Trebbiano negatively correlate with the measure of variability between evaluations ( $r = -0.466$ ;  $p < 0.01$ ). This means that Trebbiano wine is more appreciated by more expert assessors and this trend parallels real consumption. This may pinpoint the idea of Trebbiano as a specialist's wine.
- Another regressor enters both analytical models: the regular consumption of wine at meals. This variable shows significantly higher scores in the evaluation of Trebbiano ( $r = 0.328$ ;  $p < 0.05$ ) but not with the variability index ( $r = -0.084$ ;  $p > 0.10$ ). This may mean that regularly assuming wine at meals does create a feeling with Trebbiano wine but not necessarily is related to a technical expertise concerning wine tasting.
- Finally, let us consider the (positive) disposition towards Trebbiano consumption. This variable brings forward the actual consumption. In fact, there is a significant correlation between disposition towards and consumption of Trebbiano:  $r = 0.288$  ( $p = 0.047$ ). Though, the disposition is uncorrelated with all other possible regressor. In the multiple regression analysis, the disposition towards Trebbiano wine enters the model as an alternative to a positive evaluation of Trebbiano tastes. The two concepts are so tightly related to allow stating that the disposition towards the consumption of Trebbiano represents the assessors' dimension of their own wine culture.
- A regressor that does not show up in the final model is the self-perceived expertise. Indeed, it was selected at a certain step of the analysis but was rejected after the selection of the three described regressors. This means that the assessors' expertise correlates with the custom to drink wine at meals ( $r = 0.348$ ;  $p < 0.05$ ) and with sensorial skills that develop in people for whom wine is a part of the daily diet.

Crossing some characteristics of the assessors with their wine consumption habits and the general opinion about Trebbiano quality we obtain Table 2. The results summarised in table help understanding some apparent inconsistencies in our data. Indeed, there is a gap between the regular consumption of Trebbiano by the people involved in our tasting experiment and the reputation for that wine perceived by the assessors in the public opinion.

Trebbiano wine is present with a certain regularity on the dining tables of two thirds of the involved consumers. Also, the people self-rating as wine experts, the regional sommeliers and the regular consumers of wine at meals consume it at a rate above 80% and are prepared to

suggest Trebbiano as a wine alternative to match food at an outdoor dinner. No doubt that those who better know it, have a superior opinion of Trebbiano. The same categories showing the more positive opinions about Trebbiano perceive that the general public basically reputes Trebbiano a mediocre wine: 61% of sommeliers and 56% of those who drink wine at meals believe that ‘the others’ consider Trebbiano as a mediocre, ordinary wine. These percentages are higher than the average computed over all assessors (50%). Definitely, we can state that there is a large perception divide between the more expert assessors and the general public as Trebbiano reputation is concerned.

Nevertheless, it may be that the more expert assessors were influenced by a sort of complaisance towards that wine. In order to check for complaisance, we can evaluate two survey results: 1) the difference between the judgement of Trebbiano wine when its name was printed on the paper place mat each assessor had in front when tasting, and that when Trebbiano was served with a general label, for instance, ‘white wine’; and 2) the variability index of the two Trebbiano tastes (one explicit, the other hidden) each assessor was required to do.

The data show that regular consumers of Trebbiano assigned high scores to the Trebbiano whose name was explicitly indicated and even higher to the Trebbiano served under a general label. This may mean that people accustomed to drink Trebbiano at meals recognised and appreciated in the tasted wines the same qualities they appreciated when matching wine with food at home. In some sense, regular consumers of the topical wine expect to feel in their nose and palate sensations they feel when they drink wines at meals.

**Table 2.** Percentage proportion of Trebbiano consumers and of the perceived public opinion of Trebbiano wine, by contextual and assessors’ characteristics (n=48)

<i>Characteristics</i>	% Trebbiano consumers	Public opinion of Trebbiano wine		
		Mediocre	Just fair	Fine
Self-perceived as an expert (n=13)	92.3	46.2	30.7	23.1
Being a sommelier (n=18)	83.3	61.1	27.8	11.1
Drinking wine regularly (n=27)	81.5	55.6	37.0	7.4
Interviewee always deals with wines at home (n=27)	81.5	44.4	44.4	11.2
Olfactory skills (7/10 on; n=27)	74.1	55.6	33.3	11.1
Aroma tasting skills (7/10 on; n=29)	72.4	55.2	31.0	13.8
Role of Trebbiano in an outdoor dinner (7/10 on; n=22)	81.8	40.9	22.7	34.4
Below average evaluation of 2 Trebbiano tastes (n=20)	60.0	60.0	25.0	15.0
Above average evaluation of 2 Trebbiano tastes (n=28)	71.4	42.9	39.2	17.9
Below average variability of 2 Trebbiano tastes (n=26)	57.7	57.7	34.6	7.7
Above average variability of 2 Trebbiano tastes (n=22)	77.3	40.9	31.8	27.3
Below average evaluation of explicit Trebbiano (n=10)	40.0	50.0	40.0	10.0
Above average evaluation of explicit Trebbiano (n=22)	68.2	54.5	31.8	13.7
Below average evaluation of hidden Trebbiano (n=33)	57.6	54.5	24.2	21.3
Above average evaluation of hidden Trebbiano (n=31)	83.9	42.0	42.0	16.0
Smoker (n=14)	57.1	50.0	14.3	35.7
Gender (male; n=24)	62.5	33.4	45.8	20.8
Age 45 year-old and more (n=29)	72.4	44.8	37.9	17.3
<b>Total (n=48)</b>	<b>66.7</b>	<b>50.0</b>	<b>33.3</b>	<b>16.7</b>

Moreover, both the mean scores and the between-score variability of the two tastes of Trebbiano —one explicit, the other masked— were much higher among those for which Trebbiano is part of their diet. Both the assessors who gave a more positive evaluation of the tastes and those whose scores less differed perceived a scanty public reputation of Trebbiano. So, indirectly, the assessors who gave a better evaluation of Trebbiano’s qualities and possessed a superior capacity of discerning wines are among those who believe that public opinion is not inclined towards Trebbiano.

#### 4. Discussion and conclusion

This work was aimed to detect the characteristics of the Trebbiano wine consumers as stemming from the data collected through a tasting experiment on white wines from the Abruzzo region. The experiment was designed also to measure the possible complaisance that may affect the tasters' judgements. Our analysis illustrates that Trebbiano was judged as a good quality wine by the large majority of assessors and, in particular, by people who, for professional or dietary reasons, know it better. Thus, sommeliers and other experts knowledgeable of wines, after the tasting, scored Trebbiano in a very satisfactory way. A level of satisfaction that leads to its regular consumption both at home and outdoor meals.

We could summarise our results by stating that knowledgeable people evaluate Trebbiano as palatable as more renowned wines, despite its large consumption. Experts judged positively its intrinsic qualities and juxtaposed their judgements to that of the general public, who—according to them—associate the topical wine with the plethora of ordinary quality, even mediocre, wines. This contrast highlights the strength of experts' judgement in favour of Trebbiano: we (those who know) consider it a good wine, the others (the uninformed) consider it as ordinary, too diffused to be good. Now, we should define a good wine. We could relate a wine goodness to how its sensorial properties cross its relevance in a healthy diet. Wine experts, indeed, distinguish between a wine whose qualities are so peculiar (and non-disagreeable) to make it a wine with an own personality, one that other experts would similarly suggest in a particular occasion, and a wine that is so palatable that they themselves would drink it safely every day. Though, this issue would lead us far from our research questions and we leave it.

The feeling with Trebbiano shown by our experts went even further. Some of them instinctively expressed judgements on its qualities that went beyond their favourable position, adding complaisance in cases the wine they tasted was explicitly labelled as Trebbiano. In fact, comparing these judgements with those given when Trebbiano was instead administered as a 'white wine' or 'wine suited for meals' their judgements were rather different. This may be interpreted as such a biased disposition of the regional experts to Trebbiano to even bias their judgements in case they are called to evaluate it.

#### Acknowledgments

*The authors wish to thank Manuela Cornelli, chair of AIS-Abruzzo, the sommelier association of the Abruzzo region, for her help in survey design and data collection.*

#### References

- Ajzen, I. (1991). Theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50: 179-211.
- Bacci, A. (1596). *De naturali vinorum historia de vinis Italiae et de convivii antiquorum. Libri septem*. Roma ex officina Nicholai Mutiis.
- Fishbein, M., Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA.: Addison-Wesley.
- Hohnerlein-Buchinger, T. (1996). *Per un sublessico vitivinicolo: La storia materiale e linguistica di alcuni nomi di viti e vini italiani*. Tübingen: Max Niemeyer Verlag.
- Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression (2<sup>nd</sup> ed.)*. New York: John Wiley & Sons.
- IBM Corp. Released (2020). *IBM SPSS Statistics Version 27.0*. Armonk, NY: IBM Corporation.
- Labra, M., Winfield, M., Ghiani, A., Grassi, F., Sala, A., Scienza, A., Failla, O. (2001). Genetic studies on Trebbiano and morphologically related varieties by SSR and AFLP markers. *Vitis*, 40(4): 187-190.
- Smith, T.J., Walker, D., and McKenna, C.M. (2021). A coefficient of discrimination for use with nominal and ordinal regression models. *Journal of Applied Statistics*, 48(16), 3208–3219.

# Tourism and territorial economy: beyond satellite accounting

Fabrizio Antolini, Antonio Giusti, Francesca Petrei

## 1. Tourism and its representation through statistics

The pressing and increasingly urgent demand by policy makers, researchers and stakeholders for increasingly detailed and timely tourism statistics stems from the need to measure the economic impact on the one hand and the sustainability on the other of a sector that is considered to show resilience and adaptability, even in rapidly changing contexts, and poses a considerable challenge to producers of official statistics at international level.

The current European Regulation of 2011 (692/2011), which defines the reference areas and purposes of tourism statistics at European level, prescribes neither sustainability indicators nor economic and monetary indicators, despite the fact that both the previous directive (95/57 EC) and the current regulation have always considered tourism as a fundamental tool for the economic development of territories: *Tourism plays an important role in the EU because of its economic and employment potential, as well as its social and environmental implications. Tourism statistics are not only used to monitor the EU's tourism policies but also its regional and sustainable development policies*" (Eurostat, 2021).

Thus, although it seems to be well established that the transition towards a sustainable development of the territories is now indispensable and that certain phenomena linked to pollution and climate change could represent an obstacle to the growth of some tourist destinations, we are still far from having a shared and homogeneous definition of sustainable tourism and the carrying capacity indicators used do not seem able to represent exhaustively such a complex and multidimensional phenomenon (European Commission, 2004).

Furthermore, the elaboration of satellite accounts on tourism - even in their possible integration with the environmental module - continues to be a mere voluntary exercise for member countries, even though they are specifically provided for by the European System of National and Regional Accounts (SEC).

Finally, the need for timely statistics that also describe people's movements within the territory would require broadening the profile of their relevance by including the use of big data in the system of tourism statistics *"the arrival of big data is also changing the working environment for statisticians. Many sources of big data measure flows or transactions. Tourism statistics try to capture physical flows of people — as well as the accompanying monetary flows; big data provides promising new sources of data and previously unavailable indicators to measure these flows (and stocks)"* (Eurostat, 2017), but to date the first attempts in this respect are still experimental and at a very early stage.

As far as European tourism statistics are concerned, the first report by the European Commission was made in 2016, but it is only in the second one, in 2022, that there is talk of a possible revision of the Regulation, with additions towards a requirement for satellite accounting and sustainability indicators (European Commission, 2022).

In this paper, after an examination of the current state of Italian and European public statistics to (section 2), we make some attempts to arrive at a more comprehensive information picture regarding the contribution of tourism to regional added value (satellite accounting). The

---

<sup>1</sup> The views expressed in this paper are solely those of the author and do not involve the responsibility of Istat.



experimental verification was conducted in section 2.1 regarding the demand side using tourism density as a regional attractor and in section 3 regarding Value Added. The impossibility of having direct access to Istat Territorial Frame SBS (Structural business statistics) micro-data<sup>2</sup> places unavoidable limits on the estimation carried out. On the other hand, the objective of this work is to make it clear how important and urgent it is to have a measure of the economic contribution of tourism to the growth of territories.

## **2. How to implement tourism statistics: the possible role played by satellite accounts**

In 2010, satellite accounts on tourism (TSA) were compiled for the first time in twenty-three countries (Eurostat, 2009). This was then done every three years: in 2013, twenty-two countries participated in the compilation and in 2016, nineteen. Compared to the originally planned indications the greatest critical issue has always been the homogeneity and comparability of the data contained in the TSA produced in each country. The same indicators contained in the ten tables of the theoretical scheme have, depending on the country, different coverage of the required indicators. The only table that is compiled (T5) with full coverage with respect to the required indicators is the one concerning the "*production accounts of tourism industries and other industries*", which is also the only one compiled by all countries.

The table of "*tourism collective consumption*" (T9) is a relevant part of the transition from the aggregate of tourism expenditure to the broader aggregate of tourism consumption. The employment statistics themselves, which refer to jobs, are incomplete, being compiled by only thirteen countries.

With reference to the table of "*production accounts of tourism industries and other industries*" (T5) and "*Total domestic supply and internal tourism consumption*" (T6), a further difference concerns the statistical sources used, since only some countries use business statistics. The different use of the sources implies a different methodology used in the determination of the relevant aggregates<sup>3</sup>.

To date, therefore, not all member countries compile satellite accounts and those compiled often do not refer to the same time period or have discrepancies in the methodologies used or the data sources, making international comparability practically impossible.

On the other hand, the indicators that were more difficult to compile were "*Tourism gross fixed capital formation*" (T8) and the "*Tourism collective consumption*" table (T9), with Spain being the only country to compile both tables.

### **2.1 The importance of satellite territorial accounting**

A further weakness concerns the fact that while the successful introduction of satellite accounts in the regulation will be continued, thanks to which the current impasse can be overcome, there is no mention in the European Commission's report of the need to territorialise the satellite accounts on tourism. On the other hand, this is a fundamentally important aspect because tourism is a purely territorial phenomenon, since it is specifically linked to the specific characteristics and distinctive features of a specific place (Benassi et al., 2021), as well as being recognised as an important driver of local development. In Italy, for example, the differences at territorial level of the tourism are considerable and show a certain concentration of the occupancy in some specific areas; the tourist density (see table 1) is very different at regional level and even more so at municipal level, making an in-depth analysis at territorial level necessary, which would also need economic data that are currently lacking.

The satellite accounting tool would in fact be useful for understanding the economic effects of

---

<sup>2</sup> Istat: <https://www.istat.it/it/archivio/267573>

<sup>3</sup> In Italy, the main surveys involved in the preparation of the Satellite Accounts carried out by Istat are the survey on '*Occupancy of tourist accommodation establishments*'; the survey on '*Expenditure by Italian households*' (Tourism trips), the survey on '*International Tourism*' by the Bank of Italy.

policies implemented at local level, analysing the benefits that certain policies have produced on the entire tourism chain. Moreover, satellite accounting on tourism if integrated with environmental satellite accounting would also be a possible tool to have a measure of the anthropic pressure generated by tourism flows and therefore a measure of the sustainability of tourism. And this is because satellite accounting was born from its introduction in the System of National Accounts (SNA 1993) as a scheme flexible to the needs of the country compiling it.

However, the distinction between functional satellite accounts and integrated satellite accounts remains relevant. The former - which include the satellite accounts for tourism, the environment and social protection - are oriented towards the analysis of the economic system, with the aim of making visible flows that are not evident within the national accounts. The latter, on the other hand, defined as integrated or “external” satellite accounts, use alternative concepts and definitions to the national economic accounts and are therefore an extension of the national accounts.

**Table 1** - Tourist density

2019	Nights spent	Population	NS/P	Italy=1
Piedmont	14889951	4328565	3.44	0.47
Aosta Valley/Vallée d'Aoste	3625616	125653	28.85	3.95
Lombardy	40482939	10010833	4.04	0.55
Trentino-Alto Adige/Südtirol	52074506	1074034	48.48	6.64
Veneto	71236630	4884590	14.58	2.00
Friuli-Venezia Giulia	9052850	1210414	7.48	1.02
Liguria	15074888	1532980	9.83	1.35
Emilia-Romagna	40360042	4459453	9.05	1.24
Tuscany	48077301	3701343	12.99	1.78
Umbria	5889224	873744	6.74	0.92
Marche	10370800	1520321	6.82	0.93
Latium	39029255	5773076	6.76	0.93
Abruzzo	6176702	1300645	4.75	0.65
Molise	439645	303790	1.45	0.20
Campania	22013245	5740291	3.83	0.53
Apulia	15441469	3975528	3.88	0.53
Basilicata	2733969	558587	4.89	0.67
Calabria	9509423	1912021	4.97	0.68
Sicily	15114931	4908548	3.08	0.42
Sardinia	15145885	1622257	9.34	1.28
<b>Italy</b>	<b>436739271</b>	<b>59816673</b>	<b>7.30</b>	<b>1.00</b>

*Source: Our processing on ISTAT data*

For the estimation of the regional gross domestic product, the three methods proposed by the national accounts, i.e., production, income, and expenditure, remain relevant, although for income and expenditure at the regional level there are some methodological problems that require the direct use of data from business enterprise accounts. However, in this regard, the statistical archive prepared by the National Institute of Statistics of Italy, FRAME SBS, has considerably changed the availability of statistical information, as data from statistical business surveys have been supplemented with data from tax sources (Antolini and Grassini 2020a). On the other hand, the expenditure method is not considered reliable by ESA "10 due to the lack of statistical information on inter-regional trade and the flow of imports and exports". In the case of tourism, however, international trade is mainly in the credits and debits generated by incoming and outgoing tourist flows, on which expenditure (but not tourist consumption) is recorded monthly, quarterly, and annually by the Bank of Italy.

## 2.2 Demand-side approach to satellite accounting

Tourism is a sector that is defined in relation to the economic activity of visitors making a trip outside their usual environment. For this reason, from an economic point of view it lends itself well to being measured from the demand side (visitor activity). The operational difficulty on the demand

side is the identification of the visitor, which is crucial to have an estimate of tourists and their overnight stays. In the case of Italy, however, overnight stays are recorded both on the demand side (Tourism Trips) and on the supply side (Occupancy of tourist accommodation establishments) *“Provided an estimation of the average expenditure per overnight stay (from demand-side data, all tourism expenses included), the use of supply or demand-side figures leads to different results of the expenditure aggregate. The estimation provided by supply-side data offers indisputable advantages since it allows the production of scalable territorial data”* (Antolini and Grassini, 2020b). As far as visitors are concerned, economic activity is embodied in the expenditures made in preparation for and during the trip. Actually, the demand approach at macroeconomic level should consider the broader aggregate of tourism consumption, which evidently also takes into account the part of collective consumption from which the tourist indirectly benefits anyway.

Finally, an estimate of tourism demand should also be able to consider gross fixed capital formation, but, as illustrated above, both the investment and collective consumption tables are prepared by only a few countries. A further consideration concerns excursionists, whose increasing relevance in terms of flow would require the use of new statistical sources (big data).

Italy is currently using the demand approach, considering overnight stays recorded on the demand side: in 2019 (before the pandemic) total domestic travel was 216.7 million with 703.8 million overnight stays. Following the demand-side approach, in 2019 the Value Added of Tourism Industry (VATI) (United Nations, 2010) expressed in basic prices was 220.8 billion; if, on the other hand, we consider the contribution directly linked to tourism – Tourism Direct Value Added (TDVA) (United Nations, 2010) the amount is 99,9 billion (Istat, 2022). The distinction between these two aggregates, which refer to the production units pertaining (predominantly) to the tourism industry to produce those goods and services used by visitors, is due to the fact that, within tourism products, some services are also offered to those who are not tourists (for example, catering, restaurants or transports). It follows that each tourism product has its own tourism coefficient (Table n. 2), and it is for this reason that TDVA must be distinguished from VATI. For the time being, it remains impossible to produce estimates of this coefficient at the regional level, although at this level of detail the tourist expenditure of visitors is recorded and for domestic tourism it is also possible to reconstruct travel between regions.

### **2.3 Supply-side approach to satellite accounting**

This approach requires the availability of analytical data collected directly in units pertaining to the tourism industry. It can be divided into the characteristic industry (accommodation facilities; passenger air transport; travel agencies and tour operators) and the tourism-related industry (restaurants and bars; passenger rail transport; passenger road transport; passenger sea transport; hire of means of transport). The ATECO classification supports the “perimeter” of the tourism industry, however, there may be some critical issues concerning secondary activities which, depending on the criterion used, may lead to a change in classification and cause the local unit to move from the characteristic industry to the related industry (e.g., bathing establishments offering restaurant services). It should also be noted that the perimeter of the tourism industry identified by Eurostat differs in some items from that used in the satellite account (Antolini and Petrei, 2021).

As illustrated above, the methodology used also depends on the available statistical sources and there is no doubt that on the supply side the use of business registers, for those countries that have prepared them, is a potential. In Italy, the preparation of FRAME SBS, offers an availability of economic information that should be valorised and, in any case, used also in a perspective of balancing demand with supply. Moreover, the use of FRAME SBS would make it possible to estimate value added using the value-added method for units that have their own business accounts, being market units, while for non-market enterprises the applicable method could be that of income or personal (Barbieri et al. 2017).

### 3. A possible estimate of the tourism direct added value at a territorial level

To be able to make an attempt at a regional supply-side estimation, the first step was to identify the economic sectors contributing to the Tourist Direct Value Added (TDVA). Then, starting from the regional total added values, the percentages shown in Table 2 were applied for each economic activity.

**Table 2** – Tourist coefficient product (at national level) and ATECO 2007 classification of Tourism Industries

Air transportation	99.5%	51: air transportation
Travel agents, tour operators	99.3%	79: activities of travel agencies, tour operators and reservation services and related activities
Accommodation establishments	98.7%	55: accommodations establishments
Maritime transportation	86.3%	50: sea and water transport
Rail transportation	69.6%	49: land transport and pipeline transports
Road transportation	46.1%	49: land transport and pipeline transports
Food services	23.3%	56: food service activities
Shopping	13.2%	47: retail trade (excluding motor vehicle and motorcycle trade)
Cultural services	12.4%	91: activities of libraries, archives, museums and other cultural activities
Second homes owned	11.9%	
Sports and recreation services	9.7%	93: sports, entertainment and amusement activities
Vehicle rental services	6.6%	77: rental and operating leasing activities
Total	4.0%	
Other	0.8%	

*Source: Istat 2020, p. 4*

We applied these tourism coefficient to the total value added of tourism industries (as defined by ATECO in the table 2) at regional level (Regional Value Added - RVA). A limitation of the current estimation process is that these percentages used are fixed and do not vary from region to region. The result of the processing is shown in Table 3.

**Table 3** – Estimation of Regional tourism direct value added (RTDVA) and Tourism Index

2019	RVA	RTDVA	RTDVA/RVA
Piedmont	66268532	5446181	0.08
Aosta Valley/Vallée d'Aoste	1888682	308161	0.16
Lombardy	215527656	17490901	0.08
Trentino-Alto Adige/Südtirol	22130065	3715278	0.17
Veneto	87015539	8325185	0.10
Friuli-Venezia Giulia	18331175	3694914	0.20
Liguria	21778529	1602364	0.07
Emilia-Romagna	82793677	6734658	0.08
Tuscany	57314569	5697593	0.10
Umbria	9732612	685978	0.07
Marche	19500211	1431651	0.07
Latium	84719386	9732937	0.11
Abruzzo	13456617	954828	0.07
Molise	2200599	207690	0.09
Campania	42702463	5495819	0.13
Apulia	28327093	3050717	0.11
Basilicata	4286614	313484	0.07
Calabria	8937634	1172927	0.13
Sicily	26777530	3660223	0.14
Sardinia	11683689	1595886	0.14
<b>Italy</b>	<b>825372872</b>	<b>81317376</b>	<b>0.10</b>

*Source: Our processing on ISTAT data.*

From the data obtained emerges that RTDVA, passing through the estimated data at the level

of the individual regions, represents approximately 10% of the total (a credible value as far as current knowledge goes). However, as can be seen in the table, this value varies from region to region ranging from around 7% in various regions (Liguria, Umbria, Marche, and Basilicata), to 16% in Aosta Valley, 17% in Trentino Alto Adige and 20% in Friuli V. G. It should be noted that for some regions in southern and insular Italy, RTDVA is much higher than what would be expected from tourism density, also and above all given the low level of VA per capita. But this is one of the aspects on which further investigation is needed in the future.

#### 4. A final remark

The lack of access to Territorial Frame SBS does not allow the use of a true supply-side approach, so a flash estimate of the contribution of tourism at the regional level was not possible. Starting from the released data, we could only use the calculated tourism coefficients, as mentioned above. This represents, as mentioned, a simplification since it does not consider the variability of tourist flows, which are, however, contained indirectly in the added value of the branch of economic activity used. Having region-specific coefficients, at least for some branches, would be important.

Another possibility of intervention would be to succeed in identifying a model that would make it possible to arrive at an estimate of RTDVA at the regional level starting from historical or territorial series, even if not included in the tourism sphere, but which are thought to have an influence on the value to be estimated or to be an indicator, even indirectly, of this amount at the regional level, as is already the case, to give a simple example, for the estimation of presences at the municipal level through the weight of waste collected.

#### References

- Antolini F., Grassini L. (2020a). Methodological problems in the economic measurement of tourism. the need for new sources of information. *Quality & Quantity*, ISSN: 0033-5177, Doi: 10.1007/s11135-019-00962-x
- Antolini F., Grassini, L. (2020b). Issues in Tourism Statistics: A Critical Review, in *Social Indicators Research*, ISSN 0303-8300, 150, 1021-1042.
- Antolini F., Petrei F. (2021). Definizione del perimetro dell'industria turistica e misura del suo "impatto" economico, *Convegno della Società Italiana di demografia e Statistica*, Roma, 27 maggio 2021.
- Barbieri G., Faramondi A., Truglia F. (2017). La stima del valore aggiunto a livello territoriale: nuovi sviluppi nell'ambito delle statistiche strutturali, *Working Paper n.17 CERTeT*, Università Bocconi.
- Benassi F., D'Elia M., Petrei F. (2021). The "meso" dimension of territorial capital: evidence from Italy, *Regional Science Policy & Practice*, n. 13, pp. 159–175.
- European Commission (2004). *Methodological work on measuring the sustainable development of tourism*, Luxembourg, ISBN: 92-79-01680-6.
- European Commission (2022). *Relazione della Commissione al Parlamento Europeo e del Consiglio relativo alle statistiche europee sul turismo*, Bruxelles, 6 gennaio 2022.
- Eurostat (2009). *Tourism satellite accounts in the European Union Vol.1*, Methodologies and working Papers, Bruxelles. Working papers ISSN 1977-0375.
- Eurostat (2017). *Tourism statistics: Early adopters of big data?* Statistical working paper, Luxembourg: Publications Office of the European Union, ISBN 978-92-79-71898-4.
- Eurostat (2021). *Tourism statistics*, Statistics Explained, March 2021, online publications, [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_statistics)
- ISTAT (2020). Il conto satellite del turismo per l'Italia, Anno 2017. Roma, 17 giugno 2020.
- ISTAT (2022). Il Conto satellite del turismo in Italia, Anno 2019. Roma, 14 gennaio 2022.
- United Nations (2010). Tourism Satellite Account: Recommended Methodological Framework 2008, in *Studies in Methods*, Series F N. 80/Rev.1, Luxembourg, Madrid, New York, Paris.

# Short-term forecasts on time series for tourism in Lombardy

Andrea Marletta, Roberta Rossi, Elena Diceglie

## 1. Introduction

Data from official statistics are often available with a few months delay with respect to their collection. Tourism data collection is one of this kind and the statistics team in PoliS-Lombardia receives a lot of requests about predictions or provisional data in order to have real time insights about the tourism performance.

In these last years, because of the pandemic emergency due to Covid-19, the curiosity of public stakeholders about an economic recovery after 2020 downfall (and partially 2021) has increased and so the need to get official data as soon as possible. This paper aims at filling this need with short-term predictions in time series as temporary substitutes while waiting for official data to be published.

The context of this work is in the tourism sector, one of the most damaged economic sectors by the limitations due to Covid-19. Many contributions are already present in literature about the strategy and the estimation for the recovery of the travel sector after the pandemic emergency (Fotiadis et al., 2021; Yeh, 2021). In this context, an objective of this work is to verify the presence of a full or partial recover of tourists in provinces of Lombardy using short-term predictions for 2022. This issue has also been treated by Provenzano and Volo (2022). This contribution is the result of a collaboration with PoliS-Lombardia, a public institution of Regione Lombardia. It is included in the list of institutional units belonging to the public sector published by Istat.

PoliS-Lombardia has been instituted in 2018 and it is the regional institute for the support to the policies of Lombardy. Its mission is the implementation and the evaluation of the policies in Lombardy. The main functions of PoliS-Lombardia are: support to the integrated policies of education and labour coherently with fixed objectives by the administration; studies and research projects related to the institutional, local, economic and social processes; management of the regional statistical function in collaboration with ISTAT; management and coordination of the regional observatories; education of the regional employees. Given this scopes, it represents a very important stakeholders in the field of data management in Lombardy involved in a large amount of data, as for example in the tourism sector.

In this paper, using a short-term forecasts approach, some preliminary results will be presented for detecting a recovery in the travel sector for 2022 using the total number of presences in Lombard provinces. These short-term predictions will be obtained using a very well-known methodology in time-series literature, such as the ARIMA (Auto-Regressive Integrated Moving Average) models (Box et al., 2015; Hamilton, 2020; Wei, 2006). In these models, an exogenous variable representing the working positions in the food services and hospitality industry has been added supposing an high correlation between the two phenomena.

## 2. Methodological tools

Data from official sources on nights spent in an accommodation for tourists in Lombardy are available until 2021. These data on travel flows for 2020 and 2021 registered a clear downfall

Andrea Marletta, University of Milano-Bicocca, Italy, andrea.marletta@unimib.it, 0000-0002-4050-5316  
Roberta Rossi, PoliS-Lombardia, Italy, roberta.rossi@polis.lombardia.it, 0000-0003-4586-9044  
Elena Diceglie, PoliS-Lombardia, Italy, elena.diceglie@polis.lombardia.it

Referee List (DOI 10.36253/fup\_referee\_list)  
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Andrea Marletta, Roberta Rossi, Elena Diceglie, *Short-term forecasts on time series for tourism in Lombardy*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.14, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 77-82, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

because of restrictions related to Covid-19.

A time-series procedure has been applied to obtain a forecast estimate for 2022 using an ARIMA model with the addition of an exogenous variable.

The ARIMA models have been introduced as mixed models composed by an Auto-Regressive (AR) part in which the single observation depends on the lagged values of the time series, a Moving Average (MA) part in which the same observation depends on the lagged values of the errors and, if necessary, an Integrated (I) part considering the original time series in differences according an integration order (Wei, 2006).

They could be represented as:

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t$$

where  $\phi_p(B)$  represents the AR part,  $(1 - B)^d Z_t$  the I part and  $\theta_q(B)a_t$  the MA part.

The hypothesis at the basis of the model is that a punctual estimate of the travel flows could be obtained using an auxiliary variable explaining the number of employees in the food services and hospitality industry. Statistically speaking, this means to introduce ARIMAX models, that is to say, ARIMA models with an exogenous variable with the following notation:

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t + \beta_i x_i$$

where  $\beta_i x_i$  is the X part of the model. This auxiliary variable is represented as the difference between the number of starting work contracts and the contract terminations. These data are available thanks to the Informative system of mandatory communications provided by the Italian Minister of Labour. The availability of this information is daily guaranteed at level of single municipality but for the purpose of this paper, data have been aggregated at province level.

The short-term predictions obtained for 2022 have been used to verify the presence of a recovery respect to the pandemic emergency of Covid-19 using a double growth rate. A first growth rate has been computed comparing the number of estimated tourists respect to the 2021 measuring the existence of a rebound after the restrictions. A second growth rate measured the estimates for 2021 respect to the presences of 2019 to monitor the trends in Lombardy compared to the before Covid-19 period.

Data used for the prediction model refers to the total number of travel presences expressed in terms of nights in accommodation from 2017 to 2021. About the auxiliary variable, data refers to the balance expressed as the difference between the activations and the terminations of the job contracts until March 2022. All the elaborations have been computed using R following the approach proposed by Hyndman and Athanasopoulos (2018).

The approach to obtain this short-term forecasts is based on a two-step procedure: firstly, data about employees are predicted for the interval from April to December 2022; secondly, predictions for tourism presences are obtained for the entire 2022.

The time series of the COB (Comunicazioni OBbligatorie) related to activations and terminations of job contracts for the food services and hospitality industry is updated until March 2022. Since PoliS-Lombardia is interested in predicting the entire year 2022, before applying the ARIMAX model, the values for this variable for the remaining months of 2022 have been obtained using a well-known approach choosing the best model among different time-series predictors as ARIMA models and ETS (Error, Trend, Seasonality) models. The model was selected minimizing the Mean Squared Error (MSE).

Once obtained the extended time series on the balance of the job contracts, this can be used as auxiliary variable for predicting the 2022 observations for the travel indicator using an ARIMAX model.

### 3. Application and results

Data source used for the prediction about the total number of travel presences from 2017 to 2021 has been achieved from 2 different surveys. From 2017 to 2020, data are the official statistics released by Istat, for 2021 data are from Istat but they are obtained in a different way and they are still provisional.

The integration of data using provisional information about 2021 has been necessary to obtain plausible forecasts. Without this operation, data about 2020 would have deeply conditioned the predictions in a negative trend. The 2020 data have been influenced by the restrictions due to the pandemic emergency due to Covid-19. Since the Lombard tourism is characterized by seasonality (above all in the mountain provinces), the predictions take into account this aspect underlining different trends for each territory.

Data about start and end of the job contracts are sourced to the COB system provided by the Italian Minister of Labour. Since they are computed as a difference, they could assume positive and negative values. They are only referred to positions in the food services and hospitality industry. In particular, the hypothesis behind this choice is that an increase of the balance (and therefore of the activations) of the employees in this sector is a symptom of a higher request due to an increase of the travel presences. If these two series are highly correlated, it makes sense to use this variable as exogenous in explaining the travel indicator.

All data are available monthly and from a geographic point of view, they referred to Lombard provinces. In Lombardy, 12 provinces are present, they are: Bergamo, Brescia, Como, Cremona, Lecco, Lodi, Mantova, Milan, Monza-Brianza, Pavia, Sondrio, Varese. In Figure 1, a time series plot with real (in black) and predicted values (in blue) is displayed as an example for Bergamo and Varese provinces.

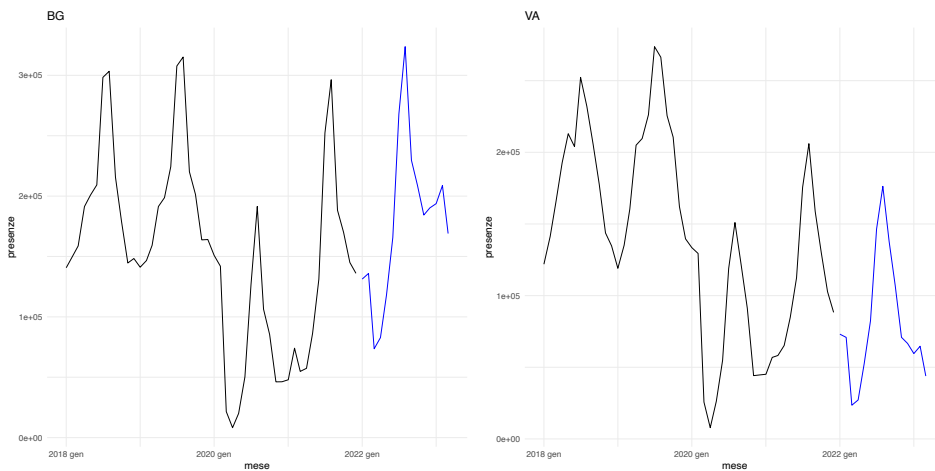


Figure 1: Time series plot for total presences for Bergamo and Varese provinces



As mentioned in the previous section, the research question of the paper is two-fold: firstly, to evaluate the plausible upswing for predicted values for 2022 respect to 2021 and secondly, to compare this predictions with the pre-Covid19 period such as 2019. The answer to this research question could be obtained using two simple growth rates:

$$t_1 = \frac{\text{predicted presences}_{2022}}{\text{official presences}_{2021}} * 100$$

$$t_2 = \frac{\text{predicted presences}_{2022}}{\text{official presences}_{2019}} * 100$$

The results of the model predict a substantial recovery of the Lombard tourism compared to 2021 for almost the 12 provinces with  $t_1$  growth rate higher than 40% in Como, Cremona and Sondrio provinces. Complete results for  $t_1$  are displayed in Figure 2.

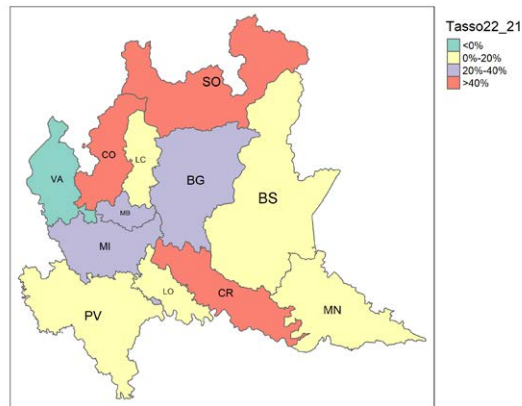


Figure 2: Growth rate per total presences in Lombardy provinces between 2022 and 2021

From the map, it is possible to note that  $t_1$  is positive for all provinces except than Varese. The highest values for  $t_1$  is for Sondrio, where the model estimated a doubling of the presences, but this is due to the fact that Sondrio is a mountain province in which 2021 has been strongly conditioned by the limitations in the winter season. Bergamo, Milan and Monza-Brianza have a growth rate between 20% and 40%. For other provinces it has been registered a moderate growth.

On the other hand, there is not a complete recovery respect to the pre-Covid19 period. Only 4 provinces have positive values for  $t_2$ : Como, Cremona, Monza-Brianza and Sondrio. Complete results for  $t_2$  are displayed in Figure 3.

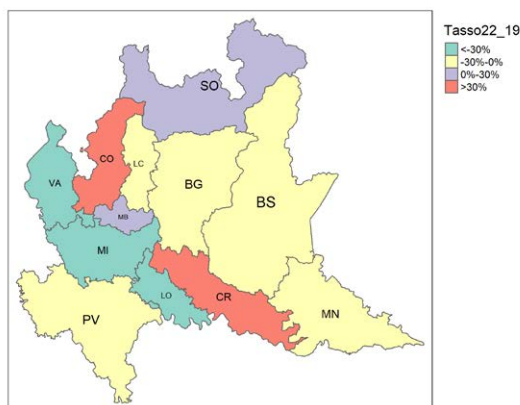


Figure 3: Growth rate per total presences in Lombardy provinces between 2022 and 2019

All the other provinces of the East Lombardy registered a light decline respect to 2019, but for some provinces as Brescia and Lecco, this decrease is only about 3%, hoping for a complete recovery in 2023. Negative growth rates more stressed are obtained for Lodi, Milan and Varese where the predicted values for presences are still 30% less than 2019, symptom of a slowest recovery.

#### 4. Summary and conclusions

The aim of this paper was to obtain short-term predictions about total presences in tourism sector in 2022 for Lombard provinces using an ARIMAX model considering data from labour market as auxiliary variable. This variable has been used hypothesizing a high correlation between the activations of contracts in food and hospitality sector and the increase of the travel presences. Preliminary results showed an evident upswing respect to 2021 and a partial recovery respect to 2019 for the majority of Lombard provinces. In particular, Sondrio is the province with the highest growth rates and Varese the province with the lowest growth rates.

Future works could focus the attention on other exogenous variables to add in the ARIMAX model hypothesizing other possible influences on the phenomena of the Lombard tourism. The same model could be also replicated for single municipalities or particular industrial districts. Finally, from a methodological point of view, some other prediction techniques could be added as comparison like for example the VAR (Vector Auto-Regressive) models and the relation between presences and workers could be enhanced through a co-integration analysis.

## References

- Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
- Fotiadis, A., Polyzos, S., Huan, T. C. T. (2021). The good, the bad and the ugly on COVID-19 tourism recovery. *Annals of tourism research*, 87, 103-117.
- Hamilton, J. D. (2020). Time series analysis. Princeton university press.
- Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.
- Provenzano, D., Volo, S. (2022). Tourism recovery amid COVID-19: The case of Lombardy, Italy. *Tourism Economics*, 28(1), 110-130.
- Wei, W. W. (2006). Time series analysis: univariate and multivariate. Methods. Boston, MA: Pearson Addison Wesley.
- Yeh, S. S. (2021). Tourism recovery strategy against COVID-19 pandemic. *Tourism Recreation Research*, 46(2), 188-194.

# Interventions for non-self-sufficiency – Focus on care and social policies in South Tyrol

Giulia Cavrini, Nadia Paone, Evan Tedeschi

## 1. Introduction

Current demographic trends and changes in family structures (increasing divorce rates, lower birth rates, a higher number of one- and two-person households, as well as high mobility, especially of the younger generations) point to a social change that poses new challenges to society as far as the care of older people is concerned (Petrini et al., 2019).

Because households are getting smaller and family structures are changing, care and support can no longer necessarily be provided within family circles (Oris et al., 2021; Quesnel-Vallée et al., 2016). However, most older people would like to stay in their own homes or familiar surroundings and neighbourhood as long as their health permits (Turjamaa et al., 2019). Thus, there is a need for enabling structures that take social change into account and ensure the long-term and continuous care of the elderly population (Plöthner et al., 2019).

The current social assistance system to support home care in Italy has several weaknesses. These include the rigidity of care hours and days, different procedural processes and, above all, a widespread lack of coordination and integration of the different interventions (Menghini & Tidoli, 2019). To date, alternative services and social policy re-examinations are marginal compared to the practical need. The bulk of the care burden in Italy continues to fall on families. Especially in rural communities, there is a dilution of the provision of local infrastructures and social networks. These developments call for new strategies that address complex needs, transform outpatient services into a care structure close to home and on time, and ensure self-determined living in one's own home.

This work stems from the doctoral thesis of Nadia Paone, and further statistical analyses were carried out by Evan Tedeschi as part of his work at the Competence Centre for Social Work and Social Policy.

This paper focuses on the older age groups, the "young old" and the very old. The research interest focuses on the living space and the immediate living environment of the target group, including relations with the neighbourhood. The basic assumption here is that the living environment opens the scope for activities outside the home (Bonaccorsi et al., 2020; Rautio et al., 2018). The following contribution analyses different forms of social support in the home environment, promoting equality and social cohesion.

A mixed-methods approach was used for the following study. Specifically, the study is based on a sequential and explorative design. The qualitative part of the research is exploratory and serves to collect elements that form the basis for the quantitative part of the research (Cohen et al., 2018). For the qualitative part, semi-structured interviews were conducted with experts (actors working in public and private institutions of elderly care).

In the qualitative part of the study, the first part of the guideline comprised general questions on age-appropriate housing and housing needs in old age; the second part concerning future approaches in the field of housing for the elderly in South Tyrol and support options to ensure that they remain in their own homes for as long as possible. Finally, the experts were asked for their views on the care situation in South Tyrol and on the gaps they perceived in the existing offer. Overall, the interviews with the experts make it clear that the immediate living environment is crucial for the subjective well-being of older people. The interviews suggest that there is a need for interdisciplinary cooperation between social and health care institutions. In summary, it can be

Giulia Cavrini, Free University of Bozen-Bolzano, Italy, Giulia.Cavrini@unibz.it, 0000-0002-9084-3081  
Nadia Paone, Free University of Bozen-Bolzano, Italy, Nadia.Paone@unibz.it  
Evan Tedeschi, Free University of Bozen-Bolzano, Italy, evan@email.it

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Giulia Cavrini, Nadia Paone, Evan Tedeschi, *Interventions for non-self-sufficiency – Focus on care and social policies in South Tyrol*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.15, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 83-88, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

concluded from the interviews that older people should be offered small everyday aid and pre-care services in addition to professional services.

The results of the qualitative interviews served as a basis for constructing the quantitative questionnaire.

Assuming that most older people want to remain in their own homes as long as possible, the question remains how to simultaneously ensure and promote dignified and participative ageing. Based on these assumptions, we aim to identify the following salient points: what supportive possibilities favour the elderly in living in their own homes as long as possible and what might need to be added to the previous services; what is or what might be the role of neighbourhood/voluntary work; what supports the social space and the living environment might offer; what features of the live environment act as resources and what as barriers for the elderly in South Tyrol.

## 2. Methodology and description of the sample

The sample comprised men and women aged between 60 and 101 and comprised 536 respondents. The sample of the quantitative part includes persons aged 60 and over who live in their own homes and reside in South Tyrol. The following considerations guided this decision: The research group should reflect the diversity of life situations of the elderly and include the entire Province of Bolzano.

The survey was conducted between June 2020 and April 2021 within the framework of a dissertation. Exclusion criteria are non-residence in the Province of Bozen-Bolzano and persons in an in-patient or semi-inpatient facility at the time of the study.

Concerning the outcome variables, based on the hypotheses identified above, we selected the following items: home satisfaction, satisfaction with the neighbourhood, time spent outside the home, and perceived health.

Using a Latent Class model, it is possible to test the conceptualisation of the idea that older people can live longer in their own homes as a latent categorical indicator, in which each option reflects a specific category that originates from the intersection of the factors we obtained above.

A possible way forward would be to consider our outcome variables.

However, this route is partially problematic, as we have ascertained that most of these indicators do not distribute usually and are characterised by strongly skewed distributions<sup>1</sup>. Consequently, we suggest dichotomising the variables into four indices to remedy the problems listed, with 0 representing an inadequate state and 1 as a good state.

The variables chosen for analysis are as follows: perceived health ("How is your health in general?": 0 = bad; 1 = good), time spent away from home (cross-reference of two questions: "How much time do you spend away from home?" and "Do you sometimes not leave the house for a few days in a row?": 0 = a little; 1 = a lot), satisfaction with the neighbourhood (how satisfied do you feel with your relationship with the neighbourhood?: 0 = not satisfied; 1 = satisfied), home satisfaction: (0 = not satisfied; 1 = satisfied).

It is now possible to consider a Latent Class Model, using the manifest variables illustrated above, which will be related to the model's latent concept to be analysed (Moisio, 2004). In other words, latent class analysis (LCA) allows us to specify the latent factor categories related to the possibility of being part of a specific category of observable variables.

The starting assumption is that local independence exists between the manifest variables, i.e., the observed association between them is zero within the different categories (McCutcheon, 1987).

Specifically, if we consider a given category of the latent factor X ( $X=t$ ), the probability of combining a particular set of responses ( $A=k; B=i; C=j$ ) is represented by an individual's chance of taking part in  $t$  of X, for the conditional probability of stating  $k$  in the case of A,  $i$  for B and  $j$  for C:

$$\pi_{tkij}^{XABC} = \pi_t^X \pi_{kt}^{A/X} \pi_{it}^{B/X} \pi_{jt}^{C/X}$$

<sup>1</sup> We performed the Shapiro-Wilk and Komogorov-Smimov tests to test the hypotheses of normality (Razali & Wah, 2011).

Where  $\pi_t^X$  denotes the probability of being a member of the latent class  $t = 1, 2, \dots, T$  of the latent variable  $X$  (Zhou et al., 2018);  $\pi_{kt}^{A/X}$  denotes the conditional probability of having the answer  $k$  within the variable  $A$  by the members of class  $t$ ; while  $\pi_{it}^{B/X}$  and  $\pi_{jt}^{C/X}$  represent the same probabilities for items  $B$  and  $C$ . Starting with the four observable items and  $X$ , representing the latent factor to be estimated, our formula becomes:

$$\pi_{ta-n}^{XSF1-SF4} = \pi_t^X \pi_{at}^{SF1/X} \pi_{bt}^{SF2/X} \pi_{ct}^{SF3/X} \pi_{dt}^{SF4/X}$$

The analysis – implemented using Latent Gold software (Van der Nest et al., 2020) – shows that the four latent class model is the one that best fits the data, as it shows an increase in explained variance and, at the same time, the lowest value on the BIC.

### 3. Results

We can verify the magnitude of the different classes, which can be given a noun meaning, from the results of the conditional probabilities (Table 1).

**Table 1 Results of latent class analysis: conditional probabilities**

	Cluster1	Cluster2	Cluster3	Cluster4
Latent classes	0.26	0.27	0.37	0.09
<b>Indicators</b>				
Satisfaction for the neighbourhood				
0	0.13	0.45	0.44	<b>0.75</b>
1	<b>0.87</b>	<b>0.55</b>	<b>0.56</b>	0.25
Housing satisfaction				
0	0.03	0.24	<b>0.74</b>	<b>0.99</b>
1	<b>0.97</b>	<b>0.76</b>	0.26	0.01
Perceived health				
0	0.02	<b>0.51</b>	0.13	<b>0.86</b>
1	<b>0.98</b>	0.49	<b>0.87</b>	0.14
Probability of going out				
0	0.01	0.34	0.01	<b>0.78</b>
1	<b>0.99</b>	<b>0.66</b>	<b>0.99</b>	0.22

Specifically, we have identified four latent classes: Cluster 1 (all indicators have very good values), Cluster 2 (indicators are good and have average values), Cluster 3 (the first indicator is average, the second has low values, and the others are very good), Cluster 4 (all indicators have low values).

**Table 2: Descriptive statistics for the main variables (N. = 536)**

Age	%	Employment status	%
60/64	18.10	Open-ended contract	8.40
65/69	19.59	Fixed-term contract	5.22
70/74	16.60	Pension: worker	19.59
75/79	13.99	Pension: employee	35.07
80/84	14.18	Pension: executive	9.89
85/89	7.84	Pension: entrepreneur	7.09
90 +	9.70	Pension: housewife	14.74
Total	100.00	Total	100.00
<b>Marital status</b>	<b>%</b>	<b>Education</b>	<b>%</b>
Single	10.45	Low	26.12
Married/cohabiting	56.34	Media	24.81
Separated/divorced	6.34	High	24.25
Widowed	26.87	Total	100.00
Total	100.00		

In conclusion, we have identified four classes that follow a conceptual structure in which the first and fourth clusters differ markedly and represent two very different types of individuals.

**Table 3 Multinomial logistic model: estimates for the probability of belonging to a particular cluster rather than the first cluster**

	Cluster 2 (medium)		Cluster 3 (medium)		Cluster 4 (low)	
	Coeff.	St. Dev.	Coeff.	St. Dev.	Coeff.	St. Dev.
<b>Age</b>						
60/64	(basis)	-	(basis)	-	(basis)	-
65/69	0.54	0.63	0.09	0.372	-1.53	1.09
70/74	1.20	0.63	0.01	0.410	-0.34	0.98
75/79	0.89	0.65	-0.38	0.446	-0.39	1.01
80/84	1.18	0.70	-0.09	0.495	-0.08	0.97
85/89	1.18	0.79	-0.73	0.633	-0.81	1.07
90+	2.07*	0.86	0.13	0.739	-0.13	1.12
<b>Gender</b>						
Male	(basis)	-	(basis)	-	(basis)	-
Female	0.16	0.37	0.17	0.28	0.44	0.54
<b>Marital status</b>						
Single	(basis)	-	(basis)	-	(basis)	-
Married/cohabiting	-0.24	0.58	-0.18	0.43	-0.44	0.77
Separated/divorced	-0.01	0.88	0.08	0.61	1.04	1.08
Widow/widower	-0.05	0.62	-0.16	0.48	-0.54	0.83
<b>Education</b>						
Low	(basis)	-	(basis)	-	(basis)	-
Medium	0.17	0.40	0.42	0.36	-0.41	0.56
High	-0.92	0.53	-0.26	0.42	-1.93*	0.86
<b>Employment status</b>						
Permanent employment	(basis)	-	(basis)	-	(basis)	-
Precarious employment	-1.79	1.25	-0.52	0.62	-12.03	1.60
Pension: blue-collar worker	-0.15	0.77	0.19	0.53	0.43	1.41
Pension: Employee	-0.03	0.73	0.34	0.46	0.65	1.35
Pension: Executive	0.57	0.79	0.18	0.53	0.00	1.69
Pension: Entrepreneur	-0.01	0.86	-0.15	0.64	-0.70	1.65
Pension: Housewife	-0.76	0.81	-0.70	0.59	-0.76	1.45
<b>Friends frequency</b>						
Every day	(basis)	-	(basis)	-	(basis)	-
2-4 times a week	0.72	0.49	0.62	0.33	0.37	1.00
1 time a week	1.35**	0.49	0.97**	0.35	2.58**	0.92
1 time a month	2.19**	0.67	1.23*	0.58	3.26**	1.02
2-4 times a year	1.74*	0.75	0.48	0.68	3.25**	1.08
<b>Architectural barriers</b>						
barriers	-1.07	1.13	3.41***	0.79	3.82**	1.25
<b>Home security</b>						
Not safe	(basis)	-	(basis)	-	(basis)	-
Safe	-0.23	0.34	-1.40***	0.26	-2.17***	0.49
<b>Physical activity</b>						
No	(basis)	-	(basis)	-	(basis)	-
Yes	-0.64	0.36	-0.37	0.29	-1.38**	0.52
<b>Attends parties</b>						
rarely	(basis)	-	(basis)	-	(basis)	-
often	-0.70*	0.36	-0.66*	0.29	-1.68**	0.54
<b>Does volunteer work</b>						
No	(basis)	-	(basis)	-	(basis)	-
Yes	-0.70	0.37	-0.45	0.30	-1.91**	0.68
<b>Intercepts</b>	-0.62	1.07	1.13	0.74	0.51	1.82

N = 536

LL = -482.007

Pseudo R2 = 0.27

\* p < 0.05. \*\* p < 0.01. \*\*\* p < 0.001)

The first cluster concerns individuals with above-average values, while the fourth has to do with individuals with the lowest scores overall. Between these two contrasting categories are

two groups of individuals who achieved intermediate values, albeit closer to the first group than the fourth. Potential confounding factors could be correlated with perceived health and the other variables seen above (Tab. 2).

We, therefore, introduced age as a categorical variable, employment status (composed, given the older age, of predominantly retired individuals), marital status and level of education. The sample comprised 326 women (61%) and 210 men (39%). The variables we use to test our hypotheses are frequency of meeting friends, the presence of architectural barriers in one's home (indicator obtained utilising factor analysis on a series of items), perceived housing safety, carrying out physical activity, and participation in neighbourhood festivals and performing voluntary work.

We aim to analyse the impact of certain variables on the clusters (Table 3). If we consider the frequency of seeing friends, we can see that as the probability of seeing friends decreases, the probability of being part of cluster 4 increases compared to cluster 1.

The index for architectural barriers follows a significantly decreasing trend in clusters 3 and 4 compared to cluster 1.

At the same time, as the probability of feeling safe at home decreases, the chances of being part of cluster 4 increase compared to cluster 1.

The same result can be observed in the case of physical activity: those who do not regularly engage in physical activity are more likely to be part of cluster 4 than cluster 1. Respondents who rarely participate in village festivals are likelier to be part of the last group, i.e. cluster 4.

#### 4. Conclusions

The above analysis highlighted the following points:

- *doing household chores independently produces a positive impact;*
- *positive association with frequent attendance of friends;*
- *architectural barriers have a significantly negative impact;*
- *the perceived safety of one's home produces a positive impact;*
- *positive association with engaging in physical activity;*
- *Participating regularly in parties organised in one's community has a positive impact;*
- *doing voluntary work has a positive impact.*

Considering the salient points in the introduction, it is undoubtedly essential to ensure the elimination of architectural barriers in the home and, simultaneously, guarantee greater safety, especially for those with serious health problems and need aids such as wheelchairs. Frequently mentioned barriers in the home are stairs or steps and the lack of a lift. The reasons for a low sense of security in one's home are architectural barriers in the surroundings, burglaries, a poor state of health and the lack of contact persons in an emergency.

As emerged from the results, the role of neighbourhood and friendship relations is central in ensuring that most elderly people remain in their homes as long as possible. Suitable meeting places include one's own home, the homes of others and public spaces such as cafes and parks. Likewise, the active voluntary work experience is essential in this respect.

The social space and living environment must play a central role in ensuring activities and opportunities for older people to meet and socialise, as this is a crucial resource. Barriers, on the contrary, are all those elements that do not guarantee the elderly to move freely, especially for those with obvious health problems.

These findings also confirm that as the radius of action in old age is or becomes smaller, the home and the living environment (Barth & Olbermann, 2012) are becoming increasingly important. The importance of the home and the living environment increases to the same extent that the radius of movement decreases in old age, and it is reduced for physical, psychological, and social reasons (Saup, 1999).



However, it must be considered that a larger number of retrospective, pre-treatment and contextual variables would certainly have facilitated a greater identification and control of unobserved heterogeneity. For this reason, we believe that it would be desirable to supplement the results with data that consider a longitudinal approach, more extensive and richer in retrospective indicators. Therefore, further theoretical and empirical investigations are indispensable to refine the proposed model and conduct complementary analyses that partially weigh essential factors and elements that we have only been able to consider.

## References

- Barth, C., & Olbermann, E. J. (2012). Wohnen im Alter. Potenziale einer nachhaltigen Quartiersentwicklung für die Lebensqualität älterer Menschen. *Blätter der Wohlfahrtspflege*, 159(4), 138-140.
- Bonaccorsi, G., Manzi, F., Del Riccio, M., Setola, N., Naldi, E., Milani, C., . . . Lorini, C. (2020). Impact of the built environment and the neighborhood in promoting the physical activity and the healthy aging in older people: An umbrella review. *International journal of environmental research public health*, 17(17), 6127.
- McCutcheon, A. L. (1987). *Latent Class Analysis*: Thousand Oaks (CA), Sage.
- Menghini, V., & Tidoli, R. (2019). La presa in carico dell'anziano non autosufficiente a domicilio I luoghi della cura. *Rivista online*, (1-2019).
- Moisio, P. (2004). A latent class application to the multidimensional measurement of poverty. *Quality and Quantity*, 38(6), 703-717.
- Oris, M., Baeriswyl, M., & Ihle, A. (2021). The Life Course Construction of Inequalities in Health and Wealth in Old Age. *Handbook of Active Ageing. Quality of Life*, 97-109.
- Petrini, M., Cirulli, F., D'Amore, A., Masella, R., Venerosi, A., & Venerosi, A. (2019). Health issues and informal caregiving in Europe and Italy. *Annali dell'Istituto Superiore di Sanita*, 55(1), 41-50.
- Plöthner, M., Schmidt, K., de Jong, L., Zeidler, J., & Damm, K. (2019). Needs and preferences of informal caregivers regarding outpatient care for the elderly: a systematic literature review. *BMC geriatrics*, 19(1), 1-22.
- Quesnel-Vallée, A., Willson, A., & Reiter-Campeau, S. (2016). Health inequalities among older adults in developed countries: Reconciling theories and policy approaches. In *Handbook of aging and the social sciences* (pp. 483-502): Elsevier.
- Rautio, N., Filatova, S., Lehtiniemi, H., & Miettunen, J. (2018). Living environment and its relationship to depressive mood: a systematic review. *International journal of social psychiatry*, 64(1), 92-103.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of statistical modeling and analytics*, 2(1), 21-33.
- Saup, W. (1999). Alte Menschen in ihrer Wohnung: Sichtweise der ökologischen Psychologie und Gerontologie. In *In H.W. Wahl, H. Mollenkopf, & F. Oswald (Hrsg.), Alte Menschen in ihrer Umwelt (S. 43-52). VS Verlag für Sozialwissenschaften*: Springer.
- Turjamaa, R., Pehkonen, A., & Kangasniemi, M. (2019). How smart homes are used to support older people: An integrative review. *International journal of older people nursing*, 14(4),
- Van der Nest, G., Passos, V. L., Candel, M. J., & Van Breukelen, G. (2020). An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research*, 43,
- Zhou, M., Thayer, W. M., & Bridges, J. (2018). Using latent class analysis to model preference heterogeneity in health: a systematic review. *Pharmacoeconomics*, 36(2), 175-187.

# The territorialisation of the 2030 Agenda: a multilevel approach

Raffaele Attanasio, Manlio Calzaroni, Alessandro Ciancio, Federico Olivieri,  
Giovanni Siciliano

## 1. Introduction

The concept of sustainable development has evolved over time, involving the international and global communities (Sachs, 2015). The 2030 Agenda for Sustainable Development, approved on 25 September 2015 by the General Assembly of the United Nations, has shaped the concept of sustainability in its most concrete definition, establishing the multidimensionality nature of sustainable development: the environmental dimension is associated with the economic, social and institutional ones (UN, 2015). It commits the governments of the 193 UN member States to work together to transform our world. However, the Agenda is divided into 17 Sustainable Development Goals (SDGs) which have a universal character, as long as they are aimed for all the countries in the world, without income nor geographical distinctions.

The monitoring process of sustainable development has acquired fundamental importance. At the international level, this process translates into an annual review at the UN Economic and Social Council, a four-year review at the General Assembly, and with the presentation of voluntary national reviews. Despite the leverage on the accountability of countries and the encouragement of initiatives aimed at raising awareness on issues related to sustainable development, the achievement of the 2030 Agenda still struggles to find concrete and rapid implementation. For example, Italy with its National Sustainable Development Strategy (SNSvS) has not yet defined quantitatively what its commitments are for achieving the 17 SDGs.

On this basis, the UN "Decade of Action" was launched in September 2019 to accelerate efforts to achieve the SDGs (UN, 2019). United Nations Secretary General António Guterres called on all components of society to mobilize for change: from world leaders to coordinate global action, to local leaders to define national, regional and city policies and strategies (Guterres, 2019).

If at national level the main common action of the member States is the definition of a national strategy for sustainable development, the international community has integrated the SDGs also in its supranational, regional or sectoral conformations. For example, the Organization for Economic Cooperation and Development (OECD) has adopted an action plan to contribute to the SDGs (OECD, 2016), while organizations such as the Security and Cooperation in Europe (OSCE) and the Council of Europe (CoE) have integrated the SDGs into their policies and activities.

In November 2016, the European Commission presented the EU strategic approach to the SDGs with the communication of "The sustainable future of Europe: next steps" (European Commission, 2016), which places sustainable development as the guiding principle of all political strategies and inaugurates a high-level multistakeholder platform level to support cross-sectoral exchange of best practice practices. To date, the SDGs are included in all six Commission priorities 2019-2024 (European Commission, 2019).

The transition process towards a more sustainable development model cannot ignore the contribution of local policies as stated by the European Commission for Economic Policy "Indeed, 65% of the 169 targets can only be reached through coordination and inclusion of local and regional governments" (Commission for Economic Policy, 2019). This is crucial for the development of the local context to contribute to the achievement of the SDGs on a global scale. The purpose of this paper is to present the methodological framework defined by the Italian Alliance for Sustainable Development (ASviS), based on the experiences developed with local administrations to support

Raffaele Attanasio, ASviS - Alleanza per lo Sviluppo Sostenibile, Italy, raffaele.attanasio@asvis.net  
Manlio Calzaroni, ASviS - Alleanza per lo Sviluppo Sostenibile, Italy, manlio.calzaroni@asvis.net, 0000-0003-1262-7815  
Alessandro Ciancio, ASviS - Alleanza per lo Sviluppo Sostenibile, Italy, alessandro.ciancio@asvis.net  
Federico Olivieri, ASviS - Alleanza per lo Sviluppo Sostenibile, Italy, federico.olivieri@asvis.net, 0000-0002-2049-7283  
Giovanni Siciliano, ASviS - Alleanza per lo Sviluppo Sostenibile, Italy, giovanni.siciliano@asvis.net, 0000-0003-3657-587X

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Raffaele Attanasio, Manlio Calzaroni, Alessandro Ciancio, Federico Olivieri, Giovanni Siciliano, *The territorialisation of the 2030 Agenda: a multilevel approach*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.16, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 89-94, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

the implementation of a "Multi-level sustainable development strategy", which makes planning territorial coherent with the national one and with the European programming.

## **2. The territorialization of UN 2030 Agenda**

By “territorialization” of the UN 2030 Agenda we mean the process of defining, implementing, and monitoring sustainable development strategies at the local level created in order to contribute to the achievement of global, national, regional and provincial objectives and targets. The approach developed for the definition of the Territorial Strategies is based on the experiences that ASviS has developed in accompanying the Italian Regions and local institutions in the development of their sustainable development Strategies.

The model is carried out in four phases which will be explored in the following paragraphs:

1. Assessment of the positioning of the territory with respect to the UN 2030 Agenda SDGs;
2. Identification of the quantitative targets that the regional/territorial administration wants/must achieve;
3. Designing of policies that should favour the achievement of local quantitative targets;
4. Involvement and dialogue with all stakeholders in sharing "specific" objectives, actions and projects.

### *2.1 Regional and local positioning*

The positioning makes it possible to assess the level of sustainability of the territory with respect to the 17 Sustainable Development Goals of the UN 2030 Agenda. This territorial analysis is carried out through specific composite indices calculated for each SDG.

The data source is the National Institute of Statistics (Istat), or institutions belonging to the National or European Statistical System. The indicators are for the monitoring of the Sustainable development and the sustainable well-being, following these criteria:

- Comparability that allows the comparison between the different territorial levels;
- Availability of information in time series;
- Polarity of the indicator. The meaning of the indicator must be clear, i.e. the interpretation in the case of its increase or decrease must be unambiguous: an increase/decrease in the value of the elementary indicator must correspond to a clear judgement, either positive or negative.

For the calculation of the composite index we have followed the systematic process proposed by (Nardo et al., 2005). Each SDG has been associated with a list of specific indicators, able to represent the characteristics of the Goal. The list of simple indicators, which form the basis of the 17 composite indices, is available on the ASviS website (*I Numeri Della Sostenibilità - Alleanza Italiana per Lo Sviluppo Sostenibile*, 2021). Subsequently, the chosen indicators were normalized using the methodology proposed by (Mazziotta & Pareto, 2015). Then, for the aggregation we have chosen the Adjusted Mazziotta - Pareto Index (AMPI), a composite index also used by Istat (Mazziotta & Pareto, 2017) deciding to attributes equal weight to all the basic indicators.

The indices show the improvement or worsening of the situation compared to the starting value recorded in the base year (for ASviS 2010). If a composite index shows an improvement, this does not necessarily mean that the Region is on a path that will allow it to meet the Goals in 2030, but simply that, on average, it is moving in the right direction, giving policy makers an assessment of where their territory stands in relation to the 17 Goals of the 2030 Agenda. Table 1 shows an example of analysis of the composite index for three SDGs calculated for the Emilia Romagna region. The SDGs chosen represent three of the four spheres in which the concept of sustainable development is articulated. Specifically, Goal 4 refers to the social sphere, Goal 8 to the economic sphere and Goal 15 to the environmental sphere. However, the calculation of the composite takes place for all the SDGs and for all the Italian regions.

**Table 1.** Values of the AMPI indices calculated for the Italian region of Emilia Romagna for SDGs 4 (quality education), 8 (economic growth) and 15 (life on earth). The values are given for 2010 and 2020.

Emilia Romagna	AMPI 2010	AMPI 2020
SDG 4	108.1	117.7
SDG 8	108.9	110.4
SDG 15	92.8	86.3

Table 1 shows that the Emilia Romagna region, between 2010 and 2021, improves in Goal 4, with an increase in the index value of 9.6 points, remains almost constant in Goal 8, showing an increase of 1.5 points and gets worse in Goal 15, highlighting low performances.

## 2.2 Identification of quantitative targets

Since the UN 2030 Agenda is an action plan for all the countries in the world, only in few cases it defines quantitative targets, delegating this task to national and local governments. It is therefore crucial for local sustainable development strategies to concretise quantitatively the targets of the 2030 Agenda. The quantitative targets values, associated with the UN 2030 Agenda, are defined according to the following hierarchy:

- A. Target values are defined by the higher institutional levels (UN, European Union and Italian government);
- B. In the absence of a defined value as in point A), the assessment of the target value is based on the judgment of the experts of the ASviS working groups;
- C. If the aforementioned methodologies are not applicable, a benchmark analysis is carried out defining the value recorded by the territorially most similar best performer as the target value.

If none of the above criteria allows to define the target values, the Eurostat methodology is used (EUROSTAT, 2021). This type of analysis allows to evaluate the performance of the regions, and, more generally, of the territory with respect to the achievement of the quantitative objectives of sustainable development defined at national and/or supranational level. This preliminary analysis attributes the same quantitative target between the different levels (national, regional and metropolitan) and within them (for example between the different Italian regions), without taking into consideration the geomorphological, social and economic characteristics of the territory. The assessment of the target is generally based on the ‘compound annual growth rate’ (CAGR) formula, which assesses the pace and direction of the evolution of an indicator (EUROSTAT, 2021). This formula uses the data from the first and the last years of the analysed time span and is used to calculate the average annual rate of change of the indicator (in %) between these two data points.

In the presence of a quantified political target (for example, the target in Table 2 which is defined by the circular economy package, published in the Official Journal of the European Union on 14 June 2018), the actual rate of change of the indicator is compared with the theoretical rate of change that would be required to meet the target in the target year.

If the actual rate is:

- 95 % or more of the required rate, the indicator shows a significant progress towards the EU target;
- Between 60 % and 95 %, the trend shows moderate progress towards the EU target;
- Between 0 % and 60 %, progress towards the EU target is insufficient;
- Below 0 % mean that the trend is moving away from the EU target.

As far as possible, indicator trends are assessed over the long-term trend, which is based on the evolution of the indicator over the past ten-year period, and the short-term trend, which is based on the evolution of the indicator during the past five-year period.

**Table 2.** Example of territorialisation of a quantitative target. The target, the reference SDG, the most updated value of the indicator associated with the target and the short (S.T.) and long-term (L.T.) CAGRs are presented.

SDG	Target	Territory	Last value	CAGR. S.T.	CAGR. L.T.
12	By 2030, reduce the share of municipal waste produced per capita by 26% compared to 2004	Italia	488,5 kg/ab.*year (2020)	-5%	9%
		Emilia Romagna	639,9 kg/ab.*year (2020)	4%	21%
		Bologna	562,8 kg/ab.* year (2020)	-8%	16%
		Monte S. Pietro (BO)	439,9 kg/ab.* year (2020)	-159%	95%

Table 2 shows an example of a quantitative target applicable to different territories, which in this case are Italy, Emilia Romagna region, the Metropolitan city of Bologna and the small municipality of Monte San Pietro (inside the Metropolitan city of Bologna district). The study of the trends shows that in the short term (2015 - 2020), apart from the Emilia Romagna regione, all the territories considered increase the per capita production of waste, while in the long term (2010 - 2020), only the municipality of Monte San Pietro has a growth rate which, if maintained, would guarantee the achievement of the objective.

Beyond the targets defined at a higher institutional level, it is necessary to define a set of specific targets linked to the institutional activities of the local and regional governments and consistent with the SDGs, to generate an information and monitoring system useful for measuring any gaps and redirecting political actions for the concrete achievement of sustainability. Implementing territorial strategies for sustainable development means defining a programmatic document based on quantitative targets to be achieved, which take into account the peculiarities of the territory and the political will of the administration, and which are consistent with the SNSvS. These must always consider the objectives set at the higher level (national and international), but at the same time they must take into account the specificities and starting conditions of the territory identified through the positioning described in par. 2.1. Targets must make clear the expected change and be measurable. By way of example, the Emilia-Romagna region has indicated in its regional Strategy all the strategic targets it intends to achieve by 2026. Many are in line with national targets but, for some areas, the region has identified specific targets. In particular, the regional administration has proposed a different target regarding the maximum number of days in which it is possible to exceed the limit concentration of fine particles established by law (35 days / year for Emilia-Romagna vs. 3 days / year for Italy). Because the fact that - due to the morphological aspect of the Po Valley - the 3 days / year target is unrealistic for the region. At the same time it has set a more ambitious target than the national and European one for early exit from the training system (8.5% vs. 9% by 2030) (Emilia Romagna, 2020).

### *2.3 Individuation of the policies and actions associable to the specific targets*

The strategic targets must subsequently be included in the planning tools of the local authorities. In this way it is possible to plan the achievement of the quantitative objectives. In the Economic and Finance Document for the Regions (DEFR) and in the Single Programming Documents for the Local Authorities (DUP), it is necessary to specify the strategic objectives that the Government intends to achieve during the legislature, indicating, for each objective, the expected results annually. The aim is to embed the territorial Strategy into the programming and monitoring tools. The quantitative objectives implemented in the DEFR or in the DUP must be correlated with the national strategic areas and choices, and through them with the global objectives of the 2030 Agenda. Increasingly integrated monitoring and evaluation of regional and local policies, in order to create a coherent and effective multilevel system are necessary. To measure performance, it is necessary to introduce outcome and/or output objectives and consequently impact indicators that

are strictly related to the defined targets. For example, the Parma municipality is one of the first in Italy to implement a direct link between the quantitative targets, defined as in paragraph 2.2, and the Strategic and operational Objectives that define the policy action of the municipality (Comune di Parma, 2020).

#### *2.4 Involvement and dialogue with all stakeholders*

A change of impact requires bringing together the contribution of different actors: public, private and civil society. The achievement of the 2030 Agenda, in fact, strongly depends on the action and collaboration of all the players in the territorial, institutional and socio-economic system.

No public administration can be considered the *deus ex machina* of the implementation of policies and the response to needs in its reference territory. The role of territorial and regional governments has changed, passing from having a predominant function in the provision and direct management of services to having the function of "directing", guiding, and controlling local development. In fact, the subsidiarity network involves public and private entities, for profit and non-profit, who collaborate with the administration in achieving policies and objectives.

Public-private partnerships therefore bring together public bodies, private companies and the third sector, with the aim of contributing to the implementation of projects and initiatives capable of generating positive impacts for the community, often called upon to actively participate in dialogue between the parts.

Local authorities must therefore equip themselves with suitable tools for participatory activities, to obtain shared governance for the entire process. Failure to participate is the cause, in fact, of the inefficiency of choices and actions, which, without the support of the beating heart of the territories (citizenship, universities, third sector, private sector, etc.), struggle to function. As an example, in the Metropolitan city of Milan ASviS, with the contribution of the *Politecnico di Milano*, has organized a co-creation laboratory, involving public and private stakeholders in the discussion of the quantitative targets defined during the process, as well as the discussion of the policy action to achieve the quantitative targets.

### **3. Conclusions**

The rising need to measure and monitor sustainable development for subnational administrations urges the development of a shared framework of goals, targets, and indicators in a systemic way. Consistently with this reflection and with the multidimensional nature of the concept, ASviS developed a "Multilevel approach", which declines the national and supranational programmatic targets on the territorial scale.

According to ASviS, the basis for correct "Multilevel approach" programming provides for a mapping of the local context with respect to the 17 SDGs through the calculation of composite indices. This to summarize the degree of sustainability of the individual territories for each Goal and to compare the performance between the different realities belonging to higher or lower levels, and through the measurement of the distance from the international targets related to the UN 2030 Agenda.

Based on these results, the public and private stakeholders are involved in identifying quantitative territory-based targets, needed to define the commitments of the territories and to monitor the impact of policies with respect to the achievement of the SDGs.

The relevance of this innovative approach is to promote a new type of territorial programming based on quantitative targets and indicators, which can support local public decision-makers in their decision-making process.

## References

- Better Policies for 2030 An OECD Action Plan on the Sustainable Development Goals.* (n.d.). <http://oe.cd/sdg>
- Böhringer, C., & Jochem, P. E. P. (2007). Measuring the immeasurable — A survey of sustainability indices. *Ecological Economics*, 63(1), 1–8. <https://doi.org/10.1016/j.ecolecon.2007.03.008>
- Commission for Economic Policy.* (n.d.). <https://doi.org/10.2863/11396>
- Comune di Parma. (2020). *2021-2023 DUP Parma.*
- Decade of Action - United Nations Sustainable Development.* (n.d.). Retrieved July 15, 2022, from <https://www.un.org/sustainabledevelopment/decade-of-action/>
- Emilia Romagna. (2020). *We build the future together REGIONAL STRATEGY “2030 AGENDA FOR SUSTAINABLE DEVELOPMENT.”*
- European Commission. (2016). *Next steps for a sustainable European future European action for sustainability.*
- European Commission. (2019). *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS The European Green Deal.* [https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF)
- EUROSTAT. (2021). *Assessment of indicator trends Assessment of indicator trends How are trends assessed? How are the assessment results presented?*
- Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., & Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81, 491–502. <https://doi.org/10.1016/J.ECOLIND.2017.05.068>
- I numeri della sostenibilità - Alleanza Italiana per lo Sviluppo Sostenibile.* (n.d.). Retrieved November 20, 2022, from <https://asvis.it/I-numeri-della-sostenibilita/>
- Mazziotta, M., & Pareto, A. (2015). *Comparing Two Non-Compensatory Composite Indices to Measure Changes over Time: a Case Study.*
- Mazziotta, M., & Pareto, A. (2018). Measuring Well-Being Over Time: The Adjusted Mazziotta-Pareto Index Versus Other Non-compensatory Indices. *Social Indicators Research*, 136. <https://doi.org/10.1007/s11205-017-1577-5>
- Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for Composite Indicators Building.* <http://europa.eu.int>
- Nations, U. (2015). *70/1. Transforming our world: the 2030 Agenda for Sustainable Development Transforming our world: the 2030 Agenda for Sustainable Development Preamble.*
- Remarks to High-Level Political Forum on Sustainable Development | United Nations Secretary-General.* (n.d.). Retrieved July 15, 2022, from <https://www.un.org/sg/en/content/sg/speeches/2019-09-24/remarks-high-level-political-sustainable-development-forum>
- Sachs, J. D. (2015). *The Age of Sustainable Development.* Columbia University Press. <https://doi.org/10.7312/sach17314>
- Tarantola, S. (2016). *State-of-the-art report on current methodologies and practices for composite indicator development.* <https://doi.org/10.13140/RG.2.1.1505.1762>

# Sustainable development goals: classifying European countries through self-organizing maps

Cristina Davino, Nicola D'Alesio

## 1. Introduction

Environmental sustainability, despite being the subject of different interpretations (Hueting & Reijnders, 1998; Goodland, 1995), involves the preservation of things and qualities valued in the environment (Sutton, 2004). To achieve this goal, the United Nations (Brundtland et al., 1997) included three goals about environmental sustainability among the proposed 17 Sustainable Development Goals (SDGs). The SDGs related to environmental sustainability are the following: number 13, which refers to climate change and its impacts; number 14, which refers to the conservation of water and marine resources; and number 15, which refers to the preservation of forests. Each of these goals is measured through a set of indicators. An important question is understanding what Europe has achieved in terms of environmental sustainability. In this paper, a mapping of the environmental sustainability within the European territory is proposed using Machine Learning techniques. In particular, Self-Organizing Maps (SOMs), an unsupervised clustering method in the framework of artificial neural networks, are exploited to identify and visualize European countries into a low-dimensional grid (Kohonen, 1982a, 1982b). The analysis considers the indicators related to the three SDGs of environmental sustainability (SDG 13, 14, and 15) and aims to identify groups of countries with similar characteristics through a dimensionality reduction, representing them in a two-dimensional map. The reference year was 2019, except for two indicators updated in 2018 and 2020. To ensure the stability of our results, we built several SOMs with different grids and chose the best one using accuracy measures and a Leave-One-Out procedure. The paper is divided as follows: Section 2 shows the concept of environmental sustainability and the different methods of measurement. In Section 3 there is a description of the data and methodology. Section 4 provides the presentation of the results. All the computations are realized using the R packages *kohonen* (Wehrens & Buydens, 2007), *aweSOM* (Julien et al., 2021), *factomineR* (Husson et al., 2016), and *Factoextra* (Kassambara & Mundt, 2017).

## 2. Literature review

Sustainability has a long and complex history. It was discussed at the end of the eighteenth century as a "derivation from the noun sustenance" (Jenkins & Schröder, 2013). A key point on sustainability is the perspective for the future: it is necessary to manage resources to guarantee them also for future generations (Hueting & Reijnders, 1998). Because of the difficulties to define sustainability, environmental sustainability has also been subject to different interpretations and discussions over time (Goodland, 1995). A proper definition is the following: "the ability to maintain things or qualities that are valued in the physical environment" (Sutton, 2004). This definition seems more appropriate as it allows us to include the sustenance of all facets of physical capital. The definition of environmental sustainability is crucial to provide policymakers with precise information on its development, but an important step of this process is also to understand how to measure it. Efforts to build indicators to measure environmental sustainability have led to the creation of several evaluation exercises. Among the best known there are the SDGs proposed by the United Nations which cover all fields of sustainability (economic, social, and environmental). They are not exempt from

Cristina Davino, University of Naples Federico II, Italy, cdavino@unina.it, 0000-0003-1154-4209

Nicola D'Alesio, University of Campania Luigi Vanvitelli, Italy, nicola.dalesio@unicampania.it

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Cristina Davino, Nicola D'Alesio, *Sustainable development goals: classifying European countries through self-organizing maps*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.17, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 95-100, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



criticism, as they are recent and, according to experts, must be integrated and updated constantly (Hak et al., 2016). Notwithstanding this, they provide an accurate framework of indicators to measure sustainability. In particular, SDGs n°13, 14, and 15 consider indicators aiming to measure environmental sustainability: climate change and its impacts (Climate Action - SDG 13), conservation and sustainable use of the oceans, seas, and marine resources and reduce marine pollution and water acidification (Life Below Water - SDG 14), protection, restoration, and sustainable use of terrestrial, inland and mountain ecosystems (Life on Land - SDG 15).

### 3. Data and methods

#### 3.1 Data

Data for the three considered SDGs are available on the Eurostat website. We used 2019 as the base year (just two indicators of the SDG-15 are updated to 2018 and 2020). A subset of 14 indicators from the set of 21 indicators was used for the analysis because some of them are not available at the national level for each country and/or because they contained more than 80% of missing values. The units of analysis are represented by the 31 countries<sup>1</sup>. Table 1 shows the list of considered indicators, divided by SDGs, with the acronym used in results figures and tables and with some descriptive statistics<sup>2</sup>. The asterisk (“\*”) denotes indicators with negative polarity with respect to the concept of environmental sustainability. Missing data and outliers have not been treated because the algorithm of the SOMs can impute a value for the missing data and isolate the effect of the outliers in the extreme regions of the network. All the considered indicators have been standardized before applying the SOM algorithm.

#### 3.2 Methods

Self-Organizing Maps (SOMs) are artificial neural networks that produce a low-dimensional representation of the input space, allowing a dimensionality reduction (Kohonen, 1982a, 1982b, 1990). They use a neighborhood function to preserve the topological properties of the input space. The SOM algorithm is divided into two phases: the competitive phase and the cooperative phase. In the competitive phase for each input vector, the neuron with the minimum distance from the input is selected and it represents the winner. Although several distance measures are available, the Euclidean distance is the most used (Miljković, 2017). The neurons within a grid interact with each other using a neighborhood function such as the Gaussian function. In the cooperative phase, on the other hand, the weights are modified as topologically related subsets on which similar weight updates are performed. During learning, not only the weight vector of the winning neuron is updated, but also those of its reticular neighbors and, therefore, that end up responding to similar inputs. This is achieved with the neighborhood function, which is centered on the winning neuron and decreases with the distance of the grid from the winning neuron. Once the units (the weights) have been initialized, the training phase starts. SOMs training is done through unsupervised learning that can be realized in a sequential formation (or online algorithm: a single statistical unit is inserted into the network at a time) or in batch modality (or batch algorithm: all statistical units are inserted into the network at once) (Matsushita & Nishio, 2020). In our case, it was preferred the online algorithm. We chose the Euclidean distance as a distance measure and the Gaussian function as a neighborhood function.

---

<sup>1</sup> Belgium, Bulgaria, Czechia, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, the Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden, Iceland, Norway, Switzerland, and the United Kingdom.

<sup>2</sup> VC means variation coefficient.

Table 1: SDGs Indicators.

<b>SDG – 13 Indicators</b>	<b>Acronym</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>VC</b>	<b>Skewness</b>
Net Greenhouse gas emissions*	Net_GHG_Emission	34.80	156.20	80.43	0.34	0.49
Net Greenhouse gas emissions of the LULUCF sector*	Net_GHG_Land	-137.60	172.00	-33.28	2.19	0.97
Contribution to the international 100bn USD commitment on climate related expending	Contr_Intern_commitment	1.00	27.00	14.00	0.57	0.00
Population covered by the Covenant of Mayors for Climate	Population Covered	7.30	91.10	44.09	0.47	0.18
Share of renewable energy in final energy consumption by sector	Share_Ren_Energy	7.05	78.61	25.69	0.70	1.54
Average CO2 emissions per km from new passenger cars*	Average_CO2	59.90	133.00	119.81	0.12	-2.54
<b>SDG – 14 Indicators</b>	<b>Acronym</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>VC</b>	<b>Skewness</b>
Surface of Marine Protected Areas	Surface_Marine_Protected_Area	2.30	45.90	16.99	0.69	0.85
Bathing sites with excellent water quality by locality	Bathing_Sites	12.00	4894.00	648.66	1.66	2.41
Marine waters affected by eutrophication*	Waters_Eutrophicated	0.00	5856.00	616.68	2.45	2.61
<b>SDG – 15 Indicators</b>	<b>Acronym</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>VC</b>	<b>Skewness</b>
Share of forest area (2018)	Forest_Area	10.40	69.90	39.71	0.41	-0.04
Surface of the terrestrial protected areas (2020)	Protected_Area	13.20	51.50	27.27	0.38	0.39
Soil Sealing Index*	Soil_Sealing_Index	0.07	17.08	2.53	1.30	3.00
Biochemical oxygen in rivers*	Oxygen_In_Rivers	0.75	3.60	1.95	0.42	0.55
Phosphate in rivers*	Phosphate_in_Rivers	0.01	0.22	0.06	0.96	1.31

The most widespread accuracy measures used in the SOM framework are the following:

- Quantization error: Average distance squared between the data points and the nodes in which they are inserted. The lower the value, the more accurate the network will be.
- Percentage of explained variance: it expresses the percentage of variance explained by the model. The higher the value, the more valid the model will be.
- Topographic error: measures how the topographic structure of data is preserved on the map. Assuming values between 0 and 1: 0 indicates an excellent topographic representation (all the best corresponding nodes and best seconds are close), and 1 is the maximum error (the best nodes and the best seconds are never close).
- Kaski-Lagus error: It is the sum of the average distance between the points and their best matching prototypes, and the average geodesic distance between the points and their second-best corresponding prototype. The smaller the error, the more accurate our network will be.

SOMs prove to be a useful and innovative tool for our study, being able to reduce dimensionality and provide a two-or three-dimensional representation of European countries in

the different facets of environmental sustainability. There are many studies of the application of these networks in environmental contexts, also in Italy (Carboni et al., 2015).

## 4. Results

After the indicator selection described in Section 3.1, the analysis is carried out through the following steps: identification of the best SOM through the estimation of several SOMs and accuracy evaluation, clustering of countries, visualization, and interpretation of the results.

### 4.1 Identification of the best self-organizing map

It is well known that one of the main drawbacks of neural networks is the selection of the architecture. We decided to train several networks with different numbers of neurons and with a grid compatible with the sample size and to select the best SOM by comparing the accuracy measures. The results in Table 2 showed that SOMs with grids 3x5 and 5x4 have very similar performance.

Table 2 - SOMs trials: evaluation with accuracy measures

Grid	Quantization error	% Explained Variance	Topographic Error	Kaski-Lagus error
3 x 3	4,07	64,28	0,16	5,26
3 x 4	3,28	71,21	0,16	4,92
3 x 5	2,38	79,1	0,06	4,53
4 x 3	2,93	74,24	0,16	5,25
4 x 4	2,97	73,96	0,03	4,15
4 x 5	2,76	75,75	0,06	3,78
5 x 4	2,5	78,04	0,06	3,91
5 x 5	2,54	77,69	0,06	3,47

The choice of the best network between these two SOMs was made taking into account the stability of the results in terms of sensitivity to the specific statistical units (countries). The two networks were trained using a leave-one-out procedure, i.e., they were estimated n-1 times by excluding one country each time. The aim is to assess how sensitive the results shown in Table 2 may be to the exclusion of even one country. Results are shown in Figure 1 where we plot the percentage of variability explained and the quantization error of the 3x5 (left-hand side) and 5 x 4 (right-hand side) networks trained excluding each time a country. We decided to use these two measures because the other two accuracy measures give the same information about the topographic qualities of a SOM. The red lines represent the values of the reference network (with all statistical units and shown in Table 2). Observing the two graphs, it results that the accuracy of the 3x5 SOM improves (quadrant in the bottom right part) by removing 5 statistical units, while the 5x4 SOM is much more unstable as it improves by removing more than half of the observations.

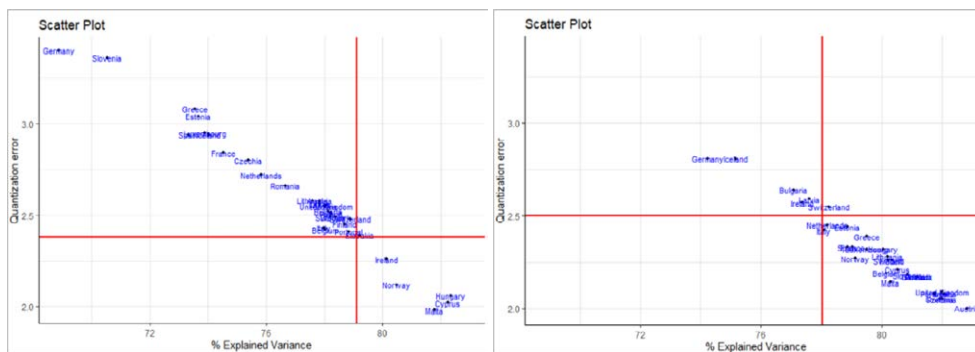


Figure 1 - Scatter Plot of the accuracy measures for the two SOMs (grid 3x5 - left; grid 5x4 - right)

Although of the two selected networks, the 3x5 network is more stable, it is necessary to find its optimal configuration by trying to figure out which of the five countries displayed in the bottom right-hand quadrant is appropriate to eliminate. The proposed procedure proceeds one step at a time starting from the elimination of the statistical unit that provides the most benefit (Hungary) to the one that provides the least benefit (Iceland). Table 6 shows the accuracy measures of these 3x5 SOMs and highlights that the best compromise is obtained just by eliminating Hungary because all the accuracy measures worsen if two or more countries are removed from the analysis.

Table 3 – Grids comparison

Countries	Quantization error	% Explained Variance	Topographic Error	Kaski-Lagus Error
<b>Hungary</b>	2,06	82,34	0,1	4,54
Cyprus	2,83	75,67	0,1	4,54
Malta	2,79	74,73	0,04	3,82
Norway	2,7	74,29	0,11	4,15
Iceland	2,32	77,57	0,08	3,73

## 4.2 Classification of countries

Once a stable SOM has been achieved, it is possible to identify the best partition of countries by applying a clustering procedure. The SOM built without Hungary is shown in Figure 2 where colors highlight the four groups identified using the Ward criterion.

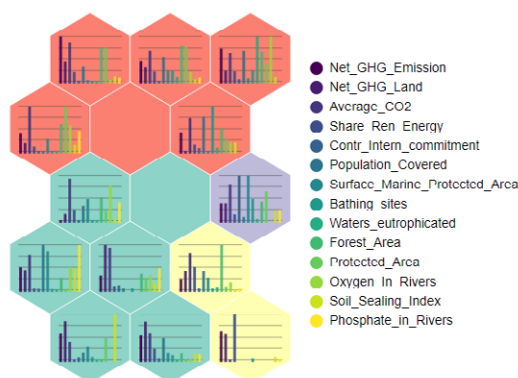


Figure 2 - Visualization of the SOM 3x5 and the partition in four groups

The characterization of the clusters is typically done by comparing, for each indicator, the group averages with the averages on the total sample. Due to lack of space, we report the result of this comparison and the countries belonging to each cluster directly below:

- Group 1, in blue, consisting of Lithuania, Romania, Belgium, the Czech Republic, the United Kingdom, Malta, the Netherlands, Denmark, and Ireland (mainly countries in the continental area), has high net emissions in land use (SDG-13), phosphate in rivers (SDG-14), and land cover index (SDG-15). It can be tagged as the group of "Countries far from achieving all SDGs".
- Group 2, in yellow, consisting of Estonia, Latvia, Finland, Sweden, Iceland, Norway, and Switzerland, (almost all countries in the northern area) has high renewable energy use in the energy sector (SDG-13) and forest areas (SDG-15). It can be tagged as the group of "Countries close to achieving SDG-13 and SDG-15".
- Group 3, in purple, consists of Germany and France and has high marine protected areas (SDG-14) and the highest values of climate change contributions (SDG-13). It can be tagged as the group of "Countries close to achieving SDG-13 and SDG-14".

- Group 4, in red, is composed of Italy, Spain, Portugal, Greece, Croatia, Cyprus, Austria, Slovenia, Bulgaria, Poland, Slovakia, and Luxembourg (these are mainly countries in the Mediterranean region). These countries have a high number of protected areas (SDG-15) but high net emissions (SDG-13). It can be tagged as the group of "Countries close to achieving SDG-15 but far from achieving SDG-13".

The previous classification separates countries closer to achieving a goal and those which are very far from some or all SDGs. This information could help policymakers in assessing what has been achieved so far, what policies need to be implemented to achieve, and which policies in the countries furthest from attainment have either not been implemented or have not been implemented appropriately. The main limitation of this paper is the typical black box effect of neural networks even if the SOMs provide at least a visualization of the grid. A possible future development could be a comparison with other techniques such as cluster analysis, although it will be necessary, in this case, to address the problem of missing data that SOMs are capable of handling. A further problem is the small sample size which has been faced proposing a study of the stability of the results through a leave-one-out procedure.

## References

- European Commission (2020). Sustainable development in the European Union — Overview of progress towards the SDGs in an EU context — 2020 edition.
- G.H. Brundtland et al. (World Commission on Environment and Development, 1987) in *Our Common Future*, Oxford University Press, Oxford.
- Goodland, R., (1995). *The concept of environmental sustainability*. Annual review of ecology and systematics, **26**(1), pp. 1-24.
- Hák, T., Janoušková, S., Moldan, B. (2016). Sustainable Development Goals: A need for relevant indicators. *Ecological indicators*, **60**, pp. 565-573.
- Huetting, R., Reijnders, L. (1998) Sustainability is an objective concept. *Ecological economics*, **27** (2), pp. 139-148.
- Husson, F., Josse, J., Le, S., Mazet, J., & Husson, M. F. (2016). Package ‘factominer’. An R package, **96**, 698.
- Jenkins, I., Schröder, R. (2013). Sustainability in Tourism A Multidisciplinary Approach, *Springer Science & Business Media*, 1.
- Boelaert J., Ollion, E., Sodoge, J. (2021). aweSOM: Interactive Self-Organizing Maps. R package version 1.2. <https://CRAN.R-project.org/package=aweSOM>
- Kassambara, A., Mundt, F. (2017). Package ‘factoextra’. Extract and visualize the results of multivariate data analyses, **76**(2).
- Kohonen, T. (1982a). Clustering, Taxonomy, and Topological Maps of Patterns in *Sixth International Conference on Pattern Recognition*, Munich, Germany, Oct. 19-22, pp. 114-128.
- Kohonen, T. (1982b). Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, pp. 59–69.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, **78**(9), pp. 1464-1480.
- Matsushita, H., Nishio, Y. (2010). Batch-Learning Self-Organizing Map with Weighted Connections avoiding false-neighbor effects in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-6.
- Miljković, D. (2017). Brief review of self-organizing maps in the *40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE.
- Sutton, P. (2004). A perspective on environmental sustainability. *Victorian Commissioner for Environmental Sustainability*, pp. 1-32.
- Wehrens, R., Buydens, L.M. (2007). Self-and super-organizing maps in R: the Kohonen package. *Journal of Statistical Software*, **21**(5), pp. 1-19.

# Individual and social aspects of after-Covid-19 pandemic depression

Pasquale Anselmi, Daiana Colledani, Simone Di Zio, Luigi Fabbris, Egidio Robusto

## 1. Introduction

The Covid-19 pandemic proved to be a social shock. Although its main sanitary effects are going to vanish, many people still struggle to recover their previous normality. All over the world, an over-than-normal incidence of headaches, fatigue, nervousness, and a generalized feeling of bewilderment were found that make it difficult to complete daily tasks. These physical and psychological ailments are sometimes named long-term, or long-Covid effects, not only because they are late consequences of the pandemic, but also because they may last for long.

In this paper, we focus on people's depression. Scholars highlighted signs of mental ailments in people who were infected with the virus, especially among those who showed severe or just temporary inflammation symptoms. Mental ailments were often classified as a post-traumatic stress disorder. Though, other experts observed symptoms such as anxiety, insomnia, and food disturbances also in other people who crossed the pandemic without showing any, or just light sanitary symptoms. Moreover, it is puzzling why so many people showed depressive symptoms even when the pandemic was close to end.

That is why we conducted a social survey on the Italian population to 1) estimate the prevalence of depression feelings among adults; 2) reveal its possible causes; and 3) try to suggest a viable way to get out. The survey was conducted in the second half of 2021 when vaccines had cooled down but not extinguished the infection rate. This suggested that the virus would not definitely vanish even if its effects were "under reasonable control" and normal life could start again.

The research hypotheses of our study were as follows.

H1: *The rate of depressed people in the pandemic time is larger than that reported in the literature for the general Italian population before the pandemic.*

H2: *Depression was related to the disease on people and their families.*

Indeed, psychiatric disturbances have been observed on patients after Covid contagion (Ellul et al., 2020; Pezzini and Padovani, 2020; Iadecola et al., 2020).

H3: *Depression was related to the psychological stress caused by the pandemic.* People who were hospitalised after Covid contagion showed, on top of neuro-physical symptoms, higher levels of post-traumatic stress disorders, anxiety, sleep disturbances, irritability and rarer neuropsychiatric symptoms (Rogers et al., 2020; Mattioli et al., 2021; Mazza et al., 2021; Taquet et al., 2021). We hypothesise that the long lasting pandemic was related to psychological distress and depression also on people uninfected or with lighter contagion symptoms. Studies show that these conditions may come from emotional and mental stress, including: the stigma related to a COVID-19 infection, concerns about infecting other people, the psychological threat of a severe and potentially fatal illness, and social isolation. Also, people who stayed in hospital and in places where they could not interact with others showed higher social isolation and loneliness.

H4: *The pandemic impact was particularly high on population categories that are normally*

Pasquale Anselmi, University of Padua, Italy, pasquale.anselmi@unipd.it, 0000-0003-2982-7178  
Daiana Colledani, University of Padua, Italy, daiana.colledani@unipd.it, 0000-0003-2840-9193  
Simone Di Zio, University of Chieti-Pescara G. D'Annunzio, Italy, s.dizio@unich.it, 0000-0002-9139-1451  
Luigi Fabbris, Tolomeo studi e ricerche, Padua and Treviso, Italy, fabbris@stat.unipd.it, 0000-0001-8657-8361  
Egidio Robusto, University of Padua, Italy, egidio.robusto@unipd.it, 0000-0002-7583-2587

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Pasquale Anselmi, Daiana Colledani, Simone Di Zio, Luigi Fabbris, Egidio Robusto, *Individual and social aspects of after-Covid-19 pandemic depression*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.18, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 101-106, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

*exposed to depression.* In the following, we test the pandemic effects on females, youngsters and higher educated people.

H5: *Ceteris paribus, depression is lower among people who benefitted of personal and social resources and is higher among those who have faced individual and social burdens.*

## 2. Method

### 2.1. Data and participants

A sample of 817 Italian adults was surveyed through a CAWI – Computer Assisted Web-based Interviewing technique. The data collection lasted from June to November 2021. The period can be considered close to the end of the official pandemic in Italy. The participants were recruited from different Italian regions and their participation in the study was anonymous and voluntary. Participants were approached through mailing lists and social networks. Following a snowball sampling procedure, each participant was asked to invite other persons to fill out the survey. The questionnaire was implemented on the LimeSurvey platform and all items were mandatory so that there were no missing data.

The majority of participants (mean age 38.87,  $SD = 18.87$ ) were females ( $N = 464$ , 56.8%), workers (46.4%), students (44.5%), not occupied (9.1%), and with a medium to high education level (basic education .9%, high school diploma 42.8%, university degree 35.9%, post lauream degree 20.4%).

### 2.2. Measures

All participants answered a questionnaire including the following measures, arranged in seven blocks.

**Y:** The Patient Health questionnaire-9 (PHQ-9; Kroenke et al., 2001). Based on the DSM-IV criteria for major depression, it is one of the most used instruments for screening and diagnosing depression. The PHQ-9 consists of 9 items that evaluate the frequency with which people experienced depression symptoms over the last two weeks (4-point scale from 0 “not at all” to 3 “nearly every day”). The instrument has been validated in several contexts and languages, showing good validity, reliability, and diagnostic accuracy (Costantini et al., 2021). A sum score of 10 or larger is usually taken to be indicative of major depression (Manea et al., 2012), with sensitivity between 0.66 and 0.85, and specificity between 0.79 and 0.90 (Manea et al., 2015)

**X<sub>A</sub>:** Health effects of the pandemic. The block includes the following descriptors: having been infected by Coronavirus ( $X_1$ ), showing psychological ( $X_2$ ) or physical ( $X_3$ ) consequences of contagion.

**X<sub>B</sub>:** Personal resources against social shocks. This block includes: possessing a higher education degree ( $X_4$ ), living as a single ( $X_5$ ), living in a couple ( $X_6$ ), clearness of future vision ( $X_7$ ), and resilience ( $X_8$ ), which is a continuous variable obtained by adding up the responses obtained on a 5-point Likert scale to a set of 9 items related to individual self-effectiveness and resilience. These items were selected from the 25-item Connor-Davidson Resilience Scale (CD-RISC scale; Connor and Davidson, 2003) and translated in Italian by authors.

**X<sub>C</sub>:** Personal or familial problems related to social shocks. This block included: having a pre-existing psychic disease ( $X_8$ ), having worked or learned from remote ( $X_9$ ), belonging to a broken family ( $X_{10}$ ), and being scared for viral infection to themselves ( $X_{11}$ ) or to Italy as a whole ( $X_{12}$ ).

**X<sub>D</sub>:** Social resources to face pandemic effects. In this application, the block includes just one variable: trust in scientists during the pandemic ( $X_{13}$ ).

**X<sub>E</sub>:** Social problems caused by the pandemic. The block contained two variables: income ( $X_{14}$ ) and work ( $X_{15}$ ) during the pandemic.

**Z:** Control variables. This block involved the following variables: Gender=Male ( $Z_1$ ), age ( $Z_2$ : three large classes: till 34, 35-64, and 65 and over), and working as an employee ( $Z_3$ ).

### 2.3. Analytic approach

The relations between the considered variables were explored by estimating a path model. In the analysis, the dependent variable was the dichotomized score at the PHQ-9 test (1 = depression diagnosis, 0 = no diagnosis), the exogenous variables were three control variables (gender, age, occupation; see section 2.1), the first level predictors were the four sets of variables labelled as  $X_B$ ,  $X_C$ ,  $X_D$  and  $X_E$  in Section 2.1, while the second level predictors were the variables included in the block  $X_A$ , being hypothesized to be causally closer to  $Y$ .

The model was run using the maximum likelihood (ML) estimator (logistic regression was applied to estimate the paths linking the binary outcome to its predictors). In the analyses, all the direct paths were estimated, and the significance of direct and indirect effects was evaluated employing bootstrapping procedures (5,000 resamples) and the 95% bias-corrected confidence interval. All analyses were performed using Mplus 7.4 (Muthén and Muthén, 2012)

### 3. Results

The analysis of the collected data is reported in Tables 1 and 2. Table 1 shows how depression was diffused in Italy: at the time of the survey, which can be considered close to the end of the pandemic, the rate was very high: 29.6%. Among the 242 individuals obtaining a score of 10 or larger to the PHQ-9, 18 (7.44%) reported having a psychiatric diagnosis.

Table 2 shows the main relations between the criterion and the  $X_A$  variables, on the one hand, and the other regressors, control variables included, on the other hand. The main results are commented as follows.

**Table 1. Frequency distribution of respondents and depression rates in Italy in the second half of the Coronavirus pandemic by characteristics of Italians**

<i>Characteristic</i>	<i>Depression rate (%)</i>	<i>Sample frequency</i>	
		<i>N</i>	<i>%</i>
<i>Overall</i>	29.6	817	100.0
Female	37.9	464	56.8
Age: 18-34	44.4	428	52.4
“ 35-64	14.2	296	36.2
“ 65 or more	10.8	93	11.4
Employee	13.6	272	33.3
Suffered Covid infection	33.7	95	11.6
Suffered psychological problems	55.1	265	32.4
Suffered physical problems	49.0	100	12.2
Family: single	30.3	234	28.6
“ marital couple	29.1	494	60.5
“ presence of children	22.7	331	40.5
“ broken	64.3	14	1.7
Had a psychic disease	72.0	25	3.1

- First, the model showed to be highly useful in the prediction of the outcome variable ( $R^2 = 0.423$ ,  $p < .001$ ). It means that the selected variables enable to thoroughly predict depression diagnosis according to the results of the PHQ-9 score. The regressors can be classified into four groups, according to their social meaning: 1) variables describing the socio-economic conditions of individuals and families; 2) effects of Covid infection and its consequences; 3) chronic diseases and other problematic conditions of individuals; and 4) peoples' psychological strength.



**Table 2. Estimates and significance of the regression coefficients between  $Y$  and  $X_A$  variables and the other predictors included in the model (\*\* $p. < 0.001$ ; \* $p. < 0.01$ ; \* $p. < 0.05$ ;  $^{\circ} p. < 0.10$ ; NS= Not significant; AIC: 12126; Adjusted BIC: 12453; RMSEA < 0.001)**

<i>Predictor</i>	<i>Y: PHQ</i>	<i>X<sub>1</sub>: Infect</i>	<i>X<sub>2</sub>: Psychol.</i>	<i>X<sub>3</sub>: Physical</i>
Z1: Male	-0.06**	0.002 <sup>NS</sup>	-0.094**	-0.044 <sup>°</sup>
Z2: Age (class))	-0.045**	-0.023 <sup>°</sup>	-0.046**	-0.008 <sup>NS</sup>
Z3: Working as an employee	-0.096**	0.006 <sup>NS</sup>	-0.104**	-0.040 <sup>NS</sup>
X1: Infected during the pandemic	0.075 <sup>°</sup>	=	=	=
X2: Showed psychological consequences	0.238***	=	=	=
X3: Showed physical consequences	0.039 <sup>NS</sup>	=	=	=
X4: Higher education degree	-0.052 <sup>°</sup>	-0.014 <sup>NS</sup>	-0.026 <sup>NS</sup>	0.026 <sup>NS</sup>
X5: Living single	-0.142**	0.015 <sup>NS</sup>	-0.169**	-0.071*
X6: Living marital couple	-0.121**	-0.030 <sup>NS</sup>	-0.193***	-0.067 <sup>°</sup>
X7: Resilience score	0.013***	0.007***	-0.005*	0.002 <sup>NS</sup>
X8: Had a psychic disease	0.266**	0.026 <sup>NS</sup>	0.078 <sup>NS</sup>	0.093 <sup>NS</sup>
X9: Active in remote learning or working	0.062 <sup>°</sup>	0.020 <sup>NS</sup>	0.068 <sup>°</sup>	0.036 <sup>NS</sup>
X10: Lived in a broken family	0.232*	-0.064 <sup>NS</sup>	0.466***	0.122 <sup>NS</sup>
X11: Feared for his own infection	-0.059 <sup>NS</sup>	0.058 <sup>°</sup>	0.001 <sup>NS</sup>	0.036 <sup>NS</sup>
X12: Feared for infection in Italy	0.099**	-0.064*	0.031 <sup>NS</sup>	-0.035 <sup>NS</sup>
X13: Trust in scientists	-0.059 <sup>°</sup>	-0.031 <sup>NS</sup>	0.078*	0.045 <sup>°</sup>
X14: Income reduced during pandemic	0.029 <sup>NS</sup>	-0.112*	0.004 <sup>NS</sup>	0.021 <sup>NS</sup>
X15: Work reduced during pandemic	-0.095 <sup>°</sup>	0.130**	0.044 <sup>NS</sup>	-0.028 <sup>NS</sup>

- Female gender and young age highly predicted depressive symptoms: females, at all ages, are more exposed to depression than males; youth are more exposed to depression than the older population. These aspects are socially relevant but are not new in the literature (Elbay et al, 2020), so we add just marginal comments in the following.
- Possessing a higher education title and belonging to a robust family setup showed to be protective factors against depression. The vast majority of the people involved in the present work had a high educational level. However, the positive role of education against depression has also been supported by other studies involving both high- and low-educated individuals (Bjelland et al., 2008; Chen et al., 2020). Also, having had an active role as a remote worker or remote learner was a protective factor. These results are commonsense because a feeling of strength and the active participation in the mass innovation put to the test during the pandemic showed, in various contexts, a positive effect on people's mental health.
- Though, being single is as significant as living in a couple as depression is concerned and having children is unrelated to depression. This may sound counterintuitive because scholars associated living single with loneliness and then with mental disease risks. Instead, it may be conjectured that the pandemic improved singles' trend to leave in a connected community and this protected them against loneliness and, consequently, depression.
- Symmetrically, bad health and a problematic familial situation are risk factors for depression: people who had a full-blown psychic disease or belonged to a broken family showed significant depression risk. Besides, the percentage of the psychically impaired was about 3% and those belonging to a broken family were 1.7% of the sample, while depressed people were close to 30%. Hence, these problematic groups of people are at top risk of depression but do not compose the depressed mass. In a similar vein, reduction of income and unemployment due to the pandemic are negative conditions but income reduction was not significant and unemployment was negatively related to depressive symptoms. As a whole, these results suggest that cogent sources of such a diffused malaise should be searched by scouting other non-conventional social aspects. Fear of viral infection may be one of them, even though it did not relate to one's own

infection but to that of other Italians. This may mean that depression does not follow worries for an own infection but for the vulnerability threat that the virus can hit anytime, now and in the future. Moreover, it could be observed that the percentage of depressed people (about 30% according to the PHQ-9 score) is much higher than that reported in the literature for the Italian population before the pandemic (i.e., 6%; <https://www.epicentro.iss.it/mentale/epidemiologia-italia>).

- Psychological strength descriptors were resilience attitude and trust in scientists. These indicators and a higher education counterbalanced the risk of depression for the majority of Italians. Resilience, in particular, showed to be the mental habit of people is able to push away the malaise.
- Another relevant outcome corroborates the general feeling stemming from our results that social and psychological aspects neatly overcome the infective ones in determining people's mental disturbances at the end of the pandemic. Having been infected mildly correlated with mental disturbances while the physical consequences are irrelevant. Instead, the new and disconnected time people have passed and they fear could continue in a vague future is the determinant factor of depression.
- We could add that reporting psychological damages is quite the same as reporting depressive symptoms. Indeed, the clinical picture obtained from the parallel analysis of (computed) depression and (self-reported) psychological damages (columns 2 and 4 of Table 2, respectively) shows identical risk and protection sources; even the significance levels of the main variables coincide. Also, the proportion of people reporting psychological damages is just a bit larger than that of depressed individuals, though the correlation between the two variables is 0.387, significant but not so cogent. This means that, in Italy, there still is a diffused malaise that the PHQ depression test was not able to capture.

#### **4. Discussion and conclusion**

In this work, we aimed to estimate the depression rate among the Italians at the end of the Coronavirus pandemic and to highlight the correlates of the depression. We have found a rate of 29.6% depression, which is dramatically high. It is much higher among females, the youth, and broken or unstructured families. Similar elevated depressive symptoms and similar risk groups were measured in many other countries at more or less the same time (Klaser et al., 2021; Taquet et al., 2021; Medda et al., 2022), and after the previous COVID pandemic (see the survey in Vindegaard and Benros, 2020).

Clinical follow-ups show that survivors of COVID-19 appear to be at increased risk of psychiatric sequelae, while a psychiatric diagnosis may be an independent risk factor for the disease (Santomauro et al., 2021). Though, the general population studies show just marginal cases of influence of the disease over mental health.

It is to be mentioned that the depression rate varied over time according to emergency situations. It was lower in the early months of 2020 when the pandemic blew but, hoped it, lasted a few months. If we apply the same rationale, the rate should decrease now that people are less afraid of the virus. Though our data showed that the health threat was important at the beginning of the pandemic, when the Coronavirus busted into people's lives, in the long run, it was something else that caused such a generalized malaise and depression. Maybe, it was the threat of hidden long-run consequences of the disease, the risk the virus would recur at any cold season, the lack of socialization and the loss of the sense of community while keeping physical distancing, the perception that the virus is changeable enough to puzzle for long time scientists and governments, the financial concerns for future employment and financial defaults, a never-ending emergency, or a combination of all these sources that may have grown people's insecurity and rendered ineffective their psychological resources. It is certain that, either one was affected by the virus or not, the pandemic has affected everybody in some way.

## References

- Bjelland, I., Krokstad, S., Mykletun, A., Dahl, A.A., Tell, G.S., Tambs, K. (2008). Does a higher educational level protect against anxiety and depression? The HUNT study. *Social Science & Medicine*, 66(6): 1334–1345.
- Chen, F., Zheng, D., Liu, J., Gong, Y., Guan, Z., Lou, D. (2020). Depression and anxiety among adolescents during COVID-19: A cross-sectional study. *Brain, Behavior, and Immunity*, 88: 36–38.
- Connor, K.M. Davidson, R.T. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC), *Depression and Anxiety*, 18: 76–82.
- Costantini, L., Pasquarella, C., Odone, A., et al. (2021). Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review. *Journal of Affective Disorders*, 279: 473–483.
- Elbay, R. Y., Kurtuluş, A., Arpacıoğlu, S., & Karadere, E. (2020). Depression, anxiety, stress levels of physicians and associated factors in Covid-19 pandemics. *Psychiatry Research*, 290, 113-130.
- Ellul, M.A., Benjamin, L., Singh, B., Lant, S., Michael, B.D., Easton, A, Kneen, R., Defres, S., Sejvar, J., Solomon, T. (2020) Neurological associations of COVID-19. *Lancet Neurology*, 19: 767–783.
- Iadecola, C., Anrather, J., Kamel, H. (2020). Effects of COVID-19 on the nervous system. *Cell*,183:16–27.
- Klaser, K., Thompson. E.J., Nguyen, L.H., et al. (2021). Anxiety and depression symptoms after COVID-19 infection: results from the COVID Symptom Study app. *Journal of Neurology Neurosurgery & Psychiatry*, 92: 1254–1258.
- Kroenke, K., Spitzer, R.L., Williams, J.B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9): 606–613.
- Manea, L., Gilbody, S., McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ*, 184: E191–E196.
- Manea, L., Gilbody, S., McMillan, D. (2015). A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. *General Hospital Psychiatry*, 37(1): 67–75.
- Mattioli F., Stampatori C., Righetti F., Sala E., Tomasi C., De Palma G. (2021). Neurological and cognitive sequelae of Covid-19: a four month follow-up. *Journal of Neurology*, 268(12):4422-4428.
- Medda, E., Toccaceli, V., Gigantesco, A., Picardi A., Fagnani, C., Stazi, M.A. (2022). The COVID-19 pandemic in Italy: Depressive symptoms immediately before and after the first lockdown. *Journal of Affective Disorders* 298: 202–208.
- Muthén, B.O., Muthén, L.K. (2012). *Mplus Version 7: User's Guide*. Muthén & Muthén, Los Angeles, (CA).
- Pezzini, A., Padovani, A. (2020). Lifting the mask on neurological manifestations of COVID-19. *Nature Review Neurology*, 16: 636–644.
- Rogers, J.P., Chesney, E., Oliver, D., Pollak, T.A., McGuire, P., Fusar-Poli, P., Zandi, M.S., Lewis, G., David, A.S. (2020). Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. *Lancet Psychiatry* 7: 611–627.
- Santomauro, D. and COVID-19 Mental Disorders Collaborators (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398: 1700-1712.
- Taquet, M., Luciano, S., Geddes, J.R., et al. (2021). Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62354 COVID-19 cases in the USA. *Lancet Psychiatry*, 8: 130–140.
- Vindegard, N., Benros, M.E. (2020). COVID-19 pandemic and mental health consequences: systematic review of the current evidence. *Brain, Behavior, and Immunity*, 89:531–542.

# Spread of Covid-19 epidemic in Italy between March 2020 and February 2021: empirical evidence at provincial level

Fabrizio Antolini, Samuele Cesarini, Francesco Giovanni Truglia

## 1. Introduction

The literature on the determinants of Covid-19 contagion is evidently rather recent and does not always draw generally accepted conclusions in identifying the factors that may explain the differences between territorial areas in the severity of Covid-19 impact (Moosa and Khatatbeh, 2021). The rate of contagion is a phenomenon that depends on many and varied factors that are not easy to interpret and must be analysed considering their spatial component (Cutrini and Salvati, 2021).

To this end, convergence models were used, in which the initial level and growth of observed infections in a certain province were related to the level of infections and the relative growth rate of all other provinces. This model was implemented for all three waves that occurred in Italy from March 2020 to February 2021. The proposed convergence model was constructed by also including environmental (Azuma et al, 2020; Copat et al, 2020) and demographic (Goumenou et al, 2020) factors as controlling elements of a conditional  $\beta$ -convergence (Truglia, 2021).

In the literature, spatial regression models have been widely used in many epidemiological studies (Guo, G. et al., 2020; Liu, X. et al., 2020; Zhao, et al., 2020). To date, however, only a few studies are available that have investigated the close association between sociodemographic and environmental determinants and the spatial convergence of Covid-19 infection incidence. Therefore, this study aims to address the mentioned research gap.

This work further contributes to the study and understanding of the impact of demographic and environmental parameters on the spread of Covid-19 cases by adopting a spatial regression approach.

The work is divided into four sections. The first describes the construction of the panel of data used and their recoding into indicators and indices. The second part circumscribes the spatial approach in the implementation of the conditional  $\beta$ -convergence model to investigate any convergence processes observed in the transmission of contagion between the spatial areal units under study. The third part presents the results obtained. Finally, the fourth part proposes a discussion of the findings and introduces some final considerations and possible implications for future studies.

## 2. Data

In the following analysis, a balanced panel of data referring to the 107 Italian provinces was used. The data on contagion were retrieved and processed from the Civil Protection repository in the 'data-provinces' section. From these, for each of the 107 Italian provinces, the contagion rates for the three waves and their respective durations and distances (in days) were calculated. The spatial context data were collected from the ISTAT data warehouse and the ISPRA environmental data yearbook.

As for the infection rate, this was measured as the simple ratio of the total number of registered cases of Covid-19 infection at period  $t$  - where  $t$  represents the first (I), second (II) and third wave (III) respectively - to a standard reference population of 100,000 individuals.

The other indices relating to contagion (duration and distance), calculated for each province,

Fabrizio Antolini, University of Teramo, Italy, fantolini@unite.it, 0000-0002-3112-524X  
Samuele Cesarini, University of Teramo, Italy, scesarini@unite.it, 0000-0001-7062-1580  
Francesco Giovanni Truglia, ISTAT, Italian National Institute of Statistics, Italy, truglia@istat.it

Referee List (DOI 10.36253/fup\_referee\_list)  
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Fabrizio Antolini, Samuele Cesarini, Francesco Giovanni Truglia, *Spread of Covid-19 epidemic in Italy between March 2020 and February 2021: empirical evidence at provincial level*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.19, in Enrico di Bella, Luigi Fabbri, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 107-112, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

do not require statistical formalisation, and represent, the first, the number of days elapsed between the beginning of one wave and its end, and the second, the number of days between the end of one wave and the beginning of the next. On the other hand, the indices that are assigned the role of explanatory variables and that will be the controlling factors for the convergence of infection are:

- The old-age index (*old\_index*): defined as the simple ratio of the population aged 65 and over to the total population.
- The average temperature (*temp\_av*): defined as the average annual temperature (expressed in C°).
- The density: defined as the simple ratio between the total population and the land area (expressed in number of inhabitants per square kilometre).
- Pollution (*pollution\_av*): defined as the total daily average observed values of particulate matter pm10 and pm2.5 and nitrogen dioxide (NO2), expressed in µg/m3.

As these control variables have different units of measurement, they are standardised for use in the convergence model.

### 3. Method

There are various procedures for analysing territorial convergence. In the present study, the most well-known convergence concepts were used to which reference is made in the bibliography (Barro e Sala-I-Martin, 1992; Mankiw, 1992; Arbia, 2005), including  $\beta$ -convergence. In short, in the literature, this approach originates directly from the neoclassical theory of economic growth theorised by Solow-Swan (Solow et Swan, 1956). This type of convergence describes an economic environment in which a poorer country develops faster than a richer country, in terms of per capita income level. Unlike formal models that require a measure of physical and/or human capital, greater freedom is granted by informal models that are not required to be traceable to the variables brought into play by growth accounting (Alexiadis, 2010). The conditional  $\beta$ -convergence model can therefore be rewritten as follows (equation 1):

$$\ln(Y_{i,t}/Y_{i,0}) = \beta_0 + \beta_1 \ln(Y_{i,0}) + \gamma Z_i + \varepsilon_i \quad (1)$$

Where,

*i*, and *t* denote respectively, the spatial unit and the time reference in which the observation *Y* is measured

$\beta_0$  is the intercept

*Z* is the matrix of the *n* control variables that are assumed to influence the growth rate

$\varepsilon_i$  is the error term at zero mean and variance  $\sigma^2$

$\ln(Y_{i,t}/Y_{i,0})$  is the natural logarithm of the growth rate

$\ln(Y_{i,0})$  is the natural logarithm of the initial level

The  $\beta_1$  coefficient, if statistically significant and of negative sign, indicates the existence of the  $\beta$ -convergence hypothesis.

The  $\beta$ -convergence model thus captures whether territorial gaps, in relation to a specific aspect, increase or decrease over a certain time span (in our study the beginning and end of the three successive waves). This research adopts a method that differs from the conventional convergence strategy by instead focusing on the spatial convergence aspect. In fact, an interesting issue to consider in the territorial convergence analysis is the recognised need to introduce elements that consider functional relationships between provinces. For these reasons, it is therefore appropriate to make use of specific procedures capable of considering the structure of

connections between the units of analysis (Guliyev, 2020). Translated into other terms, the  $\beta$ -convergence model can be transformed in such a way that it considers the spatial proximity of the  $N$  observations by means of a proximity matrix  $W$  consisting of elements  $w_{ij}$  that take on value 1 or 0, respectively in the case that units  $i$  and  $j$  are contiguous or non-contiguous.

The spatial methods that can be constructed from this common basis are many and varied depending on the spatial effects to be investigated. Below we propose the conditional  $\beta$ -convergence model (in matrix form) in the case of spatial autoregressive lag of the dependent variable (SAR) (equation 2).

$$y = \rho W y + \beta X + \gamma Z + \varepsilon \quad (2)$$

Where,

$y$  is the matrix containing the natural logarithm of the growth rate at time  $t$  and province  $i$

$X$  is the matrix containing the natural logarithm of the initial level

$Z$  is the matrix of the  $n$  control variables that are assumed to influence the growth rate

$\rho$  (**Rho**) denotes the spatial autoregressive coefficient

$W$  represents the contiguity matrix of the provinces

$\beta$  and  $\gamma$  are the coefficients to be estimated

$\varepsilon$  is the error term with zero mean and variance  $\sigma^2$ .

It was decided to use a  $W$  contiguity matrix of the queen contiguity type. In this typology, provinces that share at least one side or vertex are considered contiguous (LeSage, 1998).

## 4. Results

Table 1 show the results obtained through the estimation of the spatial autoregressive SAR model implemented for the conditional  $\beta$ -convergence model.

**Table 1.** Results conditional  $\beta$ -convergence (SAR): (a) first wave; (b) second wave; (c) third wave

		Estimate	Std. Error	z value	Pr (> z )
(a)	$B_0$	-3.747	0.414	-9.048	< 2.2e-16 ***
	$\beta_1$	-0.489	0.042	-11.429	< 2.2e-16 ***
	old_index	-0.013	0.047	-0.282	0.777
	temp_av	0.027	0.047	0.585	0.558
	density	0.126	0.049	2.550	0.010 *
	pollution_av	0.212	0.057	3.692	2.22e-04 ***
	gg_I	0.011	0.004	2.347	0.018 *
(b)	$B_0$	-0.327	0.062	-5.217	1.811e-07***
	$\beta_1$	-0.109	0.010	-10.545	< 2.2e-16 ***
	old_index	-0.021	0.010	-2.093	0.036 *
	temp_av	0.002	0.010	0.252	0.800
	density	0.041	0.011	3.713	2e-04 ***
	pollution_av	0.021	0.013	1.649	0.099 .
	gg_II	0.001	0.000	4.253	2.1e-05 ***
(c)	$B_0$	-0.581	0.092	-6.307	2.841e-10 ***
	$\beta_1$	-0.093	0.022	-4.077	4.562e-05 ***
	old_index	-0.016	0.016	-0.998	0.318
	temp_av	0.029	0.017	1.745	0.080 .
	density	0.002	0.019	0.150	0.880
	pollution_av	0.043	0.020	2.158	0.030 *
	gg_III	0.003	0.000	4.288	1.8e-05 ***

Signif. codes: 0 <= '\*\*\*' < 0.001 < '\*\*' < 0.01 < '\*' < 0.05 < '.' < 0.1 < " < 1

Source: author's elaboration of collected data

The regression results show that the coefficient of the initial level of the infection rate  $\beta_1$  is less than 0 and significant for all three waves analysed in this study. This implies the existence of the convergence hypothesis (Baumol, 1986).

Since the spatial regression parameters, unlike with the OLS method, were estimated using the maximum likelihood (ML) method, this does not allow the  $R^2$  index to be used to assess the goodness of fit of the model. In this case, therefore, the goodness of fit of the model is assessed by comparing the AIC statistics (Akaike, 1974) calculated for the OLS and SAR models (Table 2).

**Table 2.** Goodness of fit conditional  $\beta$ -convergence (SAR): (a) first wave; (b) second wave; (c) third wave

		Estimate	p-value
<b>(a)</b>	<i>Rho</i> ( $\rho$ )	0.412	-
	<i>LR test value</i>	27.52	(1,555e-07)
	Wald statistic	35.288	2.843e-09
	AIC	149.06 (OLS: 174.58)	-
	LM test for residual autocorrelation	0.019	0.888
<b>(b)</b>	<i>Rho</i> ( $\rho$ )	0.291	-
	<i>LR test value</i>	12.436	0.0004
	Wald statistic	13.497	0.0002
	AIC	-180.52 (OLS: -170.09)	-
	LM test for residual autocorrelation	1.101	0.294
<b>(c)</b>	<i>Rho</i> ( $\rho$ )	0.263	-
	<i>LR test value</i>	5.254	0.021
	Wald statistic	6.042	0.014
	AIC	-72.22 (OLS: -68.973)	-
	LM test for residual autocorrelation	0.272	0.601

Source: author's elaboration of collected data

The AIC calculated for SAR is always lower than the same measured for OLS. The Rho ( $\rho$ ) is statistically significant as is its relationship to the dependent variable (Wald test). Therefore, the spatial model best fits the data and most accurately interprets the observed convergence process.

## 5. Discussion

The results obtained are robust and consistent with the established body of literature in previous medical studies suggesting that poor air quality creates chronic exposure to respiratory disease. On the other hand, population density, the old-age index and average temperature were not always found to be conditional elements of the observed convergence processes, varying in significance depending on the wave taken as the period of observation, and thus partly confirming what emerged from the reference literature. As far as the spatial delays are concerned, the spillover effects recorded by the parameter  $\rho$  (Rho) for all three waves are significant and are respectively equal to 0.41 for the first wave, 0.29 for the second, and 0.26 for the third. According to these results, therefore, it is possible to state that increases and decreases in the average growth rate in the  $i$ -th province can also be attributed to changes in growth levels in its neighbouring provinces. According to the estimated SAR model, *spillover effects* calculated for population density (0.12) and pollution (0.21) for the first wave are also significant. It would thus appear that provinces with a high population density over the available surface area and above-average presence of substantial air pollutants are directly responsible for the growth of contagions in neighbouring areas. Density retains its spatial influence even during the second wave by

significantly reducing its magnitude (0.04). Pollution (0.02) becomes slightly significant (p-value just under 10%) and decreases its influence in exerting an effect on the growth of contagions in neighbouring provinces. During the second wave there emerges a restraining effect due to the old index ( $\text{old\_index} = -0.02$ ) according to which in provinces in which there is a high presence of individuals aged 65 years or over, relative to the resident population, there is a negative relationship with the growth rate of contagions in the contiguous provinces. Finally, as regards the third wave, a weak (p-value of just under 10%) positive spatial relationship emerges between the observed temperature (0.02) and the level of contagions in the neighbouring areas. Confirmed, on the other hand, is the significance of pollution (0.04) in producing an increase in contagions in provinces sharing a border with a province characterised by high levels of this variable. Finally, all three waves share the significance of the observed durations, respectively 0.01 the first, 0.001 the second and 0.003 the third wave, showing, however, a weak spatial influence on the average rate of contagion growth.

Although consistent with the initially hypothesised framework, however, the results obtained have several limitations and implications for future research. Firstly, some critical elements should be noted in the nature of the dependent variable used. These reflections arise from the fact that it is not possible to know the true population that has been exposed to the virus. A further investigation could examine the actual number of people tested. These data are currently not available at the provincial level, and those at the regional level suffer from multiple counting due to repeated testing of positive cases. Secondly, there are some provinces that have reallocated some positive cases to other provinces due to health facility capacity or registration errors. To address these concerns, the paper proposes an analysis on aggregated wave-level data, but possible biases may still exist. Future studies could implement estimation control procedures, potentially including some dummy variables and retesting the model. A further possible source of bias may be introduced by potential outliers. Results could potentially be driven by a few provinces showing several new cases that are exceptionally far from the average. In addition to all this, it must be remembered that the Covid-19 testing policy in Italy, especially at the beginning of the pandemic, was different over time and in the various provinces. Initially, the tests were performed on suspected patients who presented themselves in hospital and/or on persons who had been in contact with positive cases, later only patients with severe symptoms were tested, and finally the tests were also performed on suspects without severe symptoms. Finally, it should be added that the statistical significance of conditional factors does not necessarily imply causality in the recorded convergence process and based on the characteristics of the data, there is no possibility of testing causality by means of a suitable counterfactual trend (in fact, it is impossible to construct a suitably randomised control group for a phenomenon that is already occurring at the time of the evaluation).

## References

- Alexiadis, S. (2010) Regional Convergence-clubs and Dynamic Externalities. *Italian Journal of regional Science*, 9(1).
- Akaike, Hirotugu (1974). A new look at the statistical model identification. In *IEEE Transactions on Automatic Control* 19(6), pp. 716–723.
- Arbia, G. (2005). Convergence in Per-capita GDP across European Regions using Panel Data Models Extended to Spatial Autocorrelation Effects. *ISAE Working Paper*, No.51. SSRN: <http://ssrn.com/abstract=936327>.
- Azuma, K., Yanagi, U., Kagi, N., Kim, H., Masayuki, M., & Hayashi, M. (2020). Environmental factors involved in SARS-CoV-2 transmission: effect and role of indoor environmental quality in



- the strategy for COVID-19 infection control. *Environmental Health and Preventive Medicine volume*, **25**(1).
- Barro, R. J., Sala-I-Martin, X. (1992). Convergence. *Journal of Political Economy*, **100**, pp. 223-251.
- Baumol, W. J. (1986). Productivity Growth, Convergence and Welfare: What the Long-Run Data Show. *American Economic Review*, **76**, pp. 1072-1085.
- Copat, C., Cristaldi, A., Fiore, M., Grasso, A., Zuccarello, P., Signorelli, S. S., Conti, G., & Ferrante, M. (2020). The role of air pollution (PM and NO2) in COVID-19 spread and lethality: A systematic review. *Environmental Research*, **191**.
- Cutrini E., Salvati L. (2021). Unraveling spatial patterns of COVID-19 in Italy: Global forces and local economic drivers. *Regional Science Policy and Practice*, **13**(S1), pp. 73-108.
- Goumenou, M., Sarigiannis, D., Tsatsakis, A., Anesti, O., Docea, A. O., Petrakis, D., and Calina, D. (2020). COVID 19 in Northern Italy: An integrative overview of factors possibly influencing the sharp increase of the outbreak (Review). *Molecular Medicine Reports*, **22**, pp. 20-32.
- Guliyev, H. (2020). Determining the Spatial Effects of COVID-19 Using the Spatial Panel Data Model. *Spatial Statistics*, **38**.
- Guo, G., Wu, Z., and Chen, Y. (2020). Evaluation of spatially heterogeneous driving forces of the urban heat environment based on a regression tree model. *Sustainable Cities and Society*, **54**.
- LeSage, J.P. (1998). *Spatial Econometrics*. Department of Economics. University of Toledo.
- Liu, X., Sun, T., and Feng, Q. (2020). Dynamic spatial spillover effect of urbanization on environmental pollution in China considering the inertia characteristics of environmental pollution. *Sustainable Cities and Society*, **53**.
- Mankiw, N.G. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, **107**(2), pp. 407–437.
- Moosa, I. A., Khatatbeh, I. N. (2021). Robust and fragile determinants of the infection and case fatality rates of Covid-19: international cross-sectional evidence. *Applied Economics*, **53**(11), pp. 1225-1234.
- Solow, R. M., T. W. Swan (1956) Economic Growth and Capital Accumulation. *Economic Record*, **32**, pp. 334-361.
- Truglia, F.G. (2021). La nuda città. Spillover territoriali dei contagi da Covid-19 a Roma nella “seconda ondata”, in *La metropoli continua Storia e vita sociale del quadrante Sud di Roma*, Materiali e Documenti n.78 ed., La Sapienza University Press, pp. 403-437.
- Zhao, R., Zhan, L., Yao, M., and Yang, L. (2020). A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2. 5. *Sustainable Cities and Society*, **56**.

# New perspectives for the quality of sub-municipal data with the Italian permanent population and housing census

Giancarlo Carbonetti, Stefano Daddi, Giampaolo De Matteis, Marco Di Zio,  
Davide Fardelli, Raffaele Ferrara, Fabio Lipizzi, Enrico Orsini

## 1. Introduction

Over the years, official statistics have shown increasing attention to the strong need for statistical information referring to sub-municipal territorial levels and, in this sense, the Population and Housing Census has always ensured the availability of sub-municipal data useful for territorial analyses, for business objectives and for social, economic and environmental decision-making processes.

Istat modernisation programme introduced the Permanent Census that, differently from the traditional decennial census essentially based on collecting data from people, is strongly based on the integration of administrative and sample data, and planned for providing yearly census results (Falorsi, 2017). This change required the adoption of new methodological and IT architectures with the aim of providing accurate and consistent figures at the various territorial levels.

In this framework, sub-municipal data derives from the integration of the Base Register of Individuals (BRI) and the Base Register of Places (BRP) (Crescenzi and Lipizzi, 2020; Fardelli et al., 2021). The quality of data depends on the quality of the registers and the procedures adopted to integrate and elaborate input data. In this regard, Istat is working to improve the result of the linkage task between the two registers to allocate individuals that, for various reasons, could not be geocoded.

This paper describes the strategy for the Permanent Census of Population and Households (PC) in Italy, with particular reference to the process of determining data at the sub-municipal level, the main criticalities and the solutions proposed for the production of quality information. The results of an experimental study conducted for the imputation of the enumeration area to non-geocoded units and for the production of the first sub-municipal census data are also reported.

## 2. The permanent census strategy and the production of sub-municipal data

Since 2018, ISTAT has been conducting the Permanent Census of Population and Housing. The traditional census has been replaced by a census based on a system of registers supported by sample surveys. Every year, counts at municipal level are disseminated according to the BRI, the BRP and a Population Coverage Survey (PCS). BRI contains information on some demographic variables such as gender, place and date of birth, citizenship, place of residence, derived by administrative data. BRP contains addresses, Enumeration Areas (EAs) and if possible, geographical coordinates.

All other census variables not present in the registers are collected with the traditional census questionnaire each year on household samples on representative sets of municipalities. From the integration of the data in the registers and the data collected on the sample households, census results are produced for different information details down to the municipal level.

The production of sub-municipal data in the Permanent Census is based on the integration of BRI and BRP (henceforth FRAME) which allows to locate individuals and households on the territory and enumeration areas. From the FRAME corrected for coverage errors, population

Giancarlo Carbonetti, ISTAT, Italian National Institute of Statistics, Italy, carbonet@istat.it, 0000-0003-1073-9813

Stefano Daddi, ISTAT, Italian National Institute of Statistics, Italy, daddi@istat.it

Giampaolo De Matteis, ISTAT, Italian National Institute of Statistics, Italy, dematteis@istat.it

Marco Di Zio, ISTAT, Italian National Institute of Statistics, Italy, dizio@istat.it, 0000-0002-6648-6934

Davide Fardelli, ISTAT, Italian National Institute of Statistics, Italy, fardelli@istat.it

Raffaele Ferrara, ISTAT, Italian National Institute of Statistics, Italy, rferrara@istat.it, 0000-0001-7777-3835

Fabio Lipizzi, ISTAT, Italian National Institute of Statistics, Italy, lipizzi@istat.it

Enrico Orsini, ISTAT, Italian National Institute of Statistics, Italy, eorsini@istat.it, 0000-0002-3472-4344

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Giancarlo Carbonetti, Stefano Daddi, Giampaolo De Matteis, Marco Di Zio, Davide Fardelli, Raffaele Ferrara, Fabio Lipizzi, Enrico Orsini, *New perspectives for the quality of sub-municipal data with the Italian permanent population and housing census*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.20, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 113-118, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

counts for sub-municipal domains can be produced. Such integration may fall giving rise to units without an enumeration area. This is mainly due to the quality of address information in administrative sources, problems in identifying and classifying addresses and to linkage errors between the BRI and BRP<sup>1</sup>.

### 3. Process improvement actions

In order to overcome the criticalities of the archives and to make the calculation of sub-municipal data, ISTAT is working on different methodological solutions to improve the FRAME in order to deal with mismatches due to problems in the archives:

- *ex-ante: actions to improve the recognition and linkage of addresses between the BRI and BRP components;*
- *ex-post: deterministic and probabilistic procedures for recovering the enumeration area code of non-geocoded units.*

Ex-ante solutions will be implemented in the FRAME definition process, while ex-post solutions will be used for estimation purposes.

### 4. Procedures for improving address recognition and linkage

The following paragraph describes the techniques of processing addresses not recognized in the BRP entities as part of the construction of the integrated system of registers. The goal is to improve the quality and coverage of the geocoding of the resident population in Italy starting from the administrative archives of the Municipal Registry Lists (MRL).

To this end, new processing processes have been applied based on the use of different address recognition algorithms. Algorithms (normalizers) process input addresses by providing their output recognition according to their own normalized form. The address is characterized by four attributes: location; street; house number; address exponent. Failure to recognize an address is due either to under-coverage of the database on which the comparison is made or to systematic errors in the address string. In particular, systematic errors are treated according to two independent methodologies:

- *Machine learning algorithm for the deterministic parsing of systematic errors;*
- *Probabilistic Record linkage algorithm for matching the street.*

#### 4.1 Machine learning algorithm for the deterministic parsing of systematic errors

The machine learning algorithm is used to have a tool capable of predicting the address string in its locality, street, house number and address exponent in order to identify the systematic error and then, where possible, clean the address.

The probabilistic record linkage algorithm is applied to have a tool to allow the recognition of addresses even in cases where the deterministic process with parsing fails, or to recognize addresses regardless of systematic error.

In detail, address parsing is performed using Conditional Random Fields (CRF), a probabilistic algorithm that allows the construction of a model for the segmentation and labelling of data sequences (Comber and Arribas-Bel, 2019). In the specific case reported here, it is a question of predicting the constituent parts of the address by assigning the corresponding labels to locality, street, house number and address exponent, dividing the individual words of the address into tokens. For labelling, the IOB (Inside-Outside-Beginning) format is adopted which provides for the affixing of positional prefixes to the various tokens. In order to recognize and classify each address token, the sequence of attributes that can formally compose an address must be provided as input to the model. Using NLP terminology, these attributes are called Part Of Speech (POS) and as in a grammar of any language, an attribute indicates the role the word

---

<sup>1</sup> 4.4% of BRI units (as at 31/12/2019), about 2.6 million, were non-geocoded.

plays in the sentence / address. The machine learning algorithm, after training the model, was applied to predict and unpack the address keywords, allowing you to remove systematic errors and have a new string to normalize.

#### 4.2 Probabilistic Record linkage algorithm for matching the toponym

The procedure processes the distinct Street considering the specific form of an Italian address, consisting in the fact that a street is divided in Generic Urban Designations named DUG and Official Urban Designations named DUF.

The procedure compares separately set of DUG and DUF of the address (street) not recognized in the basic statistical register of places, using the form obtained by parsing through CFR described above.

The probabilistic matching algorithm compares the variables of the Street of the unrecognized address with the variables of the Street of the addresses recognized in BRP. In particular, the variable DUG is compared by means of a distance of type Cosine with q-grams equal to 1, the variable DUF is compared by means of a Jaccard distance with q-grams equal to 3 (Fortini and Tuoto, 2020).

The result, obtained by processing the individual provinces, and blocking the Street at the level of the municipality, generates a Cartesian product of combinations. The Cartesian product of the combinations is subject to a probabilistic procedure in order to determine the likelihood ratio ( $w$ ) and the analogous posterior probability ( $m.d$ ) that a pair of Street is a match. It was chosen to make the probability of concordance on the DUG dependent on the distance between the DUFs ( $dnc$ ) to favour the choice of couples with concordant DUG among those with a distance  $dnc$  lower than the given threshold. The posterior probability  $m.d$  for each pair is determined by Bayes' rule and, similarly, the log-likelihood ratio  $w$  is given by:

$$w(dnc, dug, s) = \ln \left( \frac{P(dnc|M)P(dug = 1|M, dnc, s)}{P(dnc|U)P(dug = 1|U, dnc, s)} \right).$$

The threshold level “ $s$ ” is a pre-set parameter as a proportion of the number of roads to be combined on the set of pairs of the Cartesian product. The selection of the candidate pairs is carried out by ordering the pairs of the Cartesian product by decreasing  $w$  value and then choosing, for each street, the pair with the largest  $w$  value.

Subsequently, the associated Street are supervised by revision activities, which consists in identifying for each province the highest value of the probability  $w$  where the matching is doubtful. All Streets above this threshold value are considered to be recognized correctly, so we proceed with the reconstruction of the complete address by adding all the civics and exponents of the Street data and we proceed with the reprocessing in the basic statistical register of places.

### 5. Imputation procedures for non-geocoded units

The following imputation procedures were defined for the treatment of FRAME units not placed in any enumeration area (residual units):

- Family reconstruction: if one of the members is geocoded, we assign to the other non-geocoded members the enumeration area assigned to the geocoded one.
- Spatial approximation (SA): when the coordinates of the address are known, the EA of the nearest geocoded house number is assigned (distance criteria: “ $\leq 10$ ” or “ $> 10$ ” house number).
- Address strings from the 2011 Census (AD2011): the retrieval of the EA is done by searching for the address of residence in the municipality among the municipality addresses in the 2011 Census.

- Real Estate Property (REP): EA retrieval occurs through the real estate property unit owned by the individual.
- Real Estate Rentals (RER): EA recovery occurs through the real estate unit of which the individual is a tenant.
- Probabilistic imputation. It is composed of a sequence of donor imputation steps mainly characterised by different imputation cells. The statistical unit is the household; the imputed value is the EA.

The sequence of imputations steps is:

1. Donor imputation with imputation cells: Street, EA in the 2011 Census.
2. Donor imputation with imputation cells: Street.
3. Random choice of an EAs belonging to the street of non-geocoded household.
4. Donor imputation with imputation cells: EA in the 2011 Census.
5. Donor imputation through random choice of an EA attached to an observed household.

The characteristic of those methods is that of reproducing the observed distributions of the EA with respect to the imputation cells (Little and Rubin, 2019). For example, in step 1, for an household that is in a specific street and that was in a specific EA in the 2011 census, the method reconstruct the behavior (the frequency distribution) of the units that are in the same street and that were in the same EA in 2011. A discussion on geo-imputation can be found in Henry et al. (2008), Dilekli et al. (2018) and Curriero et al. (2010).

## 6. Experimental study of the imputation procedures

The experiment for the assessment of the imputation procedures was divided into 3 phases:

- 1) The different deterministic procedures are applied independently on the same database, so that a comparative evaluation is possible, which is also useful for choosing a possible sequence of methods;
- 2) Based on the evidence of phase 1, an integrated imputation procedure is defined;
- 3) Application of the integrated procedure for assigning EAs to an updated Frame and over a larger set of municipalities. 63 municipalities are selected by different geographical area, population size, level and quality of geocoding; this set includes all major municipalities.

For deterministic imputation methods AD2011, REP, RER, an empirical evaluation is carried out on a subset of data with an observed EA considered highly reliable. The imputed EA is compared with the observed EA. The percentage of concordant EAs is an indicator of the performance of the methods. Similar evaluations are made when considering Administrative Areas (AdminA), each of which consists of the aggregation of neighbouring EAs (Table 1).

Table 1: Percentage of concordant EA/AdminA imputed by AD2011, REP, RER.

Deterministic methods	Frequency and percentage of concordant EA		Frequency and percentage of concordant AdminA	
RER	179,232	55.06%	255,195	78.40%
REP	1,468,139	82.83%	1,612,729	90.98%
AD2011	2,884,880	98.72%	2,899,427	99.21%

For the SA method, the same assessment cannot be followed, but a similar approach is adopted. SA, AD2011, REP, RER are applied independently. Units having at least two methods imputing the same EA are selected (prevalence criterion); the idea is that this EA is enough reliable. The frequency of times the EA imputed by SA is included in the prevalent EA is considered as an indirect evaluation of the performance of SA (Table 2).

Table 2: Percentage of concordant EA/AdminA imputed by SA method (2 distance criteria).

Possible inclusion of the SA method	SA with Distance $\leq 10$ house number		SA with Distance $> 10$ house number	
	EA	AdminA	EA	AdminA
SA included	93.6%	100.0%	77.7%	99.1%
SA not included	6.4%	0.0%	22.3%	0.9%
Total	100.0%	100.0%	100.0%	100.0%

We notice a general good performance, especially referring to Administrative Area level.

## 7. Assessments of the accuracy of sub-municipal counts

For the probabilistic imputation methods, a replication approach is adopted for evaluating the uncertainty of the EA counts. The probabilistic imputation is repeated 100 times. The results are used to compute the Coefficient of Variation (CV) and Confidence Interval (CI) for each Enumeration Area and for each Administrative Area. In addition to the number of individuals in BRI and the percentage of non-geocoded units (NG), the average CV% of EAs by some municipality and the Width of the 95% confidence interval (CI) are shown below (Table 3).

Table 3: Average CV% of EAs by municipalities and Width of the 95% Confidence Interval.

Municipality	Livorno	Genova	Torino	Milano	Venezia	Roma	Trento	Verona
Pop. in BRI_31/12/19 (thousands)	157.3	573.8	871.4	1,394.7	259.3	2,839.4	119.3	259.5
NG%	0.07%	0.10%	0.08%	0.07%	0.43%	0.29%	0.17%	0.15%
CV%_average	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.3%	0.3%
Width_CI_average	0.2	0.4	0.7	0.8	0.2	0.7	0.9	0.9

Municipality	Firenze	Bologna	Catania	Cagliari	Monza	Napoli	Bari	Messina
Pop. in BRI_31/12/19 (thousands)	371.9	392.0	311.1	153.2	124.2	962.7	322.4	229.9
NG%	5.64%	0.39%	0.34%	0.53%	0.36%	2.90%	5.17%	38.84%
CV%_average	0.4%	0.5%	0.6%	0.6%	0.8%	0.9%	1.9%	4.1%
Width_CI_average	1.3	1.6	1.2	0.9	2.1	4.5	8.2	13.1

We notice a general high precision of estimator and very narrow confidence intervals. Only two municipalities have an average error above 1%: “Bari” and “Messina”. They are affected by a high level of units with missing EAs: they have 5.17% and 38.84% missing EAs respectively, while the average of missing EAs in all municipalities considered is around 2%.

## 8. Census data produced at sub-municipal level

After the allocation of the non-geocoded units of the municipalities involved in the experiment, the sub-municipal data referring to the 2019 Census were determined by applying a corrector for under and over coverage errors to the FRAME population. Data for EA and AdminA were obtained as a weighted sum of individuals residing there. The variables or combinations of variables produced at the sub-municipal level are:

- ✓ Population by gender and age group;
- ✓ Employed by gender and age group;
- ✓ Population by educational attainment;
- ✓ Foreign population by age group.

These data are not official but have a provisional character. The data for administrative areas have been sent to the statistical offices of municipalities with more than 100,000 inhabitants that have such areas, and those for enumeration areas only to a few large municipalities that have high quality spatial archives. The municipalities will use these data to carry out spatial analyses and to provide Istat with feedback on the level of accuracy.

## 9. Future developments

The definition processes of the BRI and BRP registers are continuously evolving and, together with the improvement of the quality of the information entering these registers, a higher accuracy of the geo-coding operation of individuals and a reduction of non-geocoded residual units are expected. Further quality improvement is expected from the spatial integration in BRP of dwellings and buildings with individuals and households.

The whole process of enumeration area code imputation will have to become structural in the process of producing sub-municipal census estimates.

Finally, the approach for the validation of the final data will have to be defined, also with the indications coming from the municipalities to which the data have been sent. In addition, the impact on the dissemination possibilities of the final results<sup>2</sup> will have to be assessed.

## References

- Comber, S., Arribas-Bel, D. (2019). Machine learning innovations in address matching: A practical comparison of word2vec and CRFs. *Transactions in GIS*, Vol. 23, Issue 2, pp. 334 – 348.
- Crescenzi, F., Lipizzi, F. (2020). The integration of geographic and territorial data sources into the base register of territorial and geographical entities. *Statistical Journal of the IAOS*, Vol. 36, no. 1, pp. 143–149.
- Curriero, F.C., Kulldorff, M., Boscoe, F.P., Klassen, A.C. (2010) Using imputation to provide location information for nongeocoded addresses. *PLoS ONE*, 5(2).
- Dilekli, N., Janitz, A. E., Campbell, J. E., de Beurs, K. M. (2018). Evaluation of geospatial imputation strategies in a large case study. *International journal of health geographics*, 17(1), pp. 1-13.
- Falorsi, S. (2017). Census and Social Surveys Integrated System. Note by the National Institute of Statistics of Italy, presented at the UNECE/Eurostat Group of Experts on Population and Housing Censuses, Nineteenth Meeting, Geneva, Switzerland, 4–6 October 2017.
- Fardelli, D., Orsini, E., Pagano, A. (2021). The address component of the Statistical Base Register of Territorial Entities. In *Book of Short Papers SIS 2021*, pp. 1206-1211. Pearson.
- Fortini, M., Tuoto, T. (2020). Probabilistic record linkage with less than three matching variables, in *Book of Short Papers – SIS 2020*, pp. 3-8. Pearson.
- Henry, K. A., Boscoe, F. P. (2008). Estimating the accuracy of geographical imputation. *International journal of health geographics*, 7(1), pp. 1-10.
- Little, R. J., Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

---

<sup>2</sup> It is expected that data per enumeration area referring to the 2021 census wave will be released in spring 2024.

# The Land Cover/Use Code of the new Istat Census cartography<sup>1</sup>

Stefano Mugnoli, Alberto Sabbi, Fabio Lipizzi

## 1. Introduction

The renewal process of the Italian National Institute of Statistics (ISTAT) provides the data production through the new Integrated Statistical Registers system (SIR). One of the four SIR registers is the Base Register of the Site (RSLB) that will make it possible to uniquely locate all SIR information. For this reason, ISTAT has planned the implementation of the enumeration areas layer called “microzones”. Therefore, the new microzones layer constitutes the base map to realize the new Census Maps, which represents the reference layer to disseminate SIR data and information (Mugnoli et al., 2018). This paper aims to briefly set out the methodology used to realize the new ISTAT microzones and enumeration areas layers; some legend details will be provided in order to better understand the way in which each polygon is represented on the map. The name ‘microzones’ is related to the fact that the layer is a further subdivision of the ISTAT enumeration areas layer; the latter is divided into very small polygons, homogenous in their Land Cover (LC) and Land Use (LU) aspects; this creates a kind of a plot made up of many micro-areas.

The ISTAT census enumeration areas vector layer, in fact, represents the cornerstone to analyse the Italian territory from a statistical point of view. All the data collected during census surveys are linked to each of about 740.000 enumeration areas drawn on Italy. This dense plot helps us to describe the entire national territory in a very detailed way, particularly in urban areas.

Therefore, in order to improve LC/LU statistics and to better characterize each enumeration area, the ISTAT ATA (Environment Territory) Unit, planned to produce a sort of a microzones mosaic layer described by a land cover/use definition compatible with the LUCAS (Land Use/Cover Area frame Survey) legend. This certainly allows both to define more clearly the homogenous areas contour for the future and to optimize the geo-localization of all census variables. With regard to the above, it is important to remember that again this year ISTAT has just planned a continuous population Census survey. It is thus fundamental to have a very detailed reference cartography for this survey.

Census geographical datasets are essentially used for classifying and characterizing national territory in relation to resident population, buildings, services and industry. Supplementing this information with land cover and land use data, it can be possible not only to produce comprehensive data on land cover/use, but better to calculate some statistical parameters (i.e. population density by masking all the uninhabitable or uninhabited areas) at local and global level too. But not just that: in fact, statistical information at this level of detail can be used to evaluate other important phenomena like soil consumption, urban sprawl (European Environment Agency, 2006), accessibility to territory and the demographic change in population distribution. In short, our product can be considered a sort of “Land Cover/Use (LC/LU) Synthetic Layer”, in the sense of getting together geo-statistical information derived from many different geographic datasets. Therefore, its main use is to support statistical surveys since it is the result of integration and harmonization of different kinds of thematic archives such as administrative, demographic, infrastructure (road, railway, ports, airport, etc.), agricultural Census data and environmental maps.

---

<sup>1</sup> Even if the paper was devised together by the Authors, F. Lipizzi wrote paragraph 1. ‘Introduction’; S. Mugnoli wrote paragraphs 2, Microzones and Enumeration areas LC/LU Legend and References and paragraph 5 ‘Future update’; A. Sabbi wrote paragraph 3. ‘Topology rules and accuracy assessments’ and paragraph 4. ‘Conclusions’



Moreover, the peculiar legend of the map is undoubtedly useful in better understanding the synthesis process. In Italy CISIS<sup>2</sup> (Centro Interregionale per i Sistemi Informatici Geografici e Statistici) has contributed to harmonising geographical and statistical data. One of the most important results is the release of the database “DB Prior 10K” at national level. The database developed by CPSG (Comitato Permanente per i Sistemi Geografici), provides some layers (i.e. streets, railways, hydrography) with the same data structure. Furthermore, in order to implement the INSPIRE<sup>3</sup> directive, the ‘Consulta Nazionale per l’Informazione Territoriale e Ambientale (CNITA) was established.

Therefore, to align with from the above, every geographical ISTAT data is designed to pursue the same purpose: to provide standardised information for the entire national territory.

The final geo-statistical microzones layer was developed through collaboration of many people and after the review of many different intermediate products. In the end, the activity is the sequel of many ISTAT experimentations (Lipizzi and Mugnoli, 2010; Chiochini and Mugnoli, 2014; Mugnoli et al., 2011; Lombardo et al., 2017).

## 2. Microzones and Enumeration areas LC/LU legend

ISTAT enumeration are as are described by a lot of attributes that identify each polygon from an administrative and statistical point of view. There are some codes that can be useful to frame each area in a sort of LC/LU classification. Since 2011 each enumeration area had been identified according to a key related to its main “vocation”. This sort of legend was focused especially on human activities, uses or services for the citizen.

Having considered the need to define a clear and useful LC/LU legend to uniquely describe the entire national territory, the choice has fallen upon LUCAS (Land Use and Cover area frame Survey) because this is a “*survey that provides harmonized and comparable statistics on land use and land cover across the whole of the EU’s territory*”<sup>4</sup>. And not just for this reason, the microzones and enumeration areas class legend has been based on the LUCAS one because it is based on two LC and LU pure legends; moreover, all the map layers at our disposal make it possible to identify each polygon by a LUCAS class. Upon completion of the two layers description, it is easy to transfer the classification to the microzones layer since the latter is a sort of summary of the former. The first draft provides a 45 LC class, mostly at LUCAS level 1. But classifying each microzone is not always simple, especially in the case in which polygons can be referred to LU rather than to LC. For example, it is very difficult to characterize the “green urban areas” on the basis of the LC pure legend, as LUCAS is. Usually green areas are classified on the basis of their use (i.e. amusement parks, community gardens, etc.). Attempts have been made to separate grasslands and woodlands from “green artificial” ones. So, in these cases a specific code, which comes from the fusion by LUCAS LC and LU codes, was created and named *COD\_MZ* for the microzones layer; then each enumeration area has been identified by a single code *COD\_TIPO\_S* that represents a simplification of the *COD\_MZ*.

## 3. Topology rules and accuracy assessment

When different geographical databases are merged into a unique layer, some overlay errors inevitably occur. It is therefore essential to define very strict topology rules upstream. First of all, you have to decide the overlay order of the layer. In our case, in addition to enumeration areas cartography, the basic layer is represented by water (river, lake, lagoons, etc.) and wetlands; above this, railways, streets and buildings in this order; then, agricultural and natural area layers; finally, and if it’s possible, the polygons derived from the vegetation indices calculated starting from

---

<sup>2</sup> For more information regarding CISIS activities: <http://www.cisis.it/>

<sup>3</sup> <https://www.mite.gov.it/pagina/inspire>

<sup>4</sup> For more information regarding LUCAS survey: [http://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS\\_-\\_Land\\_use\\_and\\_land\\_cover\\_survey](http://ec.europa.eu/eurostat/statistics-explained/index.php/LUCAS_-_Land_use_and_land_cover_survey).

ortophotos.

Of course, in so doing, it is necessary to deal with the overlay areas (bridge, road crossings, etc.). Using some simple ArcGIS® 10.7.1 by ESRI analysis algorithms (Intersect and Symmetrical difference), (Law and Collins, 2018; Bolstad P., Manson, 2019), different layers can be merged automatically without topology errors.

It is only thanks to the fact that the topology is correct that it is possible to evaluate the land cover of each class. In Table 1 is shown a summary of land cover surfaces for each Italian region (in percentages) related to the LUCAS legend at level 0.

X,Y<sup>5</sup> tolerance is set at 1m, the same as the enumeration area layers.

An additional benefit in using the LUCAS legend is the possibility to assess the accuracy of the microzones layer by LUCAS points themselves. Class accuracy varies from 72.02% for the woodland to 33.33% for the grassland.

The real problem is due to the number of LUCAS points of the less represented classes. In our case, for example, we have very few points for the “Bare land and lichens/moss” and for the “wetlands”. Moreover, it is clear from the error matrix that there are clear overlaps between natural grasslands (pastures) and agricultural ones.

The microzones layer is completed for all Italian regions and it is now in the pipeline to transfer information to the Census 2021 enumeration area layer.

In Figure 1 a focus on the Census 2021 enumeration layer (Municipality of Florence) at the second LUCAS level; different colours represent different LC classes.

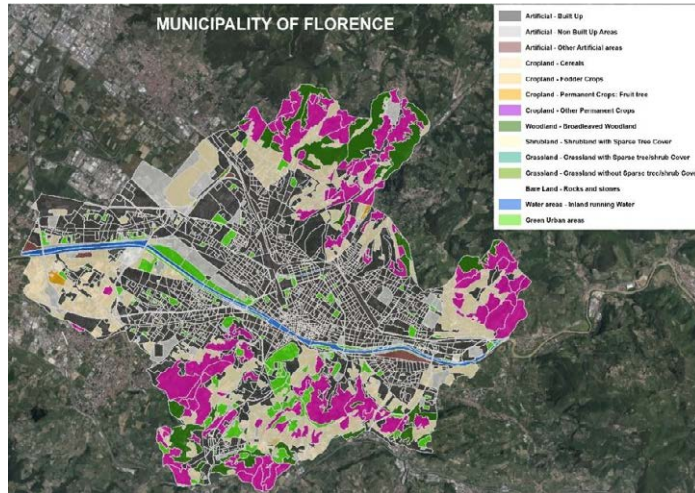
Table 1 – Summary of land cover surface for each Italian region (in percentages)

Regione	Artificial	Cropland	Woodland	Bare Land	Grassland and Shrubland	Water and Wetlands	Green Urban Areas <sup>6</sup>
Piemonte	8,97	40,72	37,55	7,66	3,98	0,99	0,14
Valle d'Aosta	2,11	4,99	31,13	50,74	1,80	9,18	0,04
Lombardia	14,17	42,54	28,16	9,20	1,77	3,82	0,35
Trentino-Alto Adige	2,95	18,84	54,22	15,33	6,79	1,82	0,05
Veneto	14,12	48,68	24,01	2,98	4,33	5,56	0,31
Friuli-Venezia Giulia	9,84	33,10	40,46	6,32	5,56	4,62	0,10
Liguria	10,64	8,55	74,65	1,51	4,55	0,01	0,10
Emilia-Romagna	8,16	58,46	25,86	0,65	3,92	2,70	0,25
Toscana	6,29	39,48	51,69	1,17	0,75	0,53	0,09
Umbria	4,95	47,72	41,25	0,13	3,97	1,95	0,04
Marche	5,93	54,85	33,58	0,41	4,82	0,30	0,13
Lazio	11,67	46,97	29,03	2,14	8,20	1,52	0,47
Abruzzo	5,36	41,92	31,08	3,48	17,84	0,27	0,04
Molise	3,37	50,30	36,91	1,17	7,80	0,40	0,05
Campania	10,83	51,35	28,16	1,01	7,97	0,51	0,18
Puglia	5,90	78,69	5,37	0,95	7,65	1,18	0,26
Basilicata	2,72	55,10	33,30	0,50	7,19	1,18	0,01
Calabria	5,28	35,33	39,43	1,33	17,30	1,28	0,05
Sicilia	6,43	62,45	8,32	3,25	18,96	0,55	0,03
Sardegna	3,48	40,13	23,52	1,69	30,05	1,11	0,02
<b>ITALIA</b>	<b>7,16</b>	<b>43,01</b>	<b>33,88</b>	<b>5,58</b>	<b>8,26</b>	<b>1,97</b>	<b>0,14</b>

<sup>5</sup> The x,y tolerance refers to the minimum distance between coordinates before they are considered equal.

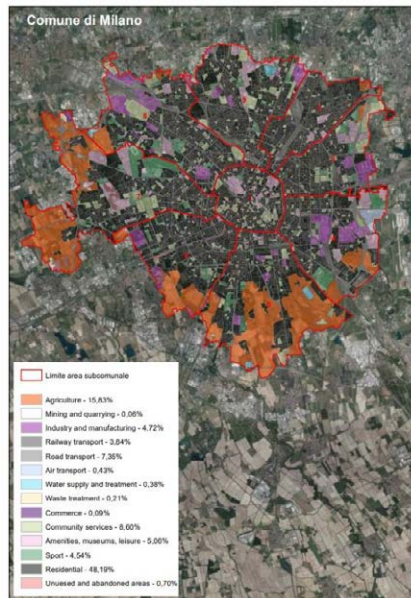
<sup>6</sup> Class not present in LUCAS legend but considered because very important for inhabited localities.

Figure 1 – Municipality of Florence (Enumeration areas 2021) at the second LUCAS level



Another advantage in using LUCAS legend is represented by the opportunity of using a Land Use pure legend too. Below, just as an example, Milan municipality represented on the base of LUCAS LU legend.

Figure 2 – Municipality of Milan (Enumeration areas 2021) according the LUCAS LU legend



#### 4. Future update

From the above, it is clear that the new Census Maps represents a fundamental benchmark for territorial analysis. However, up to now, it is a sort of something static which may lost its original meaning over time.

For this reason, in parallel to the production of the new layer, it is also thought to their dynamic update. So, some studies was carried out in this regard.

The principal of these related the inhabitant areas, which are the most important features of the layer was based on the use of deep CCN (Convolutional Neural Network) U-NET.

The U-NET was first used by Olaf Ronneberger O. (Ronneberger O. et al., 2015) in biomedical

image segmentation. The name comes from how the authors arranged their architecture in an image that resembled the letter “U”. The model implemented in our project is similar to the original model in architecture but has convolution layers that take in the 8 bands in the tiff files used.

This experimentation was sourced on python with keras. The tiff files contain 8 channels in the ortho data which requires us to define the input layer to accept an input that has dimensions of a patch (2D) times 8. A patch is a spot on the original tiff file that is randomly selected and then undergoes a random transformation to produce an analogous patch which only differs from this original patch by the transformation.

The images are 8-band commercial grade satellite imagery accessed from the SpaceNet dataset. The 8 channels are red, red edge, coastal, blue, green, yellow, near-IR1 and nearIR2. In addition to the training images, there are masks corresponding to these images which contain the true segmentation of these images; they contain information about 5 different classes: buildings, roads, trees, crops and water. The images are 16 bit resolution while the mask files are 8 bit.

The model was trained with a batch size of 10. 400 train images and 100 validation patches were generated from 24 training images with their corresponding masks. While there were only 24 images in the dataset, the code performs six random transformations including mirroring, transpose, and rotation to produce enough patches - this process is called image augmentation and increases the dataset. The validation and training losses are important parameters to understanding the fit of the model. In an ideal situation, in the long run at least, both of these quantities have identical values. If the validation loss is greater than the training loss by a large amount the model overfits; on the contrary, if the reverse happens, it is a case of underfitting.

The output of the test image and its corresponding labelled outputs are presented in Figure 2. The colour coding is as given in table 2. The test files also undergo image augmentation and the final result is the averaged out result of the independent predictions of the transformed images. In addition to the segmented images, the mask of the test image is also returned by the program. In some sense this model can be used as an extension to prepare masks for future training images once it has been perfected to a certain degree of training and validation loss.

Table 2 - Colour Coding of output

Label	R, G, B	Colour
Buildings	150, 150, 150	Gray
Roads	223, 194, 225	Pale Yellow
Trees	27, 120 55	Dark Green
Crops	166, 219, 160	Pale Green
Water	116, 173, 209	Sky Blue

## 5. Conclusions

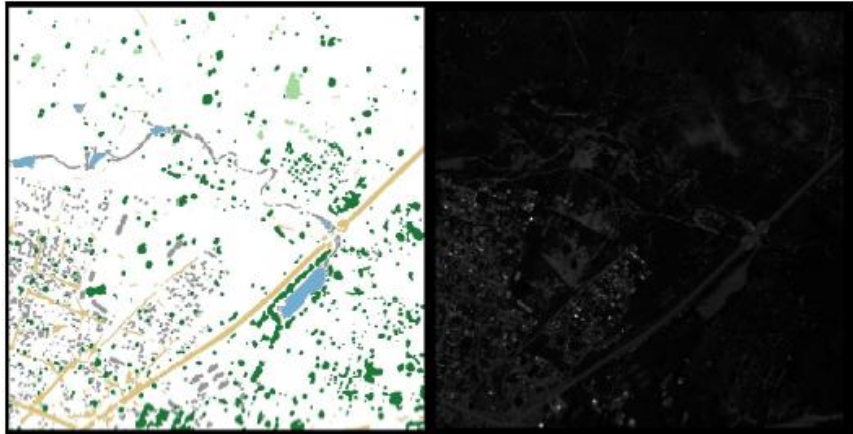
The need to have a homogeneous statistical cartography for the entire national territory is a priority not only for ISTAT but for national and local administrative authorities too. Enumeration areas layer have played a crucial role until now in describing statistical indicators in their territorial and environmental aspects.

However, until now, old enumeration areas layers was not suitable to describe LC and get territorial parameters to some important ISTAT surveys (i.e. agricultural census, transport and services surveys, etc.). So, the new ISTAT microzones and enumeration areas 2021 layer has to be seen as the base map phenomena.

Image processing activities are planned for the future to update all the ISTAT geographic databases, especially by deep learning technics.

The Authors thank all the people of the ISTAT ATA unit who daily works to implement ISTAT geodatabases.

Figure 3 - Sample Output with test image (right)



## References

- Mugnoli S., Lipizzi F. and Esposto A. (2018). New ISTAT ‘microzones’ layer: a new way to read land cover statistics. *Journal of Research and Didactics in Geography (J-READING)*. 2: 95-104.
- European Environment Agency, (2006). *Urban Sprawl in Europe – the ignored challenge*, EEA report n. 10/2006.
- Lipizzi F. and Mugnoli S. (2010). Le statistiche agricole verso il Censimento del 2010: valutazione e prospettive. In: *Proceedings of the conference: Le Statistiche agricole verso il Censimento del 2010: valutazioni e prospettive*. Cassino, Università di Cassino, 26-27 October 2006, Cassino, 2010, pp. 381-394.
- Chiocchini R. and Mugnoli S. (2014). Land Cover and Census integration geographic datasets to realize a statistics synthetic map. In: *Proceedings of European Forum for Geography and Statistics (Krakow, 22-24 October 2014), European Forum for Geography and Statistics Conference, 2014*, <https://www.efgs.info/conferences/efgs/2014-krakow/>.
- Mugnoli S., Chiocchini R., Cruciani S., Esposto A. and Lipizzi F. (2011). Integrazione di dataset geografici di copertura del Suolo e Censuari per la realizzazione di una mappa statistica sintetica. In: *Proceedings of the XV National Conference ASITA 2011 (Colorno, 15-18 November 2011)*, Parma, 2011, pp. 1633-1640.
- Lombardo G., Esposto A., Minguzzi R. and Mugnoli S. (2017). La CSS ISTAT un nuovo strumento per le statistiche territoriali. *Geomedia, XXI*, 2, pp. 26-30.
- Law M. and Collins A. (2018). *Getting to Know ArcGIS Desktop, fifth edition*, Esri Press.
- Bolstad P., Manson S. (2019). *GIS Fundamentals: A First Text on Geographic Information Systems, Sixth Edition*, XanEdu Publishing, Inc.
- Ronneberger O., Fischer P, and Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pages 234–241.

# Trusted smart statistics: new statistics for decision makers. Istat's experience

Massimo De Cubellis, Gerarda Grippo

## 1. Introduction

This paper describes the path followed by the European Statistical System and the Italian Statistical Institute to respond to changes due to the ongoing digital transformation. The digitalization of most daily activities has led to an enormous production of new data, prompting National Statistic Institutes (NSIs) to integrate these new sources within their production processes in order to: (i) enrich their information offerings; (ii) respond to the growing needs of stakeholders; (iii) support decision-making processes in a more efficient way. To achieve these results, NSIs must adapt their organizational, methodological and research paradigms to produce innovative outputs that make structured use of big data. These outputs, called Trusted Smart Statistics (TSS), represent NSIs' response to the changes taking place inside and outside the Institutes.

## 2. Trusted smart statistics: new statistics for decision makers

The last decades have been characterised by profound transformations that have led to significant changes due to the increasing availability and interaction of extraordinary technological innovations. Digitalization has given a strong boost to the data production and to the process datafication of society (Mayer-Schönberger, Cukier 2013). The spread of smart devices in many areas of daily life has led to the generation of increasingly granular data from a spatial and temporal point of view, which represent increasingly interesting sources for public and private organizations. The digital revolution has created a new environment and a new ecology; it has changed our culture in a profound and significant way. All these changes represent not only a digital but also a true ontological revolution (Floridi 2017).

The capillarity of information technology, through the spread of computers, smart devices and the development of the network and digital platforms, have consequences on people's behaviour and on the way in which they communicate, inform themselves, build their beliefs, redefine their behaviors. All these factors contribute to a change that goes beyond digitalization and brings about a transformation of meaning. This causes a significant process-change that requires a rethinking and radical redefinition of concepts, procedures, business, and management model in the sphere of social, political, and economic contexts, as well as within NSIs (Epifani 2020).

This rethinking involves statistical Institutes as institutional subjects, which make a significant contribution to the development of democracy, through official statistics to support public decisions. NSIs are essential actors within knowledge ecosystem because they offer a significant contribution to the strengthening of scientific communication, to the weakening of phenomena such as disinformation and infodemic. NSIs understood the potential of new data sources for statistical purposes: big data make our society measurable, represent a knowledge infrastructure and an opportunity to enrich its offer of information and to respond to the needs of a changing world. The world of NSIs's is mistakenly perceived as a static world with its own rules, specific processes characterized by strict quality criteria. In fact, the world changes the same as the way of doing statistics changes. We went from a world of pre-datification, where NSIs's efforts were focused on data collection, especially surveys and censuses - which have been the main source of data for years - to a world of huge data repositories (Ricciato 2020). In this data deluge,

Massimo De Cubellis, ISTAT, Italian National Institute of Statistics, Italy, decubell@istat.it, 0000-0002-8197-7280

Gerarda Grippo, ISTAT, Italian National Institute of Statistics, Italy, grippo@istat.it, 0000-0001-7568-6252

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Massimo De Cubellis, Gerarda Grippo, *Trusted smart statistics: new statistics for decision makers. Istat's experience*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.22, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 125-128, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

NSIs need to extract reliable information from many data sources, as Weinberger suggested in 2011 when he declared "Information represents to the data what wine represents to the vineyard: the delicious extract and distillate" (Weinberger, 2012) Official statistics perform a public service. Data are a collective asset, a public good; recent pandemic events have taught us that having good data allows us to arrive at more effective, timely and citizen-friendly decisions.

To make the choices of decision makers more effectively meet collective needs, new data governance and innovative business models are needed.

The traditional paradigm based on survey data statistics model does not fully meet the new information needs because, concretely, it is neither adaptable to a changing environment nor compatible with the new social infrastructure represented by digital. In this sense, we recall the words with which Mariana Kotzeva, Director General of Eurostat, opened the 13th National Statistics Conference: "Statistics follows life"; NSIs must be in the world if they want to follow and tell life through data. The so-called datafication of society has led to the spread of new data players in both the public and private sectors outside the Official Statistics system. In the pre-datafication world, the NSIs held a monopoly on statistical activity; the only alternative to official statistics was the absence of statistics. Nowadays, NSIs are one of many data-producing entities within the complex information ecosystem. Various actors in the public and private sectors are producing new data and offering alternative statistical viewpoints on emerging phenomena. Official statistics, in competing with other producers of statistics, must keep its institutional role. They must continue to produce official statistics, even with the help of new data, ensuring the same levels of quality, relevance, accuracy, and reliability the same as traditional statistics. In other words, NSIs face a twofold challenge: on the one hand, they must take advantage of the enormous availability of externally produced and collected data, and on the other hand, they must maintain the same high degree of quality in the statistical information they produce. In a context where the amount of information available to users is increasing, it is only the recognition of the quality of the data, and the institutional role of those who produced it, that can enable users and decision makers to navigate an increasingly crowded information ecosystem. The relevance of statistical information, its timeliness and usability, are crucial for building a relationship of trust with users. The response that official statistics gives to all these questions to stakeholders, decision makers, and users is Trusted Smart Statistics, an expression coined by Eurostat to indicate the mature stage of producing statistics with big data. The new model for European statistics involves greater integration of the information produced and strong use of statistical registers and big data—a holistic approach that aims to provide new, more effective and efficient tools to support decision makers. The starting point of this strategic path is the Scheveningen Memorandum, sanctioned within Eurostat in 2013. This memorandum, formalized the need for all European Statistical Institutes to consider big data sources as new sources for official statistics, launching experimental projects aimed at understanding how to exploit the big data potentiality. In fact, the European Statistical System network (ESSNet) has implemented several experimental projects such as: Essnet Big Data I, Essnet Big Data II, ESSnet Towards Trusted Smart Statistics, and Essnet Smart Surveys. The use of new data sources has been for several years at the centre of the European NSIs agenda. It has required in all NSIs an experimentation phase for studying appropriate methodologies to exploit the use of big data sources, considering the issues related with privacy constraints. Eurostat enables and contribute to these activities in both the design and execution phases, within the framework of official statistics innovation. The term Trusted Smart Statistics was proposed by Eurostat to represent the evolution of traditional statistics and was officially adopted by the European Statistical System in the Bucharest Memorandum on October 12, 2018 during the 104th Directors General of the National Statistical Institutes (DGINS) conference. The Bucharest Memorandum helped to enhance and formalize the contribution of big data in terms of validity, accuracy, and reliability of outputs. The term Smart Statistics refers to multi-source and multi-output statistical production systems that use innovative technologies aimed at flexibly integrating big new data sources into statistical production. The reliability of statistics, to which the term trust refers, is closely linked to the reliability of the institution that produces them. It is based on: (i) compliance with standards for data processing and privacy; (ii) infrastructure that enables data processing (iii) methodological characteristics; (iv) quality guarantees of the entire

production processes. Historically, NSIs have always had the full control of the entire statistical production process because it took place totally in-house. The in-house management of the statistical production process from the direct collection of data from respondents to the dissemination of the statistics produced enabled NSIs to be able to guarantee the reliability, quality, and relevance of the data collected and the methodologies applied, and all the standards necessary for the statistics produced to be called “official”. The production of statistics with the new data sources, often require the use of data collected and held by third parties (e.g.: mobile phone operators). However, it is necessary to maintain the same levels of quality and the same characteristics that make it possible to be able to ensure the official nature of the statistics produced and the trust that users have with respect to the institutional role of NSIs and their statistics. To maintain this level of confidence and ensure the same levels of quality and relevance as traditional surveys, an adjustment of the entire statistical production process is essential. If some steps in the process (data collection and processing) are external to the NSIs, they must still be designed and controlled by the NSIs themselves.

The release of the first outputs with the use of big data, experimental statistics, and the comparison between different NSIs within the EssNet projects, has made NSIs aware on the potentiality of new data sources. The use of these data not only requires strictly technical capabilities and more powerful IT infrastructure, but also requires investment in the different areas of which individual organizations are composed (methodological, organizational, legal). For example, new data sources require the following new methodological approaches: (i) to transform raw data into statistical information and concepts; (ii) to use data that were not designed and collected for statistical purposes; (iii) to overcome coverage issues; (iv) to integrate new data sources with traditional ones. The character of timeliness and temporal and spatial granularity of TSS will enable policy makers to make decisions based on much richer data than those produced with traditional statistics. New data sources have an impact both within individual organizations, in terms of organizational adjustment, and externally, in terms of the ability to produce new products to support public decision makers more effectively. TSS enable decision makers to have more timely access to data and statistics in different sectors; up-to-date statistics also enable decision makers to implement government policies with more accurate spatial detail. The official nature of TSS would give decision makers an opportunity to put new phenomena on their policy agenda. In addition, TSS helps to give a new role to citizens. We said that big data make society measurable, put humans at the centre, create a new digital humanism, and can give rise to citizen statistics as new processes institutionalized by the new social infrastructure represented by digital. TSS become the product of a trust-based exchange between citizen and NSIs. Citizens become, through their daily online and offline actions, "measurable": they become data producers and statistical users, at the same time. Through their active participation in smart surveys, they can provide smart data to support the production of TSS. In order to enhance the role of citizens, it is appropriate for NSIs to establish a "social pact" with the citizens themselves, enabling the NSIs to collect data from citizens and return it to them in the form of useful information.

### **3. Istat’s experience**

After the adoption of the Bucharest Memorandum, a reflection began in Istat on how to govern this innovation process. The release of the first outputs with the use of big data, experimental statistics, the debate between the different NSIs in the European Statistical System within the EssNet projects, has made it possible to acquire the awareness that the use of new data sources, it not only requires strictly technical skills and more powerful IT infrastructures, but requires investments in the various sectors that make up individual organizations. For Istat, the production of TSS represents a highly innovative strategic objective, both scientifically and organizationally for the following purposes: (i) enrich the supply of information in terms of timeliness and spatial granularity; (ii) efficiency, due to the automated integration of data sources and flows; (iii) the ability to capture new phenomena that cannot be measured by surveys or administrative sources alone; (iv) provide answers to stakeholders; (v) reduce the statistical burden on respondents; (vi) train or integrate new skills needed to extract information from the new data sources to contribute in a coherent way to the building of a new organizational model. All these factors are crucial to enhancing the relevance and reputation of Official Statistics through recognition of the unique role of



Istat in terms of the reliability of the statistics produced and the transparency of the production processes. Istat has established a specific Center with the purpose of guiding statistical production activities toward the Trusted Smart Statistics production system. The Center is an agile organization, whose interdepartmental character makes it possible to overcome organizational fragmentation. It represents the point of connection and monitoring of all activities aimed at building the new production system. An internal Steering Committee, consisting of the Institute's top management and responsible for the process of TSS strategic analysis, heads the Center. The strategic decisions on activities and investments, are formalized in a Roadmap. The Roadmap is a strategic document containing the planning of activities aimed at building the TSS production system. At the organizational level, the TSS Center has the task of designing a sustainable organizational structure to help and support the individual company components in working in a systemic, coherent, and synergistic way with the aim of creating the new Trusted Smart Statistics production system. This means that each organizational dimension is involved in the changing of the business model process: legal, methodological, communication, strategic planning, human resources, and skills development. All "cross-divisional" directorates must support the TSS production system to ensure its functioning and to ensure the release of new products with the same level of reliability and quality as traditional statistics. In Istat, to monitor the adjustment of individual organizational dimensions, a monitoring and guidance framework aimed at measuring the organizational maturity status of individual components of the "new" Trusted Smart Statistics has been implemented as a new production system. This tool will support the monitoring phases of the actions, implemented by the Istat individual organizational structures, on the strategic and operational levels, aimed at building the TSS system. The results of the first monitoring revealed how important the organizational component is. In addition to highly technical factors such as IT infrastructures and sound methodological systems, it is important that the new paradigm is supported by organizational changes at various levels. Communication, the legal sector for the definition of aspects related to data access and the ethical use of data, the human resources sector, the planning of objectives are all dimensions involved in the implementation of the new production system.

At the European level, there is an intense debate on this issue.

Statistical Institutes have already achieved promising results but are now facing new challenges, which require ever stronger interactions and collaborations with other public and private actors. These are paths already undertaken, but which must necessarily be followed to the end in order to ensure that the wealth of data produced that we all produce daily can be transformed into statistical information that can be trusted and become a common good.

## References

- Epifani S. (2020). *Sostenibilità digitale*, Digital Institute, Book
- European Statistical System Committee - Eurostat (2018). Bucharest memorandum on official statistics in a datafied society (Trusted Smart Statistics), Memorandum
- European Statistical System Committee - Eurostat (2013). Scheveningen memorandum on big data and official statistics, Memorandum
- Floridi L. (2017). *La quarta rivoluzione*, Raffaello Cortina Editore, Book
- Mayer-Schönberger V., Cukier K. (2013). *Big data. A revolution that will transform how we live, work and think*, Garzanti, Book
- Ricciato F., Wirthmann A., & Hahn M. (2020). Trusted Smart Statistics: how new data will change official statistics. *Data & Policy*, 2, E7, Journal Article
- Weinberger D. (2012). *Too big to know*, Basic Books, Book

# Quality of life in Health Care: focus on patients

Laura Benedan, Angela Digrandi, Paolo Mariani, Cinzia Pilo, Mariangela Zenga

## 1. Introduction

Health-Related Quality of Life (HRQoL) is a well-known concept collecting aspects of overall quality of life related to physical or mental health (Centers for Disease Control and Prevention, 2000; Selim et al., 2009). HRQoL can be defined as “an individual’s or group’s perceived physical and mental health over time” (Centers for Disease Control and Prevention, 2000). On the individual level, HRQoL includes physical and mental health perceptions and their correlates—including health risks and conditions, functional status, social support, and socioeconomic status. On the community level, HRQoL includes community-level resources, conditions, policies, and practices that influence a population’s health perceptions and functional status.

The achievement of a good HRQoL is recognised as an essential aim of health assistance, regardless of the pathology and the administered therapy (Asadi-Lari *et al.*, 2004). HRQoL is a pivotal parameter used by clinicians to evaluate how treatments and therapies influence patients’ functionality and emotional state, aiming to ameliorate interventions and their outcomes. HRQoL is determined by indices assessed by administering questionnaires that can be either generic or disease-specific (Patrick & Deyo, 1989; Rabin & de Charro, 2001; Ware, *et al.*, 2016). These questionnaires have become an important component of public health surveillance and are generally considered valid indicators of unmet needs and intervention outcomes. Currently, the majority of the HRQoL questionnaires are designed with the main contribution of clinicians and, therefore, include items that are focused on the disease rather than on its multifaceted impact on people’s life. These tools are useful for clinicians in determining the best clinical approach but may fail to truly grasp the patients’ perspective, needs, aspirations, perceptions and emotional state, resulting in a major drawback that sets medical care on clinical parameters alone. The patient’s self-assessed health status may be a more powerful predictor than many objective health measures. Unfortunately, a proper tool defining HRQoL from the patient’s perspective is missing.

The present paper aims to propose a methodology to define a bottom-up patient-designed HRQoL questionnaire.

## 2. Methodology

The demand to create an HRQoL questionnaire stemmed from the request of a rare disease patients’ association. The project’s first step consisted of examining the existing scientific literature to understand what was already known and what instruments are used nationally and internationally. After that, a pseudo-Delphi study was carried out.

The Delphi method, a flexible and iterative process, helps collect experts’ opinions in health research (Trevelyan & Robinson, 2015). It was chosen to ensure patient participation and foster the convergence of opinions through the iterative structure, i.e. the collection of experts’ opinions through multiple iterations, to allow the participants to review their evaluations at least once after a comparison with the response of the group (Pacinelli, 2008). However, in a traditional Delphi study,

Laura Benedan, University of Milano-Bicocca, Italy, laura.benedan@unimib.it, 0000-0003-0427-2487

Angela Digrandi, CNR, Italy, a.digrandi@iriss.cnr.it

Paolo Mariani, University of Milano-Bicocca, Italy, paolo.mariani@unimib.it, 0000-0002-8848-8893

Cinzia Pilo, REB Onlus Foundation, Italy, cinzia.pilo@fondazionereb.com

Mariangela Zenga, University of Milano-Bicocca, Italy, mariangela.zenga@unimib.it, 0000-0002-8112-5627

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Laura Benedan, Angela Digrandi, Paolo Mariani, Cinzia Pilo, Mariangela Zenga, *Quality of life in Health Care: focus on patients*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.23, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 129-133, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

participants are polled individually, generally via self-administered questionnaires over two (or more) rounds, and no face-to-face meeting is scheduled (Boukdedid *et al.*, 2011). In the present study, the connotation “Pseudo-Delphi” should be applied because complete anonymity of participants could not be granted as all the group discussions were organised via “face-to-face” virtual meetings. Hence, all the recruited experts could participate and contribute to the group discussion. Nonetheless, a private (and completely anonymous) evaluation of all the questionnaire’s items was granted after each meeting so that every person could critically analyse, re-consider, make suggestions, express comments and provide individual responses without any social pressure or compliance effect that may conversely arise during the group discussions. For more details on the overall study procedure, see Bartolini *et al.*, 2021, and Benedan *et al.*, 2021.

The multidisciplinary panel of experts comprised a Delphi master, six patients or patients’ caregivers, two clinicians recognised as international key opinion leaders for their disease-specific expertise, a psychologist, and a statistician.

A first group meeting was organised to discuss every step of the project, the main topics to cover, and the primary aim to be achieved. Successively, the patients and clinicians were asked to provide a list of spontaneously generated items to describe different areas of the patient’s HRQoL. The results were presented in the first roundtable session to discuss all the implications of daily living with the disease openly. On this occasion, great care was taken to ensure a comprehensive and accurate understanding of the experts’ points of view.

Seven domains were identified and endorsed by the group (see Table 1 for a description of each domain).

**Table 1: Questionnaire domains**

<b>Domain</b>	<b>Description</b>
<i>Physical</i>	It includes the most relevant aspects in terms of health and physical well-being.
<i>Functioning and autonomy</i>	It refers to self-sufficiency and includes statements about the ability to perform common routine actions.
<i>Psycho-emotional</i>	It refers to psycho-emotional well-being, including emotions, thoughts and feelings.
<i>Family</i>	It refers to the relationships with parents, siblings, or other family members such as partners and children if it applies.
<i>Relational</i>	It includes statements about relationships and frequent interactions with people who do not belong to the family (e.g., friends, classmates, colleagues, strangers on the street, etc.).
<i>Work and economic</i>	It includes statements referring to the work context and the financial implications of the disease.
<i>Medical care and assistance</i>	It refers to disease-related healthcare, including medical and nursing assistance.

After defining the domains and examining the main topics, a first questionnaire (Q1) was created. Respondents were required to rank them within each domain according to their importance.

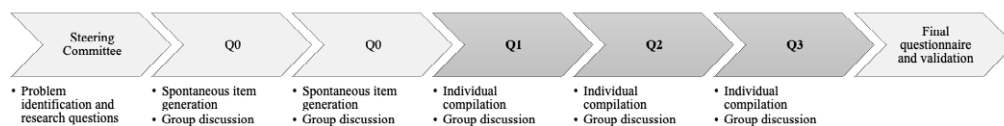
Therefore, for every domain, the rating may range from a minimum of 0 to a maximum equal to the number of items in that domain (Physical = 14; Functioning and autonomy = 15; Psycho-emotional = 13; Family = 12; Relational = 9; Work and economic = 11, Medical care and assistance = 6). They were also required to comment on the clarity and specificity of each item, to write any potential new item, and to report any missing information that might have been included. The main aim of this phase was to exclude any irrelevant items to shorten the entire set of questions and have a more manageable questionnaire. Each expert responded anonymously to the questionnaire and returned it to be discussed in the second Delphi round. All the answers were carefully examined, and a ranking was created for every item within each domain according to the degree of importance indicated by the participants. The results of this analysis were discussed in the group, and further questionnaire refinement was made. Some items were changed or rephrased for greater clarity; others were merged or removed because of their lesser importance.

A new questionnaire (Q2) was defined, considering all suggestions from the group meeting. The previously identified core domains remained unchanged, but some new items were suggested and inserted. At this stage, each participant was asked to rate both the degree of agreement and the degree of importance of each item on a four-point Likert scale (“Not at all”, “A little”, “Quite a lot”, “Very much”). This step is necessary to remove some irrelevant statements and evaluate the order in which the items are presented. In addition to the abovementioned seven domains, some specific questions were inserted about the type of the rare disease diagnosed and some socio-demographic information. Finally, an overall Quality of Life satisfaction question was asked.

The results of this phase were presented to the group to define the questionnaire structure further and prepare the new version (Q3) that each participant anonymously filled in.

Figure 1 illustrates the flow of the project from the beginning to the validation phase of the final questionnaire. For the purposes of the present study, we will focus on the Delphi rounds involving the development and refinement of the questionnaire from Q1 to Q3. The following section will provide a thorough description of how the questionnaires changed through the iterative process.

**Figure 1: Flow chart of the project**



### 3. Results and Discussion

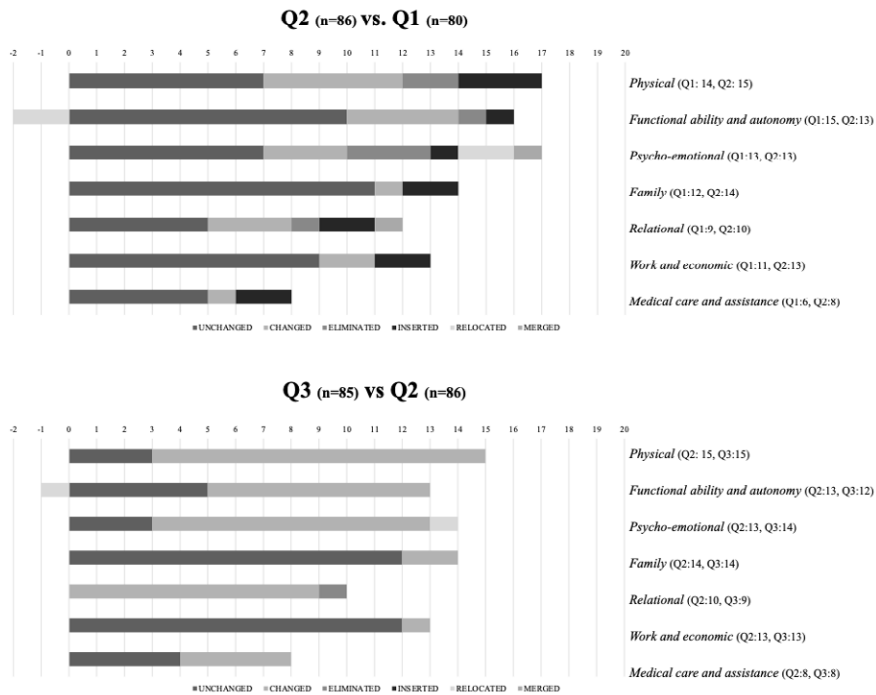
The first questionnaire (Q1) contained 80 items grouped into the seven previously identified core domains. This first version was carefully reviewed, and several changes were suggested by the panellists. After an in-depth examination of all the items, through private compilation and group discussion, many adjustments were made. From the original list of statements, 54 (68%) items remained unchanged, 19 (24%) were rephrased (e.g. “I might have children” was changed to “I can have children”), and 7 (9%) were eliminated - some were merged into one for the sake of synthesis: for instance, “I feel frustrated”, “I feel helpless”, and “I feel demoralised” were merged into a single one (“I feel helpless, demoralised/or frustrated/or”).

It should be noted that the changes concerned not only the questionnaire as a whole but also the individual domains. In fact, two items were moved from one domain to another: for instance, “I feel I’m self-reliant” was moved from the functioning and autonomy domain to the psycho-emotional domain. In addition, 13 new statements were inserted in the following version of the questionnaire. Considering all these changes, Q2 was composed of 86 items. The order in which the items were presented changed according to the importance of each statement within the domain so that the

more important items were the first, as established in the previous round. The same private examination and group discussion process aimed at reviewing the items was applied to Q2. Again, several changes were suggested, examined and, whenever approved by the group, introduced in the new version of the questionnaire. Forty-six (54%) items remained unchanged, while 39 (45%) were rephrased to be more easily understandable and clear. One of these items was also moved from the functioning and autonomy domain to the psycho-emotional domain (“I can have children”, which was also rewritten as “I worry about being able to have children”). Only one sentence was removed, and no new items were suggested.

The new version of the questionnaire (Q3) comprised 85 items. As in the previous rounds, each participant anonymously filled in the questionnaire and then the results were discussed in the group. Figure 2 shows the comparison between Q1 and Q2, and between Q3 and Q2. It can be noticed that a process of progressive refinement and definition was carried out from one iteration to the next, affecting all the domains.

**Figure 2: Comparison between Q2 vs. Q1 (n=80) and Q3 vs Q2.**



Source: elaboration of research data, collected from June to August 2021

#### 4. Conclusions

The present study is part of a more extensive research project to develop a valid and reliable questionnaire to assess the HRQoL of patients affected by a rare disease. In order to grasp the point of view and the patient’s subjective experience beyond clinical symptoms, a pseudo-Delphi study was carried out. The questionnaire’s items were progressively created, elaborated and refined through the iterations, round after round. The changes made in the wording of the items from the first version of the questionnaire to the third one were described. The result is an HRQoL

questionnaire that goes beyond the physical symptoms and the clinical evolution of the disease, encompassing functional autonomy, psycho-emotional well-being, social relations inside and outside the family context, the working field and several aspects of the medical care and assistance. The methodology proposed here may help improve patient engagement in line with the EUPATI project (Warner *et al.*, 2018) and allow the analysis of real-world data related to HRQoL, especially when the number of participants is reduced.

## References

- Asadi-Lari, M., Tamburini, M., Gray, D. (2004). Patients' needs, satisfaction, and health related quality of life: towards a comprehensive model. *Health Qual Life Outcomes*, **2**, pp. 32.
- Bartolini, B., Bertoldi, S., Benedan, L., Galeone, C., Mariani, P., Sofia, F., Zenga, M. (2021). Development of an innovative methodology to define patient-designed Quality of Life: a new version of a well-known concept in healthcare. In Bruno Bertaccini, Luigi Fabbris, Alessandra Petrucci (eds), *ASA 2021 Statistics and Information Systems for Policy Evaluation. Book of short papers of the on-site conference*, pp. 155-159, Firenze University Press
- Benedan, L., Hachem, M. El, Galeone, C., Mariani, P., Pilo, C., Tadini, G. (2022). Assessing the Quality of Life of patients with Epidermolysis Bullosa (EB): Development of a patient-centered questionnaire. In *Book of short papers - 10th International Conference IES 2022 Innovation and Society 5.0: Statistical and Economic Methodologies for Quality Assessment*. PKE s.r.l.
- Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One*, **6**(6), e20476.
- Centers for Disease Control and Prevention (2000). *Measuring healthy days: Population assessment of health-related quality of life*. Centers for Disease Control and Prevention, Atlanta, Georgia.
- Pacinelli, A., (2008). *Metodi per la Ricerca Sociale Partecipata*, Franco Angeli, Milano.
- Patrick, D.L., Deyo, R.A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Med Care*, **27**(3 Suppl), pp. S217-232.
- Rabin, R., de Charro, F. (2001). EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine*, **33**(5), pp. 337—342.
- Selim, A.J., Rogers, W., Fleishman, J.A., Qian, S.X., Fincke, B.G., Rothendler, J.A., Kazis, L.E. (2009). Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). *Qual Life Res.*, **18**(1), pp 43-52.
- Trevelyan, E. G., Robinson, N. (2015). Delphi methodology in health research: How to do it? *European Journal of Integrative Medicine*, **7**(4), pp. 423—428.
- Ware, J. E., Jr., Gandek, B., Guyer, R., Deng, N. (2016). Standardizing disease-specific quality of life measures across multiple chronic conditions: development and initial evaluation of the QOL Disease Impact Scale (QDIS®). *Health Quality Life Outcomes*, **14**, 84, pp. 1—16.
- Warner, K., See, W., Haerry, D., Klingmann, I., Hunter, A., May, M. (2018). *EUPATI Guidance for Patient Involvement in Medicines Research and Development (R&D); Guidance for Pharmaceutical Industry-Led Medicines R&D*. Front Med (Lausanne), **5**, pp. 270.



# Access to emergency care services and inequalities in living standards: some evidence from two Italian northern regions

Andrea Marino, Marco Pesce, Raffaella Succi

## 1. Introduction

The goal of this short paper is twofold. First, we want to provide an estimate of accessibility to emergency care services at a very geographically disaggregated level, namely census enumeration areas (CEAs). Secondly, we want to evaluate whether and how differences in accessibility to emergency care services relate to health inequalities and regional differences in living standards.

Quick and timely access to emergency medical services is a key factor in reducing the health implications -in terms of both mortality and disability- of adverse events. Thus, in a well-designed health system the geographical distribution of emergency care services should be able to minimize the share of people whose access time lies beyond critical thresholds. In recent years, a growing number of studies concerning different countries and/or regions have been devoted to quantify access times to emergency care services. A far from exhaustive list of recent papers includes Tolpadi et al. (2022) for the USA. Tang et al. (2021) for the Sichuan province of China. Lilley et al. (2019) for New Zealand. Kisiala et al. for Poland (2021). Silva and Padeiro for the metropolitan area of Lisbon (2020). To the best of our knowledge, the only estimates concerning Italian regions are those provided by Pesce and Succi (2016) and by Salvucci and Lombardo (2016 and 2017).

By extending Pesce and Succi (2016), this paper focuses on two Italian northern regions, Liguria and Lombardy. Regions (classified as NUTS 2 in the Eurostat nomenclature of territorial statistical units) are administrative units of particular interest for our analysis, as -starting from the early 1990s- the public responsibility to deliver health services has been increasingly decentralized towards them. An implication of this decentralization process is that health expenditures generally represent the main item in regional budgets (another implication, however, has been increasing territorial inequalities in the provision of health services: see Garattini et al. 2022).

While we plan to extend the present analysis to other areas in the future, a few words to explain on why we deal with the Liguria and Lombardy regions are in order. Actually, the work originates from a convention signed in 2016 between Istat (the Italian National Institute of Statistics) and the Regional Health Agency of Liguria. As a result, the Istat regional office located in Genoa contributed to implement a regional information system for public health, by populating the database with socio-demographical data on determinants of health in Liguria and other relevant information like estimates of ED accessibility. Regardless of these institutional arrangements, we believe that investigating emergency care in Liguria may provide useful insights on whether and how differences in accessibility to health services affects inequalities in living standards. Indeed, the region is characterized by a very elderly population, which notoriously affects the demand of emergency services (Dufour et al. 2019). A large and densely populated region like Lombardy is another interesting case study *per se* and a useful benchmark, in the light of the relatively high quality of its health services (Bruzzi et al., 2022, compute a multidimensional quality index to compare the performance of regional health systems in Italy in 2015; they find that Lombardy ranks first).

Finally, before discussing methods and results, we want to highlight that an interesting by-product of our contribution is showing how researchers, by using standard hardware resources, may rely on a free, open-source, documented and powerful software toolchain to operate on large, public geographical datasets. This allows for easy reproducibility of study results.

Andrea Marino, ISTAT, Italian National Institute of Statistics, Italy, [anmarino@istat.it](mailto:anmarino@istat.it), 0000-0002-9854-7885

Marco Pesce, ISTAT, Italian National Institute of Statistics, Italy, [pesce@istat.it](mailto:pesce@istat.it), 0000-0002-8926-7128

Raffaella Succi, ISTAT, Italian National Institute of Statistics, Italy, [succi@istat.it](mailto:succi@istat.it)

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Andrea Marino, Marco Pesce, Raffaella Succi, *Access to emergency care services and inequalities in living standards: some evidence from two Italian northern regions*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.24, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 135-140, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



## 2. Evaluating travel distances to EDs: methodology and data

Census enumeration areas (CEAs) represent our main geographic unit of analysis. Such areas are defined for statistical purposes and represent partitions of a municipality (“Comune”), the smallest administrative unit in Italy (in turn, municipalities are part of a region). Latest population data at the CEA level are currently available only from the 2011 Census and are provided by the Italian National Institute of Statistics (Istat). According to such information, in 2011 the number of CEAs in the Lombardy and Liguria regions were equal to 53,174 and 11,054 units, respectively.

Fig.1. Italy



Table 1. Regional population: distribution by estimated travel times to the nearest ED

	Population data (2011 Census)		Accessibility to emergency care services				
	Total	Density (inhab. / km <sup>2</sup> )	Population distribution (%) by travel times to the nearest ED				
			<15'	15'-30'	30'-45'	45'-60'	>60'
Liguria	1,570,694	289.9	87.2	11.1	1.5	0.2	0.1
Lombardy	9,704,151	407.0	89.0	10.5	0.3	0.1	0.1

The algorithm determining travel times to the nearest ED relies on a multi-step strategy. In the first step, it computes the minimum driving time from the CEA under scrutiny to a given ED by comparing the distances of all existing routes linking the two locations. Such a computation is done for all EDs located in the same region as the CEA. This allows singling out the closest ED (and the implied travel time). Finally, these steps are repeated for all CEAs, leading to the construction of a distance matrix containing information on travel times from each CEA to the nearest emergency care service.<sup>1</sup>

More in detail, in order to accomplish the tasks outlined above, we have drawn on a bundle of official as well as crowdsourced data and we have relied on open-source software to process them. From shapefile format maps we have computed the latitude and longitude coordinates of the centroid of each CEA. Such a centroid represents in our calculations the starting point of each travel distance from a given CEA to the existing emergency care facilities. Moreover, from open data sources we have been able to geocode a total number of 122 health facilities supplying in 2013 emergency care services, 103 in the Lombardy region and 19 in Liguria.<sup>2</sup>

Solving the routing problem (i.e. determining the fastest and/or shortest path to an emergency care facility) requires: 1) a routing graph connecting each location (CEA) to all EDs; 2) an algorithm computing (and comparing) travel distances of each possible path. The international crowdsourced project *OpenStreetMap* provides us with a routing graph. The *Open Source Routing Machine (OSRM)* engine and its related tool *OSRM Distance* allow the search of minimum road paths (documentation concerning *OSRM* may be found in Luxen and Vetter, 2011). An instance of *OSRM* backend server was built for offline processing of data extracted from the *OpenStreetMap* database. This significantly reduces computing times by avoiding limitations that even freely accessible online servers may impose upon receiving high-frequency/high-bandwidth queries (note that in the case of a large area such as the Lombardy region, the whole distance matrix contains almost 5,5 million records).<sup>3</sup>

<sup>1</sup> Our estimations of ED accessibility take into account driving times only. We lack information on the availability of alternative modes of transportation (such as helicopters). We are also aware that in many cases patients arrive at EDs by their own transport. Furthermore, they may not choose the closest emergency facility based upon subjective preferences or common information about health services quality.

<sup>2</sup> The definition of ED used throughout the paper includes only the following categories of medical centers: a) “Dipartimenti di Emergenza e Accettazione (DEA)” ; b) “Ospedali sedi di pronto soccorso”. It rules out, however, the so called “Punti di primo intervento”. These differ from the categories mentioned above in some important respect; in particular, they may be not open 24 hours a day and provide treatment for less severe emergency cases. Detailed information and definitions are available in the website of the Italian Ministry of Health: [www.salute.gov.it](http://www.salute.gov.it).

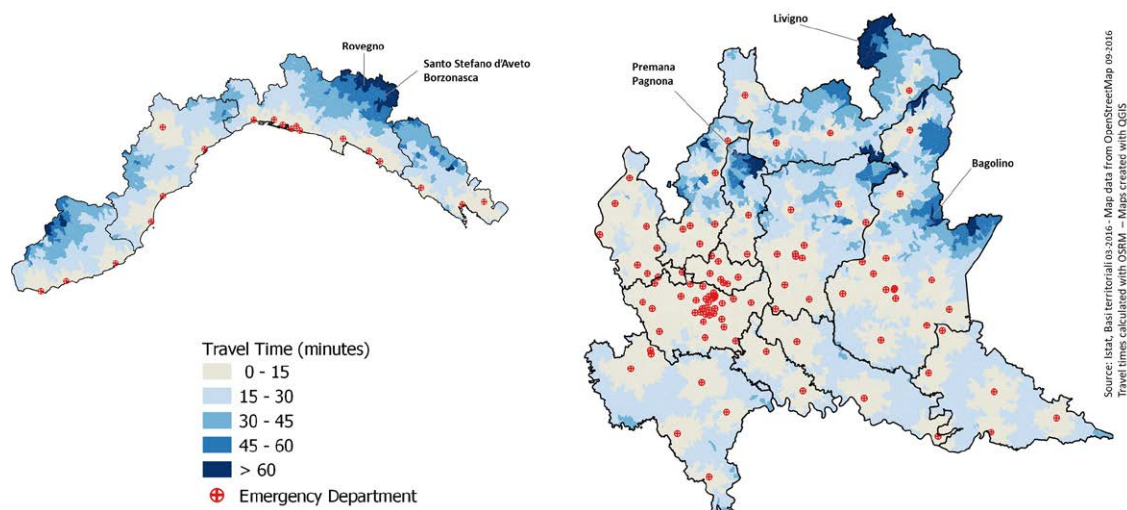
<sup>3</sup> Data from *OpenStreetMap* ([www.openstreetmap.org](http://www.openstreetmap.org)) can be obtained as file archives from multiple internet repositories: this allows for completely local, offline processing through *OSRM* (<https://github.com/Project-OSRM>). Thus

Accessibility is measured as the driving time required to travel through the fastest path from the centroid of each CEA to the nearest emergency care facility. The calculation of distance (in time units) assumes as a starting point the road junction, which is closest to the centroid. Also, travel times are computed by assuming that speed corresponds either to known speed limits (when such an information is available) or to standard speed limits for urban and non-urban roads. Moreover, the computation assumes optimal traffic conditions (no time losses due to either traffic jams or traffic lights).

### 3. Evaluating travel distances to EDs: results

Figure 2 depicts our estimates of the distribution of the population by different ranges of travel times to the nearest ED. Clearly, when emergency care is needed, arrival at ED facilities should occur in the shortest possible time. A different, but related, issue is what are the “critical” travel time thresholds, which have to be respected in order to ensure adequate treatment. From the patients’ point of view, this question can only have a case-by-case answer. When setting targets to plan or evaluate public health systems, a common threshold corresponds to 60 minutes (Lilley et al., 2019). Indeed, this is a policy-relevant threshold in the Italian case too.<sup>4</sup> Yet, there are at least two important reasons to present also results based on alternative (and more restrictive) time cut-offs. First, as Lilley et al. (2019) themselves point out, the choice of setting as a threshold the so-called “golden hour” is “not supported by strong-evidence base”. Secondly, since in many cases patients do not arrive at EDs by their own transport, a complete evaluation of driving times to the nearest ED should take into account also distances between where people live and where ground ambulance depots are located. As accurate information on this is missing, a sensitive analysis accounting also for lower time thresholds is in order.

**Fig.2. Estimated travel times to the nearest ED in Liguria and Lombardy by CEA (with ED locations and province borders)**



every step of the process remains under direct and granular control, as it is not tied to cloud services constraints (such as tariffs or usage limits) nor to undocumented, run-time variations in server-side behaviour that may affect final output. Being reliant on origin-destination tables, such a methodology may be computationally demanding in case of large areas, but it is also scalable as more efficient hardware becomes available (e.g. through multi-core processing that OSRM exploits natively).

<sup>4</sup> Italian administrative laws concerning accessibility to emergency care define as particularly disadvantaged (i.e. most remote) areas those with travel times to the nearest emergency care facility higher than 60 minutes (Ministero della Salute, “Regolamento sugli standard dell’assistenza ospedaliera” Decreto n. 70 /04/2015).

With respect to the 60 minutes threshold, the actual location of emergency-equipped hospitals in 2013 was able to yield a high population coverage rate in both Liguria and Lombardy; indeed, the share of the population living in most remote areas was about 0.1% in both regions. However, some regional differences emerge when setting lower critical time thresholds. For instance, figures reported in Table 1 imply that in the Lombardy region the share of the population facing travel times beyond 30 minutes is 0.5%, whereas the same percentage grows up to 1.8% in Liguria. Also, the share of the population whose access to the nearest ED lies within 15 minutes is 89% in Lombardy and around two percentage points lower in Liguria. These inter-regional differences seem moderate and come to no surprise given that accessibility generally grows with population density (see Table 1; see also Lilley et al., 2019, on this). Note finally that -in both regions- CEAs with low accessibility are located in mountain areas (some municipality names corresponding to these CEAs are reported in Figure 2).

Population coverage rates reported in Table 1 may not accurately describe the current situation, as the number of EDs has changed in the last years. At the time of writing (June 2022) we lack all the data required to re-run our estimation procedure on updated information. While leaving this exercise for future research, here we give a clue of how the picture may have recently evolved in Liguria, which has undergone a sizable reduction in the number of EDs (from 19 to 12, all of which are now located along the coast). To do so, we have used 2021 population data available at the municipality level and assumed that population uniformly grew within each municipality between 2011 and 2021. This provides us with an estimate of CEA-level population in 2021. By combining this with updated information on EDs and travel distances, we find that the decrease in EDs has generally implied a worsening in coverage rates; e.g., according to our estimates, the population share currently facing driving times higher than 30 minutes is about 3.5%, i.e. it has doubled with respect to the situation represented in Table 1.

#### 4. Population living standards and accessibility to emergency care

A timely provision of emergency care services throughout the national territory appears a particularly challenging goal nowadays. The tighter budget constraints the Italian NHS has to cope with impose a strong efficiency-equity trade-off. Scale economies and the high concentration of the population in urbanized areas may push regional policymakers toward a higher centralization in the supply of ED services, which comes at the cost of higher (within-region) inequalities. To understand why, it is worth recalling some mechanisms through which differences in accessibility may lead to higher inequality. To begin with, the literature has shown that differences in accessibility affect individual behavior due to a “*distance decay effect*”: compared to people living closer to EDs, those residing in more remote areas are less likely to demand certain emergency care services even when these are equally needed. Other studies point out to the existence of an “*inverse care law*”: areas characterized by low accessibility often coincide with the more socio-economically deprived ones (i.e. with those needing social and health services most). Differently stated, “the availability of good medical care tends to vary inversely with the need for the population served” (Hart, 1971).

From a normative point of view, regional emergency care services should be planned in a way to prevent the rise of *health inequities* discriminating certain population subgroups.<sup>5</sup> Such a goal requires not only accurate evaluations of physical accessibility to EDs but also a deep knowledge of some social characteristics of the population, which may contribute to give rise to health inequities (whether in combination with low accessibility or in an independent way). It is well known that socio-demographic and economic factors such as age, sex, ethnicity, education and occupational status (to mention a few) are significant social determinants of health and emergency care utilization (Marmot, 2005). Moreover, caring for more vulnerable people regardless of their numerical importance is one of main tenets underlying the Sustainable Development Goals (the so-called “*Leave No One Behind*” principle).

In order to study whether and how socio-demographical and economic factors change with

---

<sup>5</sup> *Health inequities* correspond to *health inequalities* which are “preventable and unnecessary” and thus “could be avoided by reasonable means” (Arcaya et al., 2015).

differences in accessibility, we have combined our estimates of travel times at the CEA level with some information coming from the 2011 Population Census. First, we have partitioned the territory of each region according to given thresholds of driving times to the nearest ED (such thresholds are determined by 15 minutes intervals, with the “>60 minutes” category representing a residual class). Secondly, using census data, we have computed the values of a set of socio-economic indicators referred to the subpopulations belonging to such time intervals. These variables are: the ratio of people aged 65 years and more to the total population, the population share of foreign inhabitants, the ratio of less educated people (i.e. those who do not hold at least a secondary school degree) to the population aged 6 years and more, the ratio of single-member families to the total number of families, the unemployment rate.

**Table 2. Socio-demographic characteristics and ED accessibility (percentage values)**

	Liguria						Lombardy					
	Subpopulation weight within travel time categories					Total	Subpopulation weight within travel time categories					Total
	<15'	15'-30'	30'-45'	45'-60'	>60'		<15'	15'-30'	30'-45'	45'-60'	>60'	
Age ≥ 65 years	27.4	27.0	31.9	40.5	35.7	27.4	20.8	20.4	24.6	19.0	18.3	20.8
Foreigners	7.3	5.4	5.6	4.5	2.6	7.1	9.8	9.3	6.8	10.0	3.9	9.8
Single-member family	40.3	42.2	53.7	63.6	60.8	40.9	32.2	30.0	39.7	38.4	29.4	32.0
Low education	54.8	62.2	66.6	70.6	69.7	55.9	56.7	65.8	67.4	64.5	72.5	57.7
Unemployment rate	8.0	6.9	6.2	6.7	5.4	7.8	6.9	6.6	5.1	4.2	3.7	6.8

Table 2 provides a descriptive analysis of the results obtained. As it may be observed, populations groups living in the remote (45-60 minutes) and most remotes (>60 minutes) areas in Liguria appear more vulnerable; for instance, the share of less educated people in these areas is around 70%, compared to a regional average of 55.9%. Also, the share of people aged 65 years and more achieves 40.5% and 35.7% of the total population living in remote and most remote areas, respectively; this is again clearly more than the regional average (27.4%). Something similar happens for single-member families. In Lombardy, distributions tend to be flatter. However, we observe that in the most remote areas the incidence of less educated people is higher than elsewhere, while the share of people aged 65 years and more it is only 18.3% (mainly due the upward contribution coming from mountain zones). Overall, the descriptive analysis of Table 2 seems to show that in Liguria differences in accessibility to ED services actually represent a further source of health inequalities that interacts with usual social determinants.

A straightforward question is how much distances and social determinants of health inequalities are related. Answering this question with CEA-level data is not an easy task. To see why, note that many census areas are very thinly populated, so that the socioeconomic indicators we consider may take on rather unusual values (think e.g. of unemployment rates equal to 0% or 100% in CEAs with only one inhabitant). To overcome such a problem, we have computed correlations at the municipality level (after computing population-weighted averages of travel distances measured at the CEA level). Results reported in Table 3 indicate that travel times are positively correlated to some social determinants of health inequality (like the incidence of elderly and less educated people, and the share of single-member families). Such a result (which holds for both Liguria and Lombardy) is worrying as it implies that the “inverse care law” may actually be at work and deserves further investigation in future work.

**Table 3. Correlations between socio-economic indicators and population-weighted average travel times at municipality level**

	Age ≥ 65 years	Foreigners	Single-member family	Low education	Unemployment rate
<b>Liguria</b>	0.480*** (.049)	-.043 (.054)	.588*** (.039)	.517*** (.047)	-.158** (.063)
<b>Lombardy</b>	.225*** (.025)	-.048* (.086)	.326*** (.023)	.427*** (.022)	-.064** (.028)

Bootstrap standard errors in parentheses under correlation values (9,999 replications). Significance levels: \*\*\*  $p < .01$ ; \*\*  $p < 0.05$ ; \*  $p < .1$

## 5. Summary and conclusions

Timely access to emergency care services is a relevant determinant of health inequalities; thus, a geographically detailed evaluation of accessibility is a necessary step in order to design effective policies counteracting such inequalities. In order to perform such a task, our study proposes a methodology, which should be appealing for many reasons: 1) it relies on open data and open-source software; 2) it is computationally efficient; 3) it is easily interpretable. Results show that health inequalities stemming from socio-economic differences may turn into health inequities due to differences in accessibility. An obvious direction for future research would be using updated information on EDs and extending this work to other areas. More accurate estimates of accessibility should take into account the possibility that -in some cases- people needing emergency care services be transported to EDs of other regions. Also, our analysis of how differences in accessibility affect health inequities might be extended by employing more sophisticated techniques of multivariate statistics and also by relating distances to composite indices of social deprivation.

## References

- Arcaya M.C., Arcaya A.L., Subramanian V. (2015). Inequalities in health: definitions, concepts, and theories. *Global Health Action*, **8**(1).
- Bruzzi S., Ivaldi E., Santagata M. (2022). Measuring regional performance in the Italian NHS: Are disparities decreasing? *Social Indicators Research*. **159**, pp. 1057-1084.
- Dufour I., Chouinard M.C., Dubuc N., Beaudin J., Lafontaine S., Hudon C. (2019). Factors associated with frequent use of emergency-department services in a geriatric population: a systematic review. *BMC Geriatrics*, **19**:185.
- Garattini L., Badinella Martini M., Zanetti M. (2022). The Italian NHS at regional level: same in theory, different in practice. *The European Journal of Health Economics*. **23**, pp. 1-5.
- Hart J.T. (1971). The inverse care law. *The Lancet*. **297**(7696), pp. 405-412.
- Kisiala W., Racka I., Suszynska K. (2022). Population Access to Hospital Emergency Departments: The Spatial Analysis in Public Health Research. *Int. J. Environ. Res. Public Health*, **19**(3), 1437.
- Lilley R., et al. (2019). Geographical and population disparities in timely access to prehospital and advanced level emergency care in New Zealand: a cross-sectional study. *BMJ Open*, **9**.
- Luxen D., Vetter C. (2011). Real-time routing with OpenStreetMap data. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.
- Marmot M. (2005). Social determinants of health inequalities. *Lancet*, **365**(9464), pp. 1099-104.
- Pesce M., Succi R. (2016). L'accessibilità geografica dei servizi di pronto soccorso e le condizioni socio-economiche della popolazione- Poster presented at the 2016 MiLes Conference, Milan,
- Salvucci G., Lombardo G. (2016). Molise quotidianamente resiliente, analisi demografiche in una prospettiva gis. Conferenza ESRI Italia 2016. Conference paper available at the web address: <https://www.esriitalia.it/images/pdf/conferenza2016/GianluigiSalvucci2.pdf>.
- Salvucci G., Lombardo G. (2017). Il ritardo che costa la vita: analisi dell'accessibilità ai pronto soccorso. Conferenza ESRI Italia 2017. Conference paper available at the web address: <https://www.esriitalia.it/images/pdf/conferenza2017/GianluigiSalvucci.pdf>
- Silva K., Padeiro M. (2020). Assessing inequalities in geographical access to emergency medical services in metropolitan Lisbon: a cross-sectional and ecological study. *BMJ Open*, **10**.
- Tang X., Deng Y., Yang H., Tian F., Li Y., Pan J. (2021). Spatial accessibility to emergency care in Sichuan province in China. *Geospatial Health*, **15**(2), pp. 274-284.
- Tolpadi A., et al. (2022). National travel distances for emergency care. *BMC Health Services Research*, **22**, 388

# Population ageing and sustainability in South Tyrol: measuring the economic implications of an ageing society

Giulia Cavrini, Elisa Cisotto, Alex Weissensteiner

## 1. Introduction and Background

During the twentieth century, South Tyrol has experienced a rapid and intense decline in fertility jointly with impressive achievements in extending survival, especially at older ages. Consistently low birth rates and high life expectancy have contributed to a faster ageing process of the resident population, a trend that is projected to continue until at least the middle of the twenty-first century (Christensen et al., 2009.). The implications of population ageing are pervasive and complex, and often regarded as a major cause of increased pressure on healthcare and social security systems. However, the ageing process impacts almost all spheres of society, including economy, housing, family structures and intergenerational ties (WHO, 2015; UN, 2014).

Largely, meeting the challenge of population ageing requires a better understanding of frailty and disability, and appropriate strategies to ensure the resilience of the health and social care system and long-term care spending without destabilising public finances or over-burdening the economy. Countries will face a demanding task to provide care for a heterogeneous population of older adults, finding the true balance between offering the proper social protection to people with care needs and assuring that this protection is fiscally sustainable (OECD, 2017). The long-term horizon sometimes makes it difficult to derive the necessary actions from it, but also to make the political alternatives visible. In many cases, key facts become clearer when they are broken down into a manageable geographical reality. For this reason, this paper deals with the situation of the Autonomous Province of Bozen-Bolzano. Due to the autonomy of this province within Italy, there is an implemented care system, which is well documented, but not so specific as to be considered a case study whose results can be generalised.

Within this context, we explicitly aim to assess the impact of current and future population dynamics on the sustainability of the economic, health and social system of the Province of Bozen-Bolzano. Thus, the current paper is designed to reach the following research objectives:

- (a) measure the current needs for social care in South Tyrol,
- (b) identify the local trajectories of health status, disaggregated by age, sex and severity of illness,
- (c) forecast the health care needs and the healthcare system's financial sustainability.

## 2. Data and method

Calculations are based on the population data structure by age and sex from 2009 to 2050, provided by the Italian National Statistical Institute (ISTAT). Individual health care data for administrative and billing purposes is from the Autonomous Province of Bozen-Bolzano (Department of Family, Social Affairs and Community), and used to study health care delivery, benefits, harms, and costs from 2009 to 2019 in the case of home-based care recipients, and from 2009 to 2013 for residential care receivers.

Health care local data contains all the monthly payments made by the Autonomous Province of Bozen-Bolzano for everyone receiving care allowance. For each allowance recipient, basic

Giulia Cavrini, Free University of Bozen-Bolzano, Italy, Giulia.Cavrini@unibz.it, 0000-0002-9084-3081

Elisa Cisotto, Free University of Bozen-Bolzano, Italy, elisa.cisotto@unibz.it, 0000-0001-9496-6022

Alex Weissensteiner, Free University of Bozen-Bolzano, Italy, alex.weissensteiner@unibz.it, 0000-0002-8600-0516

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Giulia Cavrini, Elisa Cisotto, Alex Weissensteiner, *Population ageing and sustainability in South Tyrol: measuring the economic implications of an ageing society*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.25, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 141-144, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

demographic and health status information is available, such as sex, date of birth, date of death, citizenship, entitlement to an attendance allowance, native language, and area of residence. Besides, based on this data, we calculate on an annual basis: the health care level of classification, whether the provision of care is home-based or institutionalized, and the total amount received by each assisted person per year and the number of payments. The health care level of classification categorises the severity of the health condition for which the person receives the care allowance. They are legally defined care levels in South Tyrol, whereby level 1 provides for a care requirement of 60-120 hours, level two of 120-180 hours, level 3 180- 240 hours and level 4 more than 240 hours care requirement per month. Each level matches a precise rate for the given allowance.

According to the following formula, we calculate the annual population prevalence ( $E_t$ ) of people (P) in need of assistance by care level (l) (1 to 4, where 4 means worst health conditions), care typology (c) (home-based or residential care), sex (s) and age (x):

$$E_{tsx}^{cl} = \frac{P_{tsx}^{cl}}{P_{tsx}}$$

Thus, the forecast estimate of the number of people in need of care results from the prevalence ( $E_t$ ) (assumed to be constant over time) multiplied by the ISTAT forecast of the population, separated by sex and age, of the corresponding year. To obtain accurate and latest estimates, we use three-years average prevalence estimates from 2017 to 2019 to forecast home care recipients from 2020 ongoing, and two-years average prevalence estimates from 2012 to 2013 to forecast residential care receivers. The research, therefore, assumes that the shares of the dependent population that receive either formal care at home or institutional care are kept constant over the projection period. Therefore, this is a pure demographic scenario, as the only relevant variable is demography, through the projected population changes.

The ISTAT population forecasts are based on a set of assumptions with respect to fertility, mortality, interregional and international residence movements. The methodological approach is semi-probabilistic. The fundamental characteristic of probabilistic forecasting is to consider the uncertainty associated with predicted values, determining the confidence intervals of the demographic variables, and allowing the user to independently choose the degree of confidence to be assigned to the results. For the purposes of this paper, we rely on the variant generally identified as the most probable, typically identified as the ‘median scenario’, with a 95% confidence interval.

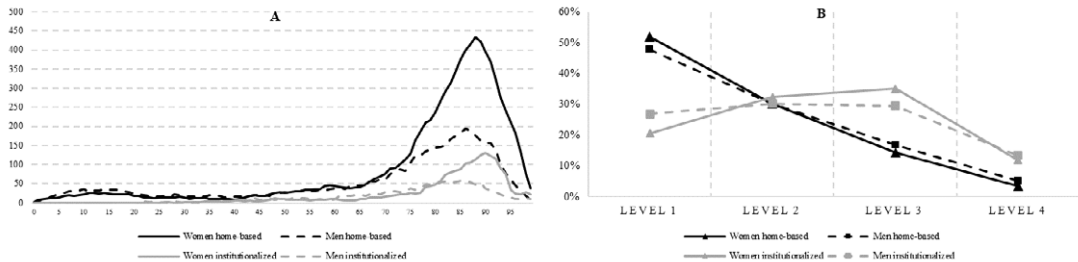
### 3. Preliminary results and discussion

Figure 1 shows the distribution of home-based assisted persons for 2017-2019 and the average number of residential assisted persons for 2012-2013. Overall, the probability of a need for care at an advanced age (65+) rises sharply compared to younger ages (Figure 1, panel A). On average, between 2017 and 2019, more than 66% of home care services were provided to over-80s and almost 85% to over-65s. Similarly, between 2012 and 2013 (the latest available data), more than 77% of facility-based services were provided to the over-80s and 90% to over-65s. Due to their higher life expectancy, women are particularly affected, so that the number of assisted women exceeds that of men, especially in old age.

Besides, the distribution by severity level of the health condition for which the allowance is received is relatively independent of age (Figure 1, panel B). Overall, greater prevalence occurs at lower levels of health condition severity (levels 1 and 2 over a four points-scale of severity). Regarding those in care at home, about 50% of those affected are in the first level of assistance, 30%, 15% and

4% in levels 2 to 4. Therefore, most home care recipients are therefore in the least severe, and least economically costly, levels of care. Differently, in the residential care structures, we find more patients in the most severe levels of assistance, ranging from 2 (31%) to 3 (32%), and 4 (13%).

**Figure 1. Average number of home-based and residential benefits by age and sex (panel A), and share by assistance level (% - panel B).**

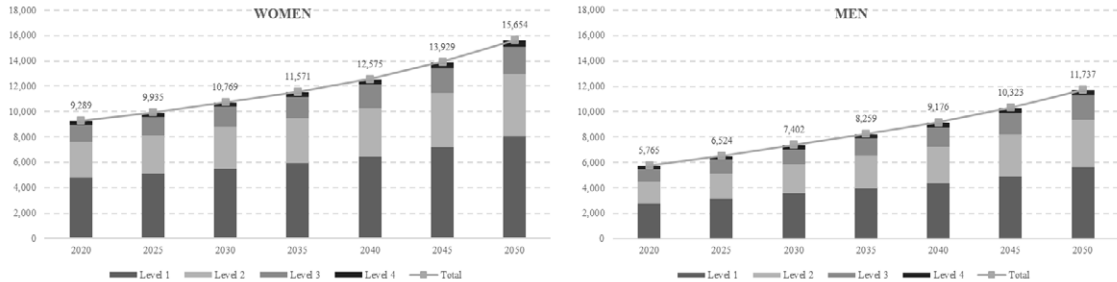


Source: Own elaborations on administrative data from the Autonomous Province of Bozen-Bolzano

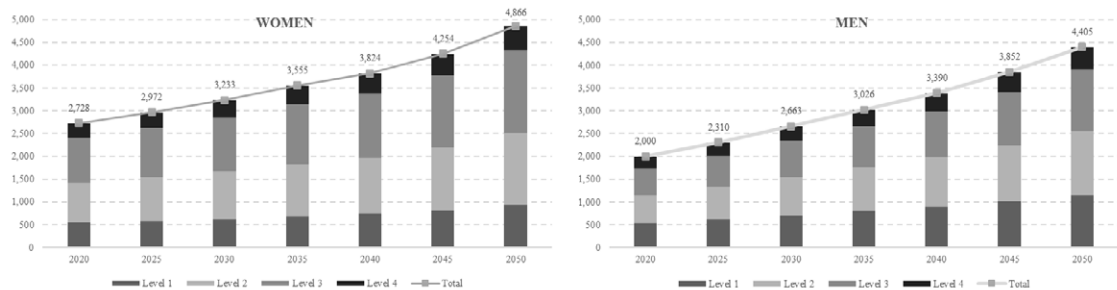
By combining the information on demographic dynamics and the care benefits prevalence by age, the weight of the home and residential assisted individuals over the next few years was estimated. Figure 2 shows how the number of home-assisted persons will grow between 2020 and 2050 by more than 68% for women and 104% for men. The same trend is expected for residential services, but with a much stronger growth of over 78% in the next 30 years for women and up to 120% for men.

**Figure 2. Home care and residential benefits by level (South Tyrol 2020-2050)**

**A – Home care benefits**



**B – Residential benefits**



Source: Own elaborations on administrative data from the Autonomous Province of Bozen-Bolzano



Our extrapolation is based on the main restrictive assumption that the population's health status will continue to correspond to that of the reference years in the future. Hence, the distribution of home care recipients and residential care receivers will remain unchanged. Nevertheless, concerning the economic impact of our preliminary results, two major drivers must be considered. First, the demographic drivers, for which the combined effect of longevity improvement and the shape of care expenditure by age will result in a projected increase in public expenditure from 2020 to 2050. However, survival at older ages may not necessarily result in an increase in the population prevalence of chronic diseases. Otherwise, it could translate into improved survival with additional years in good health, so that the future economic burden of longevity could be contained by such healthy ageing process and decreasing dependency levels. Informal and formal care is the second key driver to be considered in terms of future economic consequences of population ageing. Indeed, most care in Italy and South Tyrol is informal, provided by family and social networks. However, current changes in family structures, such as declining family size and rising female labour force participation, could lead to a decline in the availability of informal caregivers and to an increase in the need for formal aid care. These social changes, together with public spending policy and political actions on health care, can change considerably the impact of population ageing on future public expenditure, which can even become more relevant than the demographic change itself.

## References

- Christensen, K., G. Doblhammer, R. Rau and J.W. Vaupel (2009). Ageing populations: the challenges ahead, *The Lancet*, vol. 374, No. 9696, pp. 1196-1208.
- Cylus, J., Figueras, J., Normand, C. (2019). Will Population ageing spell the end of the Welfare State? A review of evidence and policy options. EU2019.FI. World Health Organisation. [www.weforum.org](http://www.weforum.org).
- OECD (2017), *Health at a Glance 2017: OECD Indicators*, OECD Publishing, Paris.
- UN (2014). Population ageing and sustainable development. *Population Facts*, 4(Rev.1), pp. 1-4.
- WHO (2015). World report on ageing and health. World Health Organization. <https://apps.who.int/iris/handle/10665/186463>.

# Job loss and financial struggle among the older age groups in 2021: Lessons from the European Union

Demetrio Panarello, Giorgio Tassinari

## 1. Introduction

The COVID-19 pandemic caused intense disruptions in the global economy. As regards Europe, the Winter 2022 European Economic Forecast projects that, following an annual gross domestic product growth rate of 5.3% in 2021, the EU economy will expand by 4.0% in 2022 and 2.8% in 2023. Ireland was the fastest-growing European economy in 2021 in comparison to the preceding year, with a growth of 13.7%, while Germany – the largest economy in the continent – was the slowest-growing one, with a 2.8% annual GDP growth. No EU countries experienced a negative growth rate compared to 2020 (European Commission, 2022).

Adults around retirement age are more likely to experience disturbances to their employment patterns (Davis et al., 2020). Indeed, older adults are in general more affected by COVID-19 than the younger ones and less comfortable with working remotely, particularly as this often implies the possession of specific technological skills. In 2021, in the EU, the unemployment rate was 7.0%, down from 7.2% in 2020, but above the rate of 6.8% in 2019 (Eurostat, 2022). Across the EU, the 2021 rates ranged from 2.8% in the Czech Republic to 14.8% in Spain. If we restrict the analysis to the older age class (55-74 years old), we can notice that unemployment rates remained unchanged between 2019 and 2020 (4.9%) but rose to 5.2% in 2021.

Here, we examine the different impacts of the pandemic crisis on the various socio-demographic groups, particularly focusing on non-retired individuals aged 50 and above who experienced an involuntary job loss in the first year of the pandemic. This is especially important in times of crisis and in the context of an increasingly ageing population (Cristea et al., 2020). We make use of the second Corona round of the Survey of Health, Ageing and Retirement in Europe (SHARE), with data collected in all continental EU countries plus Switzerland and Israel during the summer of 2021 (Börsch-Supan, 2022).

Our research focuses on European households' economic conditions, by analysing SHARE respondents' statements on the possibility of satisfying their needs through their current income. We try to identify the contextual factors that may make it particularly difficult to achieve this goal, making a distinction between retired and non-retired individuals, in a period during which a significant number of people in the sample experienced retirement or involuntary loss of employment, which translates into rising inequalities (for an analysis of the effects at the end of the first wave, see, among others, Panarello and Tassinari, 2022).

Our results rely on self-reported measures of economic well-being, measuring respondents' perceived economic vulnerability: survey respondents were hence able to portray their subjective well-being without any outside interference. Individuals' own reports of their economic circumstances allow us to capture the real distress they are forced to face in order to maintain their accustomed standard of living in times of crisis.

A relevant element in determining households' ability to cope with adverse economic situations is given by social networks (family and friends), as will be seen later, but we cannot exclude that there is an inverse relationship, for which households facing financial hardship tend to attenuate their social contacts (Gilligan et al., 2020). Moreover, we expect a direct relationship between frequency of social contacts and stated health level (Assari, 2017; Minkler et al., 1983).

The remainder of the manuscript is structured as follows. Section 2 introduces the employed data and procedures; Section 3 presents the results; finally, Section 4 offers some closing remarks.

## 2. Data and methods

For our analyses, we make use of microdata from the second Corona round of the Survey of Health, Ageing and Retirement in Europe (SHARE), with information collected in all continental EU countries plus Switzerland and Israel during the summer of 2021 (Börsch-Supan, 2022). Since 2004, SHARE regularly collects evidence on Europeans' health, socio-economic status, and social and family networks, interviewing representative samples of individuals with an age of 50 years or over, as well as their eventual cohabiting partners, even if under 50 years old. In the SHARE Corona Survey, respondents are surveyed through computer-assisted telephone interviewing (CATI), using a shortened questionnaire that was specifically developed for use in the pandemic period (Scherpenzeel et al., 2020).

To answer our research questions, we proceed with the estimation of two ordinal logistic regressions of households' ability to make ends meet during the pandemic (specifically, with regard to the approximately twelve months going from July 2020 to July 2021). We use retirement status to generate the subsamples that are used in the two estimated models.

The ordinal dependent variable in our models measures respondents' own reports of their household's ability to make ends meet, with the possible answers being: with great difficulty; with some difficulty; fairly easily; or easily.

The regressors refer to contact frequency with neighbours, friends or colleagues; eventual job loss; financial support received due to the pandemic; eventual variations in household monthly income; gender; age; rating of subjective health (excellent, very good, good, fair, or poor); household size; eventual presence of a cohabiting partner; and a country group dummy based on the United Nations Regional Groups classification (United Nations, 2021), capturing the East-West dichotomy (1: Eastern European Group; 2: Western European and Others Group).

Table 1 shows the descriptive statistics (observations, minimum value, median, maximum value, mean, and standard deviation) of the variables included in the models, based on the subsample that is not missing for any of the variables (estimation sample).

**Table 1 Descriptive statistics of the variables included in the models**

Variable	Obs.	Min	Median	Max	Mean	SD
Household's ability to make ends meet since July 2020	32665	1	3	4		
Retirement	32665	0	1	1		
Contact frequency with neighbours/friends/colleagues during last 3 months: At least weekly	32665	0	1	1		
Unemployed, laid off or business closed since July 2020	32665	0	0	1		
Received financial support due to outbreak since July 2020	32665	0	0	1		
No variations in household monthly income since July 2020	32665	0	1	1		
Male	32665	0	0	1		
Age in 2021	32665	34	71	105	71.696	9.103
Rating of subjective health	32665	1	3	5		
Household size	32665	1	2	11	1.893	0.939
Partner in household	32665	0	1	1		
Country group	32665	1	2	2		

## 3. Main results

As mentioned in the previous Section, we estimate two ordinal logistic regression models of households' ability to make ends meet, based on individuals' own reports of their economic situation with reference to the approximately twelve months going from July 2020 to July 2021 (Table 2). The first model is estimated on the sample of non-retired individuals, while the second one refers to the retired respondents.

**Table 2 Estimation results – Household’s ability to make ends meet since July 2020, estimated separately for non-retired and retired individuals**

Variable	Non-retired	Retired
	Coef.	Coef.
Contact frequency with neighbours/friends/colleagues during last 3 months: At least weekly	0.323*** (0.0420)	0.141*** (0.0244)
Unemployed, laid off or business closed since July 2020	-0.211*** (0.0783)	
Received financial support due to outbreak since July 2020	0.098* (0.0571)	-0.061 (0.0377)
No variations in household monthly income since July 2020	0.498*** (0.0583)	0.018 (0.0433)
Male	0.240*** (0.0435)	0.010 (0.0261)
Age in 2021	0.000 (0.0028)	0.024*** (0.0017)
Rating of subjective health (Reference: Excellent)		
2. Very good	-0.443*** (0.0930)	-0.373*** (0.0754)
3. Good	-0.867*** (0.0874)	-0.850*** (0.0706)
4. Fair	-1.289*** (0.0916)	-1.359*** (0.0719)
5. Poor	-2.096*** (0.1153)	-2.025*** (0.0795)
Household size	-0.216*** (0.0218)	-0.160*** (0.0163)
Partner in household	0.563*** (0.0495)	0.548*** (0.0303)
Country group: Western European and Others Group (WEOG)	0.326*** (0.0432)	0.722*** (0.0265)
Cutpoint 1	-2.022*** (0.2146)	-0.912*** (0.1534)
Cutpoint 2	-0.333 (0.2135)	1.066*** (0.1527)
Cutpoint 3	1.389*** (0.2140)	2.862*** (0.1539)
Observations	8683	23982
Pseudo-R2	0.046	0.059
Log-likelihood	-10867	-28315

Note: \* and \*\*\* stand for  $p < 0.10$  and  $p < 0.01$ . Standard errors in brackets.

Respondents stating that they have been engaging with neighbours, friends or colleagues at least weekly during the last three months, compared to those who met their acquaintances less often, are more likely to satisfactorily meet their living costs during the COVID-19 crisis.

Straightforwardly, the non-retired individuals who suffered a job loss since July 2020 are shown to be less able to make ends meet.

Having received financial support due to the outbreak since July 2020 makes non-retired people more likely to make ends meet on average, while this is not associated with significant differences when considering the retired population, maybe due to a higher monetary wealth they might be able to tap into and to a more stable economic condition.

Similarly, non-retired people who did not experience significant variations in their monthly income are more likely to get through the end of the month, while this is not significantly associated with the likelihood of meeting living costs for the retired ones.

Non-retired males are more likely than non-retired females to make their ends meet, while this association is no longer different from zero at conventional significance thresholds for the retired subsample. This result suggests that pension income provisions play a role in reducing the gender

gap in well-being during retirement.

While age does not play a relevant role for non-retired individuals, the oldest retired individuals appear to be more likely to be able to adequately make their ends meet.

For both subsamples, the lower the perceived health level, the lower the likelihood of comfortably getting to the end of the month.

A larger number of members in the household is associated with a lower likelihood of being able to make ends meet, while the presence of a partner makes it more likely to be able to adequately cover expenditure. This result suggests that intra-household sharing of resources plays a role in smoothing consumption in favour of weaker and older members.

Finally, respondents from countries belonging to the Western European and Others Group are more likely to be able to meet their expenses compared to those living in an Eastern European Group country.

#### **4. Conclusions**

In this paper, we examine the economic consequences of COVID-19 on the older European population, focusing on their ability to make ends meet since July 2020, considering retired and non-retired individuals separately.

We show the ability to adequately cover households' expenses to be associated with several factors. In particular, we reveal social networks, medical condition and family composition to be key aspects explaining the likelihood of comfortably getting to the end of the month. These features are of exceptional significance for older adults, who are commonly characterised by poorer physical health, weaker social networks and higher loneliness than younger people (Jaspal and Breakwell, 2022).

We also demonstrate the existence of remarkable differences between the eastern and western portions of the European Union.

The analysis conducted on the retired subsample shows that the ability to make ends meet is not explained by gender, income changes and provided financial assistance, highlighting a lower vulnerability – or, maybe, a higher adaptability and stability – of individuals after retirement. This fact is further bolstered by the result indicating that older retired individuals are more likely to make ends meet compared to respondents who had recently retired (of course, keeping their health status constant). These results suggest that pension income provisions are effective policies to alleviate poverty during retirement.

In essence, in light of the presented findings, we must ensure that older people feel economically safe in the face of growing social costs. Mainly, it is crucial to ensure that people continue to feel healthy and well connected to others, paying special attention to those nearing retirement.

This work does not come without limitations. First, the study does not control for individuals' place of residence, which could highlight interesting differences between capital cities and peripheral areas, or between large cities and small towns. Second, the study does not take educational level into account. Possible future waves of the SHARE Corona Survey shall allow us to assess whether the presented associations persist over time.

#### **References**

- Assari, S. (2017). Whites but not blacks gain life expectancy from social contacts. *Behavioral Sciences*, 7(4), pp. 68-89.
- Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 9. COVID-19 Survey 2. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w9ca.800.

- Cristea, M., Georgiana Noja, G., Dănăciță, D.E., Ștefea, P. (2020). Population ageing, labour productivity and economic welfare in the European Union. *Economic Research-Ekonomska Istraživanja*, **33**(1), pp. 1354-1376.
- Davis, O., Fisher, B., Ghilarducci, T., Radpour, S. (2020). A first in nearly 50 years, older workers face higher unemployment than mid-career workers (No. **2020-05**). Schwartz Center for Economic Policy Analysis (SCEPA), The New School.
- European Commission (2022). European Economic Forecast: Winter 2022 (Interim). <https://data.europa.eu/doi/10.2765/333044>
- Eurostat (2022). Unemployment by sex and age – annual data. [https://ec.europa.eu/eurostat/databrowser/view/UNE\\_RT\\_A\\_custom\\_2553687/](https://ec.europa.eu/eurostat/databrowser/view/UNE_RT_A_custom_2553687/)
- Gilligan, M., Sutor, J.J., Rurka, M., Silverstein, M. (2020). Multigenerational social support in the face of the COVID-19 pandemic. *Journal of Family Theory & Review*, **12**(4), pp. 431-447.
- Jaspal, R., Breakwell, G.M. (2022). Socio-economic inequalities in social network, loneliness and mental health during the COVID-19 pandemic. *International Journal of Social Psychiatry*, **68**(1), pp. 155-165.
- Minkler, M.A., Satariano, W.A., Langhauser, C. (1983). Supportive exchange: An exploration of the relationship between social contacts and perceived health status in the elderly. *Archives of Gerontology and Geriatrics*, **2**(3), pp. 211-220.
- Panarello, D., Tassinari, G. (2022). The consequences of COVID-19 on older adults: evidence from the SHARE Corona Survey. *National Accounting Review*, **4**(1), pp. 56-73.
- Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., ..., Börsch-Supan, A. (2020). Collecting survey data among the 50+ population during the COVID-19 outbreak: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Survey Research Methods*, **14**, pp. 217-221.
- United Nations (2021). Regional groups of Member States. <https://www.un.org/dgacm/en/content/regional-groups>



# On the use of auxiliary information in spatial sampling

Chiara Bocci, Emilia Rocco

## 1. Introduction

In many fields of application it's common to be interested in spatially-related phenomena and in particular to deal with attributes which are defined on continuous spatial domains. In this framework, if the design-based approach is assumed, the attribute is usually expressed as a function  $y(s)$  taking values on a suitable subset  $s$  of the plane. In the simplest case  $y(s)$  represents the value of an attribute at the location  $s$ . As an example, in forestal surveys  $y(s)$  could be the amount of biomass measured in sampled sites over a forestal area; in environmental studies  $y(s)$  could be the quantity of plastic materials collected by net tows in sampled areas over seas; etc... .

Technology development has led to a growing availability of low-cost spatial data ready-to-use, frequently derived from large scale observations (i.e. data from pervasive systems like GPS sensors, or remote sensing data from earth observation technologies). Oftentimes, these data can't directly answer specific questions posed by researchers and data users, or even if they can they are subject to measurement errors or self-selection bias. In both cases it is still necessary to rely, at least partially, on ad-hoc probabilistic surveys. On the other hand, the precision and quality of surveys estimates can be improved by using the data derived from these new sources as auxiliary information in the design phase and/or in the estimation phase.

Geographical data generally show a spatial pattern and an uneven spatial distribution over the population. In fact, usually spatial observations are not mutually independent and tend to be more similar to their neighbours. As stated by Tobler's first law of geography (Tobler, 1970): "everything is related to everything else, but near things are more related than distant things". This arises because nearby units interact with one another and tend to be influenced by the same set of natural and anthropogenic factors.

In such situations, it is well known that to estimate a mean or a total of a target variable selecting the units spatially best spread allows to collect more information and consequently provides better estimation. An important problem of sampling is thus to spread at best the sampled units in space. When, in addition to the spatial allocation, the value of one or more auxiliary variables is known for all the population units over the spatial domain, exploiting this information in the sampling design could further improve survey estimates.

A well-spread sample is usually said to be spatially balanced. Different types of spatially balanced sampling designs have been suggested in literature for sampling spatial population. Many, but not all of them, allow the use of auxiliary information, in a more or less simple way, during the units' selection procedure. For example various types of multi-phase systematic designs are used in different countries to produce National Forest Inventories for their forest monitoring programs. Tillé (2020, Chapter 8), Tillé and Wilhelm (2017), Benedetti et al. (2012) and Wang et al. (2012) give a review of the main spatial sampling methods. Since we are focusing on data that come from large scale observation (i.e. remote sensing data) to produce estimates at large scale, in the following we will focus on balanced sampling designs that can be easily implemented for big datasets.

We consider several sampling strategies, based on the spatially Balanced Sampling through Local Pivotal Method (LPM) introduced by Grafström et al. (2012), in order to identify the

Chiara Bocci, University of Florence, Italy, chiara.bocci@unifi.it, 0000-0001-8189-4445

Emilia Rocco, University of Florence, Italy, emilia.rocco@unifi.it

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Chiara Bocci, Emilia Rocco, *On the use of auxiliary information in spatial sampling*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.27, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 151-156, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



one which exploits geographical location and other sources of information to produce estimates for a spatially-related phenomenon in a more cost-efficient way. A strategy which could be globally applied by accounting for different areas characteristics in both the study and auxiliary variables, as well as for the differences in their relation. In all but one of the strategies under evaluation the sampling scheme consists of a different variation of the LPM, and therefore a single-phase non-informative sampling design is implemented. In addition, we propose an informative design which is based on a sequential use of the LPM and draws the final sample in two (or more) steps: (i) in the first step we collect an initial sample of observations on the target variable, which is used to investigate the relation between the auxiliary and study variables; (ii) then, this relation is exploited to target and tailor the subsequent sampling step; (iii) additional steps can be included by applying the procedure iteratively; (iv) finally, observations on the target variable collected in all the steps are used in the estimation process of the population mean.

The performance of the different strategies is investigated through Monte Carlo experiments by considering several scenarios, which differ in the distributions of the auxiliary and study variables and in their relation.

## 2. Sampling methods

Usually, in a spatial setting, the population units are plots or cells of a grid overlapping an area of interest. A value,  $y_i$ , of a variable of interest is associated with each unit  $i$  ( $i = 1, \dots, N$ ) of the population. Moreover for each unit the spatial location  $\mathbf{s}_i$ ,  $\mathbf{s} \in \mathbf{R}^2$  is known. Here, in addition we assume to know the value  $x_i$  of an auxiliary variable for each unit of the population.

For drawing a spatial sample from such a population we decided to consider as starting point the spatially Balanced Sampling through Local Pivotal Method (LPM) introduced by Grafström et al. (2012) since it is a flexible spatially balanced design that can draw equal and unequal probability samples in multiple dimensions. Unequal probability sampling can be more efficient than equal probability sampling if there is a positive correlation between the inclusion probabilities and the response values. Additional dimensions could include any auxiliary information in addition to the spatial coordinates.

The basic idea of LPM is to avoid that units close in distance appear together in the sample. First an inclusion probability  $0 < \pi_i \leq 1$  is assigned to each unit so that their sum over the population is equal to the fixed sample size. The sample is then obtained in at most  $N$  steps, where  $N$  is the population size. At each step one unit  $i$  is selected randomly from the available population and another unit  $j$  is chosen among the remaining units in the population by minimizing a distance function among them. This can be a univariate or a multivariate function that measures the distance with respect to one or more auxiliary variables, among which we can include the spatial coordinates. When all the variables are continuous the Euclidean distance is commonly used. Moreover, when multiple auxiliary variables are used, they are usually standardized or scaled in order to balance the contribution of each variable. After the selection of the unit  $i$  and  $j$  their inclusion probabilities are updated by using the following rule:

$$\begin{aligned} \text{if } \pi_i + \pi_j < 1 \text{ then } (\pi'_i, \pi'_j) &= \begin{cases} (0, \pi_i + \pi_j) \text{ with probability } \frac{\pi_i}{\pi_i + \pi_j} \\ (\pi_i + \pi_j, 0) \text{ with probability } \frac{\pi_j}{\pi_i + \pi_j} \end{cases} \\ \text{if } \pi_i + \pi_j \geq 1 \text{ then } (\pi'_i, \pi'_j) &= \begin{cases} (1, \pi_i + \pi_j - 1) \text{ with probability } \frac{1 - \pi_j}{2 - \pi_i - \pi_j} \\ (\pi_i + \pi_j - 1, 1) \text{ with probability } \frac{1 - \pi_i}{2 - \pi_i - \pi_j} \end{cases} \end{aligned} \quad (1)$$

As a result, in each step at least one unit is removed from the population frame, either because its probability becomes zero, and consequently it is definitely excluded from the sample, or because its probability becomes one and therefore is included in the sample. The procedure continues, updating at each step the probabilities of inclusion obtained in the previous step, until all units in the population are processed. LPM selects the units with the same probability  $\pi_i s$  initially assigned to them, therefore the population mean can be estimated with the usual Horvitz-Thompson estimator.

The following specific LPM based sampling designs, which differ in how they exploit location and auxiliary information, have been investigated:

1. **SpatLPM**: The original formulation of the spatially balanced sampling through LPM which produces samples that are well spread in the geographic space and is based on equal inclusion probabilities.
2. **AuxLPM**: Sampling, with equal inclusion probabilities, balanced through LPM in the space spanned by the auxiliary variable.
3. **BivLPM**: Sampling, with equal inclusion probabilities, balanced through LPM in the space spanned by both the geographical coordinates and the auxiliary variable.
4. **UneqLPM**: Spatially balanced sampling through LPM with unequal inclusion probabilities  $\pi_i s$  proportional to the auxiliary variable.
5. **StrPropAuxLPM** and **StrNeyAuxLPM**: Stratified sampling with AuxLPM design in each stratum. The area of interest is partitioned in sub-areas (strata) and then the AuxLPM is applied in each stratum. Two allocation rules are considered: Proportional and Neyman's with respect to the variance of the auxiliary variable.
6. **StrPropBivLPM** and **StrNeyBivLPM**: Stratified sampling with BivLPM design in each stratum. The same stratification designs described in the previous point, but with BivLPM applied in each stratum.
7. **SeqUneqLPM**: First an initial UneqLPM sample of size  $n_0 \leq n$  ( $n$  is the final size of the sample) is selected and used to investigate the relation between the auxiliary and study variables, specifically to estimate the parameters of a generalized additive model (GAM); then the predicted values of  $Y$  are used to draw the remaining sample units with spatially balanced sampling through LPM with unequal inclusion probabilities  $\pi_i s$  proportional to predicted values; finally an Horvitz-Thompson-type estimator is applied to produce a mean estimation that exploits the data collected in both steps.

A possible alternative to the LPM method could be the double balanced sampling of Grafström and Tillé (2012), however this sampling design is highly computationally demanding when applied to big datasets, and was unfeasible in our experiments. Conversely, LPM design has been optimized for large datasets using k-d trees (Lisic and Cruze, 2016), allowing to run our Monte Carlo experiments in a reasonable amount of time.

### 3. Simulation study

We investigate the performance of the different sampling designs through Monte Carlo experiments based on several synthetic datasets. In each of them the auxiliary ( $X$ ) and response ( $Y$ ) variables are drawn from a stationary bivariate spatial process  $[X(\mathbf{s}), Y(\mathbf{s})]$  with  $\mathbf{s} \in [0, 10]^2$  (1000 × 1000 grid). Following Diggle and Ribeiro (2007, Chapter 3), each bivariate spatial pro-

cess in turn is obtained as:

$$\begin{aligned} X(\mathbf{s}) &= f(a * Z_1(\mathbf{s}) + c * Z_2(\mathbf{s})) + k_1 \\ Y(\mathbf{s}) &= g(b * Z_1(\mathbf{s}) + d * Z_3(\mathbf{s})) + k_2 \end{aligned} \quad (2)$$

where:

- $Z_1(\mathbf{s}), Z_2(\mathbf{s}), Z_3(\mathbf{s}) \sim$  are independent univariate stationary Gaussian processes with an Exponential variogram with scale 1 and sill  $\sigma^2 = C + C_0 = 50$ , where  $C$  is the partial sill and  $C_0$  is the nugget. For  $Z_2(\mathbf{s})$  and  $Z_3(\mathbf{s})$  we assume  $C = 50$  and  $C_0 = 0$  in all cases, while for  $Z_1(\mathbf{s})$  their values vary ( $C = 50, 30, 0$ ) to change the proportion of the co-variability that has spatial structure;
- $a, b, c$  and  $d$  are constants whose values vary in order to obtain different correlation level and structure;
- $f(\cdot)$  and  $g(\cdot)$  are transformation functions, for which we consider two choices: Identity or Exponential;
- $k_1$  and  $k_2$  are adding constants to guarantee  $X(\mathbf{s}) > 0$  and  $Y(\mathbf{s}) > 0$ .

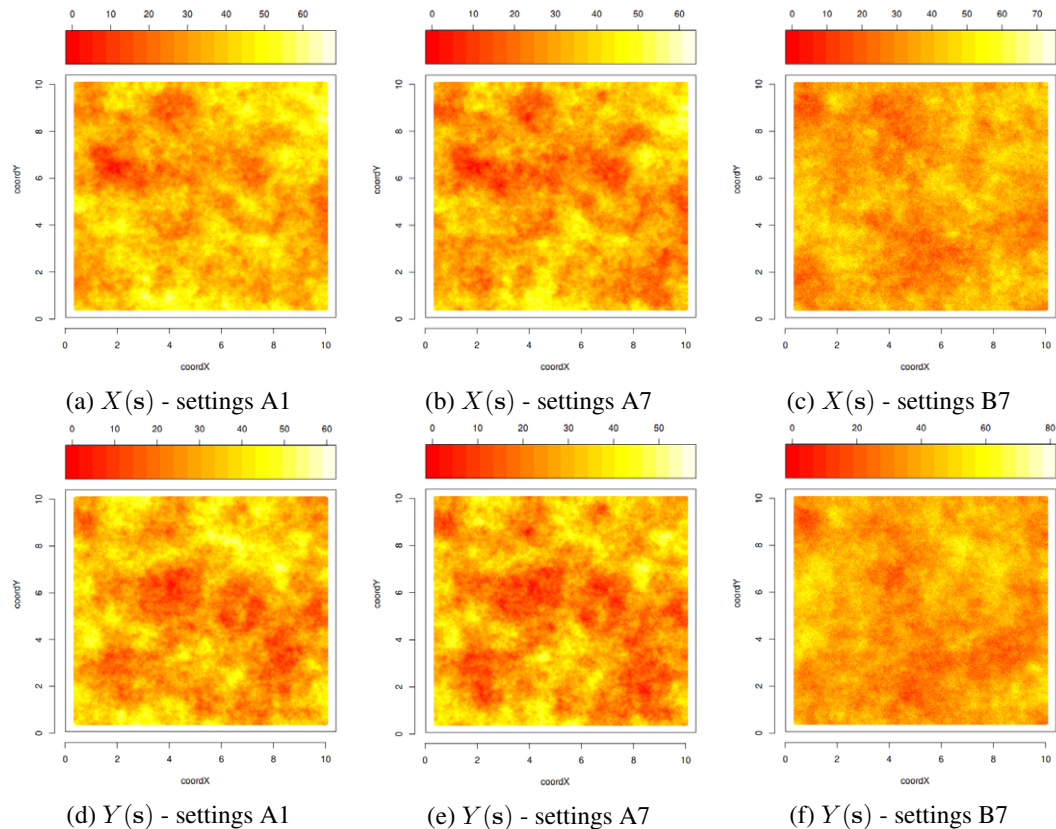


Figure 1: Variables  $X(\mathbf{s})$  and  $Y(\mathbf{s})$  simulated under settings A1, A7 and B7.

Overall, we present our results for 24 synthetic datasets which differ in the spatial distribution of both the study and auxiliary variables, as well as in their relation. The complete list of settings used to generate the synthetic datasets is presented in Table 1. To give a better idea of the different relations between  $X$ ,  $Y$  and  $\mathbf{s}$  that can be simulated in our data, Figure 1 shows

Table 1: Root mean square error, with 500 replications and sample size = 1000

Settings A:		$Z_1, Z_2, Z_3$ with $C = 50, C_0 = 0$				$f(\cdot)$ and $g(\cdot) = \text{Identity}$					
Id	a	b	c	d	Emp.corr.	SRS	SpatLPM	AuxLPM	BivLPM	UneqLPM	SeqUneqLPM
A1	0.6	0.6	1	1	0.298	0.265	0.093	0.257	0.131	0.130	0.113
A2	0.385	1	1	0.385	0.350	0.248	0.088	0.221	0.128	0.111	0.102
A3	1	0.385	0.385	1	0.361	0.242	0.085	0.244	0.115	0.143	0.103
A4	0.82	0.82	1	1	0.424	0.295	0.103	0.250	0.149	0.138	0.119
A5	1	1	1	1	0.515	0.324	0.113	0.314	0.151	0.131	0.119
A6	1	1	0.82	0.82	0.607	0.297	0.104	0.228	0.113	0.110	0.106
A7	1	1	0.6	0.6	0.738	0.269	0.095	0.178	0.095	0.078	0.082

Settings B:		$Z_1$ with $C = 30, C_0 = 20$				$Z_2, Z_3$ with $C = 50, C_0 = 0$				$f(\cdot)$ and $g(\cdot) = \text{Exponential}$	
Id	a	b	c	d	Emp.corr.	SRS	SpatLPM	AuxLPM	BivLPM	UneqLPM	SeqUneqLPM
B1	0.6	0.6	1	1	0.310	0.266	0.129	0.255	0.130	0.151	0.122
B2	0.385	1	1	0.335	0.340	0.241	0.161	0.217	0.153	0.150	0.139
B3	1	0.385	0.385	1	0.400	0.244	0.107	0.236	0.117	0.130	0.102
B4	0.82	0.82	1	1	0.437	0.295	0.156	0.264	0.147	0.140	0.136
B5	1	1	1	1	0.526	0.322	0.181	0.277	0.149	0.138	0.134
B6	1	1	0.82	0.82	0.617	0.294	0.173	0.234	0.130	0.114	0.111
B7	1	1	0.6	0.6	0.744	0.264	0.166	0.189	0.103	0.085	0.083

Settings C:		$Z_1$ with $C = 0, C_0 = 50$				$Z_2, Z_3$ with $C = 50, C_0 = 0$				$f(\cdot)$ and $g(\cdot) = \text{Identity}$	
Id	a	b	c	d	Emp.corr.	SRS	SpatLPM	AuxLPM	BivLPM	UneqLPM	SeqUneqLPM
C1	0.6	0.6	1	1	0.298	0.250	0.154	0.236	0.142	0.129	0.135
C2	0.385	1	1	0.335	0.357	0.228	0.233	0.206	0.189	0.180	0.196
C2	1	0.385	0.385	1	0.345	0.233	0.114	0.240	0.115	0.159	0.102
C4	0.82	0.82	1	1	0.429	0.275	0.200	0.303	0.158	0.138	0.142
C5	1	1	1	1	0.522	0.300	0.239	0.287	0.158	0.133	0.150
C6	1	1	0.82	0.82	0.616	0.274	0.236	0.226	0.142	0.108	0.127
C7	1	1	0.6	0.6	0.747	0.247	0.234	0.167	0.105	0.078	0.099

Settings D:		$Z_1$ with $C = 30, C_0 = 20$				$Z_2, Z_3$ with $C = 50, C_0 = 0$				$f(\cdot)$ and $g(\cdot) = \text{Exponential}$	
Id	a	b	c	d	Emp.corr.	SRS	SpatLPM	AuxLPM	BivLPM	UneqLPM	SeqUneqLPM
D1	0.12	0.1	0.08	0.09	0.508	0.038	0.022	0.033	0.018	0.015	0.015
D2	0.1	0.1	0.05	0.05	0.723	0.030	0.018	0.021	0.013	0.007	0.010

Settings E:		$Z_1, Z_2, Z_3$ with $C = 50, C_0 = 0$				$f(\cdot)$ and $g(\cdot) = \text{Exponential}$					
Id	a	b	c	d	Emp.corr.	SRS	SpatLPM	AuxLPM	BivLPM	UneqLPM	SeqUneqLPM
E1	0.1	0.1	0.05	0.05	0.763	0.026	0.013	0.018	0.011	0.007	0.008

variables  $X(s)$  and  $Y(s)$  generated under settings A1, A7 and B7: in scenario A1 we observe a weak correlation between  $X$  and  $Y$  (equal to 0.298), with both variables strongly related with space; in both scenarios A7 and B7 the correlation between  $X$  and  $Y$  is stronger (more than 0.7), but they differ with respect to the spatial structure of the data since in scenario B7 part of the co-variability (about 40%) is not spatially related.

We choose to simulate scenarios with the different settings discussed above because when the analysis concerns a phenomenon measured at global scale it is common to observe different pattern between different areas of the globe and our aim is to find a strategy which could be globally applied by accounting for the various areas characteristics.

Table 1 presents for each dataset the root mean square error (rmse) of the mean estimator for the sampling designs described above, in addition to the simple random sampling (SRS) which is included as a comparison. The results for the stratified designs are omitted for lack of space since they were in line with the other strategies but they were never the best.

Results confirm that, as expected, when we analyse spatial-related phenomena spreading the sample over the area of interest is always convenient: the SpatLPM strategy is always better than both SRS and AuxLPM. Nonetheless, the use of the auxiliary information can improve the efficiency of the estimates, in particular if it is used to calculate the inclusion probabilities in the unequal designs.

It is important to note that, in order to evaluate when it is more or less convenient to use the auxiliary variable in addition to the geographical location, it is not enough to consider the correlation between  $X$  and  $Y$ : given the same level of correlation, estimates' efficiency depends on the proportion of co-variability that has spatial structure. If the co-variability is all defined by a spatial structure (that is, when  $Z_1$  has  $C = 50$ ), the SpatLPM design (with equal selection probabilities) is enough; on the other hand when part (or all) of the co-variability is not spatially related (that is, when  $Z_1$  has  $C = 30$  or  $C = 0$ ), the additional auxiliary variable improves the estimates' efficiency, especially if used to define the unequal inclusion probabilities (UneqLPM). Moreover, UneqLPM performs better than SpatLPM even when the relation between  $X$  and  $Y$  is not linear (but still positive).

Finally, SeqUneqLPM works better than UneqLPM when the performance of the latter is worse than that of SpatLPM but it does not always manage to reach or improve the performance of the UneqLPM when this is better than the SpatLPM. These last results are very preliminary, as the experiments are still ongoing. Investigation is required on the possibility to modify the sequential procedure in order to consider more phases in which to update the inclusion probability. Moreover, experiments with more additional explanatory variables are in plan.

## References

- Benedetti, R., Piersimoni, F., Postiglione, P. (2017). Spatially balanced sampling: A review and a reappraisal. *International Statistical Review*. **85**(3), pp. 439–454.
- Diggle, P.J., Ribeiro, P.J. (2007). *Model-based Geostatistics*. Springer, New York.
- Grafström, A., Lundström, N.L.P., Schelin, L. (1986). Spatially balanced sampling through the pivotal method. *Biometrics*, **68**, pp. 514–520.
- Grafström, A., Tillé, Y. (2012). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Envirometrics*. **24**, pp. 5120–131.
- Lisic, L., Cruze, N. (2016). Local Pivotal Methods for Large Surveys. In proceedings, ICES V, Geneva Switzerland 2016.
- Tillé, Y. (2020). *Sampling and estimation from finite populations*. Wiley, New York.
- Tillé, Y., Wilhelm, M. (2017). Probability sampling designs: Balancing and principles for choice of design. *Statistical Science*. **32**(2), pp. 176–189.
- Tobler, W.R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*. **46**, pp. 234–240.
- Wang, J.F., Stein, A., Gao, B.B., Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics*. **2**, pp. 1–14.

# Measures of interrater agreement when each target is evaluated by a different group of raters

Giuseppe Bove

## 1. Introduction

Measures of interrater agreement like *kappa* of Cohen (and its weighted versions) and intraclass correlations are usually defined for ratings regarding a group of targets (subjects or objects), each rated by the same group of raters. This happens when the agreement among clinical diagnoses provided by more physicians on the same set of patients is analysed for identifying the best treatment for the patients, or when the agreement among ratings of educators who assess on a new ordinal rating scale the language proficiency of a corpus of argumentative (written or oral) texts is considered to test reliability of the new scale.

In other situations, the agreement between ratings is analysed in a group of targets where each target is evaluated by a different group of raters, like for instance when teachers in a school are evaluated by a questionnaire administered to all the pupils (students) in the classroom. In these situations, it is important to analyse the reliability of the judgments by a measure of agreement between ratings, but since the ordering of the ratings assigned to each target is irrelevant, the measure can only be defined starting from the single target level.

In this paper, an index is proposed to evaluate the agreement between raters for each single target rated on an ordinal scale, and to obtain also a global measure of the interrater agreement for the whole group of targets evaluated. The main features of the proposal will be illustrated in a study for the assessment of the behaviour of student teachers in the classroom. Data were collected in a research conducted in 2018 at Roma Tre University with students of the degree course in *Formazione Primaria*, during their experience of internship (“*tirocinio*”).

## 2. Target-specific measures of interrater agreement

When ratings provided on a quantitative (interval or ratio) scale are analysed in a group of targets where each target is evaluated by a different group of raters, a first approach available to measure the level of agreement for the whole group of targets is based on the ANOVA one-way random model (e.g., Shrout & Fleiss, 1979, McGraw & Wong, 1996). The intraclass correlation (ICC) for this model is the between-target variance divided by the sum of the between-target variance and the error variance (this sum is the ratings total variance). A high value of ICC indicates a good agreement among raters, because it is obtained when the between-target variance exceeds the error variance (that includes the within-target variance) by a wide margin. However, a low ICC value is not necessarily an indication of poor agreement, because a severe restriction in the range of ratings assigned in good agreement by the raters can cause low values of the between-target variance and low values of the ICC (the restriction of variance problem, LeBreton et al., 2003).

To overcome this problem of the ICC, target-specific measures of interrater agreement were proposed to work separately with each target  $i$  in the corresponding row of ratings in the targets  $\times$  raters data matrix. James et al. (1984) proposed the index

$$r_{WG,i} = 1 - \frac{S_i^2}{\sigma_E^2},$$

Giuseppe Bove, Roma Tre University, Italy, giuseppe.bove@uniroma3.it, 0000-0002-2736-5697

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Giuseppe Bove, *Measures of interrater agreement when each target is evaluated by a different group of raters*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.28, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 157-162, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

where  $s_i^2$  is the observed variance of the ratings in profile  $i$ ,  $\sigma_E^2$  is the variance obtained from a theoretical null distribution representing a complete lack of agreement among raters (e.g., the uniform distribution). For raters in perfect agreement, we have  $s_i^2 = 0$ , with a corresponding value  $r_{WG,i} = 1$ . For a total lack of agreement, the observed variance approaches the variance obtained from the theoretical null distribution. This leads  $r_{WG,i}$  to approach 0.

A global measure of agreement for the whole group of targets can be defined as the arithmetic average of the  $r_{WG,i}$  values ( $\bar{r}_{WG} = \frac{1}{N} \sum_{i=1}^N r_{WG,i}$ ). The accuracy of the index depends strongly on the specification of the null distribution, and negative values could be obtained. Other possible indices for quantitative scales are reviewed, for instance, in LeBreton & Senter (2008). Recently, Bove (2022) has considered the normalised standard deviation and the coefficient of variation as possible alternatives to ICC and  $r_{WG,i}$ .

All the approaches described regard quantitative scales and are not appropriate for ordinal and nominal scales. Most of the indices of interrater agreement proposed for ratings on an ordinal scale (frequently averages of the weighted *kappa* of Cohen calculated for each of the possible pairs of raters) are not suitable for ratings regarding a group of targets, each rated by a different group of raters.

In order to propose a new index of interrater agreement for ordinal scales, the representation of the profile of the ratings for target  $i$  on a  $K$ -level ordinal scale in Table 1 is considered,

**Table 1** – Profile of the ratings for target  $i$  on a  $K$ -level ordinal scale

<i>Target</i>	<i>Level 1</i>	<i>Level 2</i>	<i>..Level k ..</i>	<i>Level K</i>	<i>Total</i>
$i$	$r_{i1}$	$r_{i2}$	$\dots r_{ik} \dots$	$r_{iK}$	$R_i$

where,  $r_{ik}$  is the number of raters assigning level  $k$  to target  $i$  and  $R_i$  is the number of raters that rate target  $i$ . We propose a general approach that defines target-specific interrater agreement indices as normalised indices of variability for the distribution in profile  $i$ , according to the measurement level of the scale. A global measure of agreement can be defined as the arithmetic average of the target-specific values of the indices.

So, for ordinal scales, the following index of interrater agreement can be considered (analogous with the measure of dispersion for ordinal variables, e.g., Leti, 1983),

$$\delta_i = 1 - \frac{D_i}{D_{max}} = 1 - \frac{2 \sum_{k=1}^{K-1} F_{ik}(1 - F_{ik})}{D_{max}}$$

where  $F_{ik}$  is the cumulative proportion associated with level  $k$  of the scale in the response profile  $i$ , for  $k=1, 2, \dots, K$ ,  $D_{max}$  is the maximum of  $D_i = 2 \sum_{k=1}^{K-1} F_{ik}(1 - F_{ik})$ , and it is  $D_{max} = (\frac{K-1}{2})$  as  $R_i$  is even, and  $D_{max} = (\frac{K-1}{2})(1 - \frac{1}{R_i})$  as  $R_i$  is odd.

The index  $\delta_i$  is always nonnegative, it is  $\delta_i = 1$  in the case of maximum agreement and  $\delta_i = 0$  in the case of maximum disagreement. Some simulations and experiences with real applications suggest the following thresholds for the interpretation of the values assumed by the  $\delta_i$  index: values lower than 0.6 indicate low to moderate agreement, values between 0.6 and 0.8 good agreement, above 0.8 excellent agreement. The index allows for the identification of particular targets for which agreement is low: this is not possible with measures like *kappa* or intraclass correlations. Besides, a global measure of agreement can be defined as the arithmetic average of the  $\delta_i$  values obtained for the  $N$  targets ( $\bar{\delta} = \frac{1}{N} \sum_{i=1}^N \delta_i$ ). The index is not affected by the possible concentration of ratings in a few levels of the scale, like it happens for the measures based on the ANOVA approach or for the

*kappa*-type indices, and it does not depend on the definition of a null distributions like  $r_{WG,i}$ .

In the next section, an application will be shown in which teachers in a school are evaluated by a questionnaire administered to all the pupils in the classrooms, so each teacher is evaluated by a different group of pupils. In this situation, it is interesting to analyse the level of dispersion of the ratings in the classrooms with respect to each question of the questionnaire, in order to investigate aspects of rating's reliability. Then, a matrix  $\Delta = (\delta_{ij})$  is defined where each row corresponds to a teacher and each column to a question, and the entry  $\delta_{ij}$  is the value of  $\delta_i$  computed in the classroom of teacher  $i$  for question  $j$  (an example is provided in Table 2). Entries of matrix  $\Delta$  can be considered as similarities between teachers and questions. The values  $\delta_{ij}$  can be depicted in a diagram by the *unfolding* model (originally proposed by Coombs (1964) for rectangular matrices of preference scores). The model is

$$f(\delta_{ij}) = p_{ij} = \sqrt{\sum_{s=1}^t (a_{is} - b_{js})^2} + \varepsilon_{ij}, \quad (1)$$

where  $f$  is a monotone transformation, mapping the similarities  $\delta_{ij}$  into a set of dissimilarities  $p_{ij}$  (e.g.,  $p_{ij} = 1 - \delta_{ij}$ ),  $a_{is}$  and  $b_{js}$  are the coordinates respectively of row (teacher)  $i$  and column (question)  $j$  on dimension  $s$  in an  $t$ -dimensional space and  $\varepsilon_{ij}$  is a residual term. It is worth to notice that the Euclidean distance model usually used in multidimensional scaling for square dissimilarity matrices (e.g., Borg & Groenen 2005) is a constrained version of model (1), because for each  $j$  it is required  $b_{js} = a_{js}$ .

So, a diagram for the pattern of relationships is obtained where each row (teacher) is represented as a point with coordinates  $a_{is}$  and each column (question) as a point with coordinates  $b_{js}$ . In the planar representation ( $t=2$ ), the distance between row (teacher)  $i$  and column (question)  $j$  approximates the corresponding dissimilarity  $p_{ij}$  (so, for instance, we can detect in the diagram both the teachers and the questions with low/high levels of agreement of ratings in the classrooms). Distances within each of the two sets of the row-points and the column-points are only implicitly defined and do not have corresponding observed entries in the data matrix. Parameters in the model (1) are estimated by iterative algorithms that, starting from initial estimates of  $a_{is}^0$ ,  $b_{js}^0$  (*initial configuration*), iteratively decreases a least squares loss function moving vectors  $\mathbf{a}_i^0 = (a_{i1}^0, a_{i2}^0, \dots, a_{ir}^0)$  and  $\mathbf{b}_j^0 = (b_{j1}^0, b_{j2}^0, \dots, b_{jr}^0)$ , until convergence to a minimum. An important point is picking a good initial configuration to avoid the problem of *local minima*.

### 3. Application

A reduced version for pupils of the Teachers' Educational Practices Questionnaire (TEP-Q, Catalano et al., 2014) was administered to evaluate a group of 24 female student teachers of Roma Tre University, during their training (internship) in several primary schools of the Italian region Lazio, in school year 2018. The questionnaire consists of the following 12 questions regarding teachers behaviour in the classroom: "In the class she was relaxed" (Q1), "Before each activity, she clearly explained what we had to do" (Q2), "When someone approached her, she turn to look at him" (Q3), "She help us to repeat one thing better if we were not so clear" (Q4), "When someone of us was saying something, she interrupted him" (Q5), "When she talked to us, she also used gestures (for example, she moved her hands)" (Q 6), "She yelled at the class when she get angry" (Q7), "If someone of us needed to be consoled, she has noticed it, even if he did not tell her" (Q8), "During the activities she told us we could help each other" (Q 9), "When she was tired, she complained in class" (Q 10), "She made us do group work" (Q 11), "She praised us when we deserved it" (Q 12). Answers were provided on a 4-levels Likert scale (1=almost never, 4=almost always).



For each student teacher, ratings were obtained from the pupils in the classroom (24 school classrooms, 418 pupils, 204 females, 214 males, aged between 7 and 12 years). For each student teacher  $i$  and each question  $j$ , the  $\delta_{ij}$  value of the index was computed in order to analyse the reliability of the ratings provided by the pupils in the school classroom. Table 2 contains the matrix of the  $\delta_{ij}$  values and in addition, in the last row, the average  $\bar{\delta}_j$  for each question.

**Table 2** – Values  $\delta_{ij}$  obtained for student teachers and questions in the twenty-four school classrooms.

STUDENT TEACHER	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12
1	1.00	0.93	0.63	0.57	0.67	0.34	0.81	0.34	0.60	0.93	0.57	0.53
2	0.63	0.89	0.51	0.57	0.65	0.41	0.61	0.51	0.57	0.73	0.77	0.51
3	0.46	0.70	0.32	0.42	0.26	0.25	0.79	0.50	0.42	0.88	0.61	0.54
4	0.60	0.66	0.43	0.38	0.60	0.45	0.25	0.43	0.35	1.00	0.27	0.56
5	0.82	0.79	0.52	0.69	0.65	0.40	0.59	0.72	0.49	0.32	0.70	1.00
6	0.69	1.00	0.71	0.67	0.87	0.19	0.23	0.27	0.37	0.62	0.52	0.13
7	0.62	0.84	0.73	0.62	0.83	0.42	0.62	0.77	0.65	0.49	0.71	0.50
8	0.87	0.70	0.53	0.53	0.47	0.50	0.81	0.23	0.25	0.67	0.42	0.73
9	0.95	0.80	0.53	0.58	0.83	0.44	0.77	0.35	0.50	1.00	0.61	0.58
10	1.00	1.00	0.51	1.00	0.61	0.49	0.67	0.90	0.44	1.00	0.91	0.77
11	0.81	0.90	0.49	0.56	0.54	0.30	0.38	0.31	0.71	0.33	0.67	0.56
12	0.40	1.00	0.59	0.61	0.71	0.21	0.24	0.46	0.28	1.00	0.61	0.61
13	0.40	1.00	0.51	0.44	0.36	0.57	0.33	0.29	0.36	0.78	0.59	0.62
14	0.53	1.00	0.87	0.73	0.31	0.57	0.72	0.86	0.86	1.00	0.93	0.93
15	0.56	0.86	0.40	0.71	0.59	0.31	0.32	0.24	0.40	0.71	0.65	0.53
16	0.32	0.81	0.83	0.76	0.43	0.33	0.48	0.48	0.35	1.00	0.83	0.44
17	0.78	0.92	0.67	0.61	0.75	0.40	0.32	0.36	0.51	0.85	0.47	0.59
18	0.62	0.67	0.62	0.61	0.55	0.61	0.69	0.29	0.48	0.83	0.51	0.46
19	0.89	1.00	0.71	0.51	0.84	0.45	1.00	0.48	0.30	1.00	0.52	0.67
20	0.32	0.81	0.38	0.27	0.31	0.27	0.13	0.29	0.34	0.67	0.41	0.10
21	0.88	1.00	0.75	0.75	0.62	0.54	1.00	0.94	0.61	0.81	0.62	0.94
22	0.94	0.84	0.73	0.62	0.94	0.32	0.88	0.55	0.22	0.93	0.40	0.73
23	0.80	0.67	0.44	0.38	0.78	0.36	0.33	0.51	0.09	0.67	0.73	0.69
24	0.77	0.83	0.63	0.43	0.46	0.24	0.88	0.48	0.30	0.70	0.67	0.45
Averages $\bar{\delta}_j$	<b>0.69</b>	<b>0.86</b>	<b>0.59</b>	<b>0.58</b>	<b>0.61</b>	<b>0.39</b>	<b>0.58</b>	<b>0.48</b>	<b>0.43</b>	<b>0.79</b>	<b>0.61</b>	<b>0.59</b>

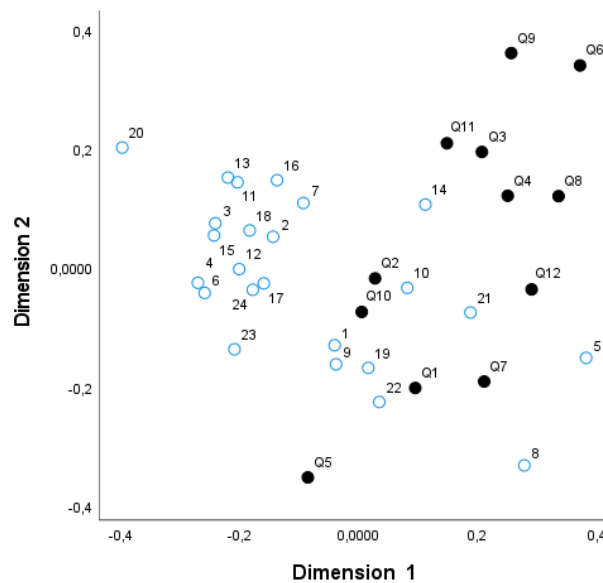
Different levels of reliability characterize the twelve questions. Questions 2 and 10 have high values of the average index (0.86 and 0.79, respectively), that means the pupils usually agree in the responses (in several classrooms it is  $\delta_{ij} = 1$ ). On the contrary, questions 6 and 9 have low values of the average index (0.39 and 0.43, respectively), that means the pupils frequently have different opinions about the aspects of teacher's behaviour considered in the two questions. The remaining questions show low to moderate levels of agreement in the pupil's responses (average values between 0.48 and 0.69).

It is also interesting to analyse the values of the index  $\delta_{ij}$  respect to each student teacher (rows of the matrix in Table 2). For instance, student teachers 10, 14, 19 and 21 have usually high levels of agreement between the pupil's responses in the twelve questions, on the contrary student teacher 20 has low values of agreement except for questions 2 and 10.

Model (1) was applied to analyse in a diagram the relationships between student teachers and questions. It is assumed  $p_{ij} = 1 - \delta_{ij}$  in model (1), this means that distances are inversely

proportional to the values  $\delta_{ij}$ .

In Figure 1, the solution for  $t=2$  dimensions is provided ( $Stress-I=0.29$ ). Distances between student teachers and questions represent the level of agreement of the responses for the questions in the classroom (the lower the distance the higher the agreement). Question 2, question 10 and, to a lesser extent, question 1 are located in the centre of the diagram, close to many points representing teachers, because they have usually high levels of agreement in the responses of the pupils in the school classrooms. Questions 6, 9 and 8 have high heterogeneity in many cases, so they are positioned far apart from many student teachers. Considering the student teachers, we observe that student teacher 20 is far from most questions because she has usually low values of agreement for the ratings obtained in her classroom. On the contrary, student teachers 10, 14 and 21 are near the centre of the diagram and close to many questions, a consequence of the homogeneity of ratings obtained on many questions.



**Figure 1:** Unfolding of the  $\delta_{ij}$  values for student teachers (empty circles) and questions (full black) in Table 2 (the higher  $\delta_{ij}$  the smaller the distance)

#### 4. Conclusion

A descriptive approach has been presented for the analysis of the agreement in ratings given to a group of targets, where each target is evaluated by a different group of raters. An index of interrater agreement defined at the single target level is proposed for ratings given on an ordinal scale, in a manner similar to the definition of the  $r_{WG,i}$  index for ratings on a quantitative scale. Besides, a measure of agreement for the whole group of targets is obtained as the average of the target-specific values. The index presents some advantages respect to the methods based on ANOVA mean squares like intraclass correlation, and respect to many *kappa*-type indices. Besides, when the index is computed for a group of targets and more questions, it is shown that an unfolding model allows to analyse in a diagram the matrix of the values of the index obtained for each target-question pair.

The index proposed is mainly considered as a measure of size of the interrater agreement, therefore developments of this research may concern: 1) an accurate definition of reliable thresholds

useful for the interpretation of the level of agreement in the applications; 2) the study of the sampling properties of the index.

## References

- Borg, I., Groenen, P.J.F. (2005). *Modern Multidimensional Scaling. Theory and Applications* (Second Edition). Springer, New York.
- Bove, G. (2022). Measures of interrater agreement based on the standard deviation, in *51<sup>st</sup> Scientific Meeting of the Italian Statistical Society, Book of short papers*, eds. A. Balzanella, M. Bini, C. Cavicchia, R. Verde, Pearson, Milano, pp. 1644-1649.
- Catalano, M.G. Perucchini, P., Vecchio, G.M. (2014). The quality of teachers' educational practices: internal validity and applications of a new self-evaluation questionnaire. *Procedia-Social and Behavioral Sciences*, **141**, pp. 459-464.
- Coombs, C.H. (1964). *A Theory of Data*. Wiley, New York.
- James, L.R., Demaree, R.G., Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, **69**, pp. 85–98.
- LeBreton, J.M., Burgess, J.R.D., Kaiser, R.B., Atchley, E.K.P., & James, L.R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: are ratings from multiple sources really dissimilar?. *Organizational Research Methods*, **6** (1), pp. 80-128.
- LeBreton, J.M., Senter, J.L. (2008). Answers to 20 questions about interrater reliability and interrater agreement, *Organizational Research Methods*, **11** (4), pp. 815-852.
- Leti, G. (1983). *Statistica descrittiva*. Il Mulino, Bologna.
- McGraw, K.O., Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, **1**, pp. 30-46.
- Shrout, P.E., Fleiss, J.L. (1979) Intraclass correlations: uses in assessing reliability. *Psychological Bulletin*, **86**, pp. 420–428

# A Natural Language Processing approach to measuring expertise in the Delphi-based scenarios

Yuri Calleo, Simone Di Zio, Francesco Pilla

## 1. Introduction

In the Futures Studies context, the Delphi method (Gordon, 1994) is a very popular and empirical approach (Dalkey and Helmer, 1963) often used in combination with the scenario method (Kosow and Gaßner, 2008). Futures scenarios, support decision-makers in a long-term planning context, helping to focus on the key projections of possible/plausible futures and on the major factors that will drive those projections (Bishop et al., 2007). Both scenario and Delphi are often combined with other methodologies, but one of the most interesting and accredited combinations involves precisely these two methods, in an approach known as Delphi-based scenario (DBS), in which the results of a Delphi study are used to develop the futures scenarios (Di Zio et al., 2021).

A crucial phase in a DBS regards the building of a panel of experts, generally formed by a group of people having comprehensive or authoritative knowledge in a particular field, therefore particularly suitable for answering very specific questions regarding the topic dealt with. An old open issue – as in any experts’ consultation – regards the measurement of the *expertise* of the panel members, because each expert has a different degree of competence, and it is very difficult to quantify that degree.

In recent years, some contributions carried out to overcome this issue, most of them proceeding with a self-assessment (or “self-rating”) of the experts, asking panellists to rate their own expertise (Mullen, 2003) on the whole subject matter, or even on each item of the questionnaire. However, this approach could solve the evaluation problem only from a general perspective, specifically, we must take into account some not trivial drawbacks: 1. Self-assessment makes the decision-making process even longer, and experts may be discouraged from participating; 2. Self-evaluation can lead to several cognitive biases which greatly distort judgments on self-competence, such as, among others, overoptimism and overconfidence biases (see, for example, Bonaccorsi et al., 2020). These aspects should not be underestimated, since if we engage experts with low knowledge in a field, this may compromise the total perspective of the survey. It is important to underline here that the measurement of the expertise degree is useful to set a suitable weighting system for the proper use of the different levels of competencies in the panel. Given these premises, with the exponential increase in the use of web-based research platforms and websites on the internet, it is possible to have valuable data and information available about experts. This paper proposes to:

1. **O1:** develop a new method to evaluate the expertise degree;
2. **O2:** implement the method with text-mining techniques and statistical analysis for experts’ evaluation;
3. **O3:** overcome the problem related to the evaluation of different types of expertise coexisting in the same panel.

To showcase our method, we selected a cohort of known experts, part of the “Smart control of the climate resilience” (SCORE) H2020 European project as this would allow us to assess the production of experts with Natural Language Processing and estimate their expertise in a specific area. This paper is organised in the following sections: in Section 2 a brief literature review with a specific statement of the problem will be conducted, in Section 3, we explain the methodology used to develop our method, and in Section 4 the results will be illustrated. In Section 5 we conclude

Yuri Calleo, University College Dublin, Italy, [yuri.calleo@ucdconnect.ie](mailto:yuri.calleo@ucdconnect.ie), 0000-0002-0190-6061

Simone Di Zio, University of Chieti-Pescara G. D’Annunzio, Italy, [s.dizio@unich.it](mailto:s.dizio@unich.it), 0000-0002-9139-1451

Francesco Pilla, University College Dublin, Ireland, [francesco.pilla@ucd.ie](mailto:francesco.pilla@ucd.ie), 0000-0002-1535-1239

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Yuri Calleo, Simone Di Zio, Francesco Pilla, *A Natural Language Processing approach to measuring expertise in the Delphi-based scenarios*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.29, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 163-168, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

with possible future implementations.

## **2. Theoretical framework and related works**

Given the variety of expertise involved in a Delphi panel, from the scientific literature, most of the attempts to evaluate the expertise degree are based on self-assessment or from an assessment made by researchers. In the environmental context, for example, Gorn et al. (2018), studying the climate change effects in the Region of Halle, divided the expertise competencies into two categories: i) expert type A and ii) expert type B. Type A is an expert who has specific competence and practical experience in regional planning and ecosystem services, type B is an expert with theoretical knowledge of spatial and environmental planning, regional geography, and ecosystem services.

Some scholars select experts based on their experience, considering their position within the organization and different variables identified by researchers. Gary and Von der Gracht (2015), for example, consider speaking roles at “futures” conferences and membership greater than six years in the area of interest. In these terms, it is interesting to understand how the range of time of experience in a field is important to evaluate since a member who manage the research context for many years should be more expert in comparison to who has low years of study. However, the previous approaches do not solve the issue of evaluating different types of expertise in the same panel.

As previously described, most of the time, researchers evaluate the experts based on a self-rating, for example, Varho et al. (2016), build a matrix where the experts can select from a series of variables, the areas where they have greater or familiar expertise. In this line of research, an interesting coefficient was developed by Barroso and Cabero (2013). The coefficient, named K-expert competence, is based on the self-assessment of experts, considering two components, one related to self-evaluated competence and another to the ability to argue on the subject.

That said, there is a need to develop an objective method that avoids self-assessment or evaluation by researchers or scholars in order to reduce cognitive errors and time-consuming, with enough flexibility to be applied on panels of different natures (for example environmental studies can include several participants with different expertise at both theoretical and practical level). To pursue the research aim, we apply web-mining and text-mining techniques to extract information, in order to obtain objective information in a short time, starting from objective criteria and taking into account a plurality of criteria which, in the mixed panels, are important to consider.

## **3. Materials and methods**

We propose a new method to evaluate the expertise degree generally applicable to all participatory decision-making processes and, in particular, to Delphi panellists. We apply the method considering a list of experts in the coastal erosion context, understanding the degree of expertise of the members in the main keywords of the H2020 SCORE project: “coastal erosion”, “sensors”, “Ecosystem-Based Approaches” (EbA), “flood risk assessment”.

The first phase starts where a list of possible experts to engage is already defined and, to showcase our method, we use a list of the H2020 SCORE project members. In the DBS, the literature does not uniformly agree on the number of experts to involve, however, there is a consensus on the range of 10-30 (see Nowack and Endrikat, 2011), for that we identify a list of  $N = 20$  possible experts to be involved as panellists.

The data on the selected experts are organized in a matrix including all the information useful to identify them and their personal pages on the web (e.g., name, surname, personal contacts, personal websites, personal portfolio etc.). In our case, we have different experts with different job roles and expertise, for that, we divide the panel using the following roles:

- $r_1$ : Academic experts;
- $r_2$ : Experts from industry and members of companies;
- $r_3$ : Local and governmental authorities.

A Delphi panel should be as varied as possible, as creativity and the differences in knowledge should be as diverse as possible. However, this opportunity turns into a challenge to be faced, as each of the categories must be evaluated with different criteria. For example, a local authority cannot be evaluated based on scientific publications, or a company manager cannot be evaluated on a social network private profile. In these terms, once we have a data repository with personal information related to each expert, we proceed to evaluate the participants on different variables of our interest, in a multi-criteria approach.

For our study, we decide to extract the number of contributions for each keyword and each expert, from publications, citations, h-index, reports, patents and policies related to the keywords. To acquire the previous information, we refer to the Google Scholar database for the publications, citations, h-index and patents, for the reports we refer to ResearchGate and personal webpages, and for the policies, we take into account the governmental webpages and portfolios of the panellists.

The procedure of data extraction cannot be carried out manually and for that we implement a Python script using the Beautiful Soup library, using text-mining in order to extract the main keywords in a webpage related to a determined topic. Beautiful Soup (Nair, 2014) is a Python library used for web scraping, it allows us to extract data from HTML and XML files obtaining a “parse tree” from the source code of the selected page. First of all, we import all the URLs acquired in the previous phase in Python, after that, we select the keywords of interest (“coastal erosion”, “sensors”, “Ecosystem-Based Approaches”, and “flood risk”) and we run the script.

The outputs show the number of times a given keyword is present on the page without repetitions, allowing us to build separate distributions of h-index, citations, publications, reports, patents, and policies for each expert.

After extracting all the data, we build a matrix, say  $X$ , with  $N$  experts on the row and  $p$  variables on the columns. The first two variables are the h-index and citation, independent of the keywords. The other four variables (publications, reports, patents, policies) are repeated within each keyword. This is because we want to take into account, for example, how many publications an expert has with “coastal erosion” as a keyword, how many reports with the same keyword, etc. Therefore, we have four variables for each of the four keywords, for a total of  $p = 18$  variables.

The main shortcoming is that the column vectors of  $X$  ( $X_{i1}, \dots, X_{ip}, i = 1, \dots, N$ ) have various locations and variabilities, so they cannot be directly combined. Therefore, the data should be made comparable by normalization and, among the various methods of normalization, here we consider the min-max:

$$Y_{ij} = \frac{X_{ij} - \min_i(X_{ij})}{\max_i(X_{ij}) - \min_i(X_{ij})}$$

To avoid computational problems, in case  $\max_i X_{ij} = \min_i X_{ij} = 0$  we set  $Y_{ij} = 0$ , and if  $\max_i X_{ij} = \min_i X_{ij} > 0$ , we set  $Y_{ij} = 1$ .

The last phase permits to have a coefficient of production for each expert (say  $K_i$ ), based on a weighted sum of the normalized variables, which represent a comprehensive measure of expertise:

$$K_i = \sum_{j=1}^p Y_{ij} w_j$$

with  $i = 1, \dots, N, j = 1, \dots, p$  and  $\sum_{j=1}^p w_j = 1$ .

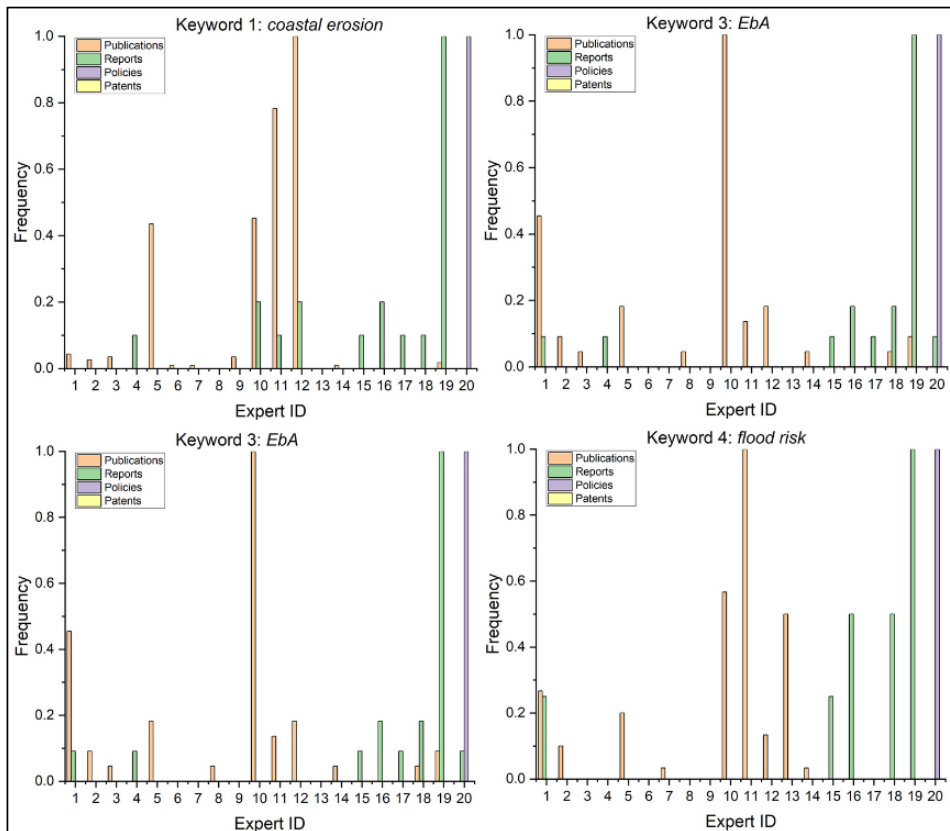
In this application, we set the weights constant to  $w_j = 1/p$ , but the method is very flexible, and the assessment of each weight is left to the team of researchers. After the normalisation of the variables, we proceeded with a weighted sum of the results with (as a first application and by way of example) constant weights  $w_j = 0.05$ . In the end, for each expert, we obtained a score for each variable and for each keyword, and a final score  $K_i$  calculated as a weighted sum, having in this way both the possibility of evaluating the experts for each keyword, understanding who has greater expertise and evaluating the degree of expertise in the macro-topic of interest.

The weighted sum at the base of the coefficient  $K_i$  is only one possible aggregation rule, but other rules can be used, such as a multiplicative one. Also, for normalization, it is possible to use other methods, such as standardization with mean and standard deviation or rank transformation. In these terms, this coefficient becomes a quantitative, flexible, and multi-criteria measure of expertise.

#### 4. Results and discussion

The results illustrated below answered the research objectives and made it possible to have an objective evaluation of a sample of experts. The method is useful for both the evaluation of a predefined panel of experts (for example to weigh their answers in a questionnaire) and to build a new panel, in order to include the people with the highest expertise. The overall results (depicted in Figure 1), demonstrate a high level of expertise in the keywords of our interests, some of which contributed to the topic with publications, reports, and policies.

Figure 1. Normalized frequency for each keyword



Academic experts ( $\#r_1 = 14$ ) contributed efficiently to the research in the field of coastal areas, sensors, EbA and flood risk (Table 1). Specifically, expert 10 is the academic who has contributed most to the areas of our interest with an expertise degree of 0.216 and an h-index of 51 with 16123 citations, for an overall of 95 publications and reports in the keywords analysed.

Experts from the industry sector ( $\#r_2 = 5$ ), contribute to the areas of interest within the publications of reports and scientific articles, however, no patents have been found. In particular, expert 18, published 6 scientific papers with an average of 194 citations. For expert 19, we have found 4 scientific publications and 25 reports submitted in the context of research projects. The only local authority expert ( $\#r_3 = 1$ ) has an overall of 17 policies, 10 in keyword 1, 5 in keyword 3 and 2 in keyword 4 with 1 report in keyword 2.

**Table 1. Expertise degree estimates**

<i>Expert Id</i>	<i>Role</i>	<i>Expertise degree (K)</i>	<i>Expert Id</i>	<i>Role</i>	<i>Expertise degree (K)</i>
1	Academic	0.113	11	Academic	0.152
2	Academic	0.028	12	Academic	0.166
3	Academic	0.040	13	Academic	0.029
4	Academic	0.016	14	Academic	0.013
5	Academic	0.132	15	Industry	0.022
6	Academic	0.007	16	Industry	0.044
7	Academic	0.010	17	Industry	0.010
8	Academic	0.002	18	Industry	0.046
9	Academic	0.003	19	Industry	0.155
10	Academic	0.216	20	Local authority	0.155

In our application, the high scores were identified in the experts' 10, 12, 19, 20 and 11, and with all other scores, we obtained a full ranking of the experts based on their degree of expertise (Table 1), demonstrating the efficiency of the approach in a different context of applications and for different work roles. With these results, it will be possible to select a subsample of more competent experts ("super experts") and/or weigh Delphi's responses/evaluations of the panel. In this way, there are no restrictions in terms of the choice of participants, as any expertise or work situation can be assessed by setting variables suited to the research work.

## 5. Concluding remarks and future works

This study proposed a new approach to evaluate the expertise degree in the participatory process, in particular in the Delphi-based future scenarios development. We applied this method to a cohort of experts' part of the "Smart control of the climate resilience" (SCORE) H2020 European project in order to estimate their expertise in the context of our interest. The results showed how the method solves one of the main problems in the decision-making process: the evaluation of participants' expertise is useful, for example, in weighing their assessments.

The method is a contribution to the objective measurement of expertise, useful in the context of panels with heterogeneous types of competencies and based on automated data retrieval.

In the application, we had no citizens that normally could be useful in the last phases of Delphi, however for the citizens we could evaluate blogs, social networks, and personal pages, referring to the main social networks (e.g., Twitter, Instagram, LinkedIn etc.).

For future work, it would be interesting to have a comparison between the objective measure described in the paper and the self-evaluation of experts. Furthermore, to consider other aggregation formulas as well as other normalization methods. Finally, to set appropriate weights for the selected variables, among the various possible approaches, we suggest the application of the Analytic Hierarchy Process (AHP), which is very efficient in generating objective weights in a multi-criteria context (Saaty, 1980).



## Acknowledgements

The work carried out in this paper was supported by the project SCORE which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003534.

## References

- Barroso O. J. M., & Cabero A. J. (2013), La utilización del juicio de experto para la evaluación de TIC: el coeficiente de competencia experta. Bordón. *Revista de Pedagogía*, **65** (2), pp. 25-38. <https://doi.org/10.13042/brp.2013.65202>.
- Bishop P., Hines A., Collins T. (2007), The current state of scenario development: An overview of techniques, *Foresight*, **9**(1), pp. 5–25.
- Bonaccorsi A., Apreda R., Fantoni G. (2020), Expert biases in technology foresight. Why they are a problem and how to mitigate them, *Technological Forecasting & Social Change*, **151**, pp. 1–17.
- Dalkey N., & Helmer O. (1963), An experimental application of the Delphi method to the use of experts. *Management science*, **9**(3), pp. 458-467. <https://doi.org/10.1287/mnsc.9.3.458>.
- Di Zio S., Bolzan M., Marozzi M. (2021), Classification of Delphi outputs through robust ranking and fuzzy clustering for Delphi-based scenarios, *Technological Forecasting & Social Change*, **173**, pp. 1211400040. [doi.org/10.1016/j.techfore.2021.121140](https://doi.org/10.1016/j.techfore.2021.121140)
- Di Zio S., Rosas, J. D. C., & Lamelza, L. (2017), Real Time Spatial Delphi: Fast convergence of experts' opinions on the territory. *Technological Forecasting and Social Change*, **115**, pp. 143-154. <https://doi.org/10.1016/j.techfore.2016.09.029>.
- Gary J. E., & Heiko A. (2015), The future of foresight professionals: Results from a global Delphi study. *Futures*, **71**, pp. 132-145.
- Gordon T. J. (1994). The delphi method. *Futures research methodology*, **2**(3), pp. 1-30.
- Gorn L., Kleemann J., & Fürst, C. (2018), Improving the Matrix-Assessment of Ecosystem Services Provision—The Case of Regional Land Use Planning under Climate Change in the Region of Halle, Germany. *Land*, **7**(2), pp. 76. <https://doi.org/10.3390/land7020076>.
- Kosow H., Gaßner R. (2008), Methods of future and scenario analysis: overview, assessment, and selection criteria, Bonn, Deutsches Institut für Entwicklungspolitik.
- Mullen P. (2003), Delphi: myths and reality, *Journal of Health Organisation and Management*, **17** (1), pp. 37–52.
- Nair V. G. (2014), Getting started with beautiful soup. Packt Publishing Ltd.
- Nowack M., Endrikat J., & Guenther E. (2011), Review of Delphi-based scenario studies: Quality and design considerations. *Technological Forecasting and Social Change*, **78**(9), pp. 1603-1615.
- Rodríguez P. O. S. (1995), Land use conflicts and planning strategies in urban fringes: a case study of western Caracas, Venezuela.
- Rowe G., & Wright G. (1999), The Delphi technique as a forecasting tool: issues and analysis. *International journal of forecasting*, **15**(4), pp. 353-375. [https://doi.org/10.1016/S0169-2070\(99\)00018-7](https://doi.org/10.1016/S0169-2070(99)00018-7).
- Saaty T.L. (1980), The Analytic Hierarchy Process. McGraw-Hill, New York.
- Sossa J. W. Z., Halal, W., & Zarta, R. H. (2019), Delphi method: analysis of rounds, stakeholder and statistical indicators. *Foresight*. <https://doi.org/10.1108/FS-11-2018-0095>.
- Varho V., Rikkonen P., & Rasi, S. (2016), Futures of distributed small-scale renewable energy in Finland—A Delphi study of the opportunities and obstacles up to 2025. *Technological Forecasting and Social Change*, **104**, pp. 30-37.

# Exploring Globalization with Cosmopolitics

Maria Serena Causo, Erika Cerasti, Fabrizio De Fausti, Monica Scannapieco

## 1. Introduction

The global economy underwent a big transformation over the last years, due to the great increase of international trade. According to World Bank and OECD national accounts data, the World export propensity, i.e. the percentage ratio of exports to the gross domestic product (GPD) has grown constantly since the mid '80s until 2008, when a maximum of 31% was reached. Despite the critical phases after 2008, in 2020 the estimated world export propensity was 26.5%, i.e. more than one fourth of total global production is exported. International production, trade and investments are increasingly entangled in the global value chains (GVCs): different production and distribution processes can be located across different countries, providing economic advantages (Surugiu and Surugiu 2015). While the interdependence between countries' economies, stemming from the GVC, increases the level of efficiency, it poses risks of instability affecting the whole production and trade system when local crises arise. This is even more true for crises on a larger scale, like the COVID-19 pandemic (Lin and Zhang 2020) or the Russian-Ukrainian conflict.

The GVC causes the transmission of shocks which can drastically disrupt the supply chains of some products, a risk that became evident in several phases of the pandemic when medical products flow was interrupted (Verschuur et al. 2021). To prevent this risk, policy makers should engage in new trade agreements to avoid disruption in products supply (Barlow et al. 2021). To this aim it would be useful to support government's decision makers with new policy tools, which can give hints about how to "re-localize" GVCs, identify key potential sources of shock exposure in GVCs and assess different policy scenarios, in terms of both economic efficiency and stability (OECD 2021).

Within this framework it is extremely important for policy makers to have appropriate tools to analyze qualitatively and quantitatively the evolving structure of GVC. A suitable tool should exploit sound quality statistical trade data, as provided by official statistics, allow dynamic multidimensional analysis, and provide a high-level, interactive, easy-to-use visualization of relevant information.

The presented dashboard was developed by Istat in the framework of the Big Data Hackathon, and it enables a general analysis of the effects of any local crisis on global world trade by both social network tools and time series analysis.

## 2. Network analysis on international trade data

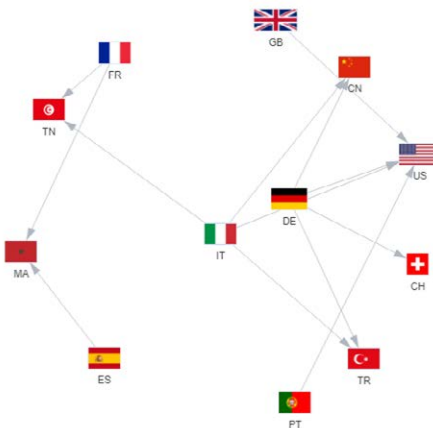
We built an integrated tool, which can provide dynamic views and interactive analysis of GVC across European and extra-European countries. The tool is based on the online available "Monthly COMEXT Data", containing all the international trades in import and export (except for trades between extra-European countries). The tool is a dashboard providing interactive views of graphs of international trade relations, in the framework of social network analysis (De Benedictis et al. 2013). Countries in the COMEXT dataset are represented as graph nodes connected to each other by arrows, edges of the graph (Wasserman and Faust 1994) that represent the traded value of products (in Euros) exchanged between the two countries in a considered time period (Figure1). The graphical visualization is useful to have qualitative information about countries holding a central role in the structure and countries serving as bridges between different areas of the network. Those insights are

Maria Serena Causo, ISTAT, Italian National Institute of Statistics, Italy, [causo@istat.it](mailto:causo@istat.it), 0000-0002-9879-013X  
Erika Cerasti, ISTAT, Italian National Institute of Statistics, Italy, [erika.cerasti@istat.it](mailto:erika.cerasti@istat.it), 0000-0002-3495-9290  
Fabrizio De Fausti, ISTAT, Italian National Institute of Statistics, Italy, [defausti@istat.it](mailto:defausti@istat.it), 0000-0002-6921-2584  
Monica Scannapieco, ISTAT, Italian National Institute of Statistics, Italy, [scannapi@istat.it](mailto:scannapi@istat.it)

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Maria Serena Causo, Erika Cerasti, Fabrizio De Fausti, Monica Scannapieco, *Exploring Globalization with Cosmopolitics*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.30, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 169-173, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



**Fig. 1 Social network of Textile products trade in January 2020**

then quantified by the centrality measures that characterize the graph and each country in the network:

- *Product spread*: a global measure corresponding to the Graph density, representing the ratio between the number of edges in a graph and the maximum number of edges that the graph can contain.
- *Vulnerability*: a local measure corresponding to  $(1 - \text{the indegree centrality})$  for each country vertex, where the indegree centrality is a normalized measure for the number of in-coming links. The vulnerability conveys the message that a country receiving a product from several countries is less dependent on singles countries for the product supply.
- *Export strength*: a local measure corresponding to the country outdegree centrality, a normalized measure for the number of out-going links.
- *Hubness*: a local measure corresponding to the closeness centrality for each country.

The tool is interactive, so the user can focus on graphs of specific products supply chain (same classification as COMEXT dataset), on import views or export views, on specific periods of time, on a percentage of the total trade flow, by selecting filters values.

Fig.1 shows an example of the social network representing 30% of the global export of Textile Yarn, Fabrics, Made up Articles and Related Product, in January 2020.

Moreover, starting from a specific supply chain graph, the tool provides the possibility to remove chosen links, both globally and for selected mode of transports, and re-compute graph indicators corresponding to the new graph configuration. This feature allows to determine if a specific trade disruption would increase country import vulnerability, or which exporting country would take advantage by increasing its export strength in the new configuration. This allows to perform scenario analysis, and to foresee if a critical trade disruption would make an importing country particularly dependent by specific geo-political areas.

### 3. Analysis insights and results

The dashboard allows the user to follow the evolution in time of a trade network, by comparing graphs associated to different time periods. It can enable to spot changes in the role played by different countries in the network of relations, allowing to detect countries playing central roles; it can give information on market contraction or expansion; it allows to detect isolated clusters or countries more vulnerable to products supply disruption; it allows to perform analysis of scenario and to evaluate the effect of political and economic agreements and strategies.

In the following we show an example of evolution analysis, comparing graphs of international trades

of all the products for the same period in different years. We consider the second trimester (T2) of 2021 (see Fig.2) and the second trimester (T2) of 2022 (see Fig.3)



**Fig. 2 Social network of all products trade of the second trimester 2021 (T2).**

The measure of Product spread (graph density) indicates the percentage of existing trading relations between countries among all the possible ones (it's not a measure of traded amounts). The product spread of all products decreased from 0.10 in T2-2021 graph to 0.087 in T2-2022 graph, meaning that some relevant commercial links between countries ceased. One possible cause could be the Russian-Ukrainian conflict.

**4. Data sources**

Data sources on international trade in goods used by the presented dashboard consist in EU official statistics data produced by the 27 Member States according to harmonized methodologies based on EU statistical regulations and available in the Eurostat COMEXT database, freely accessible at <http://epp.eurostat.ec.europa.eu/newxtweb/>. They provide trade data in monetary value and physical quantities at maximum granularity in time resolution (monthly frequency), traded product characteristics, trade partner countries, mode of transport and nature of the transaction.



Fig. 3 Social network of all products trade of the second trimester 2022 (T2).

## 5. Conclusions

The GVC presents risks of instability for international commercial trades, so assessing the country exposure to potential shocks and crises by monitoring the time evolution of indicators such as country vulnerability can be very important. The proposed interactive dashboard can be a valuable tool to support policy makers in the decision making process relative to economic strategies. It provides views, measures, and filters to analyze the structure of trading relations between countries, its evolution in times and its relevant features. It allows to perform scenario analysis, by acting on the graphs and evaluating the effects of actions on the trade structure.

The Dashboard is available at the following link: <https://www.terra.statlab.it>

## References

- Barlow P., van Schalkwyk MPH M.C., McKee DSc M., Labonté Rom, Stuckler D. (2021), “COVID-19 and the collapse of global trade: building an effective public health response”, *The Lancet Planetary Health*, Volume 5, 2, Pages 102-107
- De Benedictis L., Nenci S., Santoni G., Tajoli L., Vicarelli C. (2013), “Network Analysis of World Trade using the BACI-CEPII dataset.” *CEPII Document de travail N 2013 – 24 August*.
- OECD (2021), Global Value Chains: Efficiency and Risks in the Context of COVID-19, *OECD Tackling Coronavirus (COVID-19) Contributing to Global Effort*, 11 February 2021. OECD Publishing, Global value chains: Efficiency and risks in the context of COVID-19 - OECD ([oecd-ilibrary.org](https://oecd-ilibrary.org))
- Lin B., Zhang Y. Y. (2020), “Impact of the COVID-19 pandemic on agricultural exports”, *Journal of Integrative Agriculture*, Volume 19, Issue 12, Pages 2937 – 2945.

- M.R. Surugiu, C. Surugiu (2015), “International Trade, Globalization and Economic Interdependence between European Countries: Implications for Businesses and Marketing Framework”, *Procedia Economics and Finance*, Volume 32, 2015, Pages 131-138
- Verschuur J., Koks E. E., Hall J. W. (2021), “Observed impacts of the COVID-19 pandemic on global trade”, *Nature Human Behaviour*, 5, 305-307
- Wasserman S., Faust K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, NY, <https://doi.org/10.1525/ae.1997.24.1.219>



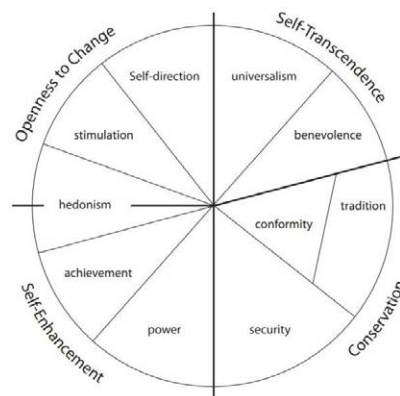
# Professional choices and personal values: Similarities and differences between Schein's career anchors and Schwartz basic values

Maria Cristiana Martini, Aldo Arra

## 1. Introduction

Values, beliefs and motivations lead every personal choice, including professional decisions. Consistency between personal values and career choices is essential to achieve job satisfaction and to attain positive career outcomes and self-realization.

Schwartz and Bilsky (1987) propose a framework of ten basic values, measured through the Portrait Value Questionnaire, related to the universal needs of existence. In their theory, the pursuit of some of these values may conflict, while others are consistent. Aiming to clarify the mutual relationships among the ten basic values, these are represented in a circular shape, according to their similarities and dissimilarities (Figure 1), with a contraposition between openness to change (values of stimulation and self-direction) and conservation (security, conformity, and tradition), and between self-enhancement (hedonism, achievement, power) and self-transcendence (universalism, benevolence). Some authors have proposed specific work values scales obtained by adapting Schwartz's basic values to the work environment (see, e.g., Pike, 1996; Porto and Tamayo, 2003; Avallone, 2009).



**Figure 1.** Circular representation of the basic values (adapted from Schwartz, 2012).

Focusing on professional goals and aspirations, Schein's Career Orientation Inventory (1990) identifies eight anchors that drive employees' career paths and orientations: general managerial competence, technical/functional competence, autonomy/independence, security/stability, entrepreneurial creativity, dedication to a cause, pure challenge, life-style. Schein affirms that a career anchor is "that one element in a person self-concept, which he or she will not give up even in the face of difficult choices" (1990). Conversely, Feldman and Bolino (1996) hypothesize that some career orientation are quite similar and complementary, while others are counterpoised and incompatible; e.g., they posit that technical competence and challenge anchors are complementary, while security and entrepreneurial creativity are mutually inconsistent (Figure 2).

Maria Cristiana Martini, University of Modena and Reggio Emilia, Italy, [cmartini@unimore.it](mailto:cmartini@unimore.it), 0000-0001-5622-9187

Aldo Arra, University of Modena and Reggio Emilia, Italy, [aldo.arra@gmail.com](mailto:aldo.arra@gmail.com)

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Maria Cristiana Martini, Aldo Arra, *Professional choices and personal values: Similarities and differences between Schein's career anchors and Schwartz basic values*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.31, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 175-180, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



However, there is no general agreement on the structure underlying career anchors (Barclay et al., 2013).



**Figure 2.** Feldman and Bolino (1996) factor structure of career anchors.

Although these two paradigms have been developed and applied in different contexts, and have rarely been compared in the scientific literature (see Abessolo et al., 2017 for an exception), they seem to share a common ground, which is worth analysing. In this paper, we aim at understanding the mutual relationship between the paradigms proposed by Schwartz and Schein, in order to enlighten how personal motivations inform career preferences and choices. Section 2 presents the survey and the preliminary analyses carried out on the two scales. Section 3 illustrates the similarities and differences between Schwartz's and Schein's theoretical frameworks that can be deduced from the data. Finally, in Section 4 we draw some conclusions and a few sparks for future research.

## 2. Data and methods

We administered the Portrait Value Questionnaire (PVQ) and the Career Orientation Inventory (COI) scales to a sample of 253 respondents through an online survey questionnaire. The respondents were a heterogeneous sample of Italians working in a wide range of fields and positions, aged between 22 and 67 (mean = 36.15; SD = 12.46); the majority are females (58%), and they are distributed in all the Italian regions (47.9% North, 13.2% Centre, 37.9% South, 2.0% abroad).

The COI consisted of eight career anchors, each measured by a set of five items, for a total of 40 items on a 7-point scale; the PVQ includes ten dimensions, measured through a number of items ranging from three to six, totalling 40 more items on a 7-point scales. We assessed each dimension of the two scales through Cronbach's  $\alpha$ , and we evaluated the structural validity of the two measurement models by means of Lisrel 8.7 (Jöreskog and Sörbom, 2004). The measurement models appeared to fit well for the COI (RMSEA = 0,028) and acceptably for the PVQ (RMSEA = 0,060) (Hu and Bentler, 1999).

If we analyse the scores<sup>1</sup> of males and females for each dimension (Table 1), and the correlations between each dimension and the age, we can see gender differences mostly affect the career anchors, while basic values are more likely to change with the age. Women score higher than men on universalism, i.e. the value of understanding, tolerance and protection, while men are more oriented to power, defined as the value of prestige, social status, and control over people and resources. Older people are more aimed at security, conformity, tradition and universalism, while

---

<sup>1</sup> We computed factor scores and average scores for each dimension, which give results completely comparable; in Table 1 we report average score for the sake of readability.

power, achievement, stimulation and hedonism score higher on younger respondents.

Looking at the career anchors, there are no age differences except for technical competences: younger people are more excited by the content of the work itself, and appreciate the feeling of being experts in their field. As for the gender differences, women value more service/dedication, and then love the idea of doing a job which in some way improves the world and helps the society; they also appreciate slightly more security, therefore they show more long-term attachment to the organization, and tend to dislike travel and relocation. On the other side, men are more led by the anchors of creative entrepreneurship and management, which implies they are attracted by the idea of leading people, creating and realising new projects, and they feel stimulated by crises. The male scores are slightly higher also for the challenge and autonomy anchors, indicating a motivation to solve difficult problems and overcome major obstacles, and a need to set own schedule.

**Table 1.** Average score of males and females, and correlation between score and age, for each dimension of PVQ and COI.

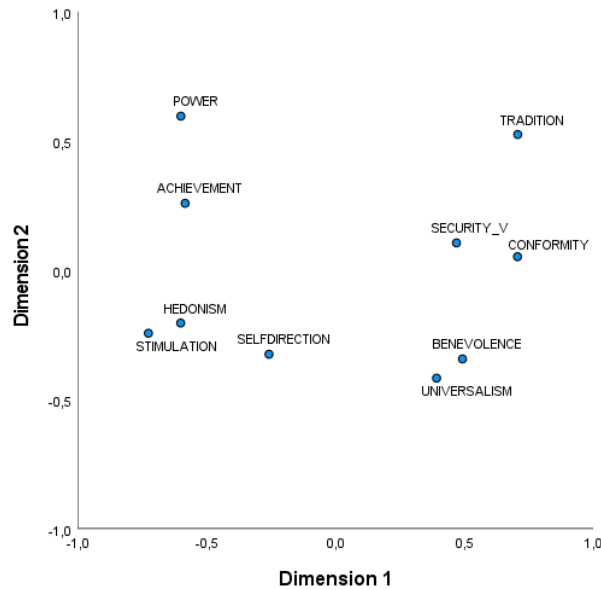
	Male average score	Female average score	Score-age correlation
<i>Portrait Value Questionnaire</i>			
Self-direction	5.70	5.69	-0.107
Power	3.95*	3.57*	-0.137*
Universalism	5.84°	6.01°	0.147*
Achievement	4.97	4.91	-0.326**
Security	5.29	5.39	0.131*
Stimulation	4.73	4.75	-0.295**
Conformity	5.38	5.47	0.230**
Tradition	4.04	4.12	0.294**
Hedonism	4.88	4.83	-0.285**
Benevolence	5.65	5.72	-0.032
<i>Career Orientation Inventory</i>			
Autonomy	5.29°	5.06°	0.012
Creative entrepreneurship	4.33*	3.95*	0.040
General management	4.11*	3.77*	-0.027
Service/Dedication	5.05*	5.35*	0.054
Challenge	5.02°	4.81°	-0.003
Security	5.19°	5.37°	-0.005
Lifestyle	5.54	5.47	-0.018
Technical competence	5.22	5.22	-0.172**

Significance level: \*\* 0.01, \*0.05, °0.10

### 3. The structure of values and career anchors

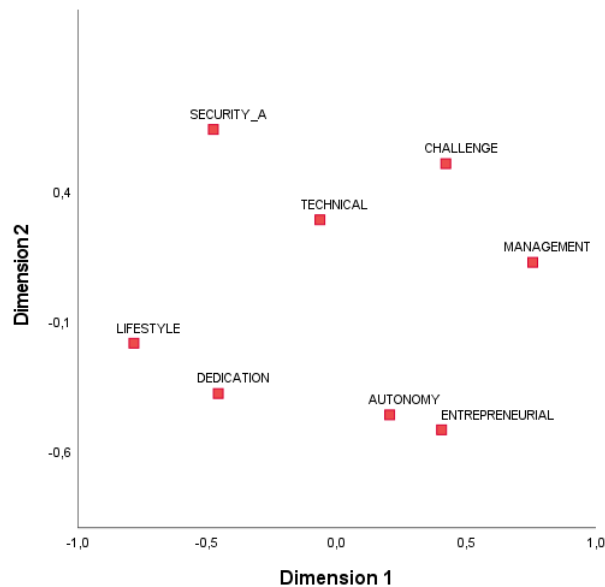
First, we aim at obtaining graphical representations of the mutual relationships among the 8 anchors and among the 10 basic values, and we perform multidimensional scaling analyses, with ordinal proximity transformations and Euclidean distance measures. As for the analysis carried out on the correlations between Schwartz's basic values, we obtain an acceptable value of 0.06 for the stress-1 measure (Schwartz and Sagiv, 1995); this solution accounts for 99.6% of the dispersion. The perceptual map is reported in Figure 3: closer points indicate higher positive correlations, while counterpoised points indicate negative correlations. The graphical representation of the basic values is perfectly consistent with the theoretical structure in Figure 1: we can recognise the openness to change area on the lower left part of the plot, the self-

transcendence on the lower right, the conservation dimension on the upper right, and the self-enhancement on the upper left.



**Figure 3.** Bi-dimensional plot of basic values.

Focusing on the representation of the career anchors, the stress measure is slightly worse but still acceptable (0.13), and the dispersion accounted for is 98.2%. In the multidimensional scaling plot of Figure 4, we can see that the structure resembles the theoretical correlation structure proposed by Feldman and Bolino (1996) and reported in Figure 2: lifestyle and service/dedication are opposed to challenge and managerial competence, while autonomy and entrepreneurial creativity are counterpoised to security.

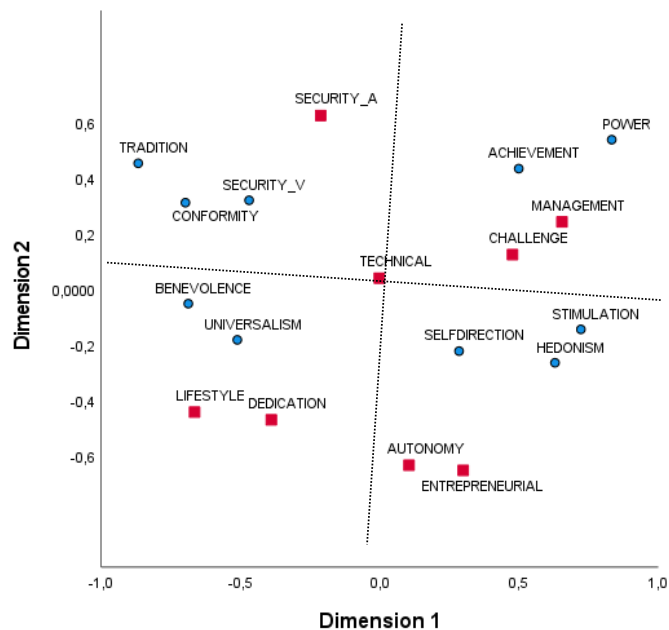


**Figure 4.** Bi-dimensional plot of career anchors.

If, instead of two separate matrices, we analyse the 18×18 matrix of similarities among all the eight anchor items and all the ten values items, besides the relationships among values and the relationships among career anchors we can also explore the mutual interconnection between the set of basic values and the set of career anchors. We can then report the whole system of correlations in a unique plot (Figure 5), and we observe that Schwartz’s and Schein’s theoretical frames show a high level of consistency. We can divide the scatterplot into four sections, corresponding to the poles of Schwartz’s main dimensions:

- those who are more oriented to the “openness to change” dimension, who appreciate independence, novelty and exploration, tend to favor careers of autonomy and entrepreneurial creativity;
- a conservative motivation leads to more safe and stable careers, through a search for commitment, respect and acceptance of social expectations and norms;
- personal values of self-enhancement, i.e. the tendency to pursue personal success, prestige and social status, support challenging careers and managerial responsibilities;
- self-transcendence pushes for dedication to a cause and work-life balance, because it values protection, understanding, and enhancing the welfare of family, friends and personal contacts in general.

The only point which do not find a clear collocation in this segmentation is the anchor of technical competences; in fact, among the career anchors this is the only one which can hardly be associated with both beliefs and motivations.



**Figure 5.** Bi-dimensional plot of career anchors (red squares) and basic values (blue circles).

#### 4. Conclusions

In this paper, we investigated the relationship between two theoretical frameworks: the Schwartz’s basic values and the Schein’s career anchors. Our study showed a clear overlap of the two schemes, and confirmed the consistency and correlation of these dimensions, as shown in Abessolo et al. (2017). This suggests that career choices are based on universal needs and beliefs, and that personal basic values should be taken into account to orientate aware professional choices, to promote a fruitful working climate, and to offer to each worker a personalised and

suitable career path, which makes the most of the individual characteristics of everyone.

We also observed some differences in the dominant anchors and in the priority values of males and females, older and younger people. Young males tend to pursue individualistic goals and materialistic recognitions, while older females are oriented toward finding their place in the society and being appreciated for their values more than for their skills.

These differences suggest that in the future it will be interesting to investigate subgroups of workers, and to assess if the underlying structure and the relationships between values and anchors are stable across age, gender and/or other characteristics. Moreover, age differences suggest that priority values are not completely stable over time, and then a longitudinal design would help finding evidence of what changes in individual values over a lifespan, and which life events or professional steps affect the change.

## References

- Abessolo, M., Rossier, J., Hirschi, A. (2017). Basic values, career orientations, and career anchors: Empirical investigation of relationships. *Frontiers in Psychology*, **8**, pp. 1-11.
- Avallone, F., Farnese, M.L., Pepe, S., Grimaldi, A. (2009). Questionario dei valori lavorativi (QVL): Uno strumento di misura mutuato dalla tassonomia di Shalom Schwartz. *Giornale Italiano di Psicologia dell'Orientamento*, **10**(2), pp. 37-49.
- Barclay, W.B., Chapman, J.R., Brown, B.L. (2013). Underlying factor structure of Schein's career anchor model. *Journal of Career Assessment*, **21**(3), pp. 430-451.
- Feldman, D.C., Bolino, M.C. (1996) Careers within careers: reconceptualising the nature of career anchors and their consequences. *Human Resource Management Review*, **6**(2), pp. 89-112.
- Hu, L., Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1-55.
- Jöreskog, K.G., Sörbom, D. (2004). *Lisrel 8.7 for Windows*. Lincolnwood, IL: Scientific Software International, Inc.
- Pike, C.K. (1996). Development and initial validation of the Social Work Values Inventory. *Research on Social Work Practice*, **6**(3), pp. 337-352.
- Porto, J.B., Tamayo, A. (2003). Escala de valores relativos ao trabalho – EVT. *Psicologia: Teoria e Pesquisa*, **19**(2), pp.145-152.
- Schein, E.H. (1990). *Career Anchors: Discovering Your Real Values*. Jossey-Bass/Pfeiffer, San Francisco (CA).
- Schwartz, S.H., Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology*, **53**, pp. 550–562.
- Schwartz, S.H., Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, **26**, pp. 92-116.

# Factors affecting tertiary education decisions of immigrants in Italy

Michele Lalla, Patrizio Frederic

## 1. Introduction

The decision to enrol in tertiary education is difficult for young people and families if the choice is made without much knowledge about the needs of society. Such decisions may be affected by individual characteristics, the socio-economic conditions of families, and the contextual background of the area. All these aspects may differ among young immigrants and non-immigrants and, in the case of the former, tertiary schooling plays an important role not only in terms of investing in human capital, the cultural formation process, and social integration, but also as an instrument of social mobility and transformation, development through attuned interactions and collective healing through cooperation (Paba and Bertozzi, 2017; De Clercq et al., 2017).

The objective of this paper is to point out the differences with respect to citizenship, a binary variable distinguishing between immigrants and non-immigrants (hereinafter also referred to as Italians), and the *tertiary* binary variable, defined as equal to one for individuals who were enrolled in a tertiary education level and equal to zero otherwise. A Bayesian model selection was performed through the Lasso method to investigate the determinants of the tertiary binary variable.

## 2. Data sources and descriptive statistics

The data were extracted from two surveys, with the reference year being 2009, carried out by the Italian National Institute of Statistics (Istat): one being the European Union Statistics (or Surveys) on Income and Living Conditions (EU-SILC) restricted to Italy, IT-SILC (Istat, 2008; Eurostat, 2009), and the other being the Italian Survey on Income and Living Conditions of families with Immigrants (IM-SILC), which is a single cross-sectional survey (Istat, 2009) that involved families with at least one immigrant component residing in Italy. The IT-SILC sample was added to the IM-SILC sample to obtain a sample with a consistent number of immigrants with respect to non-immigrants. For further details about these two data sets and about the main variables introduced in the model, see Lalla and Frederic (2020). The target sample was obtained by first selecting individuals in the age range of 20 to 25, obtaining a sample of 3,166 cases. Then, among the latter data set, the eligible cases were only those individuals whose highest attained ISCED (International Standard Classification of Education) level was equal to 3 (=upper secondary education) or 4 (=post-secondary non-tertiary education). The final target sample was made up of 2,874 individuals.

The relationship between the tertiary (binary) dependent variable and the ISCED Level Currently Attended (ILCA) showed that 55.3% of individuals, with an ISCED level equal to 3 or 4, were not enrolled in further education (termed “not-attending”), while 44.7% were currently attending a tertiary school (Table 1).

The ILCA was examined with respect to several qualitative variables and revealed many significant relationships. For the sake of brevity, only some of them are cited. The ILCA showed a significant relationship with respect to citizenship,  $CS(2) = 115.33$  ( $p < 0.000$ ), where  $CS(g)$  stands

for “Chi-Square with  $g$  degrees of freedom”, but hereinafter “(g)” is omitted because the corresponding tables do not appear here: the percentage of immigrants attending tertiary education was lower than that of Italian citizens (26.6% versus 50.0%), while the percentage of immigrants not in school was higher than that of Italians (72.4% versus 48.4%). There was a significant relationship between the ILCA and self-perceived health,  $CS= 10.87$  ( $p<0.004$ ), implying that individuals perceiving fair or bad or very bad health tended to discontinue their education with respect to those perceiving good or very good health (Ichou and Wallace, 2019). The ILCA was not related to the index of the total self-perceived health of parents, perhaps its effect operated during the upper secondary education level (Frederic and Lalla, 2021). The ILCA proved to be linked to the Italian macro-regions  $CS= 24.27$  ( $p<0.002$ ), as industrialisation and the possibility of finding employment increased, the percentage of individuals not in school increased. The ILCA was related to the maximum ISCED level attained by parents,  $CS= 198.80$  ( $p<0.000$ ). As the education of parents increased, the percentage of young individuals in school increased. The ILCA was significantly related to several variables describing the working conditions of parents, but the strength of such relationships was generally weak.

**Table 1.** Absolute and percentage frequencies of tertiary education (EDU) by the ISCED level currently attended (ILCA)

<b>Tertiary\ ILCA</b>	<b>Not-attending</b>	<b>Post-Secondary EDU</b>	<b>Tertiary EDU</b>	<b>Total</b>
Tertiary = 1			1285	1285
			100.0	100.0
Tertiary = 0	1546	43	0	1589
	97.3	2.7	0	100.0
Total	1546	43	1285	2874
	53.8	1.5	44.7	100.0

The ILCA was also analysed with respect to the main quantitative variables.

The age of fathers analysed according to the ILCA and citizenship showed that the fathers of immigrants were younger than the fathers of Italians by about twelve years. Similarly, the mothers of immigrants were younger than the mothers of Italians by about twelve years. The Disposable Family Income (DFI) per capita (in thousands of euros) is reported in Table 2 by the ILCA and citizenship. On the average, the DFI per capita for immigrants was significantly lower than that of Italians by about four thousand euros: about 35.7%.

**Table 2.** Sample size frequencies (n), means, and standard deviations (SD) of the disposable family income per capita (in thousands of euros) by citizenship and by the ISCED level currently attended (ILCA) by their children (E=Education)

<b>Citizenship\ ILCA</b>	<b>Not-attending</b>	<b>Post-Secondary E</b>	<b>Tertiary E</b>	<b>Total</b>
Italian citizen. <i>n</i>	1080	36	1114	2230
<i>Means</i>	11.389	11.508	12.543	11.967
<i>SD</i>	6.999	6.432	9.147	8.153
Foreign citizen. <i>n</i>	466	7	171	644
<i>Means</i>	7.777	5.868	7.563	7.699
<i>SD</i>	5.315	3.489	5.877	5.452
<b>Total.</b> <i>n</i>	1546	43	1285	2874
<i>Means</i>	10.300	10.590	11.880	11.011
<i>SD</i>	6.742	6.376	8.942	7.835

The other types of income considered in the models revealed various structures of relationships and levels of significance. For example, the gap between immigrant and Italian fathers' incomes amounted to about eleven thousand euros, i.e., -42.0%. The mothers' incomes also presented significant statistical differences for both marginal effects, with a gap amounting to about five thousand six hundred euros, i.e., -32.5%. However, the disposable personal income gender gaps were -35.9% for Italians and -25.3% for immigrants.

The size of immigrant families proved to be slightly lower than those of Italians, but not statistically significant. The result differed in the population involved in the transition from lower to upper secondary education (Frederic and Lalla, 2021) implying that the size of families who intended to send their children to university was similar to that of the Italians.

Citizenship was examined with respect to some other variables. Its relationship with the maximum ISCED level attained by parents was statistically significant, CS= 217.01 (p<0.000) (Bertolini et al., 2015). Citizenship was significantly related to the degree of urbanisation, CS= 19.18 (p<0.000): immigrants tended to settle in densely populated areas more than Italians (36.2% versus 35.3%) or in moderately populated areas (46.6% versus 39.6%). Citizenship also showed a significant relationship with the Italian macro-regions and yielded a significant relationship with the index summarising the total self-perceived health of parents, CS= 134.99 (p<0.000) (Ichou and Wallace, 2019). Citizenship proved to be associated with many variables describing working conditions; only the relationship with the maximum position of parents on the job, CS= 134.03 (p<0.000), is mentioned here.

### 3. Bayesian Lasso selection of regressors

Let  $Y$  be the binary variable coding if the  $i$ -th individual is or is not attending tertiary education ( $i=1, \dots, n$ ). Let  $\mathbf{x}_i$  be a vector of  $K$  regressors. Let  $\pi_i$  be the probability that  $Y=1$  given  $\mathbf{x}_i$ . Let  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_K)$  be the parameters vector of the model. The logit model is

$$\pi_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) / [1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})] \quad (1)$$

The *Lasso* method (Tibshirani, 1996) was applied to carry out the estimation and model selection. In fact, it is a procedure involving an additional penalization term,  $L_1$ , summed up to the negative log-likelihood of the model that depends on an additional parameter  $\lambda$ ,  $\lambda \geq 0$ . Many penalized methods can be interpreted as the negative logarithm of a posterior distribution in a purely Bayesian way. Let  $p(y_i | \mathbf{x}_i, \boldsymbol{\beta}) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$  be the model in the Bayesian notation and let  $p(\boldsymbol{\beta} | \lambda) \propto \exp(-\lambda \sum_{j=0}^K |\beta_j|)$  be the Laplace prior distribution on coefficients  $\boldsymbol{\beta}$ , where  $K$  is the number of regressor coefficients and  $\beta_0$  is the intercept. Then the posterior distribution is

$$\begin{aligned} p(\boldsymbol{\beta} | \mathbf{x}, \mathbf{y}, \lambda) &\propto p(\mathbf{y} | \mathbf{x}, \boldsymbol{\beta}) p(\boldsymbol{\beta} | \lambda) \\ &= \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \exp(-\lambda \sum_{j=0}^K |\beta_j|) \end{aligned} \quad (2)$$

To select  $\lambda$ , the One Standard Error Rule (1SE) procedure was applied. The estimation method consisted of two steps:

1. The model was first estimated using the *glmnet* (Friedman et al., 2010) package in R (R Core Team, 2019). Then the *optimal* lambda ( $\lambda_{1SE}$ ) and the mode estimations ( $\hat{\boldsymbol{\beta}}_{\lambda_{1SE}}$ ) were evaluated.



- Using the R package *MCMCpack*,  $N=10,000$  samples were drawn from the posterior distribution  $p(\boldsymbol{\beta}|\mathbf{x}, \mathbf{y}, \lambda_{1SE})$  to perform a full Bayesian analysis, where  $p(\boldsymbol{\beta}|\lambda_{1SE})$  was chosen to be Laplace distributed.

Note that the model matrix of the starting model consisted in 2874 rows by 880 columns, and classical methods can be affected by the *curse of dimensionality*. Instead, the Lasso method is very stable and quick, and shrinks 858 values (out of 880) of  $\hat{\boldsymbol{\beta}}_{\lambda_{1SE}}$  to zero; thus only 22 betas have a posterior distribution which is not symmetric to zero.

#### 4. Outcomes of the logistic model

The odds ratios (OR) are reported in [Table 3](#), which only presents interaction terms of the first order because the analysis of interactions orders was limited to the first order to simplify interpretation. The interactions are indicated by the symbol  $\times$ , which may be read as “by”.

Let  $\mathbf{x}_b$  be the binary variables. Let  $\mathbf{x}_c = \boldsymbol{\mu}$  be the mean values of the continuous regressors, limited to the ages of individuals, which can never be zero in practice. Note that: (1) the product of two binary variables is again a binary variable, (2) the percentage of variation of the reference probability,  $\pi_{i|\mathbf{x}_b=0 \wedge \mathbf{x}_c=\boldsymbol{\mu}}$ , is given by  $[100*(OR-1)]$  and is reported below in parentheses, (3) the corresponding value of OR may be found in [Table 3](#). The probability of having  $y=1$  (i.e., of continuing one’s education) was equal to  $\pi_{i|\mathbf{x}_b=0 \wedge \mathbf{x}_c=\boldsymbol{\mu}} = 0.120$ , calculated at the mean values of the continuous regressors ( $\mathbf{x}_c = \boldsymbol{\mu}$ ) and the binary variables equal to 0 ( $\mathbf{x}_b$ ). A binary variable having an OR greater than 1 implied that the group represented by the binary variable equal to 1 had a higher probability of having  $y=1$  than the group identified by the binary variable equal to 0; for example, for women with an  $OR=1.777$ , the probability of continuing their education was +77.7% greater than that of men. In other terms,  $\pi_{w|f} = 1.777 \times 0.120 = 0.213$ , which was +77.7% greater than the probability of men. Note that the dot in the index means keeping all other variables fixed, i.e., the binary and the continuous variables other than age equal to zero. The successive binary variable having an  $OR>1$  in [Table 3](#) was “PES (Parents’ Employment Status) is inactive” ( $x_1$ )  $\times$  “Family living in a densely populated area” ( $x_2$ ), denoted by  $x_{12}$ , which showed an  $OR=1.697$  meaning that the odds of the event  $y=1$ , when  $x_{12}=1$  (both  $x_1$  and  $x_2$  are equal to 1), were +69.7% greater than the odds of the event  $y=1$ , when  $x_{12}=0$ . Therefore,  $\pi_{x_{12}=1|f} = 1.697 \times 0.120 = 0.204$ . Similarly, significant high probabilities of continuing one’s education were observed for other interaction terms: “Father with permanent contract”  $\times$  “Only mother employed” (+95.7%), “Father with permanent contract”  $\times$  “Parents are managers or executives” (+132.1%), “Mother with permanent contract”  $\times$  “Father is limited by health” (+64.7%), “TSH (Tenure Status of Household): Subtenant”  $\times$  “Family living in a moderately populated area” (+46.6%), “TSH: Free”  $\times$  “Assets reduction for needs” (+173.3%), “Father with term contract”  $\times$  “Mother is limited by health” (+266.5%). This latter appears to be an unbelievable outcome. However, this group ( $x_{12}=1$ ) only consisted of 30 subjects and whose family income was higher than that of the group consisting of 162 subjects and having “Father with term contract” ( $x_1 = 1$ ) and “Mother not limited by health” ( $x_2 = 0$ ). In synthesis, gender, good and stable parents’ working conditions, and good actual and self-perceived health deeply affected the probability of continuing one’s education in the transition from upper secondary school to tertiary education, although this happened through interactions with other factors, consistent with the literature in any case.

The binary variables having an OR lower than 1 implied that the represented group had a

lower probability of having  $y=1$  with respect to the complementary group. In Table 3 there are six (interaction) binary variables with an OR lower than 1. For example, “Father perceives poor health”  $\times$  “Rent is burdensome” had an OR=0.440 and hence its complement to one, expressed as a percentage, was equal to  $[100*(0.440-1)] = -56.0\%$ . Therefore, the probability of continuing one’s education amounted to  $-56.0\%$  of the probability of the complementary group, which did not have fathers perceiving poor health and a burdensome rents,  $\pi_{i|x_b=0 \wedge x_c=\mu}$ . In other words, the group with  $x_{12}=1$  had a probability equal to  $\pi_{x_{12}=1} = 0.440 \times 0.120 = 0.053$ , implying that the probability of the group with  $x_{12}=1$  decreased the probability of continuing their education by an amount of  $-56.4\%$  with respect to the complementary group, which had a probability given by  $\pi_{i|x_b=0 \wedge x_c=\mu} = 0.120$ . In synthesis, unstable and unfavourable parents’ working conditions, poor actual and self-perceived health conditions, and critical and costly tenure status of the household negatively affected the probability of continuing one’s education in the transition from upper secondary school to tertiary education, although this happened through the interaction terms.

**Table 3.** Logistic regression with Lasso method and Bayesian approach: Estimated odds ratio (OR), standard errors (SE), p-values ( $p$ ), and means

<b>B=Binary/ C=Continuous Variables</b>	<b>OR</b>	<b>SE</b>	<b><math>p</math></b>	<b>mean</b>
B - Women	1.777	0.263	0.000	0.530
C - [(Individual’s age)/10]^2	0.714	0.044	0.000	5.064
C - (Father’s age)/10	1.175	0.073	0.003	4.973
C - (Mother’s age)/10	1.548	0.094	0.000	4.727
C - (Education Level of Father: years)^2	1.003	0.001	0.000	1.552
C - FDPI= (Father’s DPI)/ 10000	1.452	0.070	0.000	2.372
C - MDPI= (Mother’s DPI)/ 10000	1.285	0.062	0.000	1.248
C - FTIPC= (Family’s total income per capita)/ 10000	0.314	0.046	0.000	1.101
<i>Interactions of first order</i>				
B - (Father: poor health) $\times$ (Burdensome rent)	0.440	0.156	0.011	0.023
B - (PES= Parents’ Employment Status: pensioners) $\times$ NW*	0.494	0.188	0.031	0.017
B - (PES: inactive) $\times$ (Densely populated area)	1.697	0.428	0.048	0.043
B - (PES: part-time) $\times$ (North-West= NW*)	0.378	0.237	0.042	0.008
B - (PES: full-time employee) $\times$ immigrant	0.531	0.092	0.000	0.121
B - (Father: permanent contract) $\times$ (Only mother employed)	1.957	0.544	0.013	0.038
B - (Father: permanent contract) $\times$ (Parents: manager/ executive)	2.321	0.737	0.013	0.048
B - (Mother: permanent contract) $\times$ (Father: limited by health)	1.647	0.322	0.010	0.065
B - (Father: term contract) $\times$ (Mother: limited by health)	3.665	1.776	0.011	0.010
B - (TSH <sup>+</sup> : Subtenant) $\times$ (Moderately populated area)	1.466	0.175	0.001	0.246
B - (TSH <sup>+</sup> : Free) $\times$ (Father: poor health)	0.459	0.186	0.025	0.016
B - (TSH <sup>+</sup> : Free) $\times$ (Assets reduction for needs)	2.733	1.041	0.010	0.016
B - (TSH <sup>+</sup> [Tenure Status of Household]: Free) $\times$ Savings	0.179	0.220	0.023	0.003
Intercept	0.043	0.029	0.000	
Pseudo-R square	0.180	$n =$	2874	

**The continuous variables.** The individual’s age (range 20-25), expressed in decades, showed a parabolic and negative impact on education paths, while the ages of both parents revealed a linear positive impact on the probability of continuing one’s education. The other continuous single variables (which may be conceptually and concretely equal to 0) entering the model showed significant effects on continuing one’s education. As parents’ education levels increased, the probability of continuing one’s education increased quadratically. The father’s (FDPI) and mother’s (MDPI) disposable personal income indicated a linear positive effect, while the family’s

total income per capita (FTIPC) yielded an unexpected negative effect, but perhaps the latter balanced the effect of the former. In fact, FTIPC included both FDPI and MDPI too. However, the algebraic sum of their impacts remained positive implying the importance of welfare programmes to help families experiencing economic (and physical) difficulties, with the specific aim of reducing the number of students interrupting their education.

The main fault of the Lasso method in selecting significant explanatory variables concerns the possibility of selecting a theoretically unjustifiable variable, such as “Father with term contract” × “Mother is limited by health” (+266.5%) or of neglecting some important variables in the model.

The conclusions are similar to those explained in [Frederic and Lalla \(2021\)](#): in the applications, the interactions should be supported by social, behavioural, psychological or economic theories. Otherwise, they may be obtained automatically simply by using an adaptive procedure like the Lasso method and only as empirical findings. In fact, few models with interactions exist in the literature. The interactions may probably be easily found among binary or categorical variables, but this case is relatively interesting because they can be replaced with specific typologies. The same holds true for the interactions of a continuous variable with other explanatory binary variables, but the interaction between two continuous variables is very difficult to grasp immediately. In general, it is useful to find a theoretical justification for the existence of the interactions, instead of blindly searching for interaction terms. However, it is highly plausible that almost all phenomena are outcomes of interactions among many variables, but knowledge about and explanations of these results may become very complicated and challenging.

## References

- Bertolini, P., Lalla, M., Pagliacci, F. (2015). School enrollment of first- and second-generation immigrant students in Italy: A geographical analysis, *Papers in Regional Science*, **94**(1), pp. 141-160.
- De Clercq, M., Galand, B., Frenay, M. (2017). Transition from high school to university: a person-centered approach to academic achievement, *European Journal of Psychology of Education*, **32**(1), pp. 39-59.
- Eurostat (2009). *Description of Target Variables: Cross-section and Longitudinal*, EU-SILC 065 (2009 operation), Directorate F, Unit F-3, Eurostat, Luxembourg.
- Frederic, P., Lalla, M. (2021). Determinants of the transition to upper secondary school: differences between immigrants and Italians, in *ASA 2021 Statistics and Information Systems for Policy Evaluation – BOOK OF SHORT PAPERS of the on-site conference*, eds. Bruno Bertaccini, Luigi Fabbri, Alessandra Petrucci, Firenze University Press, Firenze, pp. 13-18.
- Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, **33**(1), pp. 1-22.
- Ichou, M., Wallace, M. (2019). The healthy immigrant effect: The role of educational selectivity in the good health of immigrants, *Demographic Research*, **40**(4), pp. 61-94.
- Istat (2008). Ceccarelli C., Di Marco M., Rinaldelli C. (Eds.). *L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*. Metodi e Norme n. 37, Istat, Rome.
- Istat (2009). Reddito e condizioni di vita delle famiglie con stranieri [electronic resource]. Istat, Rome. <https://www.istat.it/it/archivio/52405>. Accessed 3 January 2020.
- Lalla, M., Frederic, P. (2020). Tertiary education decisions of immigrants and non-immigrants in Italy: an empirical approach, *DEMB Working Papers Series*, N. 168, pp. 1-39. University of Modena and Reggio Emilia - Department of Economics.
- Paba, S., Bertozzi, R. (2017). What happens to students with a migrant background in the transition to higher education? Evidence from Italy, *Rassegna Italiana di Sociologia*, **58**(2), 315-351.
- R Core Team (2019). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna (AT). URL <http://www.R-project.org/>. Accessed 18 April 2020.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, **58**(1), pp. 267-288.

# Internet use, feeling of unacceptance and Loneliness: immigrants of first and second generation in Italy

Giovanni Busetta, Maria Gabriella Campolo, Antonia Cava

## 1. Introduction

Controversial interpretations have usually attributed by the main literature on the topic to the relationship between Internet use and loneliness. On the one side, Internet-use disorders are generally caused by depression, anxiety, and loneliness (Longstreet et al., 2019). Indeed, this is most often because Internet addiction is used as a dysfunctional strategy to face everyday life's stressful events (Brand et al., 2019; Servidio et al., 2021). On the other side, gambling, online gaming, and social media use may produce states of anxiety and depression, and loneliness (Brand et al., 2019). This second strand of literature is based on the idea that excessive specific online behaviors could be approached by people to regulate their mood, translating into Internet addiction (Blasi et al., 2019; Islam et al., 2020; King et al., 2020).

This relationship is particularly relevant for children, adolescents, and young adults because Internet represents a particularly accessible way of entertaining to escape from reality (Kwon, 2011). Conversely, high levels of Internet use are usually associated with negative psychological health conditions, including loneliness (Dong et al., 2020; Li et al., 2019; Ismail et al., 2020; Seki et al., 2019). This relationship is even more pronounced among female adolescents (Liang et al., 2016).

Following King and Delfabbro (2020), on the one hand, Internet use, and especially gaming, produces a sense of energy and an increase in self-confidence. Through these channels the use of Internet induce a reduction in levels of loneliness, feeling of acceptance and happiness. Literature have devoted increasing attention to analysing interactivity or synergy between factors contributing to increasing or reducing such emotions (Tofallis, 2020).

Indeed, gaming (Király et al., 2020) is not necessarily problematic: it appears as an adaptive behavior (Billieux et al., 2019) which could enhance people's lives (Granic et al., 2014) and reduce loneliness (Carras et al., 2017).

The Italian research concerned lifestyle and consumption of immigrants (Al-Kandari et al. 2020; Bauer et al., 2020; Biolcati et al., 2017; Gao et al., 2020; Masaeli et al., 2021; Mattioli et al., 2020), shows a significant trend. The standardisation of media fruition, especially regarding digital technologies, and between new technologies and the use of the Internet has had a remarkable development in the last years.

Considering the different aspects that the massive use of the Internet could have in the life of people, several studies focus on the relationship between the Internet and loneliness. If on the one hand, some studies have found that Internet' use has a negative impact on social relationships and for this reason is associated with increased loneliness (Kraut et al., 1998; Lavin et al., 1999)

On the other side, other studies found that the Internet uses can impact on society and on the life of persons positively, for example, removing the geographical barriers between people, or providing an ideal social environment for lonely people to interact with other persons. For this reason, lonely individuals are more likely to use the Internet excessively (Morahan-Martin and Schumacher, 2003).

Using the Survey on Social Condition and Integration of Foreign Citizens conducted by Istat

Giovanni Busetta, University of Messina, Italy, gbusetta@unime.it, 0000-0001-5843-3851

Maria Gabriella Campolo, University of Messina, Italy, mgcampolo@unime.it, 0000-0002-1075-4573

Antonia Cava, University of Messina, Italy, acava@unime.it

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Giovanni Busetta, Maria Gabriella Campolo, Antonia Cava, *Internet use, feeling of unacceptance and Loneliness: immigrants of first and second generation in Italy*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.33, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 187-192, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

in 2011-2012, we investigate the difference in using Internet between first- and second-generation immigrants in Italy. Our study wants to verify the socio-economic determinants (such as, age, gender, education level) that can affect the use of Internet. Among the explanatory variables, we included the perception of the subjects about their integration in the social framework and their feeling, such as loneliness, or the perception of unacceptance.

The rest of the paper is organized as follows. The data are presented in Section 2. In Section 3 we provide a presentation of methods and descriptive results. Section 4 contains the empirical results, and Section 5 concludes.

## 2. Data

The sample is drawn from the "Condition and Social Integration of Foreign Citizens, SCIF 2011-2012" survey provided by the Italian National Institute of Statistics (ISTAT). It represents the first national survey on immigrants. Its aims to provide information on money features of socio-economic integration of immigrants in Italy for a better understanding of the resident foreign population. It was carried out on a sample of 9,553 households residing in Italy, with at least one foreign citizen living with. In total 25,326 individuals have been surveyed: 20,379 are foreign citizens, 4,251 are native born and 696 Italian citizens for acquisition.

Behaviors, attitudes, and opinions of foreign citizens in Italy were investigated, as well as the family composition, education, migratory path, employment status, discrimination, health conditions and accessibility of health services, immigrant integration, citizen's security and victimization. Foreign citizens are identified using the principle of citizenship, instead of the place of birth. People with Italian citizenship achieved by acquisition (foreign at birth), hereafter referred to as naturalized people are also subject to the survey, as long as they cohabit in the family with a foreign person at least. Italians natives are included as part of the sampled families, but they are interviewed only with regard to their socio- demographic characteristics (gender, age, citizenship, state of birth, educational qualifications, etc).

Rumbaut (2004), distinguishes immigrants depending on the age of migration and the concerning level of socialization characterizing those ages:

- Generation 1: a person who has immigrated in a new country;
- Generation 1.25: a person who comes after age 12 but before age 18;
- Generation 1.50: a person who comes between 6 and 12;
- Generation 1.75: a person who comes before formal schooling at age 6;
- Generation 2: children of an immigrant.

In our analysis we restricted the sample to 11934 observations, mainly first- and second-generation immigrant living in Italy, without considering Italians. Following the categorization shown above (Rumbaut, 2004), we consider as First generation only persons identified as Generation 1 (78% of the sample), and as Second generation, the subjects included in the other four categories (Generation 1.25 to Generation 2). Regarding second generation (the remaining 22% of the sample), 2609 are the persons included: 10% is Generation 1.25, 7% is Generation 1.50, Generation is 1.50, 3% is Generation 1.75 and 2% is Generation 2. In general, 54% of the sample are women, and 45% live in the South or Islands. The 68% of the sample uses Internet every day. The percentage increase to 86% for second generation immigrants. In the next Section we report all sample characteristics.

## 3. Methods and descriptive results

The aim of our analysis is to investigate the difference in the use of the Internet between immigrants of first- and second-generation in Italy. In particular, through a Probit estimation model, we want to estimate the impact of socio-economic characteristics on the regularity of using

the Internet. The dependent variable “Internauta” is a dummy variable that assumes value 1, if subject use Internet every day and 0 otherwise. The independent variables include a dummy variable concerning the gender of the individuals (*Woman*: 1=yes; 0=no), the number of the household components, the level of education expressed in years of school (*Edu*), whether the subject has achieved the highest level of education in Italy (*Study\_Italy*:1=yes, 0=no), the geographical area (*South*=1 south and islands, 0=north-center), a dummy that identify the subjects as either worker or unemployed (*Work*: 1=yes, 0=no), a variable that identifies whether the subject is a first- or second-generation one (*Generation2*:1=yes, 0=no), and the age of the subjects (*Age*: 1=15-19; 2=20-29, 3=30-39, 4=40-44). Furthermore, we include two variables: *Loneliness*, which assumes value 1 if the subject feels alone, either “much” or “enough”, in Italy, 0 otherwise; and a dummy variable concerning how much the subject feels accepted in the city, in which she/he lives (*Unaccepted*: 0=Much or enough, 1=otherwise). To focus on the impact of the potential loneliness among different generations, we also include in our model interaction effects between these two last covariates and the generation of the immigrant.

In the following Table (Tab. 1) we report descriptive statistics of the variables used in our analysis.

**Table 1:** Descriptive statistics by generation

Variable	Second-generation		First-generation		Min	Max
	Mean	Std. Dev.	Mean	Std. Dev.		
Internauta	0.86	0.35	0.64	0.48	0	1
South and islands	0.41	0.49	0.46	0.50	0	1
Number of household components	4.08	1.55	3.14	1.54	1	12
Women	0.46	0.50	0.57	0.49	0	1
Study in Italy	0.69	0.46	0.10	0.30	0	1
Worker	0.37	0.48	0.69	0.46	0	1
Education (years of school)	8.98	3.35	10.69	4.36	0	17
Loneliness	0.07	0.25	0.16	0.37	0	1
Unaccepted	0.04	0.20	0.08	0.26	0	1
Age	1.70	0.68	3.01	0.69	1	4

From the descriptive statistics shown in Table 1, it emerges that, on average, 86% of Second-generation immigrants are “Internauta”, while this percentage is 64% for First-generation ones. Moreover, second-generation immigrants are characterized by a lower proportion of individuals living in south or islands and being women, employed, and a higher proportion of individuals, being older, studying in Italy, studying for more years and living in households made by a higher number of components. Finally, first-generation immigrants feel more unaccepted and lonelier compared to second-generation ones.

## 4. Empirical Results

The results of our Probit estimation model are reported in Table 2.

We can observe that the probability to use every day Internet decreases for women, for individuals living in the south of Italy and in islands, and for the first-generation immigrants. All the coefficients related to these variables are statistically significant. Moreover, the probability decreases with the ages of the individuals. An education title in Italy and the achieved education level both play an important rule. In both cases the coefficients are positive and significant.

To better understand the estimation results we have calculated also the average marginal effects, reported in Figure 1, and the predictive probabilities (Table 3). For example, as shown in Tab. 3, we can observe that the probability of being “Internauta” decreases by 10 percentage points, for subjects living in the north-center (0.73), compared to subjects living in the south or in

**Table 2: Results of the probit model**

Dep. Variable: using internet everyday	Coef.	Std. Err.	P. value
South and islands	-0.32	0.03	***
Number of household components	-0.04	0.01	***
Woman	-0.12	0.03	***
Study_Italy	0.22	0.04	***
Worker	0.05	0.03	
Education	0.08	0.00	***
Age ref. 15-19			
20-29	-0.44	0.07	***
30-39	-0.69	0.07	***
40-44	-0.99	0.08	***
Second Generation	0.42	0.05	***
Loneliness	-0.13	0.04	**
Unaccepted	-0.19	0.05	***
Second generation#Loneliness	-0.14	0.12	
Second generation #Unaccepted	-0.44	0.14	**
Constant	0.54	0.09	***

Note: p. value: \*\*\* <0.001; \*\* < 0.01; \* < 0.05

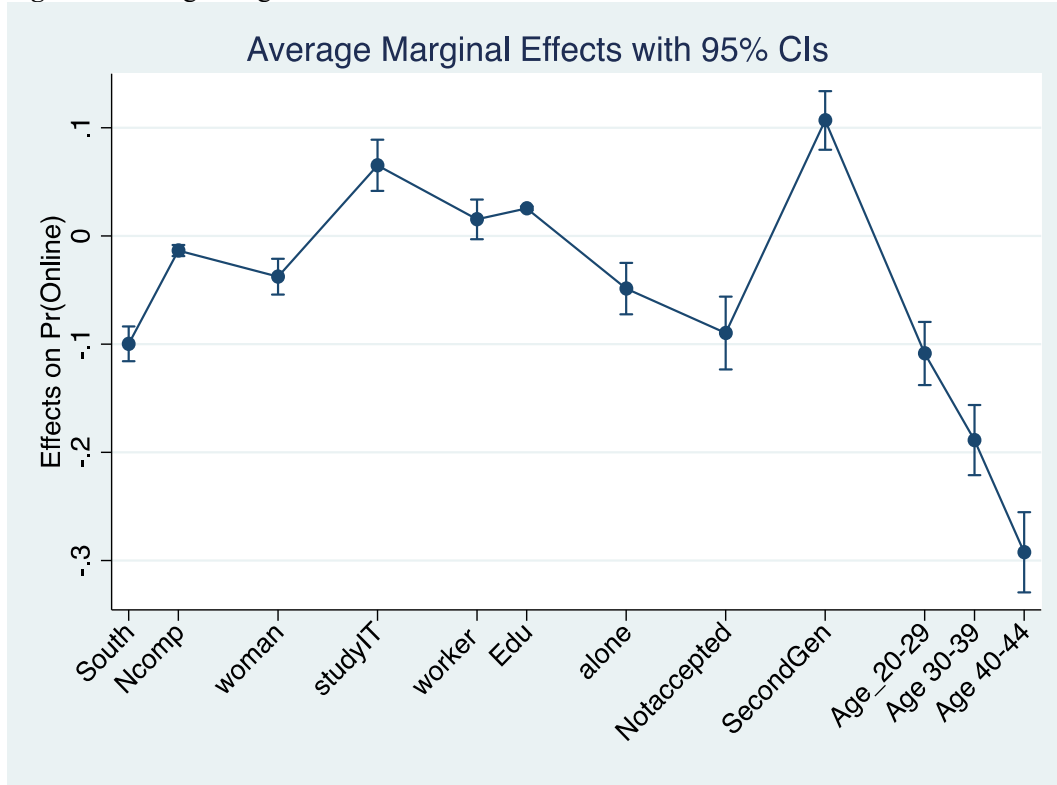
**Table 3: Estimation results of Probit model**

		Delta-method		
		Pred. Prob.	Std. Err.	P.value
South	North - center	0.73	0.01	***
	South and islands	0.63	0.01	***
Woman	Man	0.71	0.01	***
	Woman	0.67	0.01	***
study_Italy	Study abroad	0.67	0.00	***
	Study in Italy	0.74	0.01	***
Worker	Unemployed	0.68	0.01	***
	Worker	0.69	0.01	***
Age	Age 15-19	0.86	0.01	***
	Age 20-29	0.75	0.01	***
	Age 30-39	0.67	0.01	***
	Age 40-44	0.57	0.01	***
Generation	First generation	0.66	0.01	***
	Second generation	0.77	0.01	***
Alone	Feeling not alone	0.69	0.00	***
	Feeling alone	0.64	0.01	***
Unaccepted	Feeling accepted	0.69	0.00	***
	Feeling unaccepted	0.60	0.02	***
Generation#Alone	First Generation and not alone	0.67	0.01	***
	First Generation and Alone	0.63	0.01	***
	Second Generation and not alone	0.78	0.01	***
	Second Generation and Alone	0.71	0.03	***
Generation#Unaccepted	First Generation and Accepted	0.67	0.01	***
	First Generation and Unaccepted	0.60	0.02	***
	Second Generation and Accepted	0.79	0.01	***
	Second Generation and Unaccepted	0.59	0.05	***

Note: p. value: \*\*\* <0.001; \*\* < 0.01; \* < 0.05

the islands (0.63). This probability decreases also by 4 percentage points for women, moving from 0.71 (man) to 0.67 (woman). Having achieved education in Italy increases the probability of being an “Internauta” by 7 percentage points (from 0.67 to 0.74). This probability increases by 11% points (from 0.66 to 0.77) when the individual is a second-generation immigrant. The feelings of

**Figure 1:** Average marginal effects of the estimated Probit Model



loneliness or not acceptance of the subject negatively affect the probability of being an Internet user by 5 and by 9 percentage points, respectively.

Through the two-fold interaction effects, we can also calculate the different impact of loneliness and the feeling of not acceptance between and within generations. Within first-generation immigrants, the difference in probability of being an Internauta conditioned on the loneliness of the individuals is equal to 4% (0.67 for first-generation that does not feel alone and 0.63 for first-generation that feels alone). Within second-generation, the related probability is equal to 7% (0.78 for second generation not feeling loneliness 0.71 for second generation feeling it). Moreover, while between first- and second-generations the subgroup of “not alone” shows a difference of 12% (0.78-0.67) in predictive probability, this gap in the “alone” subgroup is equal to 8% (0.71-0.63).

Finally, we consider the interaction effect of feeling accepted between generation. Within first-generation, the difference due to the feeling of acceptance or unacceptance is equal to 7% (0.67 for first generation immigrants feeling accepted and 0.60 for the same generation immigrants feeling unaccepted), while the difference for second-generation is equal to 20% (from 0.79 to 0.59). Moreover, in the subgroup of the immigrants feeling “accepted”, the difference imputed to being first- or second-generation is equal to 13% (0.79-0.67), while this difference in the “unaccepted” subgroup almost collapse (0.59-0.60).

## 5. Conclusions

In this study we analyse the different behaviour in terms of frequency in the use of the Internet between immigrants of first- and second-generation. In our analysis, we controlled for socio-economic characteristics, taking into account the feeling of loneliness and of unacceptance of the



subject. Our results show that the probability of using Internet everyday increases being male and living in the north or centre of Italy. Moreover, our results show that both the feeling of loneliness and unacceptance are negatively correlated with the probability of using Internet everyday both for First- and Second-Generation immigrants. In particular, Second-generation immigrants are more likely to use the Internet everyday than the First-generation ones. The difference in predicted probability of being an Internauta is equal to 11% (0.77 and 0.66, respectively). Nevertheless, while this probability decreases to 0.59, if the second-generation immigrant feels unaccepted in the city where he/she lives, and to 0.71 if he/she feels alone.

We can conclude that new possibilities offered by “web sociability” or, in general, by the use of the Internet, is negatively correlated to the immigrants’ dissatisfaction that we identify with the perception of integration and sociability in the offline life (Loneliness and Unacceptance).

## References

- Al-Kandari, Y.Y., Al-Sejari, M.M. (2020) Social isolation, social support and their relationship with smartphone addiction. *Information, Communication & Society*, doi: 10.1080/1369118X.2020.1749698
- Bauer, L.L., Seiffer, B., Deinhart, C., Atrott, B., Sudeck, G., Hautzinger, M., Rösel, I., Wolf, S. (2020). Associations of exercise and social support with mental health during quarantine and social-distancing measures during the COVID-19 pandemic: a cross-sectional survey in Germany. *MedRxiv* doi: 10.1101/2020.07.01.20144105
- Biolcati, R., Mancini, G., Trombini, E. (2017) Brief report: the influence of dissociative experiences and alcohol/drugs dependence on Internet addiction. *Mediterranean Journal of Clinical Psychology*, doi: 10.6092/2282-1619/2017.5.1500.
- Brand, M., Wegmann, E., Stark, R., Müller, A., Wölfling, K., Robbins, T.W., Potenza, M.N. (2019). The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors, *Neuroscience and Biobehavioral Reviews*, doi: 10.1016/j.neubiorev.2019.06.032
- Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., ... Dai, J. (2020). Mental health problems and social media exposure during COVID-19 outbreak. *PLoS One* **15**(4), e0231924.
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being?. *American Psychologist Journal.*, 53, pp. 1017-1031.
- Lavin, M., Marvin, K., McLarney, A., Nola, V., Scott, L. (1999). Sensation seeking and collegiate vulnerability to Internet dependence. *CyberPsychological Behavior*, 2, pp. 425–430.
- Longstreet, P., Brooks, S., Gonzalez, E.S. (2019). Internet addiction: when the positive emotions are not so positive. *Technology in Society*. doi: 10.1016/j.techsoc.2018.12.004
- Masaeli, N., Farhadi, H. (2021). Prevalence of Internet-based addictive behaviors during COVID-19 pandemic: a systematic review. *Journal of Addictive Disorders*. **39**(4), pp. 468–488.
- Mattioli, A.V., Sciomer, S., Cocchi, C., Maffei, S., Gallina, S. (2020). Quarantine during COVID-19 outbreak: changes in diet and physical activity increase the risk of cardiovascular disease, *Nutrition, Metabolism and Cardiovascular Diseases*. 30(9), pp. 1409–1417.
- Rumbaut, R. (2004). Ages, life stages, and generational cohorts: Decomposing the immigrant first and second generations in the United States, *International Migration Review*, 38, pp. 1160–1205.
- Tofallis, C. (2020). Which formula for national happiness?, *Socio-Economic Planning Sciences*, 70, 100688. <https://doi.org/10.1016/j.seps.2019.02.003>

# A composite indicator to measure regional investment policies on R&D and innovation

Sergio Salamone, Alessandro Faramondi, Stefania Della Queva

## 1. Introduction

This work illustrates the results of the Smart Specialisation Italian enterprises classification process, aimed to support the Territorial Cohesion Agency in Italy, which is in charge of the monitoring and implementation of the European Smart Specialisation Strategy (S3).

This information policy requirement, which emphasizes the role of research and innovation as a leading factor for territorial growth and competitiveness, has resulted in the preparation of the “Statistic territorial and sectorial information for the cohesion policies 2014-2020” project from ACT, DpCoe and Istat, in which Istat has defined the enterprises S3 classification and the delimitation of national and regional areas of intelligent specialisation.

The traditional systems of classification for economic activities are often inadequate if compared to the shift from “horizontal” policies to “selective” policies (see for example place based, priority setting), and the willingness is to not turn back to traditional industrial and sectorial policies.

The potential enterprises classification S3 overcome this limit and allows to give directions on technological domains, developmental trajectories for businesses and territories.

The conceptualisation of the S3 components, derived from the original theory, aims to define a flexible and repeatable theoretical model, which could be easily adapted to different contexts.

Consequently, both the classification and its derived monitoring indicators are applicable to different domains pertaining to Smart Specialisation areas: although Smart Specialisation Strategies are not explicitly mentioned or linked in the PNRR, strong links are shown between the S3 priority areas and the Italian plan initiatives as defined, such as “Digitisation, Innovation and Competitiveness component of the productive system” and “From Research to Business”.

## 2. Composite indicator definition for regional investment policies measurement

The conceptual framework for the S3 theoretical definition recognized the Smart Specialisation Strategy as a policy guideline which emphasizes the role of research and innovation as a leading factor for territories development and competitiveness.

Furthermore, the S3 additions required to find specialization areas in order to maximize the results from research and development investments and to translate these results into new products and services.

In this scenario, the conceptual framework refers to 5 specific factors to represent the S3 enterprise general concept: “Research and Development”, “Innovation”, “Human Capital”, “the ability to foster local development” and “economic performances”<sup>1</sup>.

Based on the theoretical framework definition, the operativization of the concepts brought to the connection between elementary indicators (built by previously selected elementary variables) and sub-factors. The guidelines from the Handbook on constructing composite indicators – OECD

---

<sup>1</sup> For a complete understanding of the methodology used to build the Smart Specialisation classification, the composite indexes and the monitoring indicators, the guidelines are published at the following link: [https://www.agenziacoesione.gov.it/wp-content/uploads/2022/03/Guida-alla-lettura-degli-indicatori-S3\\_notam Metodologica-4.pdf](https://www.agenziacoesione.gov.it/wp-content/uploads/2022/03/Guida-alla-lettura-degli-indicatori-S3_notam Metodologica-4.pdf).

The statistical tables with indicators on specialization areas divided by region are at the following link: <https://www.agenziacoesione.gov.it/lacoesione/dati-statistici-sulla-politica-di-coesione/indicatori-regionali-classificazione-s3/>

2008 were applied to build the composite index.

The major data source is the 2019 Enterprises Census Survey, together with Istat statistical registries on enterprises, which allowed to give consolidated directions on about a million enterprises.

S3 theoretical framework is based on a multidimensional concept, and the S3 enterprise concept is a theoretical construct. That's why a composite index was chosen: the complexity is represented by the S3 construct multidimensionality, that requires for its measurement to overcome conceptual and definitional obstacles.

A composite index is a mathematical combination of a set of elementary indicators, which could represent the different dimensions of the examined construct.

We build a composite index for each specific enterprise, used to select the potential S3 enterprises not only based on the major economic activity but taking into account the intangible assets that represent the Smart Specialisation Strategy dimensions.

The five S3 factors described above are composed of 10 specific dimensions and 35 elementary indicators.

After numerous experiments with different methods to summarize a set of elementary indicators, two different methodologies were identified:

- a) the elementary indicators were synthesized in specific dimensions through the Wrocław taxonomic method;
- b) the Mazziotta–Pareto Index, a non-compensatory composite index, was used to summarize specific dimensions in general dimensions;
- c) the last step for the potential S3 enterprises composite index calculation was obtained by extrapolating the enterprises with scores above the median line in each one of the 5 S3 factors.

The innovation in the methodology used for the composite index definition described in this work consists in the information synthesis for each single unit of analysis, i.e. for each enterprise, and in the aggregation of qualitative and often dicotomic elementary indicators. Having a score for each enterprise allows to flexibly differentiate between economic areas of interest.

Furthermore, the composite index covers the need to be *transparent* in the calculation (compared to black box machine learning methods), *replicability* and *modularity*.

## 2. Output and results visualization

The output of the present work is composed by a set of indicators for each specialisation area, built from the potential S3 enterprises classification, both nationally and regionally.

The indicators defined through the Census data allowed the construction of 34 tables by specialization areas and are composed of: structural and economic indicators (enterprises, employees, added value, export etc.); indicators on intangible assets strategic investments (R&D, technology and digitalization, human capital, internationalisation, social and environmental responsibility); on enterprises relationships through agreements with universities, public and private research centers, Public Administration; indicators on environmental sustainability.

The output allows regional or national policy makers to compare the 12 Smart Specialisation areas as illustrated in Figure 1, which shows 2 of the 34 regional tables by specialisation area.

## Figure 1 – Regional tables for specialization area, Abruzzo Region

Tavola 2 - Imprese per area di specializzazione. Regione Abruzzo - Anno 2018  
(Valori assoluti e percentuali)

Arete di specializzazione	Imprese	% imprese sul totale imprese specializzate della regione	% imprese sul totale imprese della regione
Aerospazio	542	7,6	2,5
Agroalimentare	3.331	46,6	15,3
Economia del mare	780	10,9	3,6
Chimica Verde	612	8,6	2,8
Design Creativita e Made in Italy	1.458	20,4	6,7
Energia e Ambiente	1.521	21,3	7,0
Fabbrica Intelligente	941	13,2	4,3
Mobilita sostenibile	940	13,2	4,3
Salute	1.279	17,9	5,9
Comunità intelligenti sicure e inclusive	740	10,3	3,4
Tecnologie per gli ambienti di vita	1.228	17,2	5,6
Tecnologie per il patrimonio culturale	839	11,7	3,9
Totale imprese specializzate della Regione	7.151	100,0	32,9
Totale imprese della regione	21.756	100,0	100,0

Tavola 7 - Valore delle esportazioni per area di specializzazione. Regione Abruzzo - Anno 2018  
(Valori assoluti e percentuali)

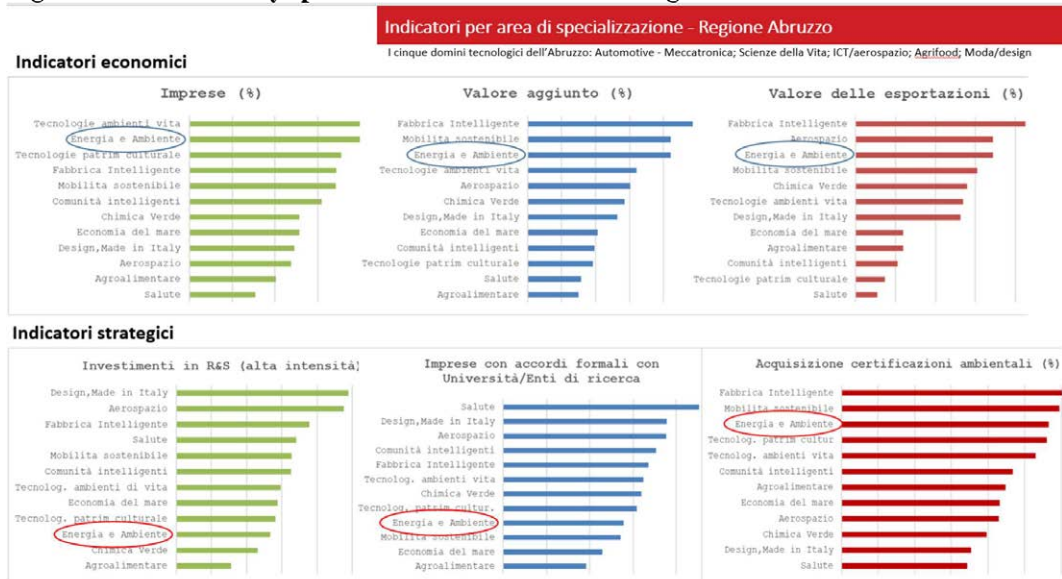
Arete di specializzazione	Valore delle esportazioni (€)	% esportazioni sul totale esportazioni delle imprese specializzate della regione	% esportazioni sul totale esportazioni delle imprese della regione
Aerospazio	1.999.104.704	59,9	34,5
Agroalimentare	815.784.124	24,4	14,1
Economia del mare	694.756.773	20,8	12,0
Chimica Verde	1.673.290.177	50,1	28,9
Design Creativita e Made in Italy	1.638.894.790	49,1	28,3
Energia e Ambiente	2.035.527.208	61,0	35,1
Fabbrica Intelligente	2.481.622.041	74,3	42,8
Mobilita sostenibile	1.772.466.786	53,1	30,6
Salute	319.283.057	9,6	5,5
Comunità intelligenti sicure e inclusive	610.439.014	18,3	10,5
Tecnologie per gli ambienti di vita	1.559.897.756	46,7	26,9
Tecnologie per il patrimonio culturale	423.095.313	12,7	7,3
Totale imprese specializzate della Regione	3.339.547.140	100,0	57,6
Totale imprese della regione	5.797.853.062	100,0	100,0

Dashboards such as the one shown in Figure 2 for Abruzzo region were built in order to simplify the learning and comparison between specialisation areas, looking at different indicators within the same territory.

The compared indicators have different nature, economic or strategic, to underline, as an example, that the specialisation area “Energia e Ambiente” in Abruzzo region performs well in economic indicators, being in the first three areas, but has some delays referring to some strategic indicators such as R&D investments, agreements with universities or environmental certifications.

At the top of the dashboard, the priority areas chosen by Abruzzo region are shown.

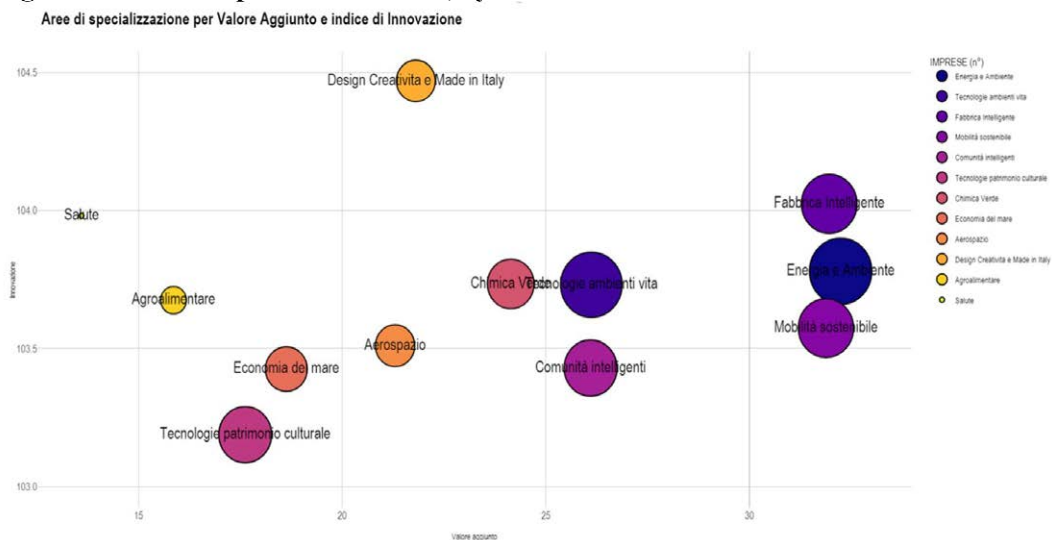
## Figure 2 – Indicators by specialisation area – Abruzzo Region



Data visualization instruments provide an observable benchmark between areas at a national level too: Figure 3 shows the 12 areas positioning regarding to the relationship between added value and innovation composite index.

The areas “Fabbrica intelligente”, “Energia e Ambiente” and “Mobilità sostenibile” show the best relationship between these two dimensions; the area “Design, creatività e Made in Italy” has an intermediate position referring to enterprises added value although it has the highest innovation level.

**Figure 3 – National specialisation areas, by added value and innovation index**



## References

- Balland, P.-A., Boschma, R., Crespo, J., Rigby, D.L., 2019. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Reg. Stud.* 53, 1252–1268. <https://doi.org/10.1080/00343404.2018.1437900>
- D’Adda, D., Guzzini, E., Iacobucci, D., Palloni, R., 2019a. Is Smart Specialisation Strategy coherent with regional innovative capabilities? *Reg. Stud.* 53, 1004–1016. <https://doi.org/10.1080/00343404.2018.1523542>
- Gianelle, C., Guzzo, F., & Marinelli, E. (2019). Smart Specialisation Evaluation: Setting the Scene. *Smart Specialisation – JRC Policy Insights*, JRC116110, March.
- Barca F. (2009), *An Agenda for a reformed Cohesion policy. A place-based approach to meeting EU Challenges and Expectations*. Independent Report to D. Hubner. Bruxelles: Commissioner of Regional Policy
- OECD (2008) "Handbook on constructing composite indicators: methodology and user guide". OECD Publications, Paris
- Lazarsfeld P.F. (1953), *Mathematical thinking in the social sciences*. Glencoe: The Free Press. Journal Article

# Assessing intimate partner violence in African countries through a model-based composite indicator

Anna Maria Parroco, Micaela Arcaio

## 1. Intimate partner violence

Violence against women has been recognized to affect all dimensions of women's lives and health, involving victims' both physical and mental conditions and their general well-being. In particular, intimate partner violence (IPV) is defined by the United Nations as a specific behavioural model of relationships, determined by either the current or former male partner perpetrating violence on women (UN, 2022). IPV is identified as either emotional, physical, and/or sexual abuse, each pertaining to one of the domains of the life of the victims.

Recent data show that 33% of ever-married women in Sub-Saharan Africa have survived this form of abuse, coming to the third-highest rate of lifetime IPV all over the world (WHO, 2021).

Many studies on this subject face IPV considering victims' and partners' characteristics, as well as the interplay between contextual and personal ones (Oyediran & Feyisetan, 2017). Gender theory has, indeed, highlighted the possible effects of contextual characteristics on abuse: for example, the ameliorative hypothesis tries to reason that as women's empowerment grows in a country, their victimization decreases, trying to connect gender equality to better living conditions overall for women (Heirigs & Moore, 2017). On the other hand, the backlash hypothesis ties equal standing for men and women to a rapid 'backlash' by men, since empowerment is seen as a threat to the existing patriarchal society (Heirigs & Moore, 2017). Moreover, maltreatment and parental-child relationships are associated with differential risks of the revictimization of children (Meinck *et al.*, 2015; Classen *et al.*, 2005).

Evidence of our previous study (Arcaio *et al.*, 2022) to investigate the determinants of physical, emotional, and sexual abuse, one independent from the other, shows that intergenerational transmission of violence – defined as witnessing parental violence – and revictimization processes – i.e., rape by a man other than her partner, and the number of past abusers in life – turned out to be crucial in predicting IPV itself. Moreover, the intensity of how justified physical violence is by women – the respondents – themselves and the number of control issues exerted by the respondents' current male partners also resulted in a significant risk factor. On the other hand, the partner's high education and higher wealth turned out to be protective factors.

However, to the best of our knowledge, the literature lacks an overall measure of violence suitable for surveys, while the Composite Abuse Scale (Revised) – Short Form (Gilboe *et al.*, 2022) captures IPV predominantly in a clinical setting. On these bases, the theoretical framework and construction of a Structural Equation Model (SEM) are proposed to create a composite indicator of IPV, also used to classify the African countries in the data to check for their levels of IPV.

## 2. Data

The Demographic and Health Survey (DHS) was used to conduct the analysis of intimate partner violence against women by their heterosexual partners. It is a nationally representative household survey, covering over 90 countries and 40 years. In particular, we focused on fifteen countries in Africa in which the module on domestic violence was administered: Angola, Burundi, Cameroon, Chad, Ethiopia, Gabon, Kenya, Liberia, Mali, Malawi, Rwanda, Senegal, Togo, Zambia, and Zimbabwe. Surveys range from 2008 to 2019.

In the survey, a sample of ever-partnered women was selected at random to collect information

on IPV in the households already involved. Respondents are asked about both current and past experiences of violence. The original sample pooled across the 15 countries accounts for over 80,000 women; however, our sample is restricted to almost 40,000 currently partnered women, due to the selection procedure for the domestic violence module in the survey. The respondents are aged 31 years on average ( $SD = 8.15$ ) – in the general survey, the women selected are aged 15-49.

**Table 1. Number of respondents per survey**

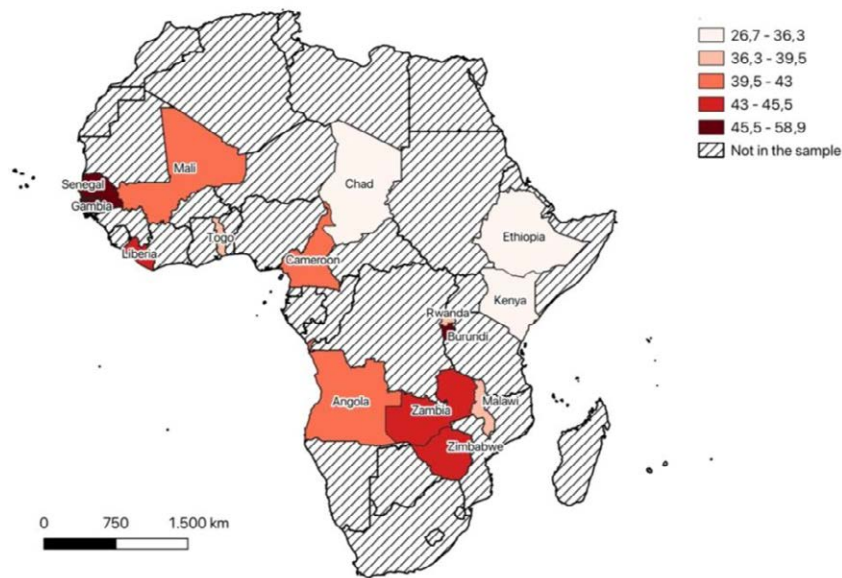
Survey	Years	Number of respondents
Angola	2015-2016	2,263
Burundi	2016-2017	2,189
Cameroon	2018	1,142
Chad	2014-2015	2,563
Ethiopia	2008	3,588
Gambia	2012	1,761
Kenya	2008-2009	1,374
Liberia	2006-2007	1,912
Mali	2012-2013	2,431
Malawi	2015-2016	3,954
Rwanda	2010-2011	1,519
Senegal	2019	2,832
Togo	2013-2014	2,255
Zambia	2018-2019	5,209
Zimbabwe	2015	4,069
Total		39,061

IPV is assessed via three indicators:

- “Physical Violence”, referring to acts of being pushed, shook, slapped, punched, or threatened at gunpoint by the partner;
- “Emotional Violence”, about humiliation, threat with physical harm or insults;
- “Sexual Violence”, indicating forced sexual acts by the respondent’s partner.

Figure 1 shows the percentage of women who have experienced any form of the three types of IPV by their partner in the country of residence. More than 40% of all respondents have experienced at least one form of abuse by their partner, and the country prevalence of IPV varies from 26.8% in Chad to 48.5% in Burundi.

Figure 1 IPV prevalence in the selected countries.



(Source: Authors' production)

### 3. Conceptual framework and methods

This work is the result of a preliminary step to build a composite indicator of intimate partner violence using a Structural Equation Model (Muthén, 1984), based on three latent variables – including IPV. This theoretical framework relies on the results of the previously estimated models that highlight the two dimensions that have an effect on IPV in all of its aspects, i.e., when it's either physical, emotional, or sexual abuse (Arcaio *et al.*, 2022).

As it is known, in SEMs, latent variables are specified as equations in the measurement model, in which a constraint is put on one of the exogenous variables to scale the latent variable. In this model, all the latent variables are built using the reflective approach.

We hypothesize that the two latent dimensions that have an effect on IPV are:

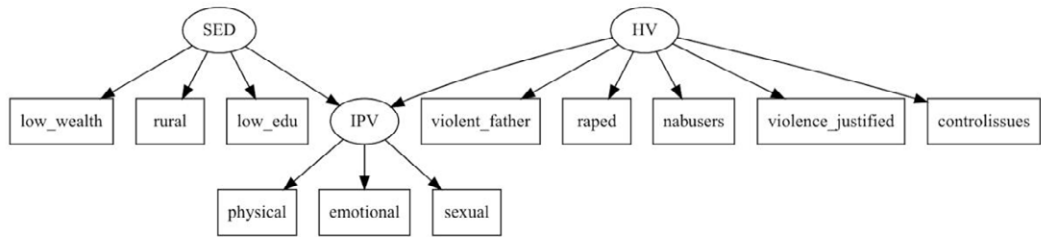
- Socio-economic deprivation (SED), which considers both personal and contextual characteristics – low household wealth (set as constraint), respondent's partner's status as a low-educated individual, and living in a rural area;
- History of violence (HV), determined by the intergenerational transmission of IPV (set as constraint), sexual violence by a man other than the current partner, number of abusers in life, number of justifications women give for physical violence, number of control issues by the partner.

On the other hand, IPV itself as a latent variable is assessed by an equation considering the presence of physical, emotional, and sexual violence by the current partner. Physical violence is used as a constraint to account for the scale of the latent variable.

The structural component of this framework checks for the association between the latent variables as specified above. The latent variables were, indeed, used in a structural model made of one equation, which checks for the socio-economic deprivation of the victims and their history of violence on intimate partner violence. A graphical representation of this model can be found in Figure 2.



Figure 2 Model of the latent variables for the Intimate Partner Violence indicator



(Source: Authors' production)

Statistics were done using R 4.2.0 (R Core Team, 2022), the lavaan (version 0.6-11; Rosseel, Y *et al.*, 2022) and the lavaanPlot (v0.6.2; Lishinski, 2021) packages.

#### 4. Results

What is presented here is the result of a preliminary analysis using a Structural Equation Model (SEM). Indeed, we think that the three components in the measurement model should be further refined. Still, the information given by the literature framework, evidence in the previously estimated models (both examined above), as well as the values of goodness-of-fit tests consent to examine these results, even if from an exploratory point of view. The Chi-Square test returned a  $p - value \approx 0$ , while the Comparative Fit Index (CIF, acceptance threshold  $> 0.9$ ) is equal to 0.926; the Root Mean Square Error of Approximation (RMSEA, acceptance threshold  $< 0.05$ ) and the Standardized Root Mean Square Residual (SRMR, acceptance threshold  $> 0.05$ ) are respectively equal to 0.046 and 0.039. All the tests point to a good fit of the model and all the coefficients in the model have  $p - value \approx 0$ . Thus, we identify an overall measure of violence to define a composite indicator of IPV.

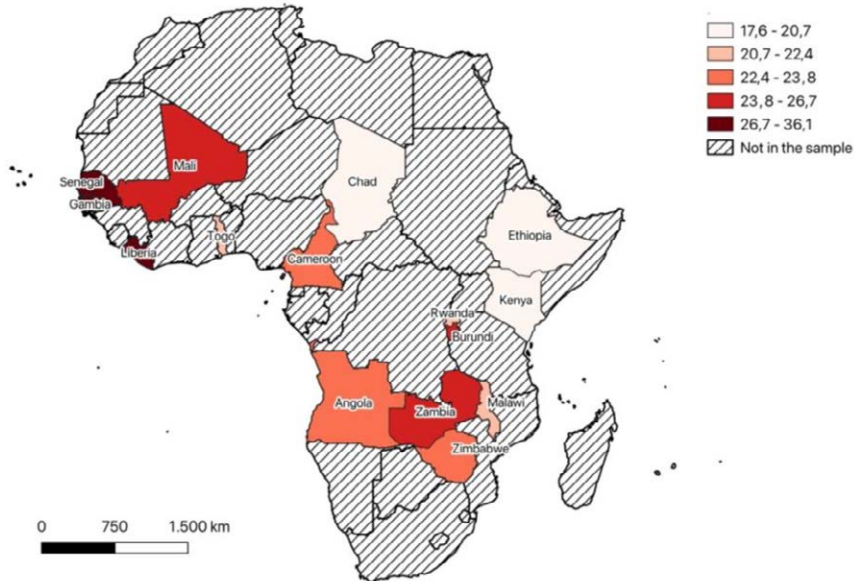
Living in a rural context has the highest standardized loading when it comes to “Socio-economic Deprivation”, while the number of control issues exercised by the partner has the highest loading for “History of Violence”. “Intimate Partner Violence” is most influenced by whether the respondent is a victim of emotional abuse or not.

The relationship between the latent variables, checked by a regression model in the SEM framework, shows that the latent variable “History of violence” (Standardised path coefficient = 0.876) has a greater positive effect on IPV than “Socio-economic Context” (Standardised path coefficient = 0.057).

Finally, a classification of the countries in the sample is built according to the value of the composite indicator of IPV, as to identify which countries are characterized by a higher level of IPV. The estimated factor scores were normalized; thus, the values are presented on a scale going from 0 (minimum levels of IPV) to 100 (maximum levels of IPV), and then the country average is computed. Senegal is the country with the highest average value of IPV, while Ethiopia is the country with the lowest average value of IPV.

All the results are shown in the map in Figure 3.

Figure 3 Map showing the IPV index.



(Source: Authors' production)

The results of this analysis show a strong association between IPV and the history of violence of victims. As stated in the literature, victims of abuse during their childhood or adolescence are more likely to fall into processes of revictimization (Meinck *et al.*, 2015; Classen *et al.*, 2005), making them more vulnerable to intimate partner violence.

When it comes to the latent variables, the standardized loadings of the indicators are examined to check for their correlation with their corresponding latent variable.

- *SED*. The highest standardised loading for SED belongs to the indicator concerning the type of residence, i.e., whether the respondent lives in a rural household: this correlation is equal to 0.774. The correlation between low wealth and SED is 0.588, while the correlation between the respondents' partner's lack of education is equal to 0.473.
- *HV*. The highest correlation for the HV latent variable can be found with number of control issues the partner exercises over the respondent (loading = 0.574). The other indicators don't check for the acceptability threshold of 0.4, with very low correlations.
- *IPV*. Here the highest correlation can be found between IPV and emotional violence, with a high correlation of 0.714. Physical violence is also well-correlated to IPV (St. loading = 0.680), while sexual violence is a bit less-so, whose standard loading is equal to 0.411.

Contextual and personal characteristics, here synthesized with the SED latent variable, although still relevant, matter less than past experiences of violence when it comes to current abuse, with 0.877 standard deviation change in IPV for a standard deviation change in HV. The data does not really support a strong effect of socio-economic deprivation on violence, with a standard deviation change of SED determining a change 0.057 change on IPV.

As for the countries considered in this analysis, it seems like Senegal, Gambia and Liberia require major interventions to fight this phenomenon with respect to the others. However, usual practices of education of women are next to futile in this particular context, where men need to be addressed for a more proactive fight against IPV.

## Conclusions

In summary, we believe that this work introduces some new elements in the study of intimate

partner violence despite the limitations related, among others, to the explorative stage of this research.

First and foremost is the idea of a composite indicator of IPV that considers the full set of relationships between the dimensions involved. The possibility of identifying the countries at greatest risk may be useful in making decisions related to choosing “where” to invest the most, to reduce its intensity.

The knowledge of explanatory variables of the phenomenon of IPV (as a whole) – such as the respondents’ partners’ educational attainment, wealth status, and the history of violence of the victims – allows the identification of those specific dimensions that need action for greater control of the phenomenon.

Both aspects are of considerable importance not only for the territorial context examined in this study but also, more generally, for developed countries, in which, as it is known, the phenomenon is equally relevant.

From a methodological point of view, further development of this study will involve refining the measurement model and the adoption of a multilevel SEM model, with the inclusion of second-level predictors to account for the nature of the data – given that they are drawn from surveys conducted in different countries and years.

However, the nature of the data themselves gives rise to several limitations. Social desirability maims the total reliability of all collected data on intimate partner violence, and these data are not exempt from this issue. Moreover, victims tend to either deny or hide their experience of violence, thus causing an underestimation of the phenomenon of intimate partner violence itself.

## References

- Arcaio, M., Mendola, D., Parroco, A.M. (2022). Domestic Violence in Africa: A Glance Through the DHS Survey, *Book of the short papers of the 51st Conference of the Italian Statistical Society*, Pearson, 1112-1117.
- Classen, C. C., Palesh, O. G., & Aggarwal, R. (2005). Sexual revictimization: A review of the empirical literature, *Trauma, Violence, & Abuse*, 6, 103–129.
- Gilboe, M., Wathen, C., Varcoe, C., MacMillan, H., Scott-Storey, K., & Mantler, T. et al. (2022). Development of a brief measure of intimate partner violence experiences: the Composite Abuse Scale (Revised)—Short Form (CAS R -SF). *BMJ Open* 2016;6:e012824. doi:10.1136/bmjopen-2016-012824
- Heirigs, M., Moore, M. (2017). Gender inequality and homicide: a cross-national examination, *International Journal of Comparative and Applied Criminal Justice*, 42(4), 273-285. doi: 10.1080/01924036.2017.1322112
- Meinck, F., Cluver, L. D., Boyes, M. E., & Mhlongo, E. L. (2015). Risk and protective factors for physical and sexual abuse of children and adolescents in Africa: A review and implications for Practice, *Trauma, Violence, & Abuse*, 16, 81–107.
- Muthén, B. (1984). *A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators*, *Psychometrika*, 49(1), 115–132. doi:10.1007/bf02294210
- Oyediran, K. and Feyisetan, B., 2017. Prevalence and contextual determinants of intimate partner violence in Nigeria. *African Population Studies*, 31(1).
- United Nations, *What Is Domestic Abuse?* <https://www.un.org/en/coronavirus/what-is-domestic-abuse>. Published 2022. Cited 13 Feb 2022.
- World Health Organisation. *Violence Against Women Prevalence Estimates, 2018: Global, Regional and National Prevalence Estimates for Intimate Partner Violence Against Women and Global and Regional Prevalence Estimates for Non-Partner Sexual Violence Against Women*. Executive Summary. Geneva: World Health Organisation; (2021)

# Students' feedback on the digital ecosystem: a structural topic modeling approach

Annalina Sarra, Adelia Evangelista, Tonio Di Battista

## 1. Introduction

In March 2020, to contain the spread of the COVID-19 pandemic, almost all educational ecosystems (school, universities and private centres) around the world were forced to cancel face-to-face classes and replace them with didactic instruction online. Various and diversified methods of teaching delivered remotely were activated quickly. These solutions have undoubtedly had the purpose of ensuring the continuity of basic education and institutional activities, but they also made it possible to experiment, on large scale, didactic solution, mediated by screen, at design and didactic mediation level and interaction. The debate around the way educational systems reacted to the emergency is probably going to be a proper theme of investigation for next years. In this respect, (14), argue that the infrastructures for digital education that have been chosen to give a reply to the pandemic crisis, will redefine public education for the future. In addition, other scholars, see for example (2) and (6), have already carried out researches on screen-mediated didactics in the pandemic context. Their studies highlighted some essential specificities for a positive teaching-learning process, mainly related to the sociality and the possibility of working in cooperative environments, the possibility of co-building knowledge in an active way, within a community of practice. Following these lines of research, in this paper, we are aimed at capturing students' perspectives and perceptions on screen-mediated didactics during the pandemic emergency. Data have been collected through a survey, which consisted of open-ended questions administrated to students attending six teaching large courses, held by four professors in two different Italian universities (Macerata and Chieti-Pescara). In particular, in the research have been involved students who attended course of Educational Sciences degree (45 from the course of "Didactics" and 48 from the course of "Special Pedagogy"). The questionnaire was also administrated to students enrolled in the Primary Education degree programme: 230 from the course of "Technologies for Education and Learning", 230 from the course "Laboratory of Technologies" and 230 from the course of "General Education". Finally, there were students who attended the course "Didactics of Training", enrolled in the Pedagogical Sciences degree. All courses refer to the year 2019/2020. To circumvent the dilemma between the benefit of having open-ended questions and the cost associated with their analysis, we adopt, in this work, an unsupervised topic modelling approach. More in detail, we focus on Structural Topic Modeling (10), which is deemed a variant of Latent Dirichlet Allocation (1), suited to address the strict statistical assumption that all texts in the modelled corpus are generated by the same underlying process. The remainder of the paper is organized as follows. Section 2 describes the unsupervised topic modelling adopted, while Section 3 presents the results. Section 4 contains an interpretation of the main findings and the conclusions.

## 2. Methodology

Topic modelling, focusing on text mining and information retrieval, has received a lot of attention and gained widespread interest among researchers, in recent years, in many research

Annalina Sarra, University of Chieti-Pescara G. D'Annunzio, Italy, [annalina.sarra@unich.it](mailto:annalina.sarra@unich.it), 0000-0002-0974-0799  
Adelia Evangelista, University of Chieti-Pescara G. D'Annunzio, Italy, [adelia.evangelista@unich.it](mailto:adelia.evangelista@unich.it), 0000-0002-7596-9719  
Tonio Di Battista, University of Chieti-Pescara G. D'Annunzio, Italy, [tonio.dibattista@unich.it](mailto:tonio.dibattista@unich.it), 0000-0003-2139-7273

Referee List (DOI 10.36253/fup\_referee\_list)  
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Annalina Sarra, Adelia Evangelista, Tonio Di Battista, *Students' feedback on the digital ecosystem: a structural topic modeling approach*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.36, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 203-208, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

fields. The core idea behind topic models is that documents are mixture of multiple topics. One of the most used probabilistic topic modelling algorithm is the Latent Dirichlet Allocation (LDA) (1). In the LDA approach, documents are generated via 3-level hierarchical Bayesian structure, under which each document  $d_m$  is modelled a finite mixture over a set of  $K$  corpus-wide topics  $z_k$  (1) and each topic is modelled as a set of  $V$  words  $w_v$ . The generative process performed by LDA on a corpus of documents can be summarized as follows: for each topic  $z$ , choose the probabilities over words  $\phi_z \sim Dir(\beta)$ , where  $\phi_z$  is drawn from a symmetric Dirichlet prior distribution with parameter  $\beta$ ; for each document  $d$ , choose the probabilities over topics  $\theta_d \sim Dir(\alpha)$ , where  $\theta_d$  is drawn from a symmetric Dirichlet prior distribution with parameter  $\alpha$ ; for each word  $w_{dn}$  in document  $d$ , choose a topic  $z_{dn} \sim Multinomial(\theta_d)$  and choose a word  $w_{dn} \sim Multinomial(\phi_{z_{dn}})$ . Being LDA a bag of words model, the order in which the words appear is disregarded. Additionally, although LDA is able to extract hidden topics from text document, it does not allow examining the relationship between document-level information and the content of a document model. This limitation can be overcome by using Structural Topic Modelling (STM), developed by (10). STM is a natural-language processing algorithm expressly designed to represent the effect of external variables on topical content (probabilities associated with words in each topic) and topical prevalence (proportion of different topics that occurs within documents). Through STM, it is possible to estimate a series of regression models that treat the prevalence of each identified topic as an outcome variable. The STM capability has been investigated in an extensive body of works, in the fields of economics, finance, political science, education, new media (see, among others, (12), (15)).

### 3. Results

The textual responses collected in this study were pre-processed using common steps for cleaning text data, including tokenization, lowercase conversions, stop-removal and lemmatizing/ stemming. Corpus preparation and cleaning were done using the *quanteda* package (4) in R (8). The final corpus contains 1354 documents. To avoid any possible inconsistencies, we carried our topic analysis on the original texts, expressed in Italian language. The most frequently 20 words of the corpus are displayed in Figure 1. To extract hidden topics from the corpus, we used a STM package in R, developed by Roberts et al. (11). As argued by Roberts, for having semantically interpretable topics, words should tend to occur within response and their top keywords should be unlikely to overlap with keywords from other themes. The first analytical step was the identification of the appropriate number of topics. By triangulating different diagnostic measures (namely, held-out likelihood, residuals, semantic coherence and lower bound), 10-topic model was settled as the best option. In the topic labelling process, to come with topic labels that reflect the main themes in a clear and concise way, high probabilities (Highest Prob) words, frequency-exclusivity (FREX) words, Lift, Score metrics and top 10 representative words of each topic were used (Figure 2 and Figure 3).

The most interpretable Topics retrieved from STM were assigned to the following dimensions: “Physical space home” (Topic 1), “Lack of direct confrontation and relationship” (Topic 2), “Building the community: use of whatsapp”, (Topic 3), “Ask question to the professor” (Topic 4), “Communication and learning tools” (Topic 5), “Feedback” (Topic 6), “Listen to the recorded lesson again” (Topic 7), “Interaction with teacher” (Topic 8) (see wordclouds displayed in Figure 4).

The top words occurring within Topic 1 (lessons, distance, face-to-face, value, added, home) stress how that topic is connected with a different reinterpretation of learning environment. In more detail, students underline two central aspects: the possibility to have more concentration at home but also some elements of distraction or linked to digital divide. Looking at the set of

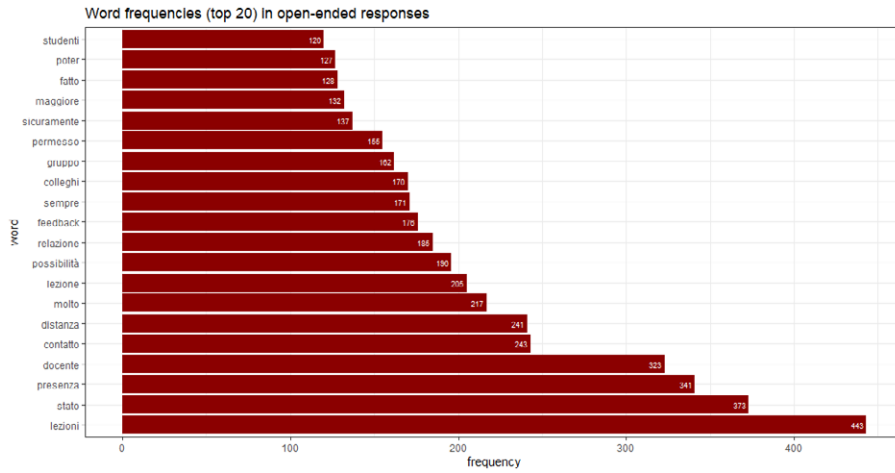


Figure 1: Word frequencies (Top 20) in open-ended responses

<b>Topic 1</b>	
Flex	valor, comodità, aggiunto, lezioni, spostamenti, dato, stato
Lift	costruzione, dovendo, fattore, persa, posto, recuperarla, riesco
Score	lezioni, valore, aggiunto, casa, seguir, comodità, spostamenti
<b>Topic 2</b>	
Flex	confronto, diretto, mancato, visivo, c'è, rapporto, altri
Lift	dispositivo, relazionarsi, visivo, confronto, diretto, adatto
Score	lezioni, valore, aggiunto, casa, seguir, comodità, spostamenti
<b>Topic 3</b>	
Flex	whatsapp, lavori, tenuti, gruppi, organizzazione, didattiche, videochiamate
Lift	accadeva, accordo, allungati, and, ansi, arricchendo
Score	whatsapp, tenuti, lavori, gruppo, gruppi, colleghi, organizzazione
<b>Topic 4</b>	
Flex	domande, dire, porre, me, disponibili, chat, secondo
Lift	accolto, adibito, all'esposizione, andata, apportato, apprendere
Score	domande, stato, maggiore, professore, dire, porre, chat
<b>Topic 5</b>	
Flex	supporto, canal, grazi, emotivo, svolto, telegram, app
Lift	aiutati, amicizie, elaborati, formativo, messaggi, organizzarci, poterci
Score	emotivo, telegram, canale, supporto, grazi, gruppo, tramite
<b>Topic 6</b>	
Flex	feedback, simile, disponibilità, minor, avvien, risponder, avviso
Lift	chiedeva, correzioni, eg, mail, ottima, preciso, riuscito
Score	feedback, simile, disponibilità, docente, stato, avviene, minor
<b>Topic 7</b>	
Flex	riascoltar, riveder, ascoltar, studentessa, registrazione, possibilità, registrare
Lift	affrontati, aggiunti, capita, dando, dell'ambiente, fondo, immediata
Score	riascoltar, riveder, registrare, registrazioni, seguir, possibilità, lezioni
<b>Topic 8</b>	
Flex	è, relazione, proprio, sì, stata, canali, diversa
Lift	abilità, abitudine, accaduta, adottata, affrontando, all'alta, all'aspetto
Score	relazione, docente, stabilir, dad, sì, diversa, cercato
<b>Topic 9</b>	
Flex	persona, dietro, crea, assolutamente, schermo, davanti, veder
Lift	rimaner, quindici, abita, accattivante, accogli, accontentare, accorgendo
Score	davanti, assolutamente, stare, schermo, persona, guardar, può
<b>Topic 10</b>	
Flex	compagni, contesto, nessuno, conoscenza, mancata, professori, pausa
Lift	adesso, compagni, contesto, correre, d'aiuto, erasmus, espormi
Score	mancata, compagni, professori, contesto, nessuno, erasmus, concetto

Figure 2: Top words for each topic according to highest probabilities, FLEX, LIFT and SCORE weighting

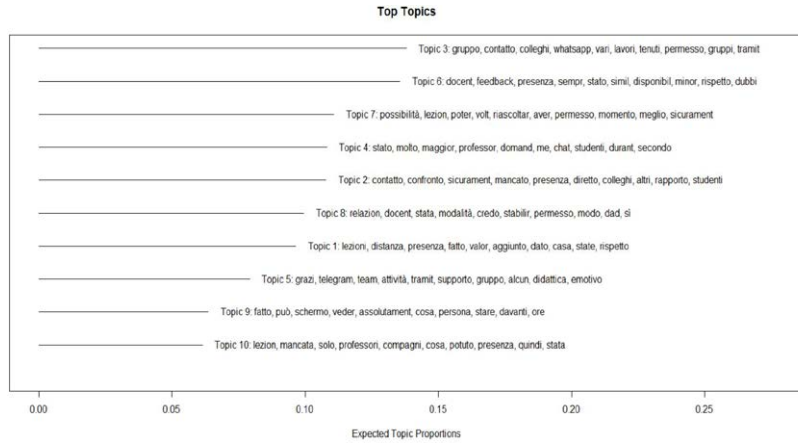


Figure 3: Top words associated with each topic resulting from structural topic modeling ( $k = 10$ )



Figure 4: Wordcloud: a) Topic 1: “Physical space home”; b) Topic 2: “Lack of direct confrontation and relationship”; c) Topic 3: “Building the community: use of whatsapp”; d) Topic 4: “Communication and learning tools”; e) Topic 6: “Feedback”; f) Topic 7: “Listen to the recorded lesson again”.

words linked to Topic 2 (contact, confrontation, absence, presence, direct), we are able to state that students think that interaction is somehow limited in the screen-mediated mode. Topic 3 focuses on the attempts made by the students of rebuilding the community or the contact with the other. Words associated to Topic 4 (questions, asking, available, greater, professor) recall the possibility for students to constantly ask questions to teacher. Terms immersed in Topic 5 refer to the online learning platform, perceived by students as essential for both supporting learning in an uncommon situation and as a space for discussion. Topic 6 captures the centrality of interaction and specifically of feedback and highlights how the teacher's feedback has not changed during the transition from face-to-face teaching methods to online mode. The top scoring words for Topic 7 clearly refer to the possibility of listening again to the lesson and of watching it more and more times, getting back to it in a recursive way. Finally the discussion in Topic 8, gives us the students' perception of having built a sound relationship with the professor. More challenging was to get insights from the last two dimensions characterized by less focused words. We also estimated the correlation between the identified topics. Except for "Interaction with teacher", the other topics are associated with at least a topic, meaning that they are likely to occur within the same documents. Finally, to complete the quantitative analysis of textual data, we incorporated the covariate information into topic modeling. Specifically, we estimated the topical prevalence by "teacher" covariate. The regression results support the causal impact of "teacher" variable that especially affects how Topic 2, Topics 5, 6 and 7 vary by document.

#### 4. Discussion and Conclusions

The purpose of this study was investigating how students, who attended courses in two Italian universities, experienced online education during the coronavirus emergency. To this end, we used an unsupervised approach, based on the identification of latent topics, to automatically analysis open-ended questions. A throughout analysis of topic modelling results allows us to draw the following conclusions. By considering the perceptions in relation to blended environments, modellized by Chang and Fisher (3), we focus on the categories of "Interaction" and "Reply", which exploring to what extent communication is achieved from students' point of view and how students had felt about using web-based medium, respectively. Topics retrieved by the structural topic modelling analysis can be aggregated into three broad themes: perceptions related to the *physically of body and space*, perceptions related to *virtual relationships* and *communication and perception related to feedback*. Topic 1 and Topic 7 fall in the category "Spatiality and corporeity". In the distance learning mode, students recognized the undeniable advantages of being free from having to move: due to distance educational technologies implementation, remote learning is available to everyone, in any place. This aspect enables to stretch the same concept of access and participation and it has to be considered as an element of inclusion. Additionally, students reported the possibility of a greater interaction and participation during the lesson and the opportunity to listening again to the lesson and of watching it more and more times, getting back to it in a recursive way along time and in different moments. Under the umbrella of "virtual relationship" theme, there are Topics 2, 3 and 5. Based on the results of the topic modelling algorithm, we found out that students expressed that the filter of the screen was perceived as a barrier. In fact, even if online learning enables them to see each other and talk each other, it interrupted the relation flow that used to be experienced in a classroom. Finally, Topics 4 and 6 are the relevant themes for the broad category "Feedback". Throughout these topics, students underlined how the emergency remote education did not compromise the possibility of giving and receiving some feedback. Overall, the results of this study suggest the fluidity of contemporary education context: in other words, we are in front of a dynamic, hybrid educational context, with a weak structure, in continuous transformation (7). This feature,



exacerbated during crisis periods for the emergences of new obstacles and constraints, requires a rethinking of learning-teaching practises. A robust pedagogically and learning environment can be guaranteed by hybridizing the educational contexts. “Vertical blended”, which provides for an alternation between moments of classroom teaching activity and remote teaching moments, must be accompanied by a “Horizontal blended”, which integrates and hybridizes real and virtual, analogical and digital in a synchronous dimension (9).

## References

- [1] Blei, D. M., Ng, A. Y. , Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, **3**, pp. 993–1022.
- [2] Carrillo, C., Flores, M.A. (2020). Covid-19 and teacher education: a literature review of online teaching and learning practices. *European Journal of Teacher Education*, **43**, pp. 466–487.
- [3] Chang, V., Fisher, D. (2003). The validation and application of a new learning environment instrument for online learning in higher education, in *Technology-rich learning environments: A future perspective*, eds. M.S. Khine, D. Fisher , London: World Scientific, pp. 1–20.
- [4] Kenneth, B., Watanabe, K., Wang, H., Nulty, P., Obeng, A. , Müller, S., Matsuo, A. (2018) quanteda: an R package for the quantitative analysis of textual data. *Journal of Open Source Software*, **3** (30), pp. 774.
- [5] Kuhn, K. D. (2018) Topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, **87**, pp. 105–122.
- [6] Pereira, S.P., Fernandes, R.L., Flores, M.A. (2021) Teacher education during the covid-19 lockdown: insights from a formative intervention approach involving online feedback. *Education Sciences*, **11** (400), pp. 1–14.
- [7] Pennycook, A. (2005). Teaching with the flow: fixity and fluidity in education. *Asia Pacific Journal of Education*, **25**(1), pp. 29–43.
- [8] R Core Team (2022). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- [9] Rivoltella, P. C. (2021). *Apprendere a distanza. Teorie e metodi*, Raffaello Cortina Editore. Milano, Italia.
- [10] Roberts, M. E., Stewart, B. M., Tingley, D., Airoidi, E. M. (2013). The structural topic model and applied social science, in *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*.
- [11] Roberts, M. E., Stewart, B. M., Tingley, D. (2019). stm: an R package for structural topic models. *Journal of Statistical Software*, **91**(2). <https://doi.org/10.18637/jss.v091.i02>
- [12] Rodriguez, M. Y., Storer, H. (2019). stm: a computational social science perspective on qualitative data exploration: using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services*, **38**(1), pp. 54–86.
- [13] Rothschild, J. E., Howat, A. J., Shafranek, R. M. , Busby, E. C. (2019). Pigeonholing partisans: stereotypes of party supporters and partisan polarization. *Political Behavior*, **41**(2), pp. 423–443.
- [14] Selwyn, N., Jandrić, P. (2020). Postdigital living in the age of covid-19: unsettling what we see as possible. *Postdigital Science and Education*, **2**(3), pp. 989–1005.
- [15] Zafari, B., Ekin, T. (2019). Topic modelling for medical prescription fraud and abuse detection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **68**(3), pp. 751–769.

# The digitization of the private sector. A non-aggregative method to monitor the NRRP agenda at macro-area level

Susanna Traversa, Enrico Ivaldi

## 1. Introduction

2020 has represented a break from the economic and social policies adopted in Europe until the SARS-CoV-2 (Covid-19) emergency spread (Grasso et al., 2021). The containment measures introduced during the health crisis have produced a series of effects, including an acceleration of the digitisation both in social and in economic sphere (OECD, 2020). Within the latter in particular, Covid-19 has had a leverage function regarding private sector innovation, leading to the adoption of measures to implement the use of digital and technologies in order to ensure continuity in the production sector of goods and services (Casquilho-Martins and Belchior-Rocha, 2022). However, bringing attention to the Italian context, it is necessary to consider the critical issues related to the still-present digital divide between the northern and south-central areas of the peninsula. The digital divide, as well as digital illiteracy and infrastructural barriers, are the main obstacles that have slowed Italy's digital transition over the past few years with respect to the European scenario (Traversa et al., 2022; European Commission, 2021). It is precisely in the promotion of digitization policies that the European Union identifies one of the main drivers for a sustainable and resilient economic recovery, through which business continuity can be ensured despite lockdown policies (European Commission, 2020a,b). Based on European investment indications, great importance has been recognized by the Italian government to the theme of digitization by reserving for it targeted interventions within the first mission of the National Recovery and Resilience Plan (NRRP) for which EUR 49.2 billion has been allocated. Specifically, the NRRP commits Component 2 of Mission 1 to the strengthening of competitiveness within the private sector to be pursued by means of greater diffusion of digitization processes, technological innovation, and the strengthening of Industry 4.0 policies. The medium - to long-term recovery goals on which the NRRP is based need tools that can assess actual effectiveness on the Italian territory. Tools that can not only express the spread of digital business integration from a geographic point of view, identifying areas that show lower performance than the national context, but that can also monitor the progress of these interventions over time. In this study, a different synthesis methodology that makes use of a non-aggregative strategy was employed. The study is divided into sections where the main opportunities for developing an index to measure the effectiveness of the policies presented in the NRRP, as well as the traits and ramifications of using a non-aggregative strategy for the temporal study of socio-economic phenomena, will be outlined. Afterward, the outcomes of the index's application to the Italian context will be discussed.

## 2. Methodology

Given the complex and multi-dimensional nature of digitization processes, it is possible to approach their study through the construction of synthetic indices. In contrast to recent literature (Traversa et al., 2022; Benecchi et al., 2021; European Commission, 2021), the choice in this case fell on non-aggregative synthesis by means of the Partially Ordered Set (POSET). Among the main advantages of using a non-aggregative method over composite index construction is the possibility, by not carrying out the aggregation and weighting of

Susanna Traversa, University of Genoa, Italy, traversa.su@gmail.com, 0000-0001-5030-2021  
Enrico Ivaldi, University of Genoa, Italy, enrico.ivaldi@unige.it, 0000-0001-6687-9378

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Susanna Traversa, Enrico Ivaldi, *The digitization of the private sector. A non-aggregative method to monitor the NRRP agenda at macro-area level*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.37, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 209-214, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

variables, to limit the loss of information due to the flattering effect, which depends on the existence of incompatibilities between variables within a multi-dimensional system. Conversely, one of the main critical issues attributable to the POSET technique is related to computational aspects, as it requires the use of advanced statistical analysis software to enable its calculation. From a theoretical point of view, a Partially Ordered Set can be defined as a set in which there is no binary relationship between all pairs of profiles of which it is composed (Davey and Priestley, 2002; Fattore, 2017). The graphical representation tool used in the literature to represent the POSET system of relations is the Hasse diagram, according to two proprieties: (1) if  $x \leq y$ , the node - i.e. profile -  $y$  is placed above node  $x$ ; (2) if  $x < y$  then a “path” links node  $y$  to node  $x$ . By considering a pair of nodes, the one placed in the upper level is defined as being connected to the lower one according to a dominance relationship. The node (or nodes) with only descending relationships is said to be “maximum” (or maxima) of the POSET (Fattore, 2008). In order to address the synthesis of the digitization index the average height approach (avh) represents the most common method for the synthesis of multi-indicator systems through the study of POSETs (Alaimo et al., 2020; Fattore, 2017; Mazziotta and Pareto, 2020).

From a practical point of view, the synthesis vector is obtained by following a stepwise procedure: (1) Extract all line extensions of  $P$  by creating  $\Omega_p$ ; (2) For each element  $p \in P$  and for each  $l \in \Omega(P \cdot)$ , assign a rank  $r_l(p)$  of  $p$  in  $l$ , which represents  $1 +$  the number of angle covers joining  $p$  to the maximum of  $l$ ; (3) The third step is the calculation of the average  $r(p)$  of  $r_l(p)$  over  $\Omega(P \cdot)$  for each  $p \in P$ .

The avr can be represented through a graph that shows the min-max avr range for each profile ( $\perp T$ ). As in the case of aggregate indices, the POSET technique also allows for the study of a phenomenon over time, thus complying with the requirements that will be presented in the course of the research design from the point of view of the selection of synthesis methods (Alaimo et al., 2020). The development of a temporal POSET consists in the merge of two (or more) POSETs related to the years under study. Each statistical unit is measured with respect to variables referring to two (or more) different  $t$  times, calculating for each year the average rank of the POSETs. From a graphical standpoint, it is possible to obtain a visual representation of the temporal POSET by merging the corresponding Hasse diagrams. Following the merging of the POSETs and rebuild of the dataset, an intertemporal POSET is obtained which must again be subjected to the calculation of average height. Continuous recourse to avr calculations could lead to a loss of information in POSET. To solve problems of comparability between nodes a possible solution should be the use of a reference system common to the whole POSET - embedded scale - which will represent the benchmark through which the evolution of the POSET can be interpreted.

## 1. Research design

As a result of the analysis of the targets enshrined by the NRRP within M1.C2. “Digitization, Innovation and Competitiveness in the Production System”, the variables were selected through the “Rilevazione sulle tecnologie dell'informazione e della comunicazione nelle imprese” (source: National Institute of Statistics – I.Stat) data-warehouse by following a formative approach (Diamantopoulos et al., 2008). Four variables expressing the digitization of the private sector were selected and presented below: (1) percentage of enterprises with fixed or mobile broadband connection (BBC); (2) percentage of enterprises using robots (ROB); (3) intra-muros research and development expenditure (thousands of euros at current prices) (R&S); (4) percent-age of enterprises that organized training courses in the previous year to develop or upgrade the ICT/IT skills of their employees (FDS). The composition of the dataset needs some specification. The unavailability of access to I.Stat data with respect to ultrabroadband deployment in the macroareas led to the use of broadband connection data as a proxy. As for the 2020 figure for R&S spending, it was imputed from 2019 data, as it was not

available at the macroarea level but only nationwide. Variables were also selected based on spatial and temporal availability requirements. Specifically, the variables are presented on the basis of a geographical breakdown into macroareas: Northwest (NO); Northeast (NE); Center (CE); and “Meridione” (ME), which includes southern regions and islands. Italy (IT) was also considered as a comparison of macro-areas against the national average. In this way, it was possible to put more emphasis on the issue related to the digital divide and how it has evolved over the three-year period 2018-2020 considered. This time period makes it possible to capture the pre-pandemic trend of digital implementation in the productive sector by comparing it with the scenario realized in 2020, following the shocks produced by the pandemic. Computation of the results was conducted by means of the statistical software Rstudio and the package “parsec” (Fattore and Arcagni, 2014).

## 2. Discussion

Following the construction of the temporal POSET, the main results obtained from the application of the non-aggregative method are presented and discussed below. First, the Hasse diagrams of the individual years examined were constructed. As can be seen from Figure 1, the graphical representation of the 2018 and 2019 POSETs exhibit the same structure. IT, NE, NO are placed as maximal nodes, which are connected by a cover relation to the two lower nodes: CE and ME.

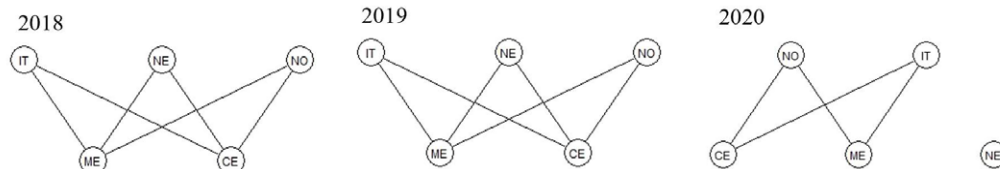


Figure 1: Comparisons between single years Hasse diagram. Period: 2018-2020.

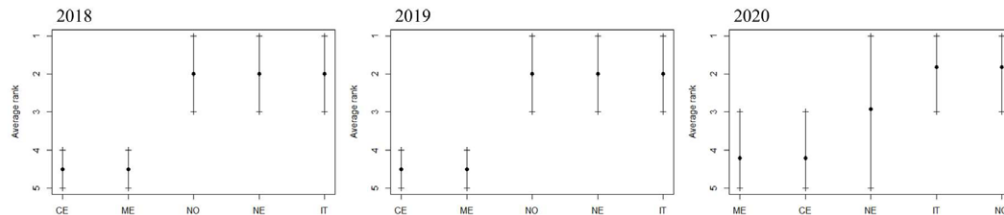


Figure 2: Comparisons between single years avr plot. Period: 2018-2020.

The dominance relationships between the profiles expressing northern and south-central regions are also confirmed following the calculation of the average ranks of the two-year pre-pandemic period (Figure 2). In the average rank plot, y-axis scale expresses the total number of observations sorted in descending order, attributing the best condition to the profile corresponding to rank 1. The points on the graph indicate the average value of the simulations obtained during the calculation of the average ranks, while the vertical bands express the variability of the profiles. A high range between the minimum and maximum value expresses the variability with respect to the identification of a unique average rank for the profile. The CE macro area confirms the worst performance in the first part of the period, presenting together with ME an average rank fluctuation range between the fourth and fifth rank. On the other hand, the situation differs for NO, NE, IT where the distance between the two whiskers is greater and ranges from 1-3. IT has better values, positioning itself on the far extreme right of the graph, followed by NE and NO. A different scenario for 2020 is reported. During the pandemic year, only IT and NO are in the maximum positions, maintaining a dominance relationship with the lower nodes of CE and ME. More peculiar, however, is the case of the

NE profile, which stands outside the Hasse as an anti-chain, reporting no comparability in its digitization to the other macro areas (Figure 1). Moving the focus to the avr plot, there is a reversal in ranking between ME and CE, with an increase in variability in the average ranks. In contrast to the previous two years, NO reports an improvement ranking ahead of IT. NE also confirms in the avr plot an “anomaly” with respect to digitization in the four variables derived from the M1.C2 presented before, with a range of variation in the avr maximum. An analysis of the original data obtained from I.Stat shows that compared to the pre-pandemic two- year period, the elementary indicators considered do not show significant changes (positive or negative), with the exception of some macroareas that have suffered more in certain dimensions from the impact of Covid-19. As can be deduced from the preliminary study of the POSETs of individual years, the Northeast macro-area experienced the greatest fluctuations during 2020. If in the case of ROB no significant - albeit positive - changes are observed, BBC percentage for NE ranks below the national average as well as being the only area that experiences a re- duction in the percentage (not significant). However, a greatest change is experienced for the FDS dimension, which loses 6.6 percentage points in 2020 compared to 2019. As for NO - along with NE - it tends to perform better over the period, trending above the national average in all variables, although there is a slight deterioration in the FDS dimension between 2019 and 2020, offset by a 3.5 percentage point improvement in BBC. Finally, the CE and ME macro areas show values below the national average for each dimension with deterioration in FDS in both macro areas and in ROB for the “Meridione” area. After reconstructing the background of digitization in the private sector in Italy, the results that emerged from the construction of the temporal POSET are, below, addressed, for which three benchmarks expressing MIN, MEDIAN and MAX were calculated. The best performances are identified for NO\_19, NO\_20, NE\_19, NE\_20 and IT\_19.

The Hasse diagram of the temporal POSET confirms an improvement despite Covid-19 for the macro-areas located in the northern part of the peninsula, although it achieves the result in ways that are not comparable with each other. Regarding national results, the nodes expressing national digitization show a deterioration for IT\_20 with implementation of digital integration not comparable to what was achieved in previous years. In addition to the effect of Covid-19 that could contextualize the worsening of the national average, the incomparability with IT\_18 and IT\_19 could be attributable to the performance of NE\_20. Lastly, as for ME and CE, the response to the shocks produced by Covid-19 has opposite effects. While ME after a positive trend in the pre covid two-year period observes a deterioration in digitization performance, CE shows an improvement in 2020 in line with what is realized for IT\_20. The impact produced by Covid-19 is evident from the average height ranking (Table 1).

The best ranking is attributed to NE\_19 followed by NO\_19 and NO\_20, underscoring on the one hand the better digital implementation within the private sector in the northern regions, and on the other hand a greater sensitivity with respect to the effects produced by Covid-19 on the NE macro area, which for the year 2020 reports a ranking lower than the IT 20 average. Coherent, on the other hand, are the performances of the southern regions, for which performance is observed to be positioned below the benchmark expressing the MEDIAN, with an improvement over the three-year period more rewarding for CE, which ranks at a higher Hasse diagram level than profiles in central and southern Italy (with the exception of NE\_19) and digitization in line with the national average IT\_20.

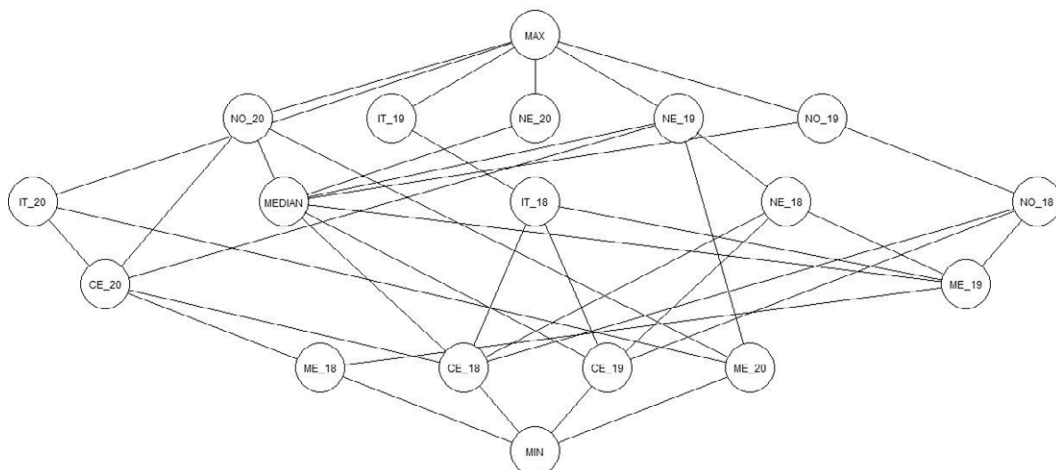


Figure 3: Temporal Hasse diagram.

<b>COD.</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
<i>IT</i>	9,792	13,902	13,463
<i>NE</i>	10,204	14,758	13,216
<i>NO</i>	10,035	14,386	14
<i>CE</i>	3,611	4,006	7,986
<i>ME</i>	2,847	4,926	6,407
<i>MIN</i>	1	1	1
<i>MEDIAN</i>	8,459	8,459	8,459
<i>MAX</i>	18	18	18

Table 1: Average height distribution of the Temporal POSET with benchmarks.

## 5. Conclusion

The development of quantitative tools to monitor the digitization goals of the private sector within the scope of the NRRP goals, is a timely issue worthy of further investigation. The still pronounced existence of the digital divide and digital illiteracy, represent a major obstacle to proper digital integration within enterprises, highlighting the negative impact produced by the rapid spread of digital in some areas of the country due to Covid-19. The POSET technique allow to contextualize rank positioning based on order relationships, which leads one to lean toward further exploration in the use of non-aggregative approach.

Indeed, by cross-referencing the information that can be obtained from the trend of basic indicators in individual regions, with the POSETs of individual years and the temporal one, it is possible to gain a greater understanding of the ways in which NRRP digitization goals are carried out in relation to territory and time. This provides an enhancing of the complexity of the phenomenon and not limiting it to a simplification as is occurring with aggregative synthesis techniques. Although the study is not free from limitations, due in part to the scarcity of available data inherent in digitization, it represents a possible starting point for the development of statistics assessing NRRP performance on a national scale. A further possibility to consider is the replicability of the study at the NUTS-2 territorial level, in order to highlight the performance of those regions that tend to show performance that is not in line with the performance of the macro-area to which they belong (Traversa et al., 2022).

## References

- Alaimo, L. S., Ciacci, A., Ivaldi, E. (2020). Measuring sustainable development by non-aggregative approach. *Social Indicators Research*, 1(22).
- Alaimo, L. S., Arcagni, A., Fattore, M., Maggino, F. (2020). Synthesis of multiindicator system over time: A poset-based approach. *Social Indicators Research*, 1(23).
- Benecci A., Bottoni C., Ciapanna E., Frigo A., Milan A., Scarinzi E. (2021), La digitalizzazione in Italia: evidenze da un nuovo indicatore regionale, in *Questioni di economia e finanza*, eds. Banca d'Italia Occasional papers, n.662.
- Casquilho-Martins, I., Belchior-Rocha, H. (2022). Responses to COVID-19 Social and Economic Impacts: A Comparative Analysis in Southern European Countries. *Social Sciences*, 11(2), pp. 36.
- Davey, B. A., Priestley, H. A. (2002). Introduction to lattices and order. *Cambridge university press*.
- Diamantopoulos, A., Riefler, P., Roth, K. P. (2008). Advancing formative measurement models. *Journal of business research*, 61(12), pp. 1203-1218.
- European Commission. (2020a). 2030 digital compass: the European way for the digital decade.(Available on Eurlex).
- European Commission. (2020b). European moment: Repair and prepare for the next generation. *Communication from the commission to the European parliament, the European council, the council, the European economic and social committee and the committee of the regions*, Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0456from=EN> (Available on line).
- European Commission. (2021). Digital economy and society index (DESI) 2021. *Thematic chapters*, Retrieved from <https://ec.europa.eu/newsroom/dae/redirection/document/80563> (Available online)
- Fattore M. (2008). Hasse diagrams, poset theory and fuzzy poverty measures. Hasse Diagrams. Poset Theory and Fuzzy Poverty Measures, *Rivista Internazionale Di Scienze Sociali*, 116(1), 63–75. <http://www.jstor.org/stable/41625201>
- Fattore M. (2017). Synthesis of indicators: The non-aggregative approach, in *Complexity in society: From indicators construction to their synthesis*, eds. Springer, pp. 193-212.
- Fattore, M., Arcagni, A. (2014). Parsec: An r package for poset-based evaluation of multidimensional poverty, in *Multi-indicator systems and modelling in partial order*, eds. Springer, pp. 317-330.
- Grasso, M., Klicperova´-Baker, M., Koos, S., Kosyakova, Y., Petrillo A., Vlase, I. (2021). The impact of the coronavirus crisis on European societies. What have we learnt and where do we go from here? Introduction to the COVID volume. *European Societies*, 23(sup.1), pp. S2-S32. Routledge.
- Italian Government. (2021). Piano Nazionale di Ripresa e Resilienza. *Presidency of the Council of Ministers*. Retrieved from <https://www.governo.it/sites/governo.it/files/PNRR0.pdf>.
- Mazziotta, M., Pareto, A. (2020). *Gli indici sintetici*. Giappichelli.
- OECD. (2020). OECD digital economy outlook 2020. *OECD Publishing*.
- Traversa, S., Ciacci, A., Ivaldi, E., Gonzalez-Relanõ, R. (2022), Measuring the Digital Gap in Italy: A NUTS-2 Level Index, in *Globalization, Income Distribution and Sustainable Development*, eds. Chandra Das, R. Emerald Publishing Limited, Bingley, pp. 265-281.

# Digital.VET: an innovative approach for teaching and training

Teresa Maltese, Maria Santarcangelo, Vito Santarcangelo, Diego Sinitò,  
Aneta Poniszewska-Marañda, Jure Šuligoj, Alcidio Jesus, Elisardo Sanchis

## 1. Introduction on context and motivation about Digital.VET

The significant changes that took place in the past decades and the big challenges posed at national and international level by the globalisation, the redefinition of the capital-labour relation and the technological revolution are bringing about a radical change of the economic, socio-political and cultural structures in European countries. VET (Vocational Education and Training) reforms (New Skills Agenda for Europe 2016) and labour market reforms have started a process aimed at filling the gap between demand and supply of competences. The demand of competences, in fact, is affected by factors requiring the constant adjustment of production and training processes as well as the greater connection between education/training system and enterprises. VET teachers must think about the training objectives required by the present innovations, taking into account the present cultural dynamic aspects and meet the students' needs by using adaptable teaching strategies, which can develop skills for inclusive participation and work independence. Digital training, included in national programmes, is essential in order to ensure effective training practices for the current VET system which is undergoing organizational and methodological change. The needs analysis carried out in early 2019 by each partner in its own territorial context has shown that out of 180 VET teachers/trainers belonging to both the private and the public sector, 91% stated that they have a poor knowledge of digital and immersive teaching methods and/or do not know how to use them effectively.

The project Digital.VET supports the objectives set out in national and European strategies for applying ICT (Information and Communication Technologies) to VET systems through teachers/trainers training. Its overall objective is to create a partnership among VET system operators aimed at the development of systematic approaches and of opportunities for the professional growth of VET teachers/trainers based on the development and innovation of education and training methods which are digital, open, innovative and effective. The partnership is made up of 5 VET and 1 IT organisations and has been implemented in 5 European countries: Poland, Italy, Portugal, Slovenia and Spain. It improves the technical knowledge as well as the expertise of VET teachers/trainers about the use of innovative and digital teaching methods by creating training pathways, training staff event and VET qualification which comply with EQF (European Qualifications Framework), ECVET (European Credit system for Vocational Education and Training) and EQAVET (European Quality Assurance Reference Framework for VET) European tools of recognition and transparency.

It has been based on the research carried out in partner countries, concerning best practices of

Teresa Maltese, Studio Risorse Srl, Italy, [teresa.maltese@studiorisorse.it](mailto:teresa.maltese@studiorisorse.it), 0000-0001-6794-9446

Maria Santarcangelo, Studio Risorse Srl, Italy, [maria.santarcangelo@studiorisorse.it](mailto:maria.santarcangelo@studiorisorse.it)

Vito Santarcangelo, ilnformatica Srl, Italy, [info@iinformatica.it](mailto:info@iinformatica.it), 0000-0003-4971-8788

Diego Sinitò, ilnformatica Srl, Italy, [diego@iinformatica.it](mailto:diego@iinformatica.it), 0000-0002-5044-0050

Aneta Poniszewska-Marañda, Akademia Humanistyczno-Ekonomiczna w Lodzi, Poland, [manielak@ahelodz.pl](mailto:manielak@ahelodz.pl), 0000-0001-7596-0813

Jure Šuligoj, Center republike slovenije za poklicno izobrazevanje, Slovenia, [jure.suligoj@cpi.si](mailto:jure.suligoj@cpi.si)

Alcidio Jesus, AFN academia formação do norte, unipessoal lda, Portugal, [elcid.justice@gmail.com](mailto:elcid.justice@gmail.com), 0000-0002-4683-1542

Elisardo Sanchis, Asociacion de innovacion emprendimiento y tecnologias de la informacion y la comunicacion inetica, Spain, [s.sancho@innetica.org](mailto:s.sancho@innetica.org)

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Teresa Maltese, Maria Santarcangelo, Vito Santarcangelo, Diego Sinitò, Aneta Poniszewska-Marañda, Jure Šuligoj, Alcidio Jesus, Elisardo Sanchis, *Digital.VET: an innovative approach for teaching and training*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.38, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 215-220, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



flipped/mobile/virtual and augmented reality learning applied to VET sector, which have then be made available in a multilingual handbook (IO1); moreover, the job analysis set out the competence profile of experts in digital and immersive teaching in VET systems (IO2). Another important output of the project is the e-learning course for expert in digital and immersive teaching in VET systems (IO3). Starting from the definition of the training plan/curriculum, the teaching materials and resources have been developed and made available on a platform, which we have developed in the form of written texts, audiovisuals, images and support material. The goal of the course is to make teachers acquire technical knowledge and teaching skills related to a teaching model based on the use of digital, mobile, virtual and augmented tools. Final outputs have been the creation of iDid (IO4), namely an application for virtual and augmented reality teaching that have been available on App Store and Google Play, and the pathway for the assessment and self-assessment of the VET teachers and trainers who adopt digital and immersive teaching methodologies (IO5).

## **2. E-learning anti-elusive platform**

The goal of IO3 is the creation of an e-learning course, with a duration of 60 hours at least, which has been based on the learning outcomes related to the competence units that have been outlined in the profile of the VET teachers and trainers who adopt digital and immersive teaching methodologies (IO2). Starting from the definition of the training plan/curriculum, the teaching materials and resources have been developed and made available on a platform, which we have developed in the form of written texts, audiovisuals, images and support material to go into detail. The required output is an on-line platform, where all the training materials can be uploaded and made available. The platform have been realised according to modern responsive and cross-platform standards. It have to be accessible on Windows, Linux and Mac operating systems. No additional software is required than a modern browser and, of course, an Internet connection. At the end of the course, a certificate will be issued to each participant to attest the success of the learning course.

The custom made Digital.VET e-learning platform was developed in order to provide the e-learning course for expert in digital and immersive teaching in VET systems, based on the training material made during the project. The goal of the course is to make teachers acquire technical knowledge and teaching skills related to a teaching model based on the use of digital, mobile, virtual and augmented tools. Trainers can upload their courses thanks to an easy content manage system (CMS) and also take the other courses uploaded by all the enrolled trainers. The e-learning platform is fully available on the website at the address <https://www.digitalvethub.com/fad>. On the platform there is the course developed during the project divided in three main modules, some of them are split in more learning units. Each learning unit is made up of a number of topics that is given by the number of hours of the learning. At the moment the course is provided in English and in the national languages of the partners.

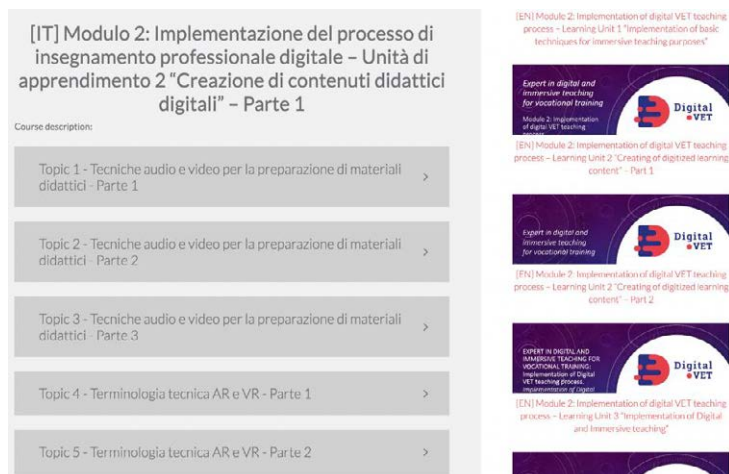


Figure 1. Screenshot of e-learning platform

Within the platform was developed a «personal area». In this section trainers can start to experiment with new technologies such as virtual reality (VR) and augmented reality (AR). Each trainer can upload 360 degrees videos and watch it in an immersive way using a cardboard or they can create ARTags associated to images, videos, GIFs that can be visualized by the provided scanner using mobile devices equipped with camera.

### 3. iDid application solution: app for digital and immersive teaching

The goal of IO4 was the creation of a mobile app for immersive teaching. iDid, is a hybrid cross-platform app accessible from personal computers (desktop, laptop) and mobile devices (available for Android and iOS mobile operating systems). Using the app, VET teachers and trainers are supported in the creation of training contents using virtual and augmented reality and digital technologies.

iDid app is a great breakthrough in the learning field thanks to the power of “i”nnovation, “i”nteraction and “i”mmersion. iDid app allows VET teachers and trainers to:

- Create, visualize and manage courses and training material;
- Upload virtual reality assets by using 360 degrees videos and making them accessible by the use of a cardboard that turns the smartphone into a viewer;
- Upload multimedia assets (images, videos, GIFs, texts) and managing and sharing of ARTags;
- Turn the smartphone into an AR scanner by the use of camera and printed ARTags;
- Manage a community-based system for sharing information and digital courses among teachers with the use of a Virtual Hub. The Virtual Hub allow the search of specific courses and the creation of a personal library by putting like in the favorite courses. As result, it is possible to set up a competition among trainers awarding the most ‘liked’ contents.

The first output produced is an app for Android and iOS smartphones. The app is available on the store of the corresponding operating system (Google Play for Android, App Store for iOS). The app is available in English and also in all partner's languages. There is no need to sign up to use the app: a guest user can navigate in the Digital Hub in order to discover the courses provided by VET trainers and download the materials. As guest user is possible to use the AR scanner using the central button in the bottom menu. When a user chooses a course a main page is showed. In the main page of the course, we can find a cover image, the title of the course and the description. Then, all the digital assets are presented. We can have different kinds of digital assets attached to the

course: documents such as PDF, AR content with the ARTag associated and VR content then can be showed using the smartphone.

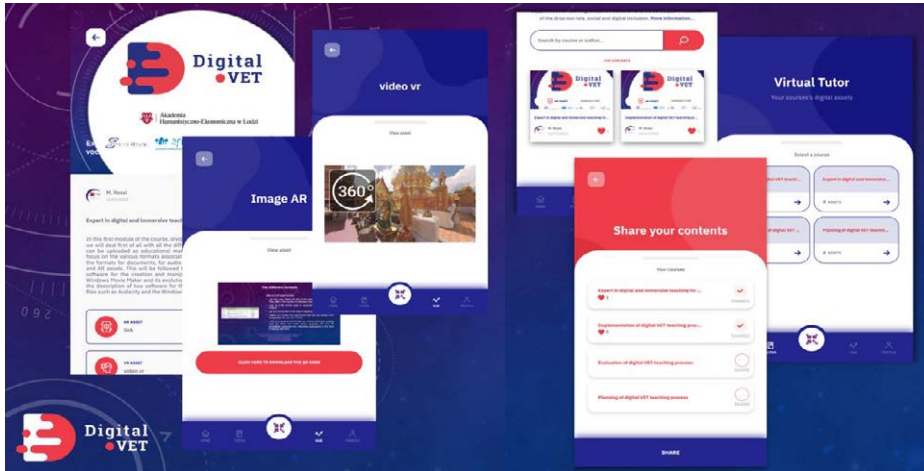


Figure 2. Examples of UI/UX of iDid

After the login, the tutor can view the «top contents» based on his/her likes. Tutor section, also, give the possibility to visualize and test the courses that are created from the tutor console. In this way, each tutor can check how other users can visualize his/her course after the sharing in the Digital Hub. Finally, the logged tutor can share his/her courses to make them public and available on the Digital Hub. Since it is a difficult work with large files for mobile devices, it is provided a tutor console for desktop access. This area can be used by teachers to create their courses and to upload all the digital materials created for the course. The console panel can be also used to manage all the courses created: it is possible to add new materials or edit the information provided within the course.

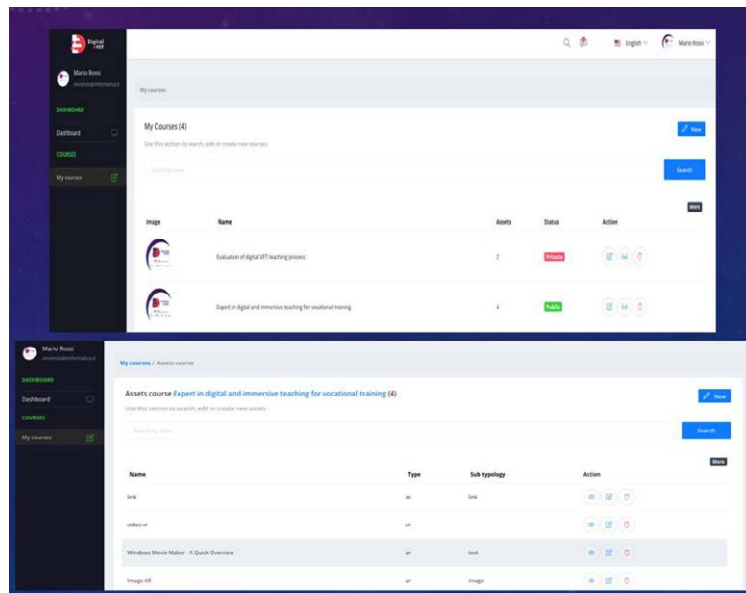


Figure 3. Screenshot of iDid tutor console

From the course page is also possible to manage all the assets such as AR, VR and other kind of documents.

In order to carry out knowledge transfer and testing of acquired skills an innovative paper board was created to experiment with the potentialities of the internet of things (2D barcode), AR tags and NFC (Near Field Communication) together with the iDid app.

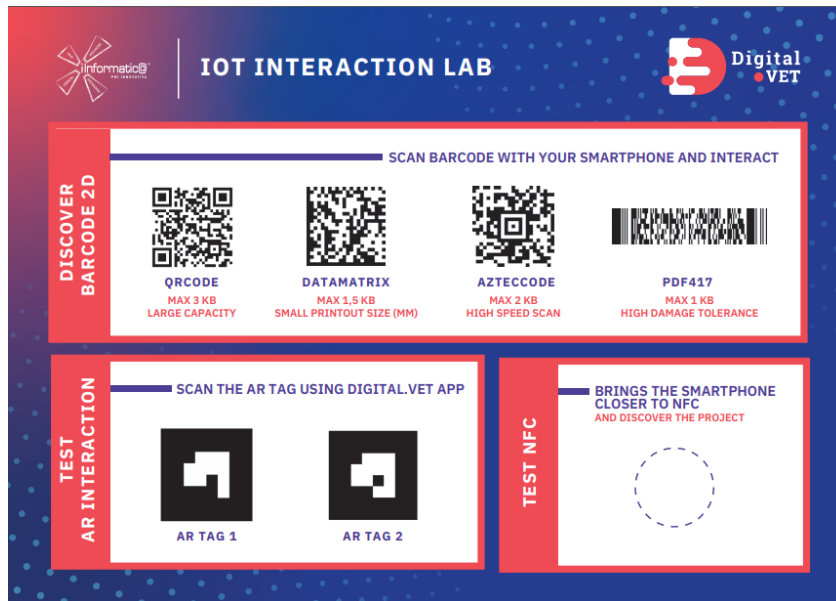


Figure 4. Screenshot of IoT Board for Interaction Lab

#### 4. Nps survey on iDid

In order to understand the level of clarity and quality of the instrument created a questionnaire was organized and administered during the concluding event presenting the project to about 100 people, with a multifaceted and distributed age range (over 18), of training professionals, teachers, IT professionals, former teachers and staff employed in institutions in the area. The results were then evaluated in NPS (Net Promoter Score) terms to understand the "word of mouth" effect expected from the event presentation.

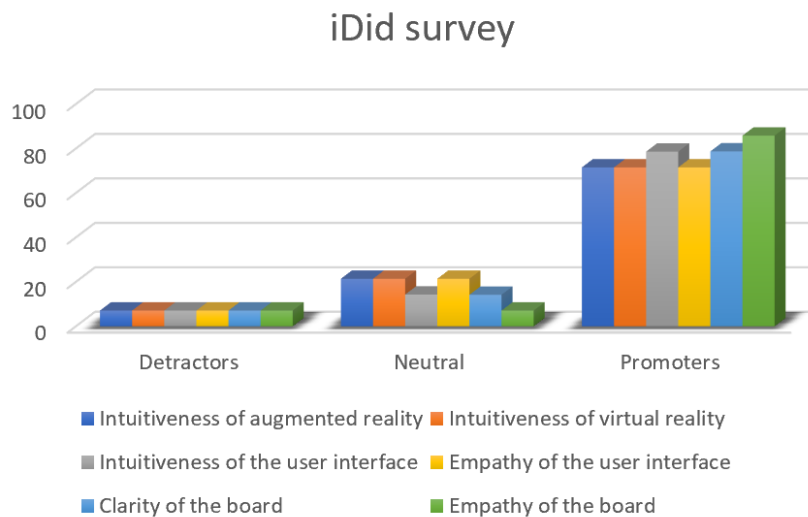
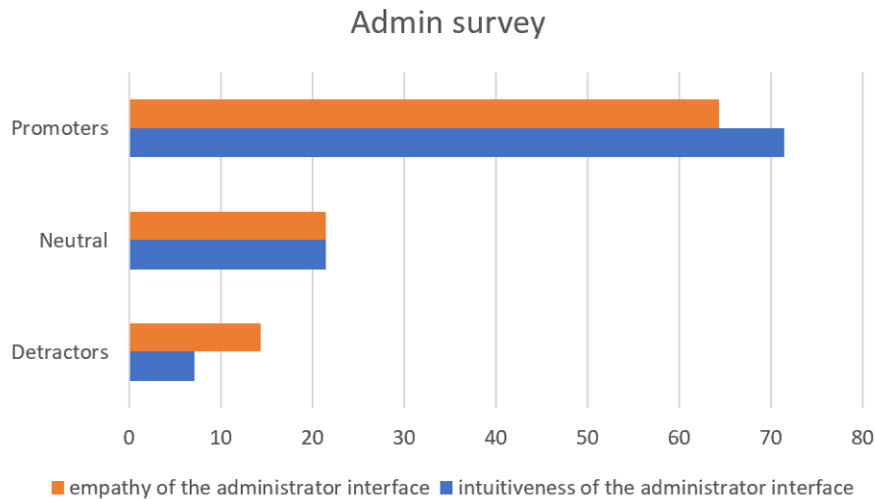


Figure 5. Data Analysis of iDid and board survey



*Figure 6. Data Analysis of admin survey*

From the analysis of the data, it appears a nearly homogeneous behavior of detractors and promoters of the application in terms of the intuitiveness and empathic design (with NPS score above a score of 64) and of the board (NPS scoring above a score of 71.5). This confirms the good performance of the iDid application. Relative to the administrator dashboard, we note an NPS score in terms of intuitiveness of 64.4 while the empathy of the technical interface is lower than the user interface (50 versus 64.4).

## 5. Conclusion

This paper introduced the concept of training and the innovative lesson approach with the use of VR and AR technologies. Digital.VET opens a new path for flipping classroom approach and for a revolution in the teaching experience. We hope that this paper can be a guide to follow for the implementation of new training courses in our countries.

## References

- Cattaneo, A., Antonietti C., Rauseo M. (2022). How digitalised are vocational teachers? Assessing digital competence in vocational education and looking at its underlying factors. *Computers & Education*, Volume 176.
- Morselli, D. (2019). *The Change Laboratory for Teacher Training in Entrepreneurship Education. A New Skills Agenda for Europe*. SpringerBriefs in Education.
- Soldi, R., Cavallini, S., Friedl, J., Volpe, M., Zuccaro, P. C. (2016). *A New Skills Agenda for Europe*. European Committee of the Regions.

# The joint estimation of accuracy and speed: An application to the INVALSI data

Luca Bungaro, Marta Desimoni, Mariagiulia Matteucci, Stefania Mignani

## 1. Introduction

In recent years, the implementation of computer based testing (CBT) has been receiving a growing interest because of its operational advantages. CBT allows to automatically collect data not only on the students' response accuracy (RA) based on item responses, but also on their response times (RT). Using the RTs, the assessment results can be further improved in terms of precision, fairness, and minimizing costs. The information obtained by RTs can be used for item calibration, test design, detection of cheating, and adaptive item selection.

The RTs used to respond to items provide information about working speed, where RA data provide information about ability. RTs are collected for estimating speed and item time-intensity (i.e., population-average amount of time needed to complete an item), to investigate relationships with speed components and accuracy, but also to investigate several issues in educational testing.

In Italy, the National Institute for the Evaluation of the Education and Training System (INVALSI) every year administers standardized tests via CBT to students attending grades 8, 10, and 13. In this study, we use the 2018 mathematics data for grade 10 to estimate the ability and speed of students and to evaluate the impact of some students' characteristics both to the performance and to the response time behaviour.

In the INVALSI test the number of involved examinees is very large and tests must be administered in multiple sessions and locations. Moreover, testing organizations need to produce several test forms to overcome security concerns, such as cheating and leaking of information. For grade 10, multiple test forms with prespecified characteristics are assembled from a Rasch item bank through automated test assembly.

The tests are administered to the whole student population, around 500,000 students. INVALSI also builds a random sample of around 41,000 units. The sampling procedure is a two-stage with Italian geographical region and school track stratification at the first stage. The units of the first stage are the schools and the units of the second stage are the classes. In this paper we analyse the results of the sample. Noteworthy, the INVALSI computer-based tests are conceptualized as power tests, not as speed tests. INVALSI imposes a time limit of 90 minutes on grade 10 tests, which is considered enough for students to read and answer all the questions<sup>1</sup>. These time constraints may have had an impact on the speed that must be considered in the results' discussion.

In the first step of the analysis, we implemented the fully Bayesian approach of Fox *et al.* (2021), following the models of van der Linden (2007) and Klein Entik *et al.* (2009). In the second step, considering the hierarchical nature of the data, we use the estimated mathematics ability and speed in a bivariate multilevel model, where the first-level units are represented by students and the second-level units are represented by classes. Covariates such as gender, school type, immigrant status, economic, social, and cultural status, prior achievement, grade retention, student anxiety, class compositional variables, and geographical area are included in the model.

## 2. Methods

The models for estimating the accuracy and speed of students and for investigating the relation

---

<sup>1</sup> Additional time is allowed to students with special needs.

between these outcome variables and a set of predictors are described in the following.

## 2.1 Models for responses and response times

In order to estimate the accuracy and speed of students, we followed the approach of Fox *et al.* (2021), who implemented in the R package `LNIRT` the models of van der Linden (2007) and Klein Entik *et al.* (2009). In particular, once the data on RA, i.e. correct/incorrect response, and RTs are collected for each item, they are modelled following a Bayesian joint model with a hierarchical structure that, at the first level, defines separate models for responses and response times. At the second level, a distributional structure is defined for the model parameters and hyperprior distributions are specified for the parameters.

At level 1, the one-parameter normal ogive (1PNO) model was used to define the mathematical relationship between the probability of response and the person and item parameters as follows

$$P(y_{ik} = 1 | \theta_i, b_k) = \Phi(\theta_i - b_k), \quad (1)$$

where  $y_{ik}$  is the binary response variable taking value 1 when the response is correct and 0 otherwise, with  $i = 1, \dots, N$  test-takers and  $k = 1, \dots, K$  items,  $b_k$  is generally known as the difficulty parameter of item  $k$ ,  $\theta_i$  denotes the ability of test-taker  $i$ , and  $\Phi(\cdot)$  is the normal cumulative distribution function.

Then, a log-normal distribution is used to model the RTs and the log RTs are stored in a  $N \times K$  matrix  $RT$ . In this way, the generic element  $RT_{ik}$  is assumed to be normally distributed as follows

$$RT_{ik} = \lambda_k - \varphi_k \zeta_i + \varepsilon_{ik}, \quad \varepsilon_{ik} \sim N(0, \sigma_{\varepsilon_k}^2) \quad (2)$$

where  $\lambda_k$  is the time-intensity parameter of item  $k$ , representing the population-average time (on a logarithmic scale) needed to complete an item,  $\zeta_i$  is the speed parameter of test-taker  $i$ , representing the constant working speed of that test-taker, as the systematic differences in RTs given  $\lambda_k$ ,  $\varphi_k$  is the time-discrimination parameter of item  $k$ , representing the sensitivity of the item for different speed levels of the test takers. Lastly,  $\varepsilon_{ik}$  is an additional error term that can model variations in RTs that cannot be explained only by the structural mean term, such as when test-takers operate with different speed values, take small pauses during the test, or change their time management.

At level 2, a distributional structure is defined for the level 1 parameters. This structure is defined for both person and item parameters. For the ability and speed, a bivariate normal distribution is defined where, without identification restrictions, the hyperprior for the covariance matrix is an inverse-Wishart distribution. In the same way, a multivariate normal distribution is specified for all the item parameters of the response and response-time models, where a normal inverse-Wishart distribution is chosen as hyperprior for the mean vector and the covariance matrix.

Model parameters are estimated through the Gibbs sampling algorithm, where parameters are divided into blocks, and the simulation procedure works by iterative sampling of the conditional posterior distributions of the parameters in each block given the previous draws for the parameters in all other blocks. To identify the model, some restrictions are imposed, both for person and item parameters. As regards the item parameters, the product of the time discrimination is fixed to one  $\prod_k(\varphi_k) = 1$ . For the person parameters, the mean of the ability is fixed to zero, as well as the mean of the speed. In this way, the `LNIRT` package is able to avoid restricting the variance of a person parameter, which would otherwise have resulted in the restriction of the covariance matrix (for the details on model estimation and identification, see Fox *et al.*, 2021).

## 2.2 Bivariate multilevel model

Predictors of students' speed and ability were investigated through bivariate multilevel modelling (MLM), which explicitly recognizes potential correlations between the outcomes and the

hierarchical data structure. Following Rasbash *et al.* (2017), bivariate MLMs were specified by treating the individual student as a level 2 unit ( $n = 35,727$ ) and the within-student measurements (Ability and Speed) as level 1 units. Students ( $n = 243$ ) with missing values in the covariates have been excluded from the MLMs data. In the INVALSI database, students are clustered into classes, which were specified in the MLMs as level 3 units ( $n = 2,273$ ). In turn, classes are nested into schools. However, since in the INVALSI national sample a maximum of two classes are sampled within each school, we preferred to not fit a four-level model also including the school level. Therefore, in our models, the class-level random effects collected the unobserved contextual factors at class and higher hierarchical levels.

To enhance the interpretability of the results, we standardized the continuous covariates and the dependent variables (Rasch ability estimate and person speed estimate from LNIRT). The following bivariate MLMs were fitted to the data by Iterative Generalised Least Squares using MLwiN version 3.05 (Charlton *et al.*, 2020).

First, we specified a bivariate random intercept empty model (M0), which allowed us to explore the correlations between ability and speed at class and student levels and to investigate how much response variables variation is present at levels 2 and 3. Level 1 existed solely to define the bivariate structure and there was no level 1 variation specified in the bivariate MLMs (Rasbash *et al.*, 2017).

In model M1, we added to M0 the fixed effects of students' sociodemographic characteristics, prior achievement (0 = the final mark at the First-cycle State Leaving Examination is equal or above the national median; 1 = the final mark is below the national median), school career (1 = student repeating one or more grades, 0 = otherwise), and mathematics test anxiety.

In model M2, the following L2 variables were included: class average ESCS and math test anxiety; the percentage of students with an immigrant background, students repeating one or more grades; students with a low final mark at the end of the First-cycle State Leaving Examination.

In the final model (M3), we added the school track (two dichotomous variables: vocational vs lyceum; vocational vs technical institute, reference category = vocational) and the geographical area (4 dichotomous variables, Center vs North-West; Center vs North-East; Center vs South; Center vs South and the Islands; reference category: Center).

The likelihood-ratio (LR) test was used to compare the nested models described above (M1 vs M0; M2 vs M1; M3 vs M2).

### 3. Results

As regards the joint modelling of RA and RTs, the main results for item parameters are summarized in Table 1, which shows mean, minimum, and maximum of the expected a posteriori (EAP) estimates.

Table 1. Item parameters

	Item Difficulty (Rasch Model)	Time Intensity	Time Discrimination	Difficulty Difference (Rasch Model)
Mean	-0.070	4.229	1.175	0.108
Minimum	-2.574	3.114	0.011	0.001
Maximum	2.726	5.151	2.288	0.281

The last column of Table 1 shows the absolute value of the difference between the parameter  $b$ , estimated by the model, and the one obtained during the calibration of the items. Note that the LNIRT package uses the IPNO model (1), while the model assumed for calibration was the Rasch model, also known as the one-parameter logistic (IPL) model. For this reason, to compare the two estimates, it was first necessary to multiply by 1.7 those provided by the package (Fox *et al.*, 2021).

For person parameters, the estimates of ability and speed are given in Table 2.



Table 2. Person parameters

	Person ability	Person speed
Mean	0.000	0.000
Minimum	-2.311	0.611
Maximum	1.946	2.283

The ability follows a normal distribution, while the speed distribution curve is slightly skewed. From the residual analysis, it turns out that the residuals of the response times violate the assumption of log-normal distribution for most items. Following several analyses, it was possible to note that this violation is due to the large number of test-takers (35,970) and the very nature of the INVALSI test.

The correlation matrices for person and item parameters are given in Table 3 and Table 4, respectively. The analysis of these results allows us to say that there is, on average, a positive relationship between the difficulty of the items and their intensity and discriminating power, in terms of time. This means that the most difficult (easy) items are also the ones that discriminate better (worse) and require more (less) time to perform. The negative correlation between time-discrimination and time-intensity, on the other hand, indicates that on average the items that require more (less) time are the ones that discriminate worse (better), but with a very low and not significant magnitude.

Table 3. Item correlation matrix

	Item Difficulty	Time Intensity	Time Discrimination
Item Difficulty	1.000	0.370 (0.000)	0.234 (0.004)
Time Intensity	0.370 (0.000)	1.000	-0.014 (0.436)
Time Discrimination	0.234 (0.004)	-0.014 (0.436)	1.000

Table 4 provides important information about the correlation between the speed and ability of the test-takers (-0.574), which is negative and significant. So, test-takers with a higher (lower) ability tends to be slower (faster).

Table 4. Person correlation matrix

	Person Ability	Person Speed
Person Ability	1.000	-0.574 (0.000)
Person Speed	-0.574 (0.000)	1.000

This result is known in the literature. In particular, it goes to consolidate that hypothesis for which those who are prepared want to engage and show their skills, even during a test that does not directly affect their school average, while those who are less prepared tend to be less interested and more hasty.

Finally, the extreme residual analysis gave the following results: around 15.54% of RT patterns are considered extreme with 95% posterior probability, while for the RA patterns the percentage is 2.19%. When considering the joint pattern (RA and RT), only 0.49% of these are extremes. The residual variance is around 0.488 and the variance in working speed and time intensities are not so small. Therefore, RT outliers only slightly affect the fit of the log-normal distribution, going to confirm what has already been anticipated about the nature of the test itself.

As for the MLM results, M0 shows that the high-ability test-takers worked slower on computer-based items than the low-ability test-takers (within-classes correlation = -.484). The between-classes correlation between speed and ability is higher than the correlation at the student level (-.779). The estimated intraclass correlation coefficients (ICCs) indicate that ability scores of students in the same classroom are correlated (ability: school ICC = .53); a similar result emerges for speed scores (speed: school ICC = .48). Therefore, a multilevel bivariate approach seems to be

appropriate for representing the structure of the data.

Table 5. Likelihood ratio test

Model	-2*Loglikelihood	Comparison	LR $\chi^2$	d.f.	p-value
M0	156145.167				
M1	148206.787	M1-M0	7938.380	14.000	<0.0001
M2	146734.413	M2-M1	1472.374	10.000	<0.0001
M3	145973.716	M3-M2	760.697	12.000	<0.0001

Table 5 summarizes results from LR tests. Results from model comparison suggest M3 as the final model. For the sake of brevity, we will discuss herein only results from M3 (Table 6).

Table 6. Final model parameter estimates

	Ability			Speed		
	Estimate	S.E.	p-value	Estimate	S.E.	p-value
Intercept	0.520	0.050	0.000	-0.330	0.069	0.000
male	0.110	0.008	0.000	0.100	0.009	0.000
student's ESCS	0.002	0.004	0.708	0.018	0.005	0.000
student_repeating_one_or_more_grades	-0.149	0.011	0.000	0.225	0.012	0.000
low prior achievement vs average and high	-0.442	0.008	0.000	0.248	0.010	0.000
math test anxiety	-0.162	0.004	0.000	-0.030	0.005	0.000
second generation immigrant vs native	-0.085	0.016	0.000	0.010	0.018	0.593
first_generation_immigrant vs native	-0.090	0.016	0.000	-0.052	0.019	0.006
Class % of stud. with low prior achievement	-0.007	0.001	0.000	0.004	0.001	0.000
Class % of immigrants	-0.005	0.001	0.000	0.006	0.001	0.000
Class average ESCS	0.211	0.029	0.000	-0.164	0.040	0.000
Class % of students repeating grades	-0.001	0.001	0.203	0.003	0.001	0.008
Class average math test anxiety	-0.046	0.026	0.075	-0.289	0.035	0.000
North West vs Center	0.210	0.028	0.000	-0.169	0.039	0.000
North East vs Center	0.251	0.028	0.000	-0.233	0.039	0.000
South vs Center	-0.259	0.027	0.000	0.159	0.038	0.000
South Islands vs Center	-0.504	0.034	0.000	0.356	0.047	0.000
Liceum vs Vocational	0.106	0.038	0.005	-0.251	0.052	0.000
Technical Inst vs Vocational	0.177	0.027	0.000	-0.371	0.037	0.000
<b>Between-class cov. Matrix</b>						
Variance	0.143	0.005		0.289	0.010	
Covariance (ability / speed)	-0.147	0.006				
<b>Within-class cov. Matrix</b>						
Variance	0.401	0.003		0.529	0.004	
Covariance (ability / speed)	-0.225	0.003				

*Ceteris paribus*, students with low prior achievement are less accurate and spend less time on mathematics items than their peers. A similar pattern of results emerged for the fixed effect of being a student who repeated one or more grades. As for gender, the unique associations with speed and ability are both positive and very similar in size: males are slightly more accurate and work slightly faster than females. Native students outperform students with an immigrant background in ability, and first-generation immigrants work slightly, albeit significantly, slower than the natives. The unique effect of students' ESCS on ability was not statistically significant, whilst a weak, albeit significant, positive effect emerged with speed. Students' self-reported anxiety before and during

the test is negatively related to ability and speed.

After controlling for relevant individual-level predictors, the contextual effect of class ESCS on ability and speed is significant: students from classes with higher ESCS spend more time on items and obtain better results in terms of ability. The percentage of students with an immigrant background is associated with lower ability and higher speed; analogous results emerged for the percentage of students with low prior achievement. Students attending classes with higher average test-related anxiety spend more time on items.

Significant differences in ability and speed also emerged by school tracks and geographical area. Students from the vocational school were less accurate and spend less time on the items than those from the lyceum and technical institute. Students from the North-East and the North-West are more accurate and work slowly on items than those from the Center of Italy, whilst those from the South and the South and Islands were less accurate and spend less time on items.

#### 4. Concluding remarks

The main results show that the ability and speed are inversely proportional, e.g. as ability increases, speed decreases. Also, differences in the students' performance by prior achievement, math test anxiety, sociodemographic characteristics, class compositional variables, school tracks and geographical area are significant for both ability and speed. The various results in this study need to be confirmed through additional research. Some further developments should also focus on the opportunity to include response information in the detection of aberrant response behaviour.

#### References

- Charlton, C., Rasbash, J., Browne, W.J., Healy, M., Cameron, B. (2020). *MLwiN Version 3.05*. Centre for Multilevel Modelling, University of Bristol.
- Fox, J. P., Klotzke, K., & Simsek, A. S. (2021). LNIRT: An R Package for Joint Modeling of Response Accuracy and Times. *arXiv preprint arXiv:2106.10144*.
- Klein Entink, R. H., Fox, J.-P., van der Linden, W. J. (2009). A Multivariate Multilevel Approach to the Modeling of Accuracy and Speed of Test Takers. *Psychometrika*, **74**(1), pp. 21-48.
- Rasbash, J., Steele, F., Browne, W.J., Goldstein, H. (2017). *A User's Guide to MLwiN, v3.00*. Centre for Multilevel Modelling, University of Bristol.
- van der Linden, W. J. (2007). A Hierarchical Framework for Modeling Speed and Accuracy on Test Items. *Psychometrika*, **72**(3), 287.

# Ammonia emissions and fine particulate matter: some evidence in Lombardy

Alessandro Fusta Moro, Matteo Salis, Andrea Zucchi, Michela Cameletti,  
Natalia Golini, Rosaria Ignaccolo

## 1. Introduction

Air quality in the Lombardy region (northern Italy) is affected by high concentrations of pollutants. One of the reasons is that Lombardy is localised in the Po Valley where air circulation is very weak because of the mountains surrounding the area. The peculiar weather conditions, the industrial development and the population density make Lombardy one of the worst European region in terms of air quality<sup>1</sup>. As a result, epidemiological studies have found that Lombardy is characterized by an elevated mortality rate related to fine particulate matter (PM<sub>2.5</sub>) exposure. It is well known that a considerable part (from 10% up to 50%) of the PM<sub>2.5</sub> is formed by the chemical reactions of the ammonia (NH<sub>3</sub>) with other precursors. In the Lombardy region, 97% of the total NH<sub>3</sub> gaseous emissions are linked to the agriculture sector (INEMAR - ARPA Lombardia, 2022). Considering that Lombardy is the leading region in Italy for agriculture production, with the highest regional density of swine and bovines, it is clear that the agriculture section has a considerable impact on air quality.

The project *Agriculture Impact On Italian Air Quality*, hereafter *AgrImOnIA*, aims to estimate the local impact of ammonia emissions on particulate matters (PM<sub>10</sub> and PM<sub>2.5</sub>). This information can be crucial for the policy-makers who have to prioritise interventions. *AgrImOnIA* is an ongoing research project, promoted and financed by Fondazione Cariplo within the framework of *Data Science for science and society*. More information on the project are available on <https://agrimonia.net/>.

In this work, we present preliminary results providing continuous spatial maps of PM<sub>2.5</sub> concentrations (with daily temporal scale) in the Lombardy region using the *AgrImOnIA dataset*<sup>2</sup>, which contains harmonised data on meteorology, emissions and land use. We implement three spatial prediction methods whose performance will be compared by using standard indexes computed with the Leave-One-Out Cross-Validation strategy (LOOCV). In particular, we consider a spatio-temporal Kriging model with external drift, and two random forest algorithms which include spatial and temporal components.

## 2. Data

The *AgrImOnIA dataset* is an open access data set containing Air Quality (AQ), Weather (WE), Land cover (LA), Emission (EM) and Livestock (LI) data with daily temporal resolution. The data are available for all the air quality monitoring stations after a pre-processing step to change the support of spatial data from area to point, when necessary. We consider the period from 2017 to 2020. The area covered by the *AgrImOnIA dataset* includes the Lombardy region

---

<sup>1</sup><https://www.eea.europa.eu/publications/air-quality-in-europe-2021>

<sup>2</sup>The *AgrImOnIA dataset* will soon be available on Zenodo, which is an open repository operated by CERN (<https://zenodo.org/>).

Alessandro Fusta Moro, University of Turin, Italy, [alessandro.fustamoro@unito.it](mailto:alessandro.fustamoro@unito.it), 0000-0003-1129-5038  
Matteo Salis, University of Turin, Italy, [matteo.salis@edu.unito.it](mailto:matteo.salis@edu.unito.it)  
Andrea Zucchi, University of Turin, Italy, [andrea.zucchi925@edu.unito.it](mailto:andrea.zucchi925@edu.unito.it)  
Michela Cameletti, University of Bergamo, Italy, [michela.cameletti@unibg.it](mailto:michela.cameletti@unibg.it), 0000-0002-6502-7779  
Natalia Golini, University of Turin, Italy, [natalia.golini@unito.it](mailto:natalia.golini@unito.it), 0000-0003-4457-5781  
Rosaria Ignaccolo, University of Turin, Italy, [rosaria.ignaccolo@unito.it](mailto:rosaria.ignaccolo@unito.it), 0000-0003-2998-1714

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Alessandro Fusta Moro, Matteo Salis, Andrea Zucchi, Michela Cameletti, Natalia Golini, Rosaria Ignaccolo, *Ammonia emissions and fine particulate matter: some evidence in Lombardy*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.40, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 227-232, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

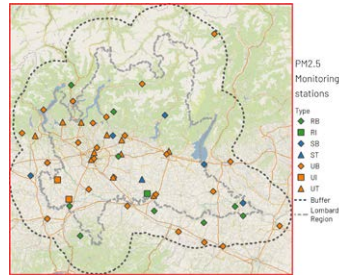


Figure 1: Area of interest with  $PM_{2.5}$  monitoring stations, classified by type: rural background (RB); rural industrial (RI); suburban background (SB); suburban traffic (ST); urban background (UB); urban industrial (UI); urban traffic (UT).

and a neighbouring area defined by applying a 0.3-degree buffer to the regional boundaries. The area and the  $PM_{2.5}$  monitoring network considered by the *AgrImOnIA* project can be seen in Figure 1.

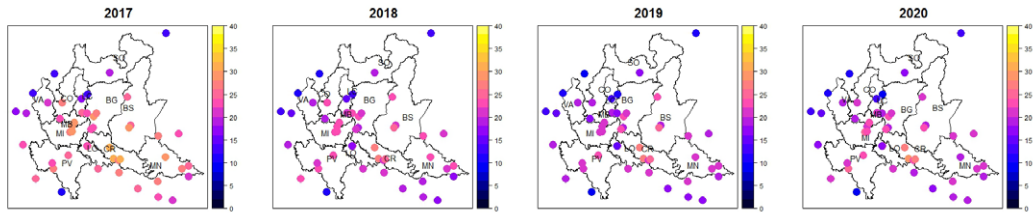


Figure 2:  $PM_{2.5}$  annual mean concentrations ( $\mu g/m^3$ ) for each station from 2017 to 2020.

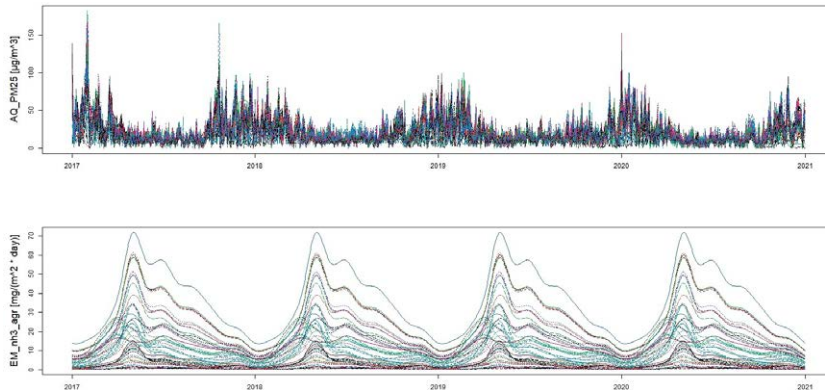


Figure 3: Time series of  $PM_{2.5}$  concentrations (top) and  $NH_3$  agriculture emissions (bottom) for all the monitoring sites

From the AQ available variables, we selected  $PM_{2.5}$  (AQ\_pm25, in  $\mu g/m^3$ ) as the response variable in logarithmic scale. Figure 2 shows the annual mean of  $PM_{2.5}$  concentrations in each monitoring station: higher values are located in the lower Po Valley, particularly in the provinces of Milan, Cremona, Lodi and Brescia. The other selected variables are described in Table 1. The overall  $NH_3$  emissions from the agriculture sector (nh3\_agr) are calculated by summing up  $NH_3$  emissions from manure management, agriculture soil and agriculture waste burning ( $EM.nh3.livestock.mm + EM.nh3.agr.soils + EM.nh3.waste.burn$ ). To generate

continuous maps as output, the covariates are also obtained on a regular grid  $0.1^\circ \times 0.1^\circ$ . Figure 3 displays the  $PM_{2.5}$  and  $NH_3$  daily time series for all the monitoring stations. As it can be seen,  $PM_{2.5}$  concentrations follow a seasonality with peaks during the winter, while ammonia shows higher values in summer, likely because of a higher uses of fertilisers.

Table 1: Description of the selected variables

Variable description	Variable name	Unit	Source
Longitude of the monitoring stations	Longitude	<i>degree</i>	AgrImOnIA dataset
Latitude of the monitoring stations	Latitude	<i>degree</i>	AgrImOnIA dataset
Date	Time	Date	AgrImOnIA dataset
Altitude of the monitoring stations	Altitude	<i>m</i>	AgrImOnIA dataset
$PM_{2.5}$ concentrations	AQ_pm25	$\mu g/m^3$	AgrImOnIA dataset
Temperature of air at 2 m	WE_2m_temperature	$^\circ C$	AgrImOnIA dataset
Mean horizontal wind speed at 10 m	WE_wind_speed_10m_mean	<i>m/s</i>	AgrImOnIA dataset
The accumulated water fallen	WE_tot_precipitation	<i>m</i>	AgrImOnIA dataset
The pressure of the atmosphere	WE_surface_pressure	<i>Pa</i>	AgrImOnIA dataset
Net solar radiation	WE_solar_radiation	$J/m^2$	AgrImOnIA dataset
High vegetation index	WE_hvi	$m^2/m^2$	AgrImOnIA dataset
Low vegetation index	WE_lvi	$m^2/m^2$	AgrImOnIA dataset
Mean horizontal wind speed at 100 m	WE_wind_speed_100m_mean	<i>m/s</i>	AgrImOnIA dataset
Maximum of boundary layer height	WE_blh_layer_max	<i>m</i>	AgrImOnIA dataset
Minimum of boundary layer height	WE_blh_layer_min	<i>m</i>	AgrImOnIA dataset
Maximum of relative humidity	WE_rh_max	%	AgrImOnIA dataset
$NH_3$ emissions - manure management	EM_nh3_livestock_mm	$mg/m^2$	AgrImOnIA dataset
$NH_3$ emissions - agriculture soil	EM_nh3_agr_soils	$mg/m^2$	AgrImOnIA dataset
$NH_3$ emissions - agriculture waste burning	EM_nh3_waste_burn	$mg/m^2$	AgrImOnIA dataset
$NH_3$ total emissions	EM_nh3_sum	$mg/m^2$	AgrImOnIA dataset
$SO_2$ total emissions	EM_so2_sum	$mg/m^2$	AgrImOnIA dataset
$NO_x$ total emissions	EM_nox_sum	$mg/m^2$	AgrImOnIA dataset
$NH_3$ emissions - agriculture sector	nh3_agr	$mg/m^2$	Own elaboration
Day of the week	day_week	Categorical	Own elaboration
Type of season	season	Categorical	Own elaboration
Type of station	type_station	Categorical	Own elaboration

(see Figure 1)

### 3. Spatial prediction techniques

In order to perform spatial prediction and to produce continuous spatial maps of  $PM_{2.5}$  concentrations, we consider two approaches: 1) spatio-temporal kriging with external drift (STKED), a classical approach in geostatistics framework; 2) a well-known machine learning method - random forest (RF) - extended to the case of data correlated in space and time.

#### Spatio-temporal kriging with external drift

Spatio-temporal kriging is a supervised parametric model which assumes that the observed  $PM_{2.5}$  data are generated by a given stochastic spatio-temporal model. In particular, we suppose that the response variable  $\log(AQ\_pm25)$  is Normally distributed with a mean changing in space and time and a variance given by the measurement error variance (i.e. the nugget). The mean of the response field is in turn defined as the sum of a large-scale trend (or external drift, which includes the linear effect of the covariates described in Table 1), and a residual spatio-temporal process with separable space-time covariance function. For the implementation of this method we use the `gstat` R-package (Gräler et al., 2016), which requires the estimation of a

spatio-temporal variogram. Once all the models parameters are estimated, spatial prediction of the expected  $\log(\text{AQ}_{\text{pm}25})$  value at any location in Lombardy region is straightforward: it is given basically by a local weighted mean of the spatio-temporal residuals plus the large scale component (evaluated using the covariate values at the new sites).

### Random forest for spatio-temporal data

Random forest is a data-driven non-parametric machine learning technique given by an ensemble of regression trees fitted using several bootstrapped version of the original data and subsets of the considered covariates. The main limitation of random forest is that it is not able to take directly into account the temporal and spatial correlation, as kriging does. In order to include in the fitting algorithm some information about the data spatial structure, we propose here two different implementations of the method: 1) the standard RF algorithm (RFbase) which includes in the covariate set, besides the variables described in Section 2, also the spatial coordinates (longitude and latitude) of each observation; 2) the spatial RF (RFsp) method proposed by Hengl et al. (2018). This method expands the set of covariates by including the buffer distances from the observation sites (i.e., if we have  $n$  monitoring stations we will have  $n$  additional columns in the training set each referring to a given station and including the distances from the remaining locations). To take into account the temporal component we consider as covariates the date of the day, the day of the week and the type of season. The two RF algorithms are implemented using the *Ranger* R-package. Prediction in a new spatial location is usually given by the averages of the, say,  $B$  predictions computed using the single trees in the forest. Indeed, we consider an ordinary spatio-temporal kriging model for the differences between observed and predicted data in order to include a term taking into account spatio-temporal correlation and predict a term for the small scale component.

## 4. Preliminary results

Starting considering the STKED technique, we subset covariates through stepwise strategy and we estimate the coefficients shown in Table 2 also referred to interaction terms between season and emissions. We can see that, among the emissions, `nh3_agr` has a larger impact on  $\log(\text{AQ}_{\text{pm}25})$  during the winter, while `EM_nox_sum` shows larger effects in the remaining seasons; this is consistent with the results in Thunis et al. (2021). The sample variogram of the residuals of the large-scale component is used to choose the exponential variogram model.

Figure 4 (right) shows the 2020 mean of the daily predicted  $\text{PM}_{2.5}$  concentrations in the area of interest. It is worth to note that higher  $\text{PM}_{2.5}$  concentrations are predicted where we observe higher  $\text{NH}_3$  emissions from the agriculture sector, as shown in Figure 4 (left).

As for the ML approach, the variable importance analysis returns similar results for RFbase and RFsp. Figure 5 shows the weather components as the most important, in accordance with the literature (Cameletti et al., 2011) together with the temporal components. Moreover, it turns out that the euclidean distances between sites (`dist_from_`) is not very important.

The comparison between the prediction capability of the three models is performed through LOOCV and the results are shown in Table 3. STKED shows higher performance compared to the two versions of RF, although it is worth to note that these results are based on a preliminary version of the *AgrImOnIA* dataset.

## 5. Discussion and further development

Further developments of this work will consider the forthcoming versions of the *AgrImOnIA* dataset and extensions of the considered techniques, always in the framework of the comparison

Table 2: Large-scale component coefficient estimates of STKED

Covariate	Dependent variable: log(AQ_pm25)	
	Estimate	SE
Altitude	-0.152***	0.003
Latitude	-0.032***	0.003
Longitude	-0.022***	0.003
WE_sum_total_precipitation	-0.119***	0.003
WE_mean_wind_speed_10m	-0.142***	0.003
WE_min_boundary_layer_height	-0.082***	0.003
seasonautumm	-0.149***	0.008
seasonspring	-0.411***	0.008
seasonwinter	0.600***	0.010
EM_so2_sum	-0.038***	0.007
EM_nox_sum	0.167***	0.007
nh3_agr	0.111***	0.008
seasonautumm:EM_so2_sum	-0.017*	0.010
seasonspring:EM_so2_sum	0.003	0.010
seasonwinter:EM_so2_sum	0.022**	0.009
seasonautumm:EM_nox_sum	0.027**	0.013
seasonspring:EM_nox_sum	-0.116***	0.016
seasonwinter:EM_nox_sum	-0.077***	0.008
seasonautumm:nh3_agr	-0.130***	0.009
seasonspring:nh3_agr	-0.078***	0.009
seasonwinter:nh3_agr	0.195***	0.016
Constant	2.774***	0.005
Observations	70,119	
R <sup>2</sup>	0.325	
Adjusted R <sup>2</sup>	0.325	
Residual Std. Error	0.664 (df = 70097)	
F Statistic	1,607.042*** (df = 21; 70097)	

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

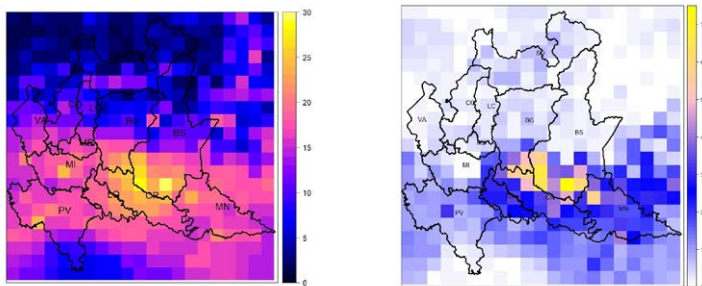


Figure 4: Mean of daily NH<sub>3</sub> emissions from agriculture over the period 2017-2020 (left); mean of daily PM<sub>2.5</sub> concentrations predicted by STKED for 2020 (right).

Table 3: LOOCV comparison of the models described in Section 3.

	MAE	RMSE	BIAS	COR
STKED	0.316	0.510	0.021	0.782
RFbase	0.352	0.543	-0.022	0.741
RFsp	0.361	0.547	-0.041	0.738



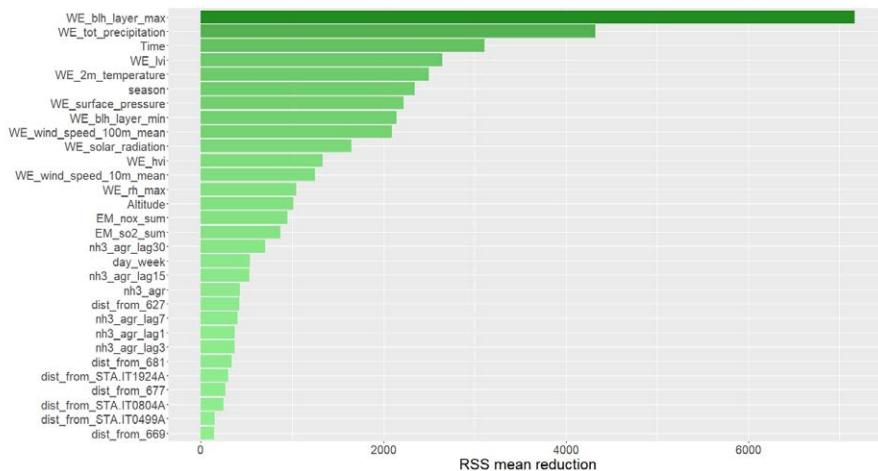


Figure 5: Variable importance of the spatial random forest (RFsp) measured by the RSS mean reduction for each variable.

between classical geostatistics and machine learning approaches.

Recent studies found that NH<sub>3</sub> emissions reductions are the most cost-effective way to reduce PM<sub>2.5</sub> concentrations (Gu et al. 2021). Scenario analysis based on spatial prediction techniques, in compliance with the Regional Plane for emissions reductions<sup>3</sup>, will allow to assess the expected impact of NH<sub>3</sub> emissions reduction's policies before their implementation.

## References

- Cameletti, M., Ignaccolo, R., and Bande, S. (2011). Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics*, 22(8):985–996.
- Gräler, B., Pebesma, E., and Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, 8(1):204–218.
- Gu, B., Zhang, L., Dingenen, R. V., Vieno, M., Grinsven, H. J. V., Zhang, X., Zhang, S., Chen, Y., Wang, S., Ren, C., Rao, S., Holland, M., Winiwarter, W., Chen, D., Xu, J., and Sutton, M. A. (2021). Abating ammonia is more cost-effective than nitrogen oxides for mitigating PM<sub>2.5</sub> air pollution. *Science*, 374(6568):758–762.
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., and Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6:e5518.
- INEMAR - ARPA Lombardia (2022). *INEMAR, Inventario Emissioni in Atmosfera: emissioni in Regione Lombardia nell'anno 2019 - versione in revisione pubblica*. ARPA Lombardia Settore Monitoraggi Ambientali. <https://www.inemar.eu/xwiki/bin/view/InemarDatiWeb/Inventario+delle+emissioni+in+atmosfera>.
- Thunis, P., Clappier, A., Beekmann, M., Putaud, J. P., Cuvelier, C., Madrazo, J., and de Meij, A. (2021). Non-linear response of PM<sub>2.5</sub> to changes in NO<sub>x</sub> and NH<sub>3</sub> emissions in the Po basin (Italy): consequences for air quality plans. *Atmospheric Chemistry and Physics*, 21(12):9309–9327.

<sup>3</sup><https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/istituzione/direzioni-generalidi/direzione-generale-ambiente-e-clima/piano-regionale-interventi-qualita-aria-pria>

# On the utility of treating a vineyard against *Plasmopara viticola*: a Bayesian analysis

Lorenzo Valleggi, Federico Mattia Stefanini

## 1. Introduction

*Plasmopara viticola* is the causal agent of the downy mildew, the most severe disease of the grapevine leading to economic damages (Wong et al., 2001). In order to prevent downy mildew, fungicide treatments are required, but they are dangerous for the environment and human health (Kab et al., 2017). Optimal scheduling and selection of treatments is the key to managing downy mildew in an eco-friendly way (Chen et al., 2020). This goal is quite difficult to achieve due to the variability shown by downy mildew among years. Indeed *Plasmopara viticola* growth mostly depends on variables like temperature and rain, plant's genotype and soil conditions. The latter are usually assumed to be homogeneous in the considered vineyard, possibly because of the difficulty in obtaining local measurements. Meteorological variables are typically measured at whole-field levels, despite that *Plasmopara viticola* growth depends on microclimate (Bove et al., 2020a). Simulations of the key steps in the biological process of the pathogen have been performed to obtain information about airborne sporangia, sporangia availability, relative severity and number of lesions in secondary infection cycles (Brischetto, et al., 2021) (Bove et al., 2020b). Unfortunately these important deterministic models do not also provide information on the variability of the above attributes describing events related to the infection.

In this work, we propose a Bayesian prior-predictive approach (Gelman, et al., 2017) where future environmental conditions and the probability of infection both depend on the selected treatment. A multi-attribute utility function taking the three most important variables as argument has been elicited to describe the utility of consequences following the decision to treat the vineyard (Lavik, et al., 2020): the expected values under alternative decisions enable the winemaker to take the optimal decision of treating the vineyard or not.

## 2. Methods

In this section the approach followed to support the decision maker is described.

### 2.1 Scenarios

In this study intervals of temperature values and of humidity promoting the disease were defined by exploiting the information available in the literature. The following scenarios were defined: (i) a temperature favorable for pathogen's growth but not for humidity, (Temperature  $> 10^{\circ}\text{C}$  and  $< 30^{\circ}\text{C}$ , Humidity  $\leq 0.8$ ) labeled as "Useful, N-Useful"; (ii) a temperature not favorable for pathogen's growth and a favorable humidity (Temperature  $< 10^{\circ}\text{C}$  or  $> 30^{\circ}\text{C}$  Humidity  $\geq 0.8$ ), labeled as "N-Useful, Useful"; (iii) a temperature and humidity both favorable for pathogen's growth, labeled as "Useful, Useful" (Temperature  $> 10^{\circ}\text{C}$  and  $< 30^{\circ}\text{C}$ , Humidity

Lorenzo Valleggi, University of Florence, Italy, lorenzo.valleggi@unifi.it, 0000-0002-8529-3046

Federico Mattia Stefanini, University of Milan, Italy, federico.stefanini@unimi.it, 0000-0003-4248-6275

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Lorenzo Valleggi, Federico Mattia Stefanini, *On the utility of treating a vineyard against Plasmopara viticola: a Bayesian analysis*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.41, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 233-237, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

$\geq 0.8$ ); (iv) neither temperature nor humidity favorable for pathogen's growth (Temperature  $< 10^\circ\text{C}$  or  $> 30^\circ\text{C}$  with Humidity  $\leq 0.8$ ), labeled as "N-Useful, N-Useful". Given that scenario  $e_j$  ( $j \in \{1, 2, 3, 4\}$ ) is realized in the vineyard, the expert must take the decision "to treat",  $a_1$ , or "not to treat",  $a_0$ .

## 2.2 States, actions, consequences

Expected values of the probability  $\pi_{i,j}$  of infection for one leaf sampled from the vineyard given each environmental scenario  $e_j$  and decision  $a_i$ ,  $i \in \{0, 1\}$ , were elicited under the assumption that all of these combinations of temperature and humidity lasted from dawn to sunset just before taking the decision. After assuming that  $(\pi_{i,j} \mid e_j, a_i) = \text{Beta}(\alpha_{i,j}, \beta_{i,j})$ , the values of model parameters  $\alpha_{i,j}$  and  $\beta_{i,j}$  were defined for each pair scenario-treatment  $i, j$  by fitting a Beta distribution to the elicited quantile 0.9 and the elicited expected value of  $\pi_{i,j}$  given  $a_i, e_j$ , i.e. pairs made by an action and a temperature-humidity scenario (Table 1). The implied credible intervals were checked by the expert (Table 1) without finding any need of refinement.

Higher levels of variability characterize the prior-predictive distribution under no chemical treatment ( $a_0$ ) in comparison to the decision of treating ( $a_1$ ). In Table 1, the expected value of the probability of infection is shown for each scenario,  $p(\pi_{t+1} \mid a_i, e_j)$ , together with other elicited quantities.

Table 1: Elicited expected values of the probability of infection in the considered scenarios; "Useful" ("N-Useful") means able (unable) to produce the infection; T=Temperature and H=Humidity.

Treatments $\{a_0, a_1\}$	Scenarios $e_1, \dots, e_4$		Probability $E[\pi_{i,j}]$	Credibility Interval: 0.8	Parameters $(\alpha_{i,j}, \beta_{i,j})$
	T	H			
0	Useful	N-Useful	0.75	(0.67296, 0.80032)	(40.50, 13.50)
0	N-Useful	Useful	0.70	(0.62413, 0.74968)	(43.17, 18.50)
0	N-Useful	N-Useful	0.06	(0.00066, 0.10263)	(0.19, 3.00)
0	Useful	Useful	0.80	(0.72362, 0.84969)	(38.00, 9.50)
1	Useful	N-Useful	0.50	(0.46957, 0.52000)	(221.50, 221.50)
1	N-Useful	Useful	0.40	(0.3696, 0.42000)	(169.33, 254.00)
1	N-Useful	N-Useful	0.10	(0.06991, 0.12001)	(14.89, 134.00)
1	Useful	Useful	0.30	(0.2696365, 0.32002)	(50, 112.50)

Two attributes were defined to quantify the impact of a selected treatment on soil and biodiversity of the vineyard at the subsequent time point  $t + 1$  (e.g. next week) after the decision-action:

- $s_{t+1}$  : a score that classifies the degree of cleanness of soil after chemical treatment (including derived side products),  $\Omega_{s_{t+1}} \in \{1, 2, \dots, 5\}$ , where  $s_{t+1} = 1$  for the worst state after 10 years from treatment, and  $s_{t+1} = 5$  for the cleanest case after 10 years;
- $b_{t+1}$  : a biodiversity score to classify the degree of biological diversity,  $\Omega_{b_{t+1}} \in \{1, 2, \dots, 5\}$ , thus  $b_{t+1} = 1$  refers to the worst state of biological diversity after 10 years from treatment and  $b_{t+1} = 5$  is the best diversity class after 10 years from treatment.

Given that the winemaker is willing to consider the two attributes on equal footing, a value function averaging and rescaling biodiversity and soil scores was considered as an environmental summary of the future state:  $f_{s,b,t+1} = ((s_{t+1} + b_{t+1})/2 - 1)/4$ , with  $\Omega_{s,b} = [0, 1]$ . In order to recognize the inherent uncertainty of  $f_{s,b,t+1}$ , a prior distribution was elicited by restricting

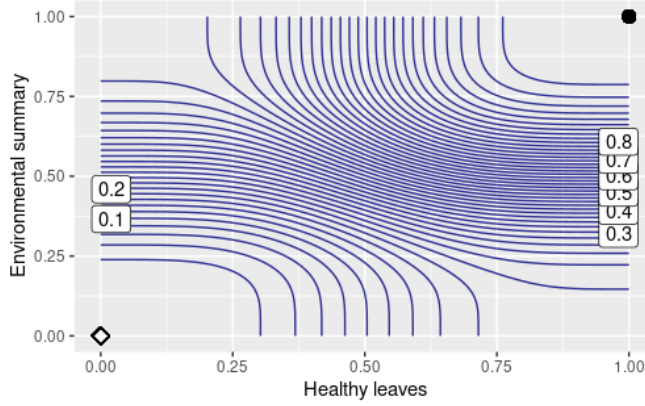


Figure 1: Contour plot of the utility function.

the attention to the decision of treating,  $p(f_{s,b,t+1} | a_1) = \text{Beta}(\phi_1, \phi_2)$ , because the decision of no treatment  $a_0$  is associated with no change of biodiversity and nor of soil: a degenerate probability distribution follows under  $a_0$ . For this reason the value of  $f_{s,b,t}$  was also calculated at the time of decision, thus  $p(f_{s,b,t+1} | a_0) = I_{f_{s,b,t}}(f)$ . The elicited value of the two parameters is  $\phi_1 = 57$ ,  $\phi_2 = 22$ , thus the treatment has a medium impact on the environment (quantile 0.1 of  $f_{s,b,t}$  is 0.6559175; quantile 0.9 of  $f_{s,b,t}$  is 0.7846756). Hereafter, the probability of healthy leaves  $\tilde{\pi}_{i,j} = 1 - \pi_{i,j}$  will be considered in the utility function.

Under conditional independence of future attributes, the prior predictive distribution is

$$p(f_{s,b,t+1}, \tilde{\pi}_{i,j} | f_{s,b,t}, \phi_1, \phi_2, e_j, a) = \text{Beta}(\tilde{\pi}_{i,j} | \alpha_{i,j}, \beta_{i,j}) \cdot [\text{Beta}(f_{s,b,t+1} | \phi_1, \phi_2) I_1(a) + I_{f_{s,b,t}}(f) I_0(a)] \quad (1)$$

thus the expected value of the utility function  $U(f_{s,b,t+1}, \tilde{\pi}_{i,j})$  is

$$E[U(f_{s,b,t+1}, \tilde{\pi}_{i,j}) | a_i, e_j] = \int_{\theta} U(f_{s,b,t+1}, \tilde{\pi}_{i,j}) p(f_{s,b,t+1}, \tilde{\pi}_{i,j} | f_{s,b,t}, \phi_1, \phi_2, e_j, a_i) d\theta$$

where  $\theta$  is the vector of all model parameters. In the following, the current value of environmental summary is  $f_{s,b,t} = 1$  under  $a_0$ , i.e. a fully unmodified environment is in place.

### 2.3 Elicitation of the utility function

An utility function was elicited with arguments the environmental summary and the probability of healthy leaves: under mutually utility independence (French et al., 2000) (Keenye et al., 1993):

$$U(f_{s,b,t+1}, \tilde{\pi}_{i,j}) = k_1 U_1(f_{s,b,t+1}) + k_2 U_2(\tilde{\pi}_{i,j}) + k_1 k_2 U_1(f_{s,b,t+1}) \cdot U_2(\tilde{\pi}_{i,j})$$

where  $k$  satisfies  $1 + k = \prod_{r=1}^2 (1 + k_r)$ ;  $U_i(x_i) = \int_0^{x_i} \text{Beta}(z | \psi_{1,i}, \psi_{2,i}) dz$ ,  $i = 1, 2$  are marginal utility functions which depend on parameters  $\psi_{1,i}$  and  $\psi_{2,i}$ ; the best  $x_i^*$  and worst  $x_i^0$  cases take value equal to 1 and 0 respectively; the weights are elicited so that  $k_1 = u(f_{s,b,t+1}^*, \tilde{\pi}_{i,j}^0)$  is the utility value associated to the best value for the environmental summary and the worst value for the probability of a healthy leaf; similarly,  $k_2 = u(\tilde{\pi}_{i,j}^*, f_{s,b,t+1}^0)$  is the utility value associated to the best value for the probability of a healthy leaf and the worst for the environmental summary. After eliciting  $U_1$  and  $U_2$  a graphical exploration was performed with the expert to check for the need of refinement (Figure 1). The optimal decision  $a^\dagger$  under condition  $e_j$  follows from the expected values of the utility function:  $a^\dagger = \arg \max_{i \in \{0,1\}} E[U(f_{s,b,t+1}, \tilde{\pi}_{i,j}) | a_i, e_j]$ .

### 3. Results

The expected values of the utility function were computed for each scenario as described in the previous section. In Table 2 the main results are shown.

By comparing the different scenarios under different decisions, it was found that for  $e_1 =$  "Useful N-Useful", the expected utility was higher in the "not treat" case ( $a = 0$ ), than "treat" case; when  $e_2 =$  "N-Useful Useful", the expected utility was higher in the "treat" case ( $a = 1$ ), than "not treat" case; for  $e_3 =$  "N-Useful N-Useful", the expected utility was higher in the "not treat" case ( $a = 0$ ), than "treat" case; finally, when  $e_4 =$  "Useful Useful", the expected utility was higher in the "treat" case ( $a = 1$ ), than "not treat" case.

### 4. Discussion and conclusion

Optimal scheduling and managing of treatments is a way to reduce the environmental impact of agriculture. This goal is quite challenging while dealing with phytopathogens that have high infectious potential and that may produce extensive and severe damage. *Plasmopara viticola*, the main enemy of viticulture, is one of these phytopathogens requiring the adoption of highly tuned prevention strategies. The wide adoption of treatments based on copper and sulphuric compounds is leading to over-accumulation in the soil, especially of copper, which causes a phytotoxic effect on the grapevine. They also have a negative impact on biodiversity by reducing the number of species and weakening the ecosystem in the long term.

The optimal decision about treatment with chemicals rests on the available (prior) information about the risk of infection at decision time, the probability of observing a healthy leaf after treatment and the expected impact on the environment. The availability of data collected in the vineyard of interest is the natural next step to improve the performance of the decision process by better calibrating expectations and beliefs: here the advent of low cost sensors for oospores could lead to decisions taken for local microenvironments. Furthermore, agronomist's preference scheme over prospects coded into the elicited utility function is crucial in order to define a trade-off between environmental sustainability and yield, both for quantity and quality. Here the four most fundamental scenarios of climatic conditions have been considered but a multi value discrete scale on more intervals for several other variables could increase the resolution of the description, when needed. Similarly, a direction for further research could be a more detailed description of both environmental changes and end products, grapes, by choosing key chemical components required to produce high valued wine.

Table 2: Expected values of the utility function for each scenario considered; "Useful" ("N-Useful") means able (unable) to produce the infection; T=Temperature and H=Humidity.

Treatments $\{a_0, a_1\}$	Scenarios T	$e_1, \dots, e_4$ H	Expected Value of Utility function
0	Useful	N-Useful	0.251
0	N-Useful	Useful	0.253
0	N-Useful	N-Useful	0.959
0	Useful	Useful	0.250
1	Useful	N-Useful	0.231
1	N-Useful	Useful	0.374
1	N-Useful	N-Useful	0.902
1	Useful	Useful	0.581

The proposed utility function was based on cumulated Beta distributions resembling to s-shaped curves. This is not the only possible choice, e.g. logistic functions could be used instead, as well as many other functions. Nevertheless, the fundamental feature that we believe should not change is the presence of high utility values only when high values are present both for the environmental attributes and for the leaves: this is quite expected in view of the increasing importance of environmental sustainability in agricultural decision-making processes.

The end-user should not take the elicited functions as a black box reference ready to be exploited. The elicitation of soil and biodiversity classes is strongly dependent on the considered vineyard and on the selected chemical, e.g. more or less impacting and more-less effective against *Plasmopara viticola*. Furthermore, our utility function could be extended to include more specific sustainability indexes, more attributes describing quality and yield of grapes, and even alternative types of chemical treatment. Any extension in the above directions should always put the individual preference scheme of the winegrower at the core of an unbiased elicitation procedure.

**Acknowledgements.** We thank prof. Silvia Bacci and all reviewers for comments that helped to improve the manuscript.

## References

- Brischetto, C., Bove, F., Fedele, G., and Rossi, V. (2021). A weather-driven model for predicting infections of grapevines by sporangia of *plasmopara viticola*. *Frontiers in Plant Science*, **12**:636607.
- Bove, F., Savary, S., Willocquet, L., and Rossi, V. (2020a). Simulation of potential epidemics of downy mildew of grapevine in different scenarios of disease conduciveness. *European Journal of Plant Pathology*, **158**, pp. 599-614.
- Bove, F., Savary, S., Willocquet, L., and Rossi, V. (2020b). Designing a modelling structure for the grapevine downy mildew pathosystem. *European Journal of Plant Pathology*, **157**, pp. 251-258.
- Chen, M., Brun, F., Raynal, M., and Makowski, D. (2020). Forecasting severe grape downy mildew attacks using machine learning. *PLOS ONE*, **15**(3):e0230254.
- French, S., Rios Insua, D. (2000). *Statistical Decision Theory. Kendall's Library of Statistics 9. Oxford University Press Inc., New York.*
- Gelman, A., Simpson, D., Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, **10**:555.
- Kab, S., Spinosi, J., Chaperon, L., Dugravot, A., Singh-Manoux, A., Moisan, F., and Elbaz, A. (2017). Agricultural activities and the incidence of Parkinson's disease in the general French population. *European Journal of Epidemiology*, **32**, pp. 203–216.
- Keeney, R.L., Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge University Press.*
- Lavik, M.S., Hardaker, J.B., Lien, G., Berge, T.W. (2020). A multi-attribute decision analysis of pest management strategies for Norwegian crop farmers. *Agricultural Systems*, **178**:102741.
- Wong, F. P., Burr, H. N., and Wilcox, W. F. (2001). Heterothallism in *Plasmopara viticola*. *Plant Pathology*, **50**, pp. 427–432.



# Trust and security in Italy

Silvia Golia

## 1. Introduction

Starting from 2018, the European University Institute (EUI) and the corporate YouGov implemented a survey aimed to study the evolution of European, transnational solidarity denoted as EUI-YouGov survey on Solidarity in Europe (Hemerijck et al., 2021). At the moment four waves (2018, 2019, 2020, 2021) are available for the analysis. The survey covers many aspects of the solidarity (issues, instruments and beneficiaries of the solidarity) plus other dimensions related to it, such as security and trust in the own government or in the European Union (EU). The survey was administered to a representative sample of citizens from 11 (2018) to 13 (2021) EU member states plus the United Kingdom, and was carried out online during the month of April. The datasets are freely available for download <sup>1</sup>.

The sections of the questionnaire evolved during the four waves, adding new questions, revising the text of some of the old questions and eliminating some other questions. Nevertheless, there were sections remained unchanged over the years, such as the ones concerning security and trust in national government and EU. The interesting thing in these three sections is that they are composed of the same 10 areas.

The data are not longitudinal, given that the subjects change at each time span, so the four waves can be considered together. This paper starts from this characteristic to investigate if and how the feeling of security and trust about the 10 areas changes over the time, and the tool used is the Differential Item Functioning (DIF) analysis across time. DIF analysis was born as a tool to assess the validity of a scale, given that it tests the invariance of an item with respect to the characteristics of the subjects (a typical example is the gender); if an item shows DIF then, in most cases, it has to be revised or deleted. Instead, in this paper the primary interest is to study the possible evolution of the items difficulty in order to get insights on what the population felt in these four years.

Moreover, given that the period of study is 2018-2021 and the administration was done in April, the answers of the first two years refer to a Covid-19 pre-pandemic period, whereas the answers collected in the following two years are referred to the pandemic period, and this is another interesting aspect of these data.

The paper is organized as follows. Section 2 reports a brief description of the tools used in the analysis whereas Section 3 the description of the main findings. Conclusions follow in Section 4.

## 2. Methods

The model used in the paper to take into account the available data and to hit the aim of the study is the Rating Scale Model (RSM) (Andrich, 1978), which belongs to the family of the Rasch models. RSM turns raw scores into linear and reproducible measures expressed in logits. Given an item  $i$  with  $m + 1$  response categories ( $c = 0, 1, \dots, m$ ), according to RSM the probability of the subject  $s$  with level of latent trait  $\theta_s$  (denoted also as the ability of the subject

---

<sup>1</sup><https://cadmus.eui.eu/handle/1814/72778>



s) to respond in category  $c$  is given by:

$$P(X_{si} = c) = \frac{\exp\{c(\theta_s - \delta_i) - \sum_{j=0}^c \tau_j\}}{\sum_{l=0}^m \exp\{l(\theta_s - \delta_i) - \sum_{j=0}^l \tau_j\}} \quad (1)$$

where  $\delta_i$  represents the difficulty of item  $i$  and the  $\tau_j$  are called thresholds ( $\tau_0 \equiv 0$  and  $\sum_{j=1}^m \tau_j = 0$ ) and they are equal for all the items. The choice to use the RSM as the measurement model, instead of other alternatives such as, for example, the partial credit model, is motivated by the fact that, in the present study, all the items forming each questionnaire make use of the same response format. Therefore, it is reasonable to assume that the test constructors, respondents, and test users all perceive the items to share the same rating scale (Linacre, 2000). All the parameters are expressed in the same scale (logit) and this allows comparisons. The difficulties of the items  $\delta$  can be compared between each other and also with the abilities distribution. The estimates of all the parameters involved in RSM are done imposing that  $\sum_{i=1}^k \delta_i = 0$ , where  $k$  is the number of items, and this implies that zero is the average item difficulty. Items with estimated difficulty below zero are easier items, that is they are items for which it is not so difficult for the respondents to score high.

DIF refers to the different functioning of a test item for comparable groups of respondents and it is formally defined as follows. An item exhibits DIF if respondents of equal ability on the construct intended to be measured by a test, but from separate subgroups of the population, differ in their expected score on that item (Roussos and Stout, 2004). The reasons for which an item exhibits DIF are various and linked to the context of analysis. With respect to the year of interview as subject characteristic, the hypothesis about the reasons for a different functioning is the change of the external conditions from one year to the next due to the presence/absence of actions implemented by the national and/or European institutions.

There is a large literature regarding methods able to investigate DIF for both dichotomous and polytomous items, focusing primarily on the two-group case; less literature regards methods for the multiple-groups case. The context to which this paper belongs, is the one of multiple-groups and polytomous scored items, so it was addressed as follows. Firstly, the difficulties of all the items and the abilities of all subjects, regardless of their membership group, were estimated under the hypothesis that there are no DIF items. The resulting estimates of the items' difficulties can be interpreted as overall measures of them. Then, for the subjects of each group, the items' difficulties were estimated by applying the anchored maximum likelihood estimation, anchoring the measure of abilities of the involved subjects at the measure previously obtained. This anchoring procedure allows the resulting estimates of the difficulty parameters to be compared. For each item, the statistic computed taking the difference between the estimate for one group and the estimate from the main analysis and dividing it by its approximate standard error, was used to verify the null hypothesis "this item has the same difficulty as its average difficulty for all groups". It corresponds to the approximate Student's t-statistic test (Linacre, 2022). Moreover, the previous group DIF statistics for each item can be summarized as a chi-square statistic, which allows one to verify whether the observed DIF within each item is due to chance alone; the null hypothesis is "this item has no overall DIF across all groups" (Linacre, 2022). The test statistic is computed summing the Student's t-statistics, previously squared and normalized applying the Peizer and Pratt transformation (Peizer and Pratt, 1968).

Moreover, given that the first set of tests compares the item difficulty of one group versus the item difficulty under the hypothesis that there is no DIF, the Mantel test (Mantel, 1963) for pairwise testing for DIF was applied.

### 3. Results

As stated in the introduction, between the sections remained almost unchanged over the years in the EU-YouGov survey on Solidarity in Europe, the ones concerning security and trust in the national government and the EU were analyzed in this paper. The common question regarding *Security* is "how secure or insecure do you feel about each of the following areas?", whereas the one regarding *Trust in the national government* and *Trust in the EU* is "how much do you trust ... to make things better in the following area?". The areas (items) considered by these three dimensions are listed in Table 1; when the formulation of an area is slightly different in the security section, it is reported in parenthesis in the table. For all the three dimensions and

Table 1: *List of the items of the security and trust sections of the survey. In parenthesis the formulation for the security section*

Item	Item
1 The economic situation	2 Climate change
3 Military defence	4 Protection against (The threat from) terrorism
5 Protection against (The threat from) crime	6 Food standards
7 Employment opportunities (in your area)	8 Your own financial situation
9 Healthcare	10 Immigration

items, there were four possible response categories; Very secure (1), Fairly secure (2), Fairly insecure (3) and Very insecure (4) for *Security* and Trust a lot (1), Trust a fair amount (2), Do not trust very much (3) and Do not trust at all (4) for *Trust in the national government* and *Trust in the EU*. They form a 4 point Likert scale. There was also an other possible response category, "Don't know enough to say", but in the analysis it was treated as a missing answer. Moreover, in order to be able to use the RSM, the response categories were reversed.

The number of citizens involved in the four waves of the survey is reported in the first line of table 2. Nevertheless, not all of them responded to all the items, so, for each of the three dimensions investigated, the citizens who responded to at least 5 of the 10 items were taken into account, in order to have a sufficient amount of information to estimate the respondents' degree of security and trust. Table 2 reports their number with respect to the dimension and the wave.

Table 2: *Number of citizens who responded to at least 5 of the 10 items*

	2018	2019	2020	2021
With "Don't Know"	1065	895	2021	2028
Security	1045	873	1953	1982
Trust in the national government	1030	867	1958	1966
Trust in the EU	1025	855	1920	1930

The analysis was conducted as follows. Firstly, the chi-square statistic, at a significance level of 0.05, for testing the hypothesis that an item has no overall DIF across all groups was considered. For the dimensions *Security*, *Trust in the national government* and *Trust in the EU* the items 3 (Military defence) and 7 (Employment opportunities in your area), the items 3 and 4 (Protection against terrorism) and the item 5 (Protection against crime) were, respectively, the only ones which did not suffer for DIF, which means that their difficulty remained stable across the years. It has to be noted that, even if the overall test rejected the null for item 3 and dimension *Trust in the EU*, the analysis of a series of pairwise Mantel tests revealed that the hypothesis of no DIF item was always accepted, so it is possible to conclude that military defence is the unique item stable across the years in common between the three dimensions. The other items did not remain invariant over the years and their trend is shown in figures 1, 2

and 3, where the blue dots correspond to the items' difficulties estimated anchoring the measure of abilities, as explained in the previous section, the red dashed line the measure of the item difficulty under the hypothesis of no DIF and the dotted line highlights the zero, which is the average item difficulty. Analyzing the three figures, it can be observed that the difficulties of

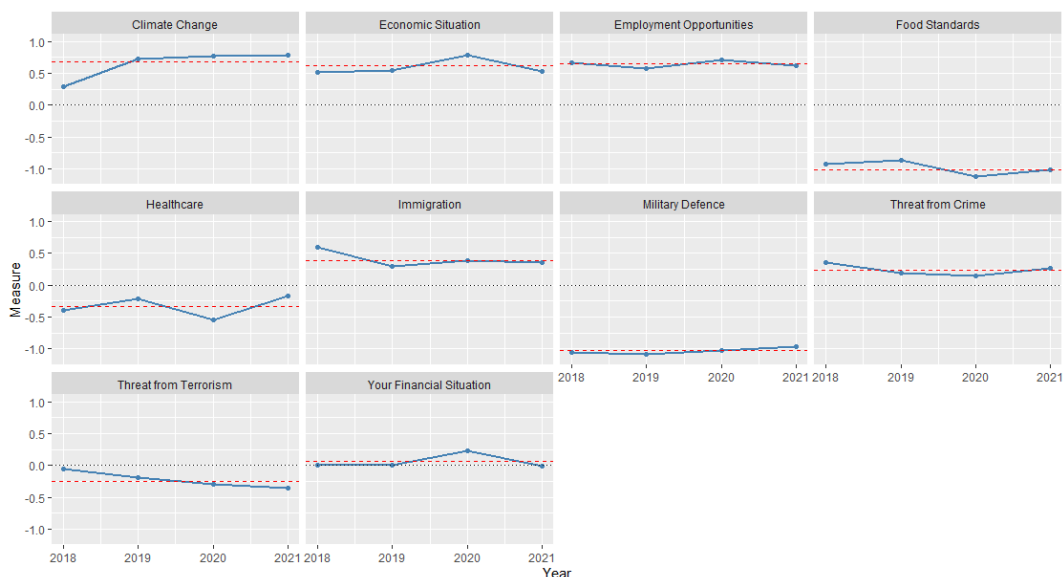


Figure 1: Security: items' difficulties across the waves; red dashed line corresponds to the measure of the item difficulty under the hypothesis of no DIF and the dotted line highlights the zero

all the items exhibit a similar trend across the three dimensions, except for items 2 and 4 for which there are some differences.

Looking at item 2 (Climate Change) it can be noted that there is a jump in the difficulty to feel secure about it moving from 2018 to 2019, and then there is a light increasing trend in the last two years. Its values indicate that from 2019 climate change became one of the themes of greater insecurity among those treated, given that its difficulties remained highly above the mean. Contextually to the 2019 peak of insecurity, there is associated a peak in the scepticism that the national government can improve the existing situation with suitable politics. Nevertheless, in the following two years the item difficulty decreased, going back to the 2018 level, even if this item has still a difficulty over the mean. It is interesting to observe the different attitude of the citizens towards what the EU can do regarding this theme. During the entire period the item difficulty remained under the mean and its value at the end of the period was lower than that of 2018. The results reveal that the citizens trust the EU more than the national government in being able to make things better regarding climate change.

Considering item 4, the feeling of insecurity regarding the threat from terrorism was decreasing along the period; this theme does not represent an issue of particular concern for the citizens, in fact the item difficulty is under the mean. Contextually, there is trust that the government and the EU are able to protect the citizens against terrorism, in fact the item difficulty remained below the mean for the entire period.

It is of interest also the behaviour of item 9 (Healthcare). One can observe that there is a drop of the difficulty of the item moving from 2019 to 2020, that is before and during the first wave of the Covid-19 pandemic, for all the three dimensions, and this drop is more pronounced for *Trust in the national government*. Despite the terrible situation experienced by the Italian

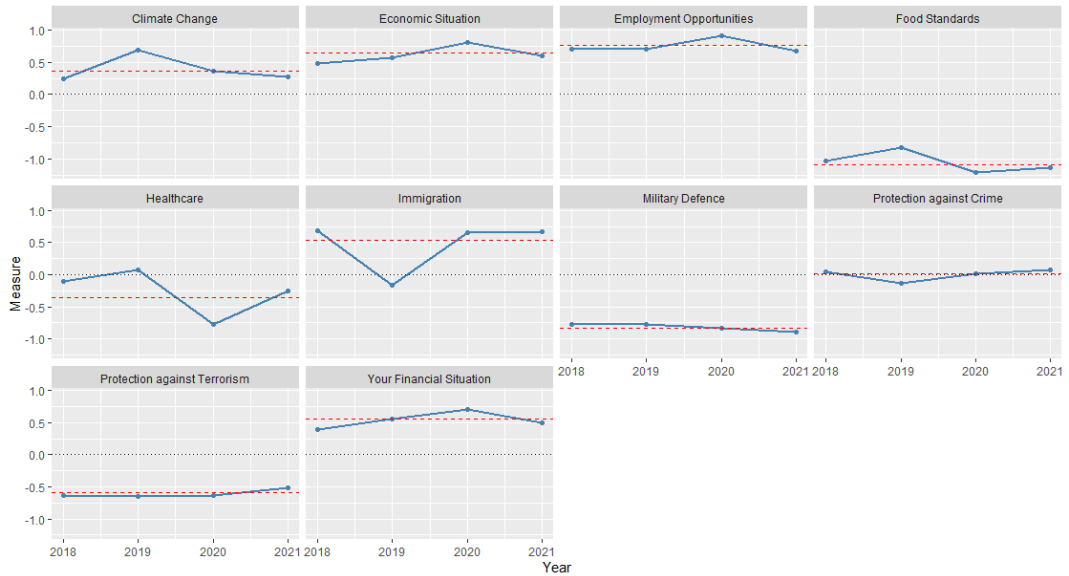


Figure 2: Trust in the national government: items' difficulties across the waves; red dashed line corresponds to the measure of the item difficulty under the hypothesis of no DIF and the dotted line highlights the zero



Figure 3: Trust in the EU: items' difficulties across the waves; red dashed line corresponds to the measure of the item difficulty under the hypothesis of no DIF and the dot line highlights the zero

citizens in 2020, they were confident in the actions of the Italian government and the EU to reduce the impact of the pandemic on the population. Moving from 2020 to 2021, the item difficulty increased, meaning a decrease in the citizen trust regarding the healthcare theme, even if this theme remains one of the themes of low concern, given that its difficulty remained

below the mean.

Regardless of the trend, it is of interest to highlight the different behaviour in the level of difficulty of items 5 and 8 between the three dimensions. During the entire period the citizens felt insecure regarding the threat from crime, given that the item difficulty remained above the mean, whereas they felt not sceptical that the national government or the EU could make things better regarding the protection against crime (the item difficulty was about the mean). A similar behavior, but reversed, can be observed considering the citizen own financial situation; the difficulties of the item 8 for *Trust in the national government* and *Trust in the EU* were above the mean, meaning that the citizens did not trust much both the national government and the EU in improving the existing situation, whereas the item difficulties for *Security* were around the mean, meaning that they did not feel particularly insecure regarding their own financial situation.

#### 4. Conclusions

The paper analyzed three sections of the EUI-YouGov survey on Solidarity in Europe concerning the dimensions of security, trust in the national government and trust in the EU. All of them are related to the theme of solidarity, which is the main focus of the survey. The intent of the study was to inspect if and how the feeling of security and trust about the 10 areas (items) covered by the questionnaire changed over time analyzing the trend of the items' difficulties by means of the DIF analysis. Most of the items exhibited DIF across time and interesting patterns.

Future developments of this analysis will concern the relations between the measures of each dimension and the time and between the three dimensions.

#### References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, **43**, pp. 561–573.
- Genschel, P., Hemerijck, A., Nasr, M., Russo, L. (2021). *Solidarity and trust in times of Covid-19*. EUI RSC PP, 2021/11, European Governance and Politics Programme.
- Hemerijck, A., Genschel, P., Cicchi, L., Nasr, M., Russo, L. (2021). *EUI-YouGov survey on solidarity in Europe trendfile and yearly datasets (2018-2021)*. EUI Research Data, Robert Schuman Centre for Advanced Studies - <https://hdl.handle.net/1814/72778>
- Linacre, J. M. (2000). Comparing and Choosing between "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). *Rasch Measurement Transactions*, **14** (3), pp. 768.
- Linacre, J. M. (2022). *Winsteps®Rasch measurement computer program User's Guide. Version 5.2.3*. Portland, Oregon: Winsteps.com
- Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, **58**, pp.690–700.
- Peizer, D.B., Pratt, J.W. (1968). A Normal Approximation for Binomial, F, Beta, and Other Common, Related Tail Probabilities, I. *Journal of American Statistical Association*, **63**(324), pp. 1416–1456.
- Roussos, L.A., Stout, W. (2004). Differential item functioning analysis: Detecting DIF item and testing DIF hypotheses, in *The Sage handbook of quantitative methodology for the social sciences*, eds. D. Kaplan, Sage Publications, Thousand Oaks (CA), pp. 107–116.

# Topic modeling for analysing the Russian propaganda in the conflict with Ukraine

Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Michelangelo Misuraca,  
Agostino Stavolo

## 1. Introduction

The conflict between Ukraine and Russia is changing Europe, which is facing a crisis destined to reshape the internal and external relations of the continent, shifting international balances. As the war in Ukraine continues, Russian propaganda about the conflict evolves. Modern propaganda can be seen as an attempt to influence opinion through the communication of ideas and values of a specific persuasive purpose (Abd Kadir et al. 2014).

A political organisation would want to convince people to concur with the message presented and accept it as their own beliefs, rejecting other point of view. It has been argued that practically all governments use forms of propaganda to bolster their support from other nations and citizenry (Pratkanis and Aronson, 1991).

For this reason, we present the preliminary results of an analysis conducted on the content of online newspapers used as propaganda tools by the Russian government. The selected newspapers create and amplify the narrative of the conflict, conveying information filtered by the Kremlin to advance Putin's campaign on the war. The goal of the work, therefore, is to understand the communication strategies that the Russian press used to motivate and justify the conflict in Ukraine and what types of information are disseminated by the selected newspapers. In this regard, through a Symmetric Non-Negative Matrix reduction factorization technique (symNMF), we extracted the main themes found in Russian newspaper articles to identify the topics used for propaganda.

## 2. Non-negative matrix factorization

Non-negative matrix factorization (NMF) is a dimension reduction method to uncover latent low-dimensional structures in high-dimensional data (Kim and Park, 2008). NMF is an unsupervised approach in that the low-rank factor matrices are constrained to have only nonnegative elements (Kuang et al. 2015). So, the basis vectors of the matrix are represented as a linear combination of vectors with positive coefficients.

Nonnegativity improves interpretations of the information extracted from a given data matrix, allowing a better understanding of the results obtained from the analysis process. This is in contrast to dimensionality reduction techniques that rely on the singular value decomposition (SVD) method, such as principal component analysis (PCA). One of the major problems with PCA is that the basis vectors have positive and negative components, and the data are represented as a linear combination of these vectors with positive and negative coefficients (Pauca et al. 2004). This is because the principal components are orthogonal, implying the presence of some negative values. Factors obtained from the NMF, on the other hand, are positive vectors and better approximate the data, but are not necessarily orthogonal (Casalino et al. 2016).

Given a  $\mathbf{X}$  matrix of size  $m \times n$ , the decomposition of  $\mathbf{X}$  into a matrix  $\mathbf{W}$  of size  $m \times k$  (called the base matrix) and a matrix  $\mathbf{H}$  of size  $k \times n$  (called the encoding matrix), such that their product approximates the matrix  $\mathbf{X}$ :

$$\mathbf{X} \approx \mathbf{WH} \quad (1.1)$$

Maria Gabriella Grassia, University of Naples Federico II, Italy, mariagabriella.grassia@unina.it, 0000-0002-7128-7323  
Marina Marino, University of Naples Federico II, Italy, marina.marino@unina.it, 0000-0002-0742-5912  
Rocco Mazza, University of Naples Federico II, Italy, rocco.mazza@unina.it, 0000-0002-4901-5225  
Michelangelo Misuraca, University of Calabria, Italy, michelangelo.misuraca@unical.it, 0000-0002-8794-966X  
Agostino Stavolo, University of Naples Federico II, Italy, agostino.stavolo@unina.it, 0000-0001-5890-2195

Referee List (DOI 10.36253/fup\_referee\_list)  
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Michelangelo Misuraca, Agostino Stavolo, *Topic modeling for analysing the Russian propaganda in the conflict with Ukraine*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.43, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 245-250, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

where  $\mathbf{W}$  e  $\mathbf{H}$  are both non-negative matrices. The product  $\mathbf{WH}$  is an approximate factorization of rank at most  $k$ . Generally, the  $k$  rank of the two matrices  $\mathbf{W}$  and  $\mathbf{H}$  is assumed to satisfy that  $k \ll \min \{m, n\}$  (Gaujoux and Seoighe, 2010). The value of the parameter  $k$  identifies the numbers of factors to be used to explain data (Casalino et al. 2016). Matrix multiplication can be implemented as computing the column vectors of  $\mathbf{X}$  as linear combinations of the column vectors in  $\mathbf{W}$  using coefficients given by columns of  $\mathbf{H}$ . Each column of  $\mathbf{X}$  can be computed as follows:

$$x_i = \mathbf{W}h_i \quad (1.2)$$

where  $x_i$  is the column vector of the product matrix  $\mathbf{X}$  and  $h_i$  is the vector of the matrix  $\mathbf{H}$ .

Suppose we have  $n$  data points represented as the columns in  $X = [X_1, \dots, X_n]$ , and try to group them into  $k$  clusters. When  $\mathbf{W}$  and  $\mathbf{H}$  are subject to nonnegativity, it is possible to interpret the dimension reduction in (1.2) as clustering results: the columns of the first factor  $\mathbf{W}$  provide the basis of latent  $k$ -dimensional space, and the columns of the second factor provide  $\mathbf{H}$  the representation of  $x_1, \dots, x_n$  in the latent space. So, the cluster assignment for each data point is made by choosing the largest item in the corresponding column of  $\mathbf{H}$  (Kuang et al. 2015).

The matrices  $\mathbf{W}$  and  $\mathbf{H}$  are found by solving an optimization problem defined with the Frobenius norm (a distance measure between two given matrices), the Kullback-Leibler (KL) divergence (a distance measure between two probability distributions), or other divergences.

The usual approach to NMF is to approximate  $\mathbf{X}$  by calculating  $\mathbf{W}$  and  $\mathbf{H}$  to minimize the Frobenius norm of the  $\mathbf{X} - \mathbf{WH}$  difference, such that (Pauca et al. 2004):

$$\sum_{i=1}^n \sum_{j=1}^m (X_{ij} - (WH)_{ij})^2 = \|\mathbf{X} - \mathbf{WH}\|_F^2 \quad (1.3)$$

The formulation in (1.3) has been applied to many clustering tasks in which the  $n$  data points are available in  $\mathbf{X}$  and are used as an input. The relationship between the data points is represented as a graph, where each node corresponds to a data point and a similarity matrix  $\mathbf{A}_{n \times n}$  contains the similarity values between each pair of nodes (Moutier et al. 2021). The NMF is not a general clustering method that performs well in every circumstance, where the limitation can be attributed to its assumption on the cluster structure (Kuang et al. 2015). As we know, the goal is to approximate the original data matrix using a linear combination of basis vectors. When the underlying  $k$  clusters have nonlinear structure, NMF cannot find any  $k$  basis vectors that represent the clusters respectively.

So, it is used the SymNMF, the symmetric variant of the NMF, that handles symmetric matrices  $\mathbf{A}$  as input. This method is based on a similarity measure between data points and factorizes a symmetric matrix containing pairwise similarity values into the product of a nonnegative matrix and its transpose (Jia et al. 2021). The factorization of  $\mathbf{A}$  will generate a cluster assignment matrix that is nonnegative and captures the cluster structure inherent in the graph representation. Given an  $n \times n$  symmetric matrix  $\mathbf{A}$  and a reduced rank  $k$ , SNMF seeks to find the best factorization so that:

$$\mathbf{A} \approx \mathbf{H}\mathbf{H}^T \quad (1.4)$$

where  $\mathbf{H}$  can be viewed as the cluster indicator and  $\mathbf{H}^T$  the transpose matrix.

Compared with NMF, SymNMF concerns only the factorized similarity matrix  $\mathbf{A}$  and doesn't consider whether the structure of the data is linear or non-linear. It can be regarded as a graph clustering method, and it is more effective for nonlinearly separable data than NMF (Kuang et al.2015). It has demonstrated to be a powerful method for data clustering (Jia et al. 2021), for learning topics in text mining (Yan et al.2013).

Also, SymNMF is related to spectral clustering, SC, and both share the same loss function only with different constraints (Ng et al. 2001) and it can directly generate the clustering indicator without post-processing, while SC needs extra post-processing, like K-means, to finalize clustering.

### 3. Methodology

The proposed work shows preliminary results. Specifically, the analysis is carried out from March 2021, when the Russian military moved weapons and equipment into Crimea, to the end of March 2022, the day of the first negotiations in Istanbul. The selection of the newspapers is based on past study: the report “Pillars of Russia’s Disinformation and Propaganda Ecosystem” produced by U.S. Department of State. According to the report, the journals cover various geographies, and they have their own target audiences. These newspapers are influenced by the Russian government and institutions, thus highlighting a Kremlin-driven information and regime interpretations given to the facts of war. The papers chosen are as follows:

- Strategic Culture Foundation: it is an online journal directed by Russia’s Foreign Intelligence Service (SVR) and closely affiliated with the Russian Ministry of Foreign Affairs.
- Global research: a Canadian website that has become deeply enmeshed in Russia’s broader propaganda ecosystem.
- News Front: it is a Crimea-based disinformation with the goal of providing an “alternative source of information” for Western audiences.
- South Front: it is an online information site registered in Russia that focuses on military and security issues.
- Katehon: a journal that plays the role of a provider of material aimed largely at a European audience, with content devoted to the "creation and defence of a secure, democratic and just international system."
- Geopolitics: a platform for Russian ultra-nationalists to spread disinformation and propaganda targeting Western and other audiences.

We extracted 3,396 newspaper articles, and two of them were withdrawn because they were not written in English; so, we had 3,394 articles. As we know, textual data are unstructured, so it’s necessary to perform some phases of pre-processing for having structured data. There are different steps:

1. Normalized the text, so convert all the letters of the texts into a lower case;
2. Tokenized the documents, obtaining a set of distinct strings (tokens) separated by spaces or punctuation marks;
3. Removed special characters, punctuations, and numbers from the dataset. Also, hashtags, symbols and stopwords are eliminated;
4. Defined a grammatical tagging, which is the process of marking a word in a text as corresponding to a particular part of speech. In this case, we considered the nouns, verbs, and adjectives.

The pre-processing phase returned a database composed by 40.360 tokens, 5010 types and 3394 documents. In this way, the term-document matrix indicates the number of occurrences of each term in the document. The dimension of term-document matrix is 5010x3394.

### 4. Preliminary results

In the final stage of the pre-treatment process we applied the documents and words matrix vector space model. Each word is considered a vector where each element  $a_i$  represents the weight of that element within the individual document. In NMF, the term-document matrix is too sparse to estimate reliable arguments, so more stable and dense data are used.



According to Yan et al. (2013), for reducing the sparsity of term-document matrix, we created a co-occurrence matrix  $\mathbf{W}$  composed of the vectors  $\mathbf{w}_i$ , whose elements  $a_i$  represent the number of times each word pair  $\langle \mathbf{w}_i, \mathbf{w}_j \rangle$  co-occurs within the same document. For each pair of vectors, we calculated the cosine measure, a similarity index that measures the similarity between two vectors of an inner product space. In this way we created the similarity matrix  $\mathbf{S}$ . On this matrix was applied the SymNMF for identifying the main topics. According to this, we found five topics:

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
vital	kill	basis	disagreement	manifestation
threat	launch	arrangement	development	modern
state	kiev	attempt	dispute	identity
ultimate	military	background	divide	individual
turn	mercenary	asset	disagreement	materialism
step	march	attention	economic	philosopher
urgent	nearby	auspice	economy	manifest
threaten	humanitarian	attack	drive	individualism
statement	operation	associate	dream	mankind
tolerate	munition	benefit	digital	liberty

**Tab.1** – Topics extracted from SymNMF

Tab. 1 shows the terms associated with the topics extracted from the SymNMF. It is possible to define the topics that the Russian media used as motivations for the Ukrainian conflict.

The first topic excerpt refers to the threat posed by Ukraine. These newspapers present the conflict as a potential problem for relations with Europe and the West. There is an increasing urgency to seek common ground with the nation, declaring that it is responsible for all the casualties that are occurring. According to the Russian government, Ukraine is exaggerating the issue by not thinking of the collective good and making decisions that only sour the relationship with the Kremlin.

In “topic 2” reference is made to the war dimension. The terms given allow identification of the main arguments the Russians used to justify the invasion. The war is presented as a “humanitarian operation” that Putin undertook to liberate Kiev and Ukraine. Russian propaganda aims to present the Ukrainian people, not as victims but as perpetrators of their crimes and murders: the term “mercenary” refers to a narrative that Ukrainian soldiers murdered Danish mercenaries. This serves to dispel the idea of Ukrainians as a subjugated people. All words referring to the dimension of war emphasize the belligerent spirit of the population, which wants to get rid of the Russian “nearby” enemy even with the use of atomic bombs and munitions received by the U.S.

Related to this, there is “Topic 3” that highlights diplomatic-international relations. Especially, the topic appears to be declined in a general way but such that the dimension described can be interpreted. According to Russian media, President Putin has repeatedly proposed talks and negotiations and set out Russia's conditions, the first is that no NATO base be installed in Ukraine. The topic of international agreements turns out to be central as on the one hand the media talk about the Russian government's willingness to mediate with Ukraine, and on the other hand, they emphasize how the same nation wants to join NATO and improve the relationship with President Biden.

“Topic 4” identifies the economic motivations of the invasion. There is a common view that the Ukrainian government initiated the conflict to increase its economic and geopolitical power to expand into neighbouring countries and counter the Russian nation. In particular, the newspapers report a series of events related to the Ukrainian economy: the emigration of citizens to Poland to

improve wage conditions, the loss of public money, and the resulting debts. The latter aspect is central as it predicts Ukrainian dependence on America. By this, it is narrated that European powers are also obliged to give money to Ukraine to pay for previous situations: for example, the Germans pay a debt to Ukraine for the occupation of Crimea.

The last topic refers to a philosophical-identarian dimension, in which Western and liberal values are criticized and exaggerated. The thoughts of numerous philosophers who present liberalism in a negative light are reported, stating that it "should be opposed to fascism" but it imposes itself on Western civilizations, taking on the same characteristics. For this reason, the U.S. is presented as a nation that does not want to assert other world powers by imposing its own economic and social vision. In this regard, the Russian media propose a vision of its nation as one that engages in the pluralism of ideas and goes to represent an alternative of freedom to the Western world. Numerous articles are referencing how the U.S. wants to change tradition and classical roots (e.g., Dante's works have been called politically incorrect and have undergone liberal cleansing) by going on to criticize the philosophers and thinkers of the time. On the other hand, Russia is presented as the guardian of "true and authentic" European values and not of "globalization" and "liberalism".

## 5. Conclusions

As discussed earlier, the work is in the preliminary stage and the aim is to identify the themes that the Russian government used to narrate the conflict. For future developments, we are expanding the list of Russian information sources in order to conduct a more comprehensive analysis.

Russian Federation invests its propaganda channels and its intelligence services to conduct activities to support their information system, and it leverages outlets on news sites or research institutions to spread these narratives.

So, the Kremlin use these tactics as part of its approach to using information as a weapon. In this regard, the Russian government has issued a series of measures, ordering all media outlets to report on the invasion of Ukraine only through official state sources, blocking numerous sites for spreading unfounded news and threats of high treason. Russia's willingness to employ this approach provides it with three advantages. First, it allows for the introduction of numerous variations of the narratives, to fine tune their information narratives to suit different target. Second, it provides plausible deniability for Kremlin officials when they peddle different information, allowing them to deflect criticism while still introducing damaging information. Lastly, it creates a media multiplier effect that boost their reach and resonance.

## References

- Abd Kadir, S., Abu Hasan, A. S. (2014). A content analysis of propaganda in Harakah newspaper. *Journal of Media and Information Warfare (JMIW)*
- Casalino, G., Del Buono, N., Mencar, C. (2016). Nonnegative matrix factorizations for intelligent data analysis. *Non-negative Matrix Factorization Techniques* (pp. 49-74). Springer, Berlin, Heidelberg.
- Jia, Y., Liu, H., Hou, J., Kwong, S., Zhang, Q. (2021). Self-supervised symmetric nonnegative matrix factorization. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Kim, J., Park H., (2008) Sparse nonnegative matrix factorization for clustering, Georgia Inst. of Technology, Tech. Rep.
- Kuang, D., Choo, J., Park, H. (2015). Nonnegative matrix factorization for interactive topic modeling and document clustering, in *Partitional Clustering Algorithms* (pp. 215-243). Springer, Cham.
- Moutier, F., Vandaele, A., Gillis, N. (2021). Off-diagonal symmetric nonnegative matrix factorization. *Numerical Algorithms*, 88(2), 939-963.

- Ng, A., Jordan, M., Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14.
- Pauca, V. P., Shahnaz, F., Berry, M. W., Plemmons, R. J. (2004). Text mining using non-negative matrix factorizations, in *Proceedings of the 2004 SIAM International Conference on Data Mining* (pp. 452-456). Society for Industrial and Applied Mathematics.
- Pratkanis, A., Aronson, E. (1991). What is news? In *Age of Propaganda: The Everyday Use and Abuse of Persuasion*, (pp. 32 – 39). New York: W. H. Freeman.
- U.S. Department of State, GEC Special Report (2020). Pillars of Russia’s Disinformation and Propaganda Ecosystem, [https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem\\_08-04-20.pdf](https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf)
- Yan, X., Guo, J., Liu, S., Cheng, X., Wang, Y. (2013). Learning topics in short texts by non-negative matrix factorization on term correlation matrix, in *Proceedings of the 2013 SIAM International Conference on Data Mining* (pp. 749-757), Society for Industrial and Applied Mathematics.

# The relationship between religiosity, religious coping, and anxieties about the future: a multidimensional analysis on the Evangelical churches of Naples

Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Agostino Stavolo

## 1. Introduction

The Covid-19 pandemic has had an impact on the social and personal lives of individuals, leading to the development of new forms of adaptation and response to crises. Extraordinary and traumatic events can have significant consequences on the way of living and practicing faith.

The research is part of the studies on Temporal Perspective (PT) concerning to religiosity, deepening the idea of temporal perspective as culturally sensitive, and therefore, also influenced by religious factors. The intent is to investigate how transcendental can relate to the perspective of individuals and the consequent way of interpreting and acting reality, especially in crises. The aim of contribute is to investigate the relationship between religiosity, religious *coping*, anxiety for the times to come, and the prospect of the transcendental future in the period of the pandemic.

The study aims to understand whether religiosity and beliefs, experiences, and practices (public and private) have affected the prospects of individuals. We referred to the concept of anxiety about the future due to the emergency in which there has been a response with an approach to faith and religious practices, using religion as a *coping* tool.

According to this, we administered a survey on a sample of subjects of the Neapolitan protestant Christian population of the Evangelical churches of the Assemblies of God in Italy (A.D.I.). Then, a Multiple Correspondence Analysis (MCA) was carried out to identify the relationship between religiosity, *coping* tools, and prospects.

## 2. Literature review: temporal perspective and religious coping

The study on Future Time Perspective has influenced much of the research on Temporal Perspective (PT). Researchers refer to future perspectives using various conceptualizations, including Future Thinking and Future Time Perspective. The former concerns plans and expectations through which potential outcomes and goals may be achieved (Aspinwall 2005); the latter refers to an individual's beliefs and convictions or perspective toward the future about temporally distant goals (Bembenutty and Karabenick 2004).

Scholars have emphasized the benefits of future-oriented thinking, which is motivational for health and well-being (Boyd and Zimbardo 2005), influences the nature of social relationships (Lang and Carstensen 2002), and promotes goal setting, motivation, and achievement efforts (Shipp et al. 2009). However, the negative effects on future events and actions need to be considered. Future Time Perspective has focused less attention on how negative futures can impact a person's overall well-being by destabilizing both physical and mental health. Zaleski (1996) introduces the concept of Future Anxiety, which is a state of apprehension, uncertainty, fear, and worry about changes. In this context, religious *coping* is introduced (Pargament 1997).

*Coping* strategies enable the development of behaviours to manage traumatic events, stressful situations, and moments of conflict. While related to sacred elements, religious *coping* also includes a wide range of coping tools for various stressors: prayer, confession, seeking spiritual support from religious organizations, and accepting circumstances as representative of God's will (Pargament 2002).

Maria Gabriella Grassia, University of Naples Federico II, Italy, mariagabriella.grassia@unina.it, 0000-0002-7128-7323

Marina Marino, University of Naples Federico II, Italy, marina.marino@unina.it, 0000-0002-0742-5912

Rocco Mazza, University of Naples Federico II, Italy, rocco.mazza@unina.it, 0000-0002-4901-5225

Agostino Stavolo, University of Naples Federico II, Italy, agostino.stavolo@unina.it, 0000-0001-5890-2195

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Maria Gabriella Grassia, Marina Marino, Rocco Mazza, Agostino Stavolo, *The relationship between religiosity, religious coping, and anxieties about the future: a multidimensional analysis on the Evangelical churches of Naples*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.44, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 251-256, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

Nowadays, the development of Temporal Perspectives with the emergence of *coping* strategies has become a much-studied issue due to the instability and unpredictability of the social situation and the growth of socio-psychological intensity. The Covid-19 pandemic has been a pressure factor for individuals: levels of depression and anxiety are increased compared to those observed in pre-pandemic surveys (Lei et al. 2020); while noting an increase in the general use of religious and spiritual practices to alleviate the negative consequences of social isolation measures during the pandemic (Luchetti et al. 2020).

### 3. Methodology

The study is exploratory and aims to analyse the relationships between religiosity, religious *coping*, anxiety for the times to come, and the prospect of the transcendental future. To reach these goals, we developed these research questions:

*RQ1*: What are the dimensions emerging from the relationship between religiosity, religious *coping*, and anxiety about the future during the Covid-19 pandemic?

*RQ2*: Are there relationships between the transcendental future and earthly future perspective?

To answer these questions, we conducted a preliminary study using a non-probabilistic sample. We referred to the population of 2555 faithful residents in Naples, belonging to the Evangelical churches of the Assemblies of God in Italy (A.D.I.). We decided to study the Evangelical church in Naples on the one hand because it is a fast-growing church, and on the other hand because its territorial and geographical proximity allowed us to be able to study the evangelical community. We used the distribution by church location neighbourhood and by gender of each individual to define quotas. We reached 279 individuals. The reason we worked with a small number of identified subjects is that, although sufficient for a preliminary analysis, they are not powerful enough to reach the entire population.

Then, we administered a survey from June 9 to July 30, 2021. The survey was carried out using a CAWI (Computer Assisted Web Interviewing) system. Thanks to this system, respondents were able to access the online questionnaire via a hyperlink disseminated through the use of the main social channels (WhatsApp, Facebook, and Instagram) and were able to answer the survey by sending their answers in real time. The survey is divided by content areas:

- a) Ascribed characteristics of the respondent: questions were asked to find sociodemographic information. The section consists of 13 categorical questions.
- b) Health experience with Covid-19: the focus is on the possible consequences of an infection. The section consists of 12 categorical questions.
- c) Religion, faith in quarantine and future: it aims to detect the respondent's spiritual and religious orientation, use of religious coping during the period of the first quarantine (March to May 2020), level of anxiety about the future, and view of the transcendental future. The section consists of 8 categorical questions.

We studied religion using the Centrality of Religion Scale (CRS) designed by Huber S. and Huber O.W. in 2012. The CRS is a 5-point validated Likert scale that measures the centrality, importance, and relevance of religion in an individual's life. The theory supporting the design and validation of this scale is Charles Glock's multidimensional model of religiosity (1968); the scale measures the intensity of religious life in five dimensions. The

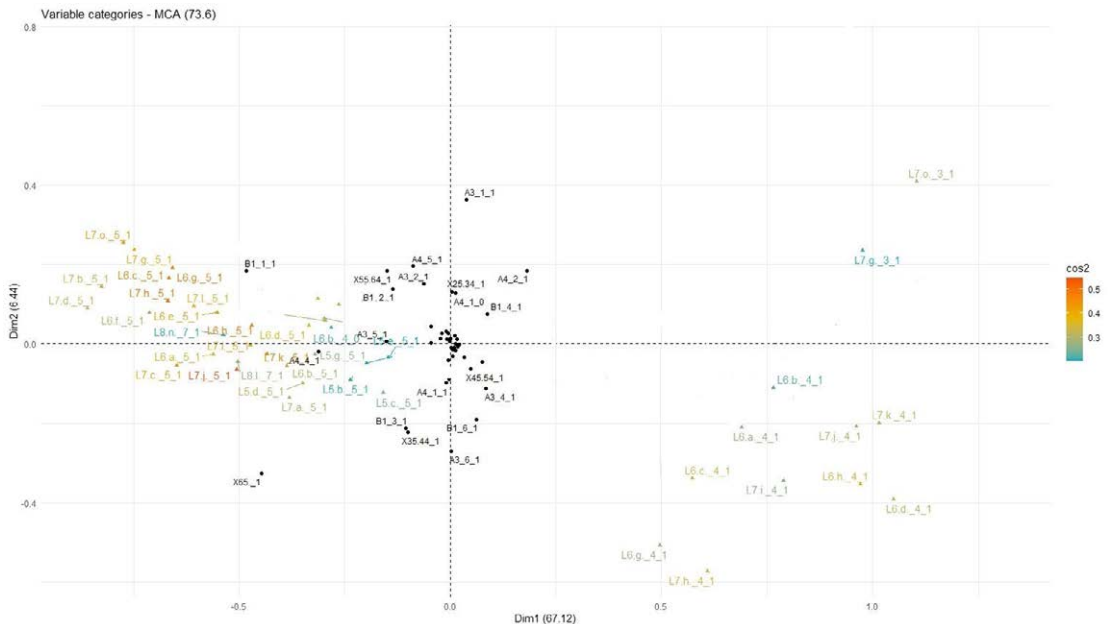
dimensions are:

1. Intellect: the intellectual dimension consists of knowledge of religious themes.
2. Ideology: the ideological dimension refers to beliefs.
3. Public practices: refer to membership in religious communities manifested through public participation in rituals and community activities.
4. Private practices: this dimension refers to the actions and rituals that individuals enact in an individual form in their private space to get in touch with transcendent reality.
5. Religious experience: it consists of the perceptions and emotions related to the relationship with the divine and transcendent.

#### 4. Preliminary results

Factor analysis techniques allow the synthesis of the information contained in the original data, through the identification of an optimal space of reduced size. The method agrees to the construction of a set of latent variables (or factors), a combination of the original variables, that express concepts not directly observable. We performed a Multiple Correspondence Analysis (MCA) for identifying the relations of the variables investigated by the survey. In the MCA analysis, each principal inertia value is expressed as a percentage of the total inertia. These values quantify the amount of variation accounted for by the corresponding dimension. We selected the first two factors, whose percentage of explained inertia is 73,6%, following the Benzecri correction formula.

Fig 1. shows the factorial map of the variables. We coded with L the modalities related to the variables on religion. The modalities of the analysed variables are presented with the cosine measure ( $\cos^2$ ).



**Fig. 1 – Factorial map**

We nominated the first factor (67% of total inertia) such as the “Intensity of religion”. The

variables that determined the construction of the first factor related to how one turned to religion, and specifically to God, during the pandemic (e.g., "God's intervention," "God present," "Frequency of prayers," "Feeling heartened after a religious message").

Especially, for contributing to the construction of the factor are the ways related to concerns about the future and how subjects deal with daily difficulties, where an active attitude and positive outlook are evident. On the left side, the variables show more emotional involvement in religion compared to the left side, where the modalities of the variables with a lower value on the Likert scale are reported.

The second factor (6% of inertia) is labelled "Dimensions of religiosity". The variables that determined the second factor, on the other hand, highlight one's relationship with religion, differentiating between a personal dimension ("Frequency of prayer," "Reading of sacred texts") and a collective dimension ("Prayer with the religious community," "Importance of having a community of reference," "Remote religious services"). In particular, the representative modalities come close to defining an active spirituality, where variables are predominantly associated with the purely spiritual dimension of religious experience. Deity is seen as present and active, that can act in human life and can relate and communicate with it. The factor divides the map into two sides: in the upper part, the private and individualistic dimension of religiosity is highlighted: there are ways concerning the attendance and use of personal prayer, the reading of sacred texts, and the relationship with the divine; in the lower part is the collective dimension of prayer, evidencing the role of the relevant evangelical community and the importance of attending church services (in attendance and remotely).

According to this, we defined the four quadrants.

In the upper left quadrant are the modalities that evidence an individual's relationship with transcendental during the pandemic. In fact, by projecting the additional dots, we notice that the respondents were infected with Covid. This highlights the use of individual religious practices as a tool to counteract the psychological and social difficulties experienced during the period. Feeling God's presence and increasing the relationship with divinity through intimate moments, such as prayer and reading sacred texts, highlights the need for the faithful to have personal times and spaces for communication. This is related to decidedly convinced view of the future as life even beyond the earthly one.

The lower left quadrant emphasizes the importance of having a religious collectivist to refer to, they are devoting assiduously to religious practices and have an active relationship with the community. The element of the evangelical community appears to be central, showing how, during the Covid, the faithful needed to attend services. The deep relationship with divinity and community through religious practices is strongly associated with not having felt abandoned by God or spiritually dejected during the pandemic period.

The bottom right quadrant refers to the use of positive religious *cop*ing tools. Indeed, the reassuring and comforting element that faith has during times of stress is emphasised. Prayer turns out to be a central element of the quadrant. The same people who purposely devote much time to personal prayer are the same people who very often pray instinctively inspired by everyday situations. The feeling that God is able and willing to communicate, relates to the believer's awareness of his presence, which helps to reassure from fears due to the emergency and to hearten through listening to religious messages.

The last quadrant in the upper right refers to a less optimistic view of the future and a lower intensity of faith than the previous ones.

## 5. Conclusions

We reported some preliminary conclusions about the analysis. Through the use of the MCA, it is possible to visualize the relationship between the variables considered. It was found that the intensity of the use of positive religious coping during the pandemic generally follows responses

to "The Centrality of Religion Scale" (CRS), which measures religiosity regardless of the historical period experienced. Therefore, it allows us to relate religious behaviours before and during the emergency period. We noticed that high participation in public religious practices in habitual situations is equivalent to high participation in religious services remotely in quarantine. Therefore, it can be said that the distress situation does not seem to have affected religious orientations and behaviours by evidencing estrangement or rapprochement of individuals toward divinity, religious practices, or the evangelical community.

According to Pargament (2011), greater religiosity corresponds to greater use of positive religious coping methods. In the relationship between positive religious coping and religiosity, we can determine elements of the association that are repeated in the observation of the factorial plan. Carry out the importance of the image and awareness of a God who is present, able to come in contact with individuals, and able to take an interest in his life, to establish a relation. It is supported concretely by religious practices (the meetings and prayer) that enable a direct connection with God to soothe fears. The more one feels that ability to concretely intervene in an individual's life, the more one turns to the entity. The importance attached to the idea of an active, present, and working God has an effect on the perception of the future and the resulting feeling of anxiety. Prayer together with the community and family, as well as religious meetings, allows people to feel heartened by the message conveyed and strengthen their faith (*RQ1*).

This dimension can also be found in the relationship that anxiety about the earthly future establishes with the transcendental future, where the possible function of ascribing a purpose to live is evidenced. Observation of the factorial plan shows that greater religiosity corresponds to greater belief and trust in life after death. For evangelicals with the highest degree of religiosity is associated with a view of trust in life beyond death seen as a new beginning (*RQ2*). Moreover, it is the transcendental future that mediates between religiosity and anxiety about the future (Boyd and Zimbardo, 2005).

## References

- Aspinwall, L. G. (2005). The psychology of future-oriented thinking: From achievement to proactive coping, adaptation, and aging. *Motivation and emotion*, 29(4), 203-235.
- Bembenuatty, H., Karabenick, S. A. (2004). Inherent Association Between Academic Delay of Gratification, Future Time Perspective, and Self-Regulated Learning. *Educational Psychology Review*, 16(1), 35-57.
- Boyd, J. N., Zimbardo, P. G. (2005). Time Perspective, Health, and Risk Taking., in A. Strathman & J. Joireman (Eds.), *Understanding behavior in the context of time: Theory, research, and application* (pp. 85-107). Lawrence Erlbaum Associates Publishers.
- Glock, C. Y., Stark, R., (1968). *American piety: The nature of religious commitment* (Vol. 1). University of California Press
- Huber, S., Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religions*, 3(3), 710-724.
- Lang, F. R., Carstensen, L. L. (2002). Time counts: Future time perspective, goals, and social relationships. *Psychology and Aging*, 17(1), 125-139.
- Lei, L., Huang, X., Zhang, S., Yang, J., Yang, L., Xu, M. (2020). Comparison of prevalence and associated factors of anxiety and depression among people affected by versus people unaffected by quarantine during the COVID-19 epidemic in *Southwestern China*. *Medical science monitor: international medical journal of experimental and clinical research*, 26.
- Luchetti, M., Lee, J. H., Aschwanden, D., Sesker, A., Strickhouser, J. E., Terracciano, A., Sutin, A. R. (2020). The trajectory of loneliness in response to COVID-19. *American Psychologist*, 75(7)
- Pargament, K. I. (1997) *The Psychology of Religion and Coping: Theory, Research, Practice*. New York, Guilford.
- Pargament, K. I. (2002), The bitter and the sweet: An evaluation of the costs and benefits of



- religiousness, *Psychological Inquiry*, 13(3), 168-181.
- Pargament, K. I. (2011). Religion and coping: The current state of knowledge. In S. Folkman (Ed.), *The Oxford handbook of stress, health, and coping* (pp. 269–288). Oxford University Press.
- Shipp, A. J., Edwards, J. R., Lambert, L. S. (2009). Conceptualization and measurement of temporal focus: The subjective experience of the past, present, and future, in *Organizational behavior and human decision processes*, 110(1), 1-22.
- Zaleski, Z. (1996). Future anxiety: Concept, measurement, and preliminary research, in *Personality and individual differences*, 21(2), 165-174.

# An application of the Agency for Digital Italy guidelines and CSA Star self-assessment: A Docustar case study

Pierluigi Calabrese, Paola Lunalbi, Vincenzo Ribaudò, Saverio Crisafulli, Antonio Ruoto, Vito Santarcangelo, Diego Sinitò, Carlo Bonelli, Giuseppe Stella

## 1. Introduction

The digital documents play a predominant role in the production of business and public administration documents; they are created through telematic tools and, in the same way, they are stored, with the aim of guaranteeing a better efficiency and lower costs of business and public authority processes, definitively replacing the use of paper.

Consequently comes the problem of uniformly regulating the way in which these documents are produced and stored, to guarantee their integrity and authenticity, so is enacted the Digital Administration Code (CAD) with the function of regulating, among other things, the validity and the effectiveness of public administration's informatic documents; subsequently, the "Agenzia per l'Italia Digitale" (AgID) adopted initial guidelines aimed precisely at giving technical application to the rules of the CAD and establishing the procedures for the production, management and storage of digital documents by public administration's and private entities.

In 2020, AgID issued new guidelines in this regard, with the aim of updating the technical rules on the formation, registration, management and storage of digital documents in application of the CAD, bringing together all the various provisions and guidelines on the subject in a single text containing, precisely, all these rules.

The structure and objectives of these AgID guidelines will be outlined below, followed by a presentation of Docustar, the platform developed by Stella All in One for managing access to digitalised versions of business documents in compliance with the General Data Protection Regulation (GDPR), and certified ISO 27001:2013.

## 2. AgID guidelines

AgID's guidelines have the dual purpose of updating the current technical rules under article 71 of the Digital Administration Code (CAD), concerning the formation, protocol, management and storage of computerised documents, and of incorporating all the technical rules and circulars on the subject into a single guideline.

The general purpose of these guidelines is to simplify the entire process of managing computerised documents through an overall vision that aggregates within a single guideline all the subjects that were previously regulated separately, highlighting the functional interdependencies between the various phases of document management, from the moment of formation to its permanent preservation.

Six documents are also attached to the guidelines, and form an integral part of them. Among these we can find the one on file formats that can be used for the formation of digital documents (annex 2) and the one on metadata related to the same documents (annex 5): with regard to usable files, the digital formats that documents must have are identified from among those used by the different software known today, such as .doc, .docx, .pdf; with regard to metadata, on the other

Pierluigi Calabrese, Stella All in One Srl, Italy, luigi@dittastella.it  
Paola Lunalbi, Stella all in one Srl, Italy, paola@dittastella.it  
Vincenzo Ribaudò, ilnformatica Srl, Italy, vincenzo@iinformatica.it  
Saverio Crisafulli, ilnformatica Srl, Italy, saverio@iinformatica.it  
Antonio Ruoto, ilnformatica Srl, Italy, antonio@iinformatica.it  
Vito Santarcangelo, ilnformatica Srl, Italy, vitho87@hotmail.it, 0000-0003-4971-8788  
Diego Sinitò, ilnformatica Srl, Italy, diego@iinformatica.it, 0000-0002-5044-0050  
Carlo Bonelli, Keylogic Srls, Italy, c.bonelli@keylogic.it  
Giuseppe Stella, Stella All in One Srl, Italy, giuseppe@dittastella.it, 0000-0002-5967-5446

Referee List (DOI 10.36253/fup\_referee\_list)  
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Pierluigi Calabrese, Paola Lunalbi, Vincenzo Ribaudò, Saverio Crisafulli, Antonio Ruoto, Vito Santarcangelo, Diego Sinitò, Carlo Bonelli, Giuseppe Stella, *An application of the Agency for Digital Italy guidelines and CSA Star self-assessment: A Docustar case study*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.45, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 257-262, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

hand, we identify the minimum set of information relating to the file/document that must be associated with the file itself, such as the ID, the producer, the date, the title, the subject, etc.

The management of digital documents is characterised by a process consisting of three distinct phases, which we will now look at in detail: the formation, management and preservation of the document.

The first aspect on which the guidelines are based deals with the formation of an electronic document, identifying four different ways in which an electronic document must be created to be considered valid:

- the creation of the document through software or cloud services that are qualified and able to guarantee that documents are produced in formats that allow interoperability between systems;
- the acquisition of an electronic document by telematic means or by storage device or the creation of a copy of an analogue document by scanning it and subsequent acquisition on an electronic medium, or the direct acquisition of an electronic copy of an analogue document;
- the storage of information in digital format on a storage device resulting from computer transactions or processes or from the submission of data via modules or forms made available to the user;
- The generation or grouping, also automatically, of a set of data or records, from one or more databases, according to a predetermined logical structure and stored in static form.

The digital document produced must be identified in a unique and persistent manner. As far as public administration is concerned, the guidelines require that identification take place by means of the document's registration, whereas in the case of any documents that are not registered, identification is entrusted to the functions of the computerised document management system. An identification system other than protocol is envisaged, which can be used as an alternative to the former by associating the document with a cryptographic fingerprint based on hash functions that are considered cryptographically secure. Subsequently, the document must be rendered unalterable: to achieve this, it is established that the document is stored on a computer medium in a digital format that cannot be altered in its access, management and preservation. The operations that must be performed to guarantee the immutability and integrity of the computer document are also established within the guidelines for each of the types of computer document formation set forth.

With regard to the computerised administrative document, the same rules apply as for the ordinary computerised document, with the difference that the immutability and integrity of this type of document can also be achieved through its registration in the entity's protocol register or in the other registers, directories, lists, archives or data collections that are contained in the entity's computerised document management system, and by the fact that the computerised file of the administrative document is associated with the set of metadata provided for protocol registration and those for classification and storage.

The guidelines then go on to regulate the stage of managing the computerised document, establishing the technical rules, criteria and specifications of the information that must be complied with when recording computerised documents. Each public administration must appoint a document management manager, as well as a document management coordinator, who have legal, IT and archiving skills. The computerised registration of documents is carried out through the application of electronic data attached or connected to the computer document that serve to uniquely identify it. Once the registration is completed, the document will be identified with the set of data in electronic format. The protocol registration, therefore, is made up of the set of metadata applied to the documents received or sent by the public administration (PA) that are stored in the protocol registry and that are associated in a permanent and unmodifiable form, a registry that must ensure that each protocol operation performed is traced, historicized and

attributed to the operator who performed it; in particular, it must be ensured that the information (subject, sender and addressee of a registered document) cannot be modified, nor cancelled, and that the only information that can be modified is that relating to internal administration assignment and classification. All modification or cancellation operations must be historicised and always visible. In addition, the system used for filing must be developed in compliance with the cyber security provisions of the guidelines, which must guarantee the unambiguous identification and authentication of users, the guarantee of access to resources only to users who are authorised and/or to groups of users according to the definition of appropriate profiles, the permanent tracking of any event of modification of the information processed and the identification of its author, sending the daily record of the protocol for the previous day to the filing system, through transmission methods that guarantee the unchangeability of the content.

Finally, the guidelines regulate the digital document preservation system, establishing that the computerised document management system must transfer closed computer files and closed computer series to the preservation system, transferring them from the current archive or from the deposit archive, and computer files and series that have not yet been closed, transferring the computer documents they contain according to the specific needs of the institution, with particular attention to the risks of technological obsolescence. The function of the preservation system is to guarantee the preservation of computerised documents and computerised administrative documents with the relevant metadata, as well as computerised document aggregations (i.e. files and series) and computer files with the relevant metadata until the eventual discarding of such computer files, through the adoption of rules, procedures and technologies in such a way as to guarantee the characteristics of authenticity, integrity, reliability, readability and retrievability of the same. In addition, the preservation system must have functions and requirements to ensure that it is possible to access the preserved documents for the entire period laid down in the owner's preservation plan and in current legislation, or for a longer period that may be agreed between the parties.

The guidelines also identify the subjects that play roles in the preservation process: the owner of the preservation object; the producer of the deposit package; the authorised user; the preservation manager and the preserver. In the public administration, the role of preservation manager is entrusted to an internal manager or official identified by the owner of the preservation object, who has legal, IT and archiving skills, or to a person outside the body, provided that he or she has the required skills and is a third party with respect to the owner of the preservation object, the preserver. His task is to define and implement the policies of the preservation system and to manage it independently under his responsibility: in particular, he defines the preservation policies and the functional requirements that the preservation system must have, manages the preservation process and ensures its constant compliance with the law, generates and signs the deposit report, monitors the proper functioning of the preservation system, carries out the periodic check, at least every five years, of the integrity and legibility of the documents contained in the preservation system, provides for the duplication or copying of computer documents as the technological context evolves, and prepares the necessary measures to ensure the physical and logical security of the preservation system.

The guidelines also provide the formation and adoption of a preservation manual (to be published on the institutional website), an IT document that specifically identifies the organisation, the subjects involved and their roles, as well as the operating model, a description of the process and the architectures and infrastructures used, the security measures adopted and all other information useful for managing and verifying the operation of the preservation system.

### **3. Analysis of solutions on AgID Cloud Marketplace**

In order to carry out an analysis of the Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) services qualified with AgID, we used the open data database of the Cloud Marketplace, taking advantage of the datasets obtained and analysing the

solutions of service offered on the marketplace, by year and by category.

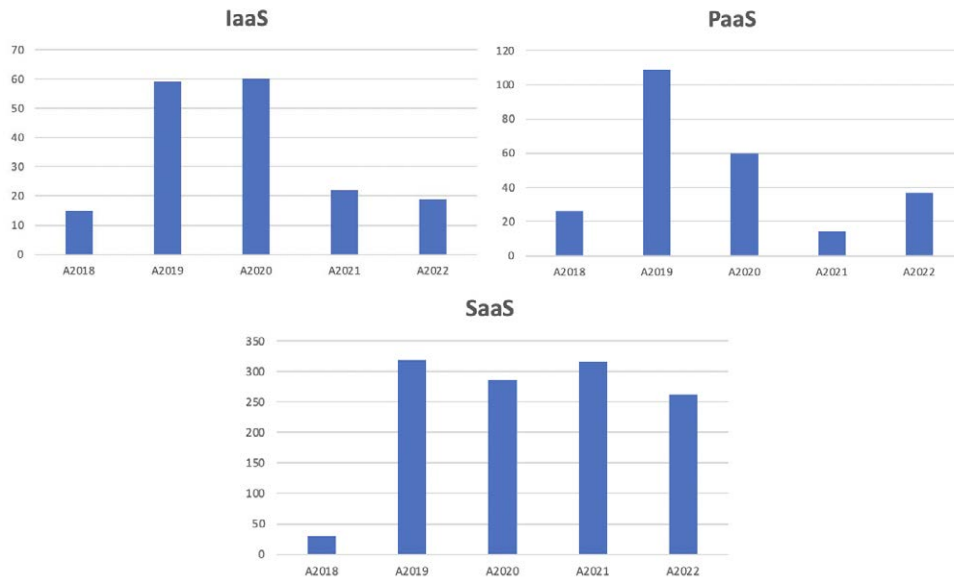


Figure 1. IaaS, PaaS and SaaS services on marketplace per year

The analysis carried out shows a greater presence of IaaS services within the marketplace in the two-year period 2019-20 (34%) as well as for PaaS services in 2019 (44%); it is also important the figure for SaaS, with an exponential growth from 2018 (2%) to 2019 (26%) that has remained constant over the years. This trend gives us evidence of how it has become necessary, as of 1 April 2019, for these services to be qualified by AgID and published in the Cloud Marketplace so that they can be acquired by public administrations.

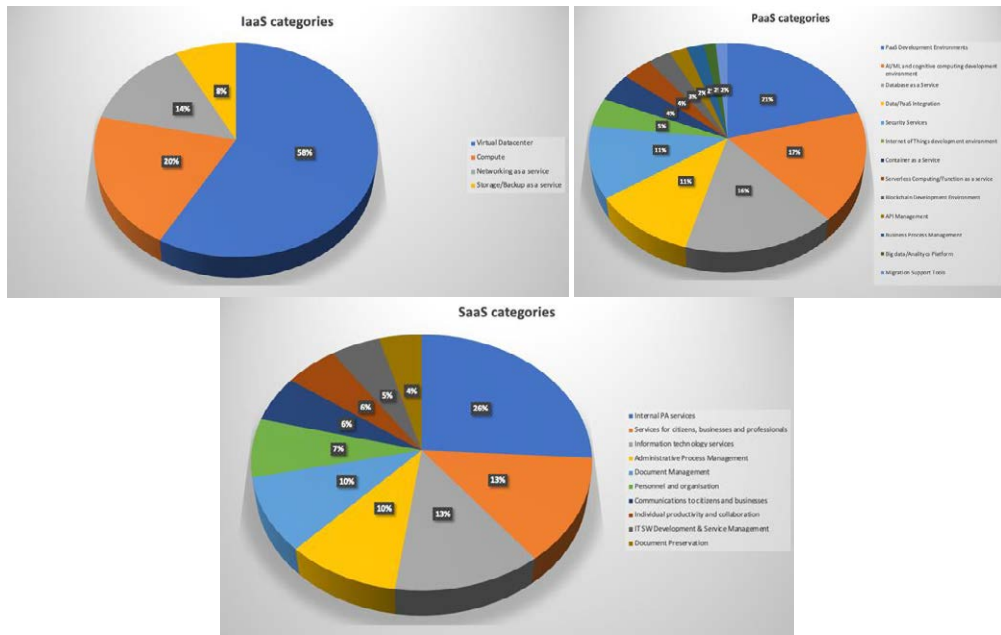


Figure 2. IaaS, PaaS and SaaS services on marketplace by category

An analysis by category, on the other hand, shows us that 58% of the IaaS services present on the market is related to virtual data centres, while in the case of PaaS services we find a 21% of PaaS development environments, 17% AI/ML and cognitive computing development environments, 16% database as a service environments, while a small slice (only 3%) concerns blockchain development environments. With regard to SaaS, most of the software relates to internal PA services (26%); 10% of the software on the marketplace relates to document management software, while only 4% of the software relates to document preservation software.

It is therefore clear that IaaS and PaaS services are the clear minority, given the considerable costs and above all the requirements involved, with mainly large accredited players starting out (such as IBM, Amazon, Oracle, Microsoft and Google), while SaaS services, for which it is sufficient to rely on an accredited Cloud Service Provider (CSP), are increasing.

#### 4. Case study: Docustar

In order to be in perfect compliance with the requirements of the AgID guidelines, innovative SME Stella All in One Srl designed and developed the Docustar software, implementation of the new DRM-related industrial privative technique and in compliance with ISO 27001 no. 10202000032405 entitled 'Method for digital document rights management for digitisation, archiving and destruction for ISO27001 compliance' and Cloud Security Alliance (CSA) STAR Cloud Assessment. The latter is a free tool and registry that documents the security controls provided by different cloud computing services, thus helping users assess the security of the cloud providers they currently use or are considering to use.

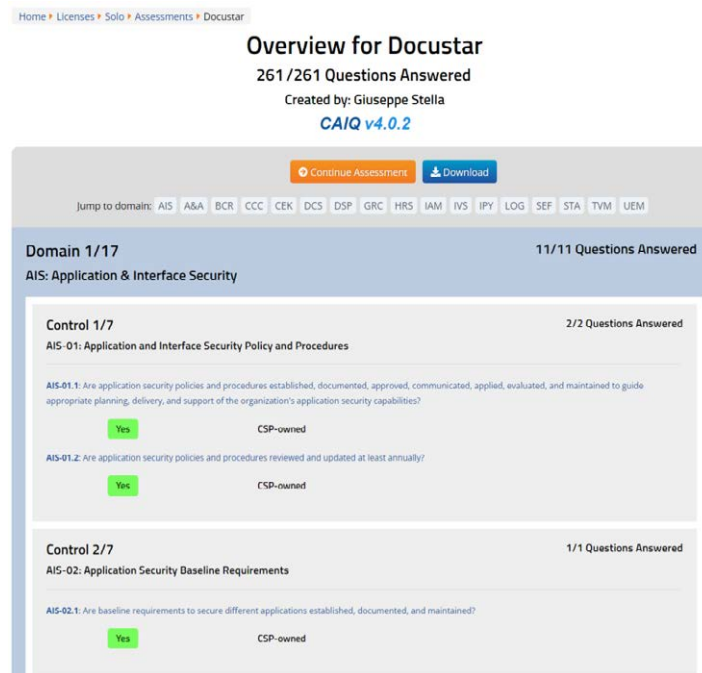


Figure 3. Overview for Docustar in CAIQ questionnaire

Observance of the RID paradigm (confidentiality, integrity and availability) and the related information security compliance is the object of the entire Docustar project. In fact, each document, in addition to being profiled and encrypted, is the subject of an appropriate workflow that traces each access to the system, the individual document request and the access to the resource, in order to guarantee appropriate confidentiality in the access and management of information resources.

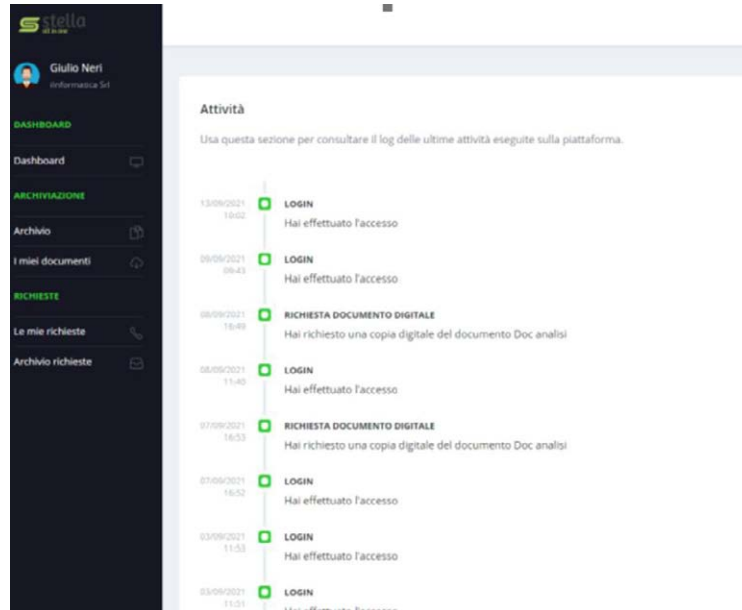


Figure 4. Docustar activities

## 5. Conclusions

This paper described the rigorous standards introduced at the Italian national level to regulate digital documents and document preservation and provided in Docustar a possible solution for complying with the relevant requirements set out, combined with a revolutionary document workflow approach with time-based Digital Rights Management (DRM). Docustar is a SaaS solution that confirms the potentiality of these applications that aims to improve PA services following AgID requirements compliance. This innovative approach that combines DRM within SaaS document management application opens the door to a new concept of data and file confidentiality by further enhancing the security of information exchanges in the cloud.

## References

- Calder, A. (2009). Information Security based on ISO 27001/ISO 27002. Van Haren.
- Lisi, A. (2009). The Digital Administration Code in Italy: Light and Shade. *Curentul Juridic, The Juridical Current, Le Courant Juridique* 1, pp. 57-63.
- Phattanateeradej, C., Twittie S. (2016). Storage and search tool for cloud provider security information in CSA STAR. 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE.
- Ziming, L. (2008). Paper to digital: Documents in the information age. ABC-CLIO.
- Voigt, P., Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer International.
- Diamantopoulou, V., Tsohou, A., Karyda, M. (2019). General Data Protection Regulation and ISO/IEC 27001: 2013: Synergies of activities towards organisations' compliance. *International Conference on Trust and Privacy in Digital Business*. Springer, Cham, pp 94–109.

# Remote working in Italy: Just a pandemic accident or a lesson for the future?

Luigi Bollani, Simone Di Zio, Luigi Fabbris

## 1. Introduction

During the Covid-19 pandemic, remote working (RW) became a way to ensure that Italians continued to perform their productivity duties while protecting human health. The government's aim was to limit the movement of workers and reduce the presence of people in offices without compromising services. During 2021 and 2022, about half of Italian workers experienced, at least partially, RW (Fondirigenti, 2020; Eurofound, 2021). RW, also known as telecommuting or telework, is an arrangement between employee and employer in which the employee's work duties are performed remotely, usually at home or in specific locations off the employer's premises, using information and communication technologies (Felstead and Henseke, 2017; Donnelly and Johns, 2021). According to Eurofound and the ILO (2017), right before the pandemic, Italy had the lowest percentage of RW employees in Europe. In 2019, Istat, the Italian Statistical Institute, estimated that, overall, less than 2.5% of Italian workers engaged in RW. Before the pandemic, RW was a *'luxury for the relatively affluent few'*, since few workers—predominantly white-collar workers and higher income earners—had the opportunity to work remotely (Desilver, 2020). The pandemic outbreak, which resulted in several times more people working remotely, was a de facto global RW experiment. For some time, working from home became the norm. Although the loosening of Covid-19 containment measures put an end to this mass experiment, things could change considerably in the medium term, with many workers—about half of workers, according to futurist scholars (Glenn et al. 2019)—working from home regularly. For this reason, we aimed to measure Italians' willingness to work remotely in the upcoming years. To that end, we analysed data collected through a survey of adult workers conducted in the second half of 2021. The survey was aimed at investigating how Italians evaluated their working experiences during the pandemic and how they perceived the possibility of working remotely in the future. Thus, we measured the frequency and intensity of the RW phenomenon, the opinions of those who practiced it and their feelings about the possibility of practicing it in the future. The analysis aimed to address the following research questions:

*RQ1: Is there a relationship between having performed RW during the pandemic and willingness to do so in the future?*

*RQ2: Did work activity and workers' individual characteristics influence their disposition towards RW?*

*RQ3: What resources and problems shape workers' disposition towards RW?*

The rest of this paper is organised as follows. Section 2 introduces the data and the model used for data analysis. Section 3 presents the main results of the statistical analysis. Section 4 discusses the results with reference to the mainstream literature on RW.

## 2. Data, models and methods

### 2.1. Data

A sample of adult Italian workers was surveyed using a computer-assisted web-based

Luigi Bollani, University of Turin, Italy, luigi.bollani@unito.it, 0000-0002-2488-3659

Simone Di Zio, University of Chieti-Pescara G. D'Annunzio, Italy, s.dizio@unich.it, 0000-0002-9139-1451

Luigi Fabbris, Tolomeo studi e ricerche, Padua and Treviso, Italy, fabbris@stat.unipd.it, 0000-0001-8657-8361

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Luigi Bollani, Simone Di Zio, Luigi Fabbris, *Remote working in Italy: Just a pandemic accident or a lesson for the future*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.46, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 263-268, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



interviewing questionnaire. The sample was formed by merging five samples selected by a pool of Italian universities. The data collection lasted from June to November 2021. A total of 817 people participated in the survey filling an electronic questionnaire. Of these, 193 were workers; three of them did not respond to a basic question and were excluded from the analysis. Thus, the analysis included 190 respondents. The data collection method lends to suspect a certain self-selection of the sample that favours more educated people. The analysis focused on descriptors of the propensity to work remotely and their possible predictors.

The variables used in the relational model were as follows:

**Y:** Propensity to work remotely in a post-pandemic future. The relevant question was as follows: “*The health emergency will end. If you continue to work after that, would you rather work from home or at your workplace?*” The four ordinal responses to this question were collapsed into two:  $Y = 1$  indicated a propensity to work remotely, and  $Y = 0$  otherwise.

**X<sub>A</sub>:** Health effects of the pandemic. The block included the following descriptors: having been infected by Coronavirus ( $X_1$ ) and facing the psychological ( $X_2$ ) or physical ( $X_3$ ) consequences.

**X<sub>B</sub>:** Personal or social resources against social shocks. This block included possessing a higher education degree ( $X_4$ ), living alone ( $X_5$ ), living with a partner ( $X_6$ ), having children ( $X_7$ ), resilience ( $X_8$ ), proactive attitude ( $X_9$ ), resorting to vaccines ( $X_{10}$ ) and trusting scientists during the pandemic ( $X_{11}$ ). Variable  $X_8$  denoted the standardised scores obtained by a factor analysis of a set of nine items related to self-efficacy and resilience selected from the 25-item Connor-Davidson resilience scale (Connor and Davidson, 2003). Variable  $X_9$  denoted the standardised scores obtained by a factor analysis of a set of eight items related to optimism–proactivity selected from the 20 items comprising the BHS (Beck et al., 1974). The variables  $X_{12}$ – $X_{16}$  referred to motives for preferring RW to office work, as described in Table 2.

**X<sub>C</sub>:** Personal or social problems related to RW. This block included chronic diseases ( $X_{17}$ ) and depression ( $X_{18}$ ). The latter was a dichotomous variable computed using the nine-item Beck Hopelessness Questionnaire proposed by Spitzer et al. (1999) and translated into Italian by Mazzotti et al. (2003). A value  $X_{18} = 1$  indicates major depression. The variables  $X_{19}$ – $X_{28}$  are motives for preferring office work to RW, as described in Table 2.

**Z:** Control variables. This block included working as an employee ( $Z_1$ : dichotomous), working in industry ( $Z_2$ : dichotomous), gender = male ( $Z_3$ : dichotomous) and age ( $Z_4$ : up to 34 years, 35–64 years and 65 years or older).

## 2.2 Analytical model

The analytical model included the propensity to work remotely in the future as a dependent variable ( $Y$ ) and two sets of regressors as control variables:  $X_1$ – $X_{28}$  selected individually through a forward stepwise selection according to their significance and  $Z_1$ – $Z_4$ . The relationship may be written as  $Y = f(X_1, X_2, \dots, X_{28} | Z_1, \dots, Z_4)$ . The logistic regression model is written as follows (Hosmer and Lemeshow, 2000):  $\text{logit}[p(Y=1)] = \beta_0 + \beta_1 X_1 + \dots + \beta_J X_J + \beta_{J+1} Z_1 + \dots + \beta_{J+4} Z_4$ , where  $\text{logit}(p) = \ln[p/(1-p)]$ , and  $\beta_j$  ( $j = 0, 1, \dots, J$ ) measures the relationships between  $Y$  and  $X_j$  ( $j = 1, \dots, 28$ ) and between  $Y$  and  $Z_k$  ( $k = 1, \dots, 4$ ) when all other variables in the model remain fixed. To select the predictors, a stepwise selection technique was adopted with a significance level  $< 0.10$ . The statistical analyses were performed in R (R Core Team, 2022). A logistic regression model with a binary response variable was performed using the *glm* function from the MASS package. The *My.stepwise* package and *My.stepwise.glm* function were used for the stepwise model selection. Finally, the *DescTools* package and *PseudoR2* function were used to measure the model’s goodness of fit.

## 3. Results

Table 1 reports the joint frequency distribution of recent RW experience and the disposition to practice it in the future. The pandemic experience allowed workers to understand the opportunities related to RW, at least with respect to pre-pandemic practices. Indeed, of the

workers who experienced RW (67.9% of the total), 52.6% reported that they would be willing to do it again if offered, at least under certain conditions. Conversely, 15.3% of the respondents were not interested in repeating the experience. Out of three workers who did not experience RW during the pandemic, two stated a preference for office work and one for RW. The difference between the number of workers who did not wish to repeat the experience and those who would be willing to do it for the first time was about 5% of the total number of respondents. Overall, the respondents who would be willing to practice RW in the future represent 63.2%.

**Table 1.** Per cent estimates of during-the-pandemic remote working and availability to do it in the future among Italian workers, 2021.

<i>During pandemic experience</i>	<i>Availability for the future</i>			
	<i>No</i>	<i>At conditions</i>	<i>Fully</i>	<i>Total</i>
Null	21.6	10.0	0.5	32.1
1 – 50% time	9.0	16.3	0.5	25.8
51 – 100% time	6.3	31.6	4.2	42.1
Total	36.8	57.9	5.3	100.0

Tables 2 and 3 report the frequency distribution of the possible predictors and the estimate of the regression coefficients of the predictors selected for the model.

**Table 2.** Mean of the variables used in the statistical analysis of Italian workers, 2021.

<i>Variable</i>	<i>mean</i>	<i>Variable</i>	<i>Mean</i>
X <sub>1</sub> : Infection: personal	0.137	X <sub>17</sub> : Number of chronic diseases	0.342
X <sub>2</sub> : Suffered psychologic damages	0.211	X <sub>18</sub> : Depression	0.137
X <sub>3</sub> : Suffered physical damages	0.100	X <sub>19</sub> : Job inadequate for RW	0.037
X <sub>4</sub> : Possessing a higher education degree	0.679	X <sub>20</sub> : Office better for teamwork	0.063
X <sub>5</sub> : Being single	0.263	X <sub>21</sub> : Office to interact with customer	0.032
X <sub>6</sub> : Living in couple	0.695	X <sub>22</sub> : Office production monitoring	0.116
X <sub>7</sub> : Having children	0.505	X <sub>23</sub> : No home isolation	0.111
X <sub>8</sub> : Resilience score	0.000	X <sub>24</sub> : Help desk inadequate for RW	0.032
X <sub>9</sub> : Proactive attitude	0.000	X <sub>25</sub> : Office better internet connection	0.200
X <sub>10</sub> : Vaccinated: Yes	0.800	X <sub>26</sub> : House workplaces inadequate	0.153
: Not yet	0.137	X <sub>27</sub> : Interferences with family life	0.074
: Never	0.063	X <sub>28</sub> : Difficult family-work balance	0.042
X <sub>11</sub> : Trusted scientists	0.684	Z <sub>1</sub> : Working as an employee	0.716
X <sub>12</sub> : Saving time and money	0.321	Z <sub>2</sub> : Working in industry	0.195
X <sub>13</sub> : Working in a more comfortable context	0.216	Z <sub>3</sub> : Gender (male)	0.532
X <sub>14</sub> : Optimizing working schedules	0.121	Z <sub>4</sub> : Age till 34	0.290
X <sub>15</sub> : Clarity in operational goals	0.132	“ 35-64	0.668
X <sub>16</sub> : Balancing family and work	0.042	“ 65 and over	0.042

- The model showed a significant fit for Italian workers' propensity for RW. The Nagelkerke pseudo-R<sup>2</sup> index was 49.8%, indicating that a high proportion of the criterion variable deviance was explained.
- Gender, age, working in industry (vs. any other sector) and working as an employee (vs. self-employment) did not predict a disposition to RW once other personal and familial descriptors were entered into the model.
- The only socio-demographic variable that correlated with RW propensity was the possession of a higher education degree, with less educated workers being more willing to work remotely than higher educated ones. This suggests that the willingness to engage in RW is stronger among employees with a secondary school education than with a university education. Moreover, given that the willingness to engage in RW in the future was greater among workers with both lower self-efficacy ( $r = -0.172$ ) and

lower proactivity ( $r = -0.138$ ), the educational profile of most people oriented towards RW was intermediate.

- Covid-19 infection also played an important role. Workers who contracted the disease were less prone to RW than those who did not. This is rather surprising, considering that during the pandemic, people were forced to work from home to reduce the risk of infection. We can conjecture that workers who avoided infection felt stronger and more open to new experiences than those who were infected.
- Other predictors were related to conditions that may have favoured or disfavoured RW. Predictors that may have favoured RW were saving commute-related time and money and the adequacy of one's home as a workplace. Predictors that may have disfavoured RW were the presence of children in the family, the partial inadequacy of RW for effective teamwork and the difficulty in supporting people through help desks. These indicators are consistent with a diffused idea of RW—a mode of working in which time management and commuting costs are optimised, while other factors, such as internet connection quality, suitability of one's home as a workplace, idea sharing and opportunities for exchanges with managers and colleagues, make working from home less effective.
- The fact that the presence of children reduced the willingness to work remotely can be considered a counterintuitive finding. Although RW was considered a way of balancing family and work lives, children seemed to be incompatible with it.

**Table 3.** Beta estimates of the logistic regression model with remote working preference as criterion variable (forward stepwise selection of regressors,  $n=190$ ; Nagelkerke  $R^2=0.498$ ; control variables and type of job were forced into the model; \*\*\*  $< 0.001$ ; \*\*  $< 0.01$ ; \*  $< 0.05$ ; °  $< 0.10$ ; NS: Not Significant).

<i>Regressor</i>	$\hat{\beta}$	<i>se</i> ( $\hat{\beta}$ )	<i>Significance</i>
<i>Intercept</i>	-0.997	0.959	NS
Gender: male	-0.414	0.421	NS
Age (classes)	-0.103	0.413	NS
Employee	0.810	0.433	NS
Industry	0.616	0.497	NS
RW experienced during pandemic	1.183	0.276	***
Infection: self	-2.150	0.699	**
House workplaces inadequate for RW	2.026	1.146	°
Children	-0.919	0.439	*
Office better for teamwork	-1.974	0.878	*
Possessing higher education title	-1.102	0.462	*
Saving time and money	1.835	0.725	*
Help desk inadequate for RW	-2.389	1.228	°

#### 4. Discussion and conclusion

This study aimed to examine how Italians experienced RW during the pandemic and whether they were willing to work remotely in the future. Our findings suggest that although RW was compulsory during the pandemic, the experience influenced workers' future interests. RW can be seen as an experiment that several workers evaluated positively and in which they showed interest, even for the future. About 63% of our respondents stated that they would consider accepting such an offer. Thus, the pandemic, along with all its negative aspects, also brought new opportunities (Willcocks, 2020; Grzegorzczuk et al., 2021). Our data show that the RW experience was also associated with negative perceptions. Indeed, the number of people who were willing to engage in RW in the future was lower than that of workers who experienced it during the pandemic. This seems reasonable, since the pandemic forced people to stay at home for a few months, while future possibilities imply consent and wider time spans.

Our analysis reveals the main characteristics of people particularly oriented towards RW. Employees with an intermediate education and low-to-medium skills or clerk positions represented the vast majority of workers willing to work remotely. Fana et al. (2020) and Sostero et al. (2020) suggested that low-skilled clerks and medium-skilled professionals favoured RW because their jobs were characterised by standardised procedures. Our results also show that many workers prone to future RW lacked proactivity and self-efficacy. These personality traits may enhance job autonomy, thereby increasing motivation, self-discipline and affect for one's own work (Parker et al., 2010), which are necessary for building trust between employee and employer. However, a risk of RW is that it may induce free-riding and other opportunistic behaviours if RW is not designed and monitored appropriately. Our findings also suggest that Italian workers were aware of the need for RW to be effective. They recognised its advantages in terms of time and money saving but also understood that self-discipline, internet connection quality, adequacy of the home as a workplace, a conflict-free dwelling and an efficient redesign of working schedules were required to make RW feasible. Work redesign must also consider the need for job humanisation (Donnelly and Johns, 2021), which includes highly valued out-of-family socialisation. An RW culture relies on a balance between workers' expectations and results-based accountability. Our survey suggests that training, investments in technology, location adaptation and an agreed system of norms and organisational factors, especially to combat isolation and improve work–life balance, personal development and career progression, are necessary before a major transition to RW. Other challenges are related to how to organise production to enhance creativity and innovation, promote employee learning, engage workers in informal exchanges with senior managers and colleagues and, ultimately, guarantee a company's productive efficiency. All this requires a wise integration of employees' and employers' perspectives (Allen et al., 2015; Wang et al., 2020; Delany, 2021). Finally, learning from the Covid-19 shock, legislators should consider not only the productivity and social acceptance of flexible RW but also the possibility of maintaining productivity during the next crisis. A limitation of our study may be the sample representativeness. In fact, the response rate to the survey questionnaire was low. This may be due to the possibility that the pandemic accelerated a falling trend of people's availability to collaborate in surveys. This could limit the possibility to generalise our level estimates, while it should not threaten the possibility to make statements about between-variable relationships. For the future, a study based on a larger sample could provide further insights: I) Since local economic and organisational conditions can lead to differences in the willingness to work remotely, a regionally based control in the regression model would be important; II) The analysis of the possible relation between the willingness to remote working and the temporal distance from the Covid experience could highlight if this willingness depend on time ; III) It would be interesting to analyse subsets of the sample, e.g. only those subjects who experienced RW during the pandemic; IV) Finally, a simulation experiment could highlight if our research results depended in particular on the adopted stepwise technique, which, as is well known (Steyerberg et al., 1999), may have limited power in selecting important covariates in small samples.

## References

- Allen, T.D., Golden, T.D., Shockley, K.M. (2015). How effective is telecommuting? Assessing the status of our scientific findings. *Psychological Science in the Public Interest*, 16(2): 40–68.
- Beck, A. T., Weissman, A., Lester, D., Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, 42(6): 861–865.
- Connor, K.M. Davidson, R.T. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC), *Depression and Anxiety*, 18: 76–82.
- Delany, K. (2021). What challenges will organisations face transitioning for the first time to the new normal of remote working?, *Human Resource Development International*, DOI:

- 10.1080/13678868.2021.2017391
- Desilver, D. (2020). Before the coronavirus, telework was an optional benefit, mostly for the affluent few. *Pew Research Center* (<https://www.pewresearch.org/fact-tank/2020/03/20/before-the-coronavirus-telework-was-an-optional-benefit-mostly-for-the-affluent-few/>).
- Donnelly, R., Johns, J. (2021). Recontextualising remote working and its HRM in the digital economy: An integrated framework for theory and practice. *International Journal of Human Resource Management*, 32(1): 84–105.
- Eurofound (2021). *COVID-19: Implications for Employment and Working Life*. Luxembourg: Publications Office of the European Union.
- Eurofound and the International Labour Office (2017). *Working Anytime, Anywhere: The effects on the World of Work*. Publications Office of the European Union, Luxembourg, and the International Labour Office, Geneva, Switzerland.
- Fana, M., Milasi, S., Napierala, J., Fernandez Macias, E., Gonzalez Vazquez, I. (2020). *Telework, Work Organisation and Job Quality during the COVID-19 Crisis: A Qualitative Study*. JRC Paper No. 2020/11, European Commission.
- Felstead, A., Henseke, G. (2017). Assessing the growth of remote working and its consequences for effort, well-being and work–life balance. *New Technology, Work and Employment*, 32 (3): 195–212.
- Fondirigenti (2020). *Quick survey Smart working: Il boom che annuncia il futuro*. Roma: Confindustria, Federmanager.
- Glenn, J.C. and the Millennium Project Team (2019). *Work/Technology 2050. Scenarios and Actions*. Washington, D.C.: The Millennium Project.
- Grzegorzczak, M., Mariniello, M., Nurski, L., Schraepen, T. (2021). Blending the physical and virtual: A hybrid model for the future of work. *Policy Contribution*, 14/21, Bruegel.
- Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression (2nd ed.)*. New York: John Wiley & Sons.
- Mazzotti, E., Fassone, G., Picardi, A., Sagoni, E., Ramieri, L., Lega, I., Camaioni, D., Abeni, D., Pasquini, P. (2003). Il Patient Health Questionnaire (PHQ) per lo screening dei disturbi psichiatrici: Uno studio di validazione nei confronti della Intervista Clinica Strutturata per il DSM-IV asse I (SCID-I). *Italian Journal of Psychopathology*, 9(3): 235-242.
- OECD (2020). *Productivity gains from teleworking in the post COVID-19 era: How can public policies make it happen?* (<https://www.oecd.org/coronavirus/policy-responses/>).
- Parker, S.K., Bindl, U.K., Strauss, K. (2010). Making things happen: A model of proactive motivation. *Journal of Management*, 36(4): 827– 856.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, AT (<http://www.R-project.org>, accessed 26 June 2022).
- Sostero, M., Milasi, S., Hurley, J., Fernández-Macias, E., Bisello, M. (2020). *Teleworkability and the COVID-19 Crisis: A New Digital Divide?*, Seville: European Commission, JRC121193.
- Spitzer, R.L., Kroenke, K., Williams, J.B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ Primary Care study. *Journal of the American Medical Association*, 282(18): 1737-1744.
- Steyerberg, E. W., Eijkemans, M. J., & Habbema, J. D. F. (1999). Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of clinical epidemiology*, 52(10), 935-942.
- Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., McIntyre, R.S., Choo, F.N., Tran, B., Ho, R., Sharma, V.K., Ho, C. (2020). A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain, Behavior, and Immunity*, 87: 40-48.
- Willcocks, L. (2020). Remote working: here to stay? (<https://blogs.lse.ac.uk/businessreview/2020/04/02/remote-working-here-to-stay/>).

# Repression of the future-oriented disposition of Italians by a never-ending pandemic

Simone Di Zio, Luigi Fabbris

## 1. Introduction

This paper was aimed at highlighting how the coronavirus disease (COVID-19) pandemic influenced the future-oriented disposition of Italians. Having a future outlook is an attitude that motivational psychologists consider a mental trait that enables people to find motivations for their future plans and behaviours. Roseman (2013) defines this attitude an ‘emotional syndrome’ for coping with the future.

Having a future time perspective (FTP) and the instrumentality to operate for its realisation creates motivation, deep conceptual learning and intensive persistence. As suggested by Van Calster et al. (1987) and Simons et al. (2004), this perspective should consider the degree of specificity and the content of future goals and the context in which goals are designed. Thus, the clarity of the future background influences the possibility to design and achieve feasible goals and plans. Persons who are hopeful and have an optimistic opinion about their future tend to generate instrumentality and energy for better outcomes, whatever their goals are.

Our research question is not limited within the perimeter of health emergency but also involves social and economic aspects. Thus, we conducted a survey among a sample of Italians in the second half of the year 2021 using a web-administered electronic questionnaire (computer-assisted web-based interviewing [CAWI]). The survey was conducted when the COVID-19 pandemic was close to its end. The questions posed were oriented to understand the consequences of the health turmoil and the possibilities for a quick return to normality.

The fundamental idea of the survey was that the pandemic was a unique, dramatic experience for most Italians and that the health, economic and social relics of this 2-year experience could teach future behaviours, which could lead to a more sustainable future. Also, as Commodari and La Rosa (2020), among others, have proposed, the COVID-19 outbreak made the future fuzzier and darker than ever. This may reduce people's energy to operate for a strategic change. Accordingly, large groups of the population started experiencing malaise and psychological distress.

Our analysis was motivated by the following hypotheses:

*H1: Did the COVID-19 infection influence the perception of Italians of their ability to master their futures?*

*H2: Which social obstacles and personal problems are at higher risk of (negatively) influencing the FTP of Italians?*

*H3: Are there social, familial and personal resources that may protect against the difficulty to perceive one's own perspective after the pandemic?*

*H4: Which socio-demographic descriptors mediate the social and individual resources and obstacles in shaping a clear view of Italians about their own future?*

The rest of the paper is organised as follows: Section 2 describes the data at hand and introduces the relational model and basic methodological aspects for data analysis. Section 3 presents the main results of the statistical analysis. Finally, Section 4 provides the interpretations of the results with reference to the mainstream literature on FTP.

Simone Di Zio, University of Chieti-Pescara G. D'Annunzio, Italy, s.dizio@unich.it, 0000-0002-9139-1451  
Luigi Fabbris, Tolomeo studi e ricerche, Padua and Treviso, Italy, fabbris@stat.unipd.it, 0000-0001-8657-8361

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Simone Di Zio, Luigi Fabbris, *Repression of the future-oriented disposition of Italians by a never-ending pandemic*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.47, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 269-274, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

## 2. Data and methods

### 2.1. The data

From June to November 2021, a sample of Italian adults was surveyed using a CAWI questionnaire, through mailing lists of (mostly) students, teachers and workers. At the end of the data collection, 817 respondents filled in the questionnaire, among which 52.4% aged between 18 and 34, 36.2% between 35 and 64, and 11.4% over 64 (other characteristics of the sample in Tab. 1). About the geographical distribution of the respondents, we have a tiny overestimation of the central-northern area of the country, being 73.6% of the sample against 63.5% of residents in this area.

The questionnaire survey was aimed at highlighting the frequency and the effects of the COVID-19 infection and how people faced the various moments of the pandemic, including isolation ('lockdown') and learning or working remotely. In this work, we focused on two descriptors of people's mentality and their possible predictors. The variables used in the relational model are described as follows:

**Y:** Having clear views about what to do after the pandemic as a measure of FTP. Even though psychometric tests were performed to evaluate FTP (among others: Zimbardo and Boyd, 1999), the question was posed dichotomously. FTP relates to the perception of time rather than to the actual physical time as it passes in the calendar (Husman and Shell, 2008). Simons et al. (2004) conjectured that the further into the future an individual's time perspective is extended, the greater the number of goals and plans to reach those goals the individual has.

**X<sub>0</sub>:** Proactive attitude. The responses obtained were classified into three ordinal categories after performing a one-dimensional factor analysis of an 8-item set. The items were selected from the 20-item Beck Hopelessness Scale (Beck et al., 1974). The first category included the standardised factor scores till -0.25 ('passive'), the second category included scores from -0.26 to 0.39 ('reactive'), and the third category included scores from 0.40 and higher ('proactive').

**X<sub>1</sub>:** Self-efficacy attitudes. This is a continuous variable obtained by factor analysing a set of 9 items related to self-effectiveness and resilience. The items were selected from a 25-item resilience scale (Connor and Davidson, 2003) and translated to Italian by the authors. Self-efficacy was defined as an individual's belief in their ability to achieve an outcome (Bandura, 1977); and resilience, as the ability to cope mentally or emotionally with a crisis or to return fast to pre-crisis status (de Terte and Stephens, 2014).

**X<sub>18</sub>:** Full-blown depression. This is a dichotomous variable computed using the 9-item patient health questionnaire, as proposed by Spitzer et al. (1999) and translated to Italian by Mazzotti et al. (2003). A cumulative response score of  $\geq 10$  identifies a person with depression.

The **X<sub>2</sub>/X<sub>17</sub>** and **X<sub>19</sub>/Z<sub>3</sub>** variables are described in Table 1.

### 2.2 The model

The model for data analysis included the dichotomous variable, *Y*, as a criterion variable; the antecedent predictor, *X<sub>0</sub>*; a selection of 26 predictors, **X**; and 3 control variables, **Z**. The relationship may be written as follows:

$$Y = f(X_0, X_1/X_{26}|Z),$$

where *X<sub>0</sub>* denotes a proactive personality, *X<sub>1</sub>/X<sub>6</sub>* represents the personal resources available to one who went through the pandemic, *X<sub>7</sub>/X<sub>12</sub>* is the available social resources, *X<sub>13</sub>/X<sub>23</sub>* is the individual problems and *X<sub>24</sub>/X<sub>26</sub>* is the social obstacles that could limit without let or hinder one's future goals or plans. As a matter of fact, *resource* is a synonym of *protective factor*, and *obstacle* is a synonym of *risk factor*. For this analysis, the possible infection of the respondents and their parents, their contact with the healthcare system and the effects of the possible infection were assimilated to individual problems. Moreover, *X<sub>0</sub>* was transformed into three dichotomous variables.

The model assumes a hierarchy of causal relationships between the criterion variable *Y*, the

main predictor  $X_0$ , the remaining  $X$  predictors and the  $Z$  control variables. Within this hierarchy, the relationships between  $Y$  and the correlates and between  $X_0$  and the remaining  $X$ 's identified the theoretical model *à la Ajzen* (Fishbein and Ajzen, 1975; Ajzen, 1991), in which blocks of positive and negative correlates altogether concur to the statistical fit of the disposition to actively participate in the post-pandemic society.

The logistic regression model can be written as follows:

$$\text{logit}[p(Y = 1)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \beta_{k+1} Z_1 + \dots + \beta_{k+3} Z_3,$$

where  $\text{logit}[p] = \ln(p/(1 - p))$  and  $\beta_i$  ( $i = 1, \dots, k$ ) measure the relationship between  $Y$  and  $X_i$  when all other variables in the model remained fixed.

Statistical analyses were performed in the SPSS environment. A logistic regression model to a dichotomous response variable was performed with the forward stepwise selection function. The control variables were forced into the model.

### 3. Results

Tables 1 and 2 summarise the survey results. Table 1 shows that 72.8% of the Italians had a clear view about what to do after the pandemic. By contrast, the remaining 27.2% were unable to imagine their future. Their difficulties may stem from their pandemic experience, health status and personality characteristics. The diffusion of mental health problems, as measured with a depression diagnosis, concerns 29.6% of the sample. Moreover, people who have claimed to have experienced psychological damages represent 32.4% of the participants. Of the respondents, 3.1% developed full-blown psychiatric diseases before the survey.

During the pandemic, approximately 21% used the social media as the main information source, and only 11.5% believed that TV programs informed correctly about the pandemic.

**Table 1.** The mean values of the variables used in the statistical analysis

<i>Variable</i>	<i>mean</i>	<i>Variable</i>	<i>Mean</i>
Y: Clearness of future perspective	0.728	X <sub>14</sub> : Infection: parents	0.201
X <sub>0</sub> : Optimistic attitude: Passive	0.327	X <sub>15</sub> : Suffered psychologic damages	0.324
“ : Reactive	0.317	X <sub>16</sub> : Suffered physical damages	0.122
“ : Proactive	0.349	X <sub>17</sub> : Had controls through swabs	0.696
<b>Personal resources</b>		X <sub>18</sub> : Full-blown depression	0.296
X <sub>1</sub> : Self-efficacy score	0.000	X <sub>19</sub> : Had a psychic disease	0.031
X <sub>2</sub> : Higher education degree	0.563	X <sub>20</sub> : Fear for infection – personal	0.296
X <sub>3</sub> : Single	0.286	X <sub>21</sub> : Fear for infection – Italy	0.410
X <sub>4</sub> : Children in family	0.405	X <sub>22</sub> : Remote learning/working	0.490
X <sub>5</sub> : Marital status: couple	0.548	X <sub>23</sub> : Belonged to a broken family	0.017
X <sub>6</sub> : Possessing own working tools	0.315	<b>Social obstacles</b>	
<b>Social resources</b>		X <sub>24</sub> : Income reduced after pandemic	0.065
X <sub>7</sub> : Vaccinated: Yes	0.745	X <sub>25</sub> : Work time reduced	0.087
“ : Not yet	0.176	X <sub>26</sub> : Lost job during pandemic	0.001
“ : Never	0.077	<b>Control variables</b>	
X <sub>8</sub> : Scientists were crucial in pandemic	0.736	Z <sub>1</sub> : Male (gender)	0.430
X <sub>9</sub> : Family doctor available during pandemic	0.408	Z <sub>2</sub> : Age: 18-34	0.524
X <sub>10</sub> : Hospitals were a source of contagion	0.127	“ : 35-64	0.362
X <sub>11</sub> : Televisions informed correctly	0.115	“ : 65 or more	0.114
X <sub>12</sub> : Social media as main information source	0.206	Z <sub>3</sub> : Employee	0.338
<b>Individual problems</b>			
X <sub>13</sub> : Infection: personal	0.116		



Table 2 shows the results of the multivariate regression analysis with the estimates of the regression betas, their significance, the estimates of the odds ratios ( $exp(\hat{\beta})$ ) and their 95% confidence intervals. The results highlight that the Italians in this study had a clear perception of their future when their self-efficacy and resilience scores were high and when working as an employee (as opposed to self-employed). Also, the coefficients relative to the variables of the optimistic attitude scale are highly significant to explain the FTP and, as expected, the strong positive relationship with proactive attitude and the strong negative correlation with passive attitude. Symmetrically, the vision of the future was blurred and made uncertain owing to psychologic damages, depression and other psychic disturbances caused or exacerbated by the pandemic.

Finally, the use of social media as a main information source entered the model with a negative coefficient. This may mean that during the pandemic, dazed people looked for health information from any source, even though they knew the risk of fake news. In addition, unreliable information about the viral threat and the long-term consequences of the infection might have led people to fear for their future. Definitely, approximate and distorted news from social media might have contributed to the imagining of dramatic future scenarios (Barua et al., 2020).

Finally, gender and age were not significant, which means that there were no gender-related differences or youth-specific difficulty as far as FTP was concerned. For the sake of precision, females and younger people showed, consistent with the literature, that the pandemic had a large negative impact (Carstensen et al., 2020; Eurofound, 2021). Notwithstanding, in the multivariate analysis, these differentials vanished because they were absorbed by significant psychological aspects. Instead, *ceteris paribus*, the self-employed have a vision of their futures that is significantly darker than those of employees.

**Table 2.** Beta estimates of the regression model with clear vision of the future as criterion variable (forward stepwise selection of regressors,  $n = 817$ ; Cox & Snell  $\chi^2 = 25.6\%$ ; Omnibus tests of model coefficients:  $\chi^2 = 237.628$ , significance  $< 0.001$ )

<i>Regressor</i>	$\hat{\beta}$	$se(\hat{\beta})$	<i>sig.</i>	$exp(\hat{\beta})$	95% CI $exp(\hat{\beta})$	
Intercept	1.476	0.275	***	4.376		
Z1: Male (gender)	0.149	0.200	NS	1.161	0.785	1.718
Z2: Age 18-34	0.007	0.237	NS	1.007	0.632	1.602
Z3: Employee	0.761	0.260	**	2.140	1.286	3.561
X12: Social media as main info. Source	-0.618	0.222	**	0.539	0.349	0.832
X1: Self-efficacy score	0.323	0.110	**	1.382	1.114	1.714
X15: Suffered psychologic damages	-0.620	0.204	**	0.538	0.361	0.803
X18: Full-blown depression	-0.824	0.213	***	0.439	0.289	0.667
X19: Had a psychic disease	-1.074	0.509	*	0.341	0.126	0.927
X0: Optimistic attitude: Passive	-0.732	0.210	***	0.481	0.319	0.725
“ : Proactive	1.265	0.285	***	3.544	2.028	6.195

\*\*\*  $0 < \alpha^{oss} < 0.001$ ; \*\*  $0.001 < \alpha^{oss} < 0.01$ ; \*  $0.01 < \alpha^{oss} < 0.05$ ; °  $0.05 < \alpha^{oss} < 0.1$ ; NS= Not significant

#### 4. Discussion and conclusion

In this work, we analysed how Italians went through the pandemic and are perceiving their futures. A main outcome of this study was that the susceptibility to and the severity of a potential viral infection were not a significant threat for FTP. Instead, the frustration from such a powerful virus in comparison with humans' vulnerability, together with the procrastination of the national government to implement measures to contain the spread of the virus and the economic, financial and occupational turmoil, affected the people's perceptions of their futures (see also Rupperecht et al., 2022). This caused malaise and depression.

Indeed, the end of the COVID-19 pandemic may be considered a time when many people feel more doubtful than hopeful. Medical researchers have correlated mental disturbances to the delayed effects of COVID-19 infection (among the others: Mattioli et al., 2021). However, it may be argued that such a diffused psychological distress mainly has a social origin.

About one-third of the respondents perceived future opportunities as decreasing and their future lives as more fragile and constrained, which may hamper their activity plans and behaviours. Greater difficulties were highlighted among young people, females and broken or unstructured families. However, a resilient and proactive attitude proved effective against post-pandemic malaise. Thus, gender and age are no longer significant if psychological variables are considered.

The socio-emotional selectivity theory (Lang and Carstensen, 2002) assumes that perceiving one's future as limited and constrained forces a selection of emotionally meaningful goals, whereas an extended FTP allows the selection of instrumental and knowledge-related goals. Conversely, a distorted perception of the future may force some people into an irrational and emotional selection of their own goals.

Ling et al. (2022) argued that a proactive and future-oriented personality is an indicator of an adaptive capacity that can favour successful changes in people's lives. The improvement of FTPs makes individuals believe that their futures are widely open and that time to realise their plans is abundant; thus, they tend to expand their horizons and widen their social circles.

After such a dramatic social shock due to the COVID-19 pandemic, it is relevant to measure people's capacity to start new life strategies in an aware and purposeful manner. Precisely, to effectively imagine one's own future, one has to frame it as clearly as possible upon a social background. If people aspire to master their futures, first, they need to determine the social background of their plans and behaviours. Many people feel as if the pandemic has cast a long shadow against the future social life.

During one's lifetime, there is a mutual feed between resilient, proactive and future-oriented attitudes, so it is difficult to state which one follows the other in a causal chain. Ideally, it is a convolution of two positive attitudes, resilience and self-effectiveness, that may strengthen people's FTP. Both attitudes, particularly resilience, are ideally dynamic in the sense that they imply, on the one hand, a situation to improve and, on the other hand, the disposition to utilize psychological resources to fulfil that aim (O'Neill et al., 2022). Indeed, they strongly correlate with each other and with FTP.

The COVID-19 breakthrough was a social event that affected both individual and collective feelings. Hence, therapies to restore individuals' mental health and their capacity to figure out their own futures could be ineffective if they anticipate that the purpose of social and political interventions was to picture a medium-term social background and empower people's resilience and self-effectiveness capacity.

Finally, the COVID-19 pandemic has emphasised the dramatic role of misinformation through social media (Barua et al., 2020). As evidenced in other studies (Elbarazi et al., 2022; Xie and Liu, 2022), the haphazard use of social media is often associated with poor well-being, negative emotions and fear of infection. Our study highlights how this issue can affect people's ability to imagine and proactively build their own future. Safe public information is one that provides the foundation upon which a clear social background and, therefore, people's future are built.

## References

- Ajzen, I. (1991). Theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50: 179-211.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review*, 84(2): 191-215.
- Barua, Z., Barua, S., Aktar, S., Kabir, N., Li, M. (2020), Effects of misinformation on COVID-

- 19 individual responses and recommendations for resilience of disastrous consequences of misinformation, *Progress in Disaster Science*, 8: 100119.
- Beck, A. T., Weissman, A., Lester, D., Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, 42(6): 861–865.
- Commodari, E., La Rosa, V.L. (2020). Adolescents in quarantine during COVID-19 pandemic in Italy: Perceived health risk, beliefs, psychological experiences and expectations for the future. *Frontiers in Psychology*, 23(11):559951. doi: 10.3389/fpsyg.2020.559951
- Connor, K.M., Davidson, R.T. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC), *Depression and Anxiety*, 18: 76–82.
- de Terte, I., Stephens, C. (2014). Psychological resilience of workers in high-risk occupations. *Stress and Health*, 30(5): 353–355. doi:10.1002/smi.2627. ISSN 1532-3005.
- Elbarazi, I., Saddik, B., Grivna, M., Aziz, F., ElSORI, D., Stip, E., Bendak, E. (2022), The impact of the COVID-19 “infodemic” on well-being: A cross-sectional study, *Journal of Multidisciplinary Healthcare*, 15: 289-307.
- Eurofound (2021). *Living, Working and COVID-19 (Update April 2021): Mental Health and Trust Decline across EU as Pandemic Enters Another Year*. Luxembourg: Publications Office of the European Union.
- Fishbein, M., Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA.: Addison-Wesley.
- Husman, J., Shell, D.F. (2008). Beliefs and perceptions about the future: A measurement of future time perspective. *Learning and Individual Differences*, 18(2): 166-175.
- Lang, F.R., Carstensen, L.L. (2002). Time counts: Future time perspective, goals and social relationships, *Psychology and Aging*, 17(1):125-39.
- Ling, H., Teng, S., Liu, X., Wu, J., Gu, X. (2022). Future work self salience and future time perspective as serial mediators between proactive personality and career adaptability. *Frontiers in Psychology*, 13: 824198. doi: 10.3389/fpsyg.2022.824198
- Mazzotti, E., Fassone, G., Picardi, A., Sagoni, E., Ramieri, L., Lega, I., Camaioni, D., Abeni, D., Pasquini, P. (2003). Il Patient Health Questionnaire (PHQ) per lo screening dei disturbi psichiatrici: Uno studio di validazione nei confronti della Intervista Clinica Strutturata per il DSM-IV asse I (SCID-I). *Italian Journal of Psychopathology*, 9(3): 235-242.
- O’Neill, E., Clarke, P., Fido, D., Correa Vione, K. (2022). The role of future time perspective, body awareness, and social connectedness in the relationship between self-efficacy and resilience. *International Journal of Mental Health and Addiction*, 20: 1171–1181.
- Roseman, I.J. (2013). Appraisal in the emotion system: Coherence in strategies for coping. *Emotion Review*, 5(2): 141-149.
- Rupprecht, F.S., Martin, K., Kamin, S.T., Lang, F.R. (2022). COVID-19 and perceiving finitude: Associations with future time perspective, death anxiety, and ideal life expectancy. *Psychology and Aging*, 37(2), 260–271.
- Simons, J., Vansteenkiste, M., Lens, W., Lacante, M. (2004). Placing motivation and future time perspective theory in a temporal perspective. *Educational Psychology Review*, 16(2): 121-139.
- Spitzer, R.L., Kroenke, K., Williams, J.B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ Primary Care study. *Journal of the American Medical Association*, 282(18): 1737-1744.
- Van Calster, K., Lens, W., Nuttin, J.R. (1987). Affective attitude towards the personal future: Impact on motivation in high school boys. *American Journal of Psychology*, 100: 1–13.
- Zimbardo, P. G., and Boyd, J. N. (1999). Putting time in perspective: a valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, 77: 1271–1288.
- Xie, J., Liu, L. (2022). Identifying features of source and message that influence the retweeting of health information on social media during the COVID-19 pandemic, *BMC Public Health*, 22: 805. <https://doi.org/10.1186/s12889-022-13213-w>.

# Monitoring and evaluation of gender equality policies

Giuliana Coccia, Emanuela Scavalli

## 1. Introduction

The 2030 Agenda for Sustainable Development and its 17 Sustainable Development Goals (SDGs) adopted by world leaders in 2015, embody a roadmap for progress that is sustainable and leaves no one behind.

The global SDG indicator framework establishes a set of measurement tools to assess country performances in a comparable way, and helps governments to identify appropriate policy interventions to achieve the SDG targets. Seven years into the implementation of the 2030 Agenda, however, still different methods are being used by leading international organisations for assessing whether the SDG targets will be achieved or not. This may lead to different results, sometimes contradictory, generating confusion among users and policy-makers, who therefore cannot base their policy decisions on solid and coherent assessments. International organisations address two distinct measurement objectives: (i) monitor the “current” status of achievement of a SDG target, i.e. the situation as pictured by the latest available data, and (ii) assess whether the SDG targets can be achieved by 2030. These distinct objectives are then translated in various methodological approaches, that often include also a way for identifying the targets when not explicitly set, and the procedure to obtain regional and global aggregates (as well as, aggregates by target and goal).

Gender inequality is one of the biggest obstacles to sustainable development, economic growth and poverty reduction. SDG 5 advocates equal opportunities for men and women in economic life, the elimination of all forms of violence against women and girls, the elimination of early and forced marriage, and equal participation at all levels. Ending all forms of discrimination against women and girls is not only a basic human right, but it also has a multiplier effect across all other development areas.

Monitor progress towards the Sustainable Development Goals (SDGs) at national level requires an appropriate set of metrics statistics and indicators on the situation of women and men are needed to describe the roles of women and men in society, the economy, and within the family, to provide the basis for the development of SMART policies and establish sound monitoring and evaluation of their effectiveness. They can help us to reflect upon the challenges strict gendered roles in society present, and demonstrate the negative or positive changes in the status of women in comparison to men in areas such as education, work, access to resources, health or decision-making.

Monitoring is defined as a continuing function that uses the systematic collection of data on specified indicators to provide management and key stakeholders of an ongoing intervention, with indications both of the level of progress and achievement of the objectives as well as the use of any allocated fund.

In this paper, after a recognition of the international indicators related to SDG 5 and the possible sources of production of statistical data, the Italian indicators are analysed in term of current production, reliability and timeless.

## 2. Indicators for monitoring SDG 5

At the international level monitoring and evaluation of 17 DSGs is based on a statistical indicators system developed by Inter Agency Expert Group on SDGs (IAEG-SDGs) and endorsed by the UN Statistical Commission (United Nations 2017).

Giuliana Coccia, ASviS - Alleanza per lo Sviluppo Sostenibile, Italy, giuliana.coccia1@gmail.com

Emanuela Scavalli, ISTAT, Italian National Institute of Statistics, Italy, scavalli@istat.it

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Giuliana Coccia, Emanuela Scavalli, *Monitoring and evaluation of gender equality policies*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.48, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 275-280, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

With respect to each target of SDG 5 the follow indicators are defined as follows:

**Target 5.1 End all forms of discrimination against all women and girls everywhere**

*Indicator 5.1.1* Whether or not legal frameworks are in place to promote, enforce and monitor equality and non-discrimination on the basis of sex. It measures Government efforts to put in place legal frameworks that promote, enforce and monitor gender equality.

*Unit measure* Percentage (%) of legal frameworks that promote, enforce, and monitor gender equality.

*Data source* The data for the indicator are derived from an assessment of legal frameworks using primary sources/official government documents, in particular laws, policies and action plans.

**Target 5.2 Eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking and sexual and other types of exploitation**

*Indicator 5.2.1* Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age.

*Indicator 5.2.2* Proportion of women and girls aged 15 years and older subjected to sexual violence by persons other than an intimate partner in the previous 12 months, by age and place of occurrence.

*Unit measure* % of victims respect total female population in the same age class.

*Data source* Specialized national surveys dedicated to measuring violence against women and administrative data from health, police, courts, justice and social services, among other services used by survivors of violence. No standard definitions and methods have been globally agreed yet to collect data on the place where the violence occurs.

**Target 5.3 Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation**

*Indicator 5.3.1* Proportion of women aged 20–24 years who were married or in an informal union before age 15 and before age 18.

*Unit measure* % of subject respect total female population in the same age class.

*Data source* National censuses, other national household surveys or administrative data.

*Indicator 5.3.2* Proportion of girls and women aged 15–49 years who have undergone female genital mutilation/cutting, by age.

*Unit measure* % of subject respect total female population in the same age class.

*Data source* UNICEF undertakes a wide consultative process of compiling and assessing data from national sources for the purposes of updating its global databases on the situation of children.

**Target 5.4 Recognise and value unpaid care and domestic work through the provision of public services, infrastructure and social protection policies and the promotion of shared responsibility within the household and the family as nationally appropriate**

*Indicator 5.4.1* Proportion of time spent on unpaid domestic and care work, by sex, age and location.

*Unit measure* Data are expressed as a proportion of time in a day.

*Data source* Time-use information collected by a specific survey.

**Target 5.5 Ensure women’s full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life**

*Indicator 5.5.1* Proportion of seats held by women in (a) national parliaments and (b) local governments

*Unit measure* % of women on total elected members.

*Data source* National Parliaments, administrative data based on electoral records.

*Indicator 5.5.1* Proportion of women in managerial positions.

*Unit measure* It is recommended to use two different measures jointly for this indicator: the share of females in (total) management and the share of females in senior and middle management (Gennari F., 2015).

*Data source* Labour force survey, administrative registers.

**Target 5.6 Ensure universal access to sexual and reproductive health and reproductive rights as agreed in accordance with the Programme of Action of the International Conference on Population and Development and the Beijing Platform for Action and the outcome documents of their review conferences**

*Indicator 5.6.1* Proportion of women aged 15–49 years who make their own informed decisions regarding sexual relations, contraceptive use and reproductive health care. Women who make their own decision regarding seeking healthcare for themselves are considered empowered to exercise their reproductive rights.

*Unit measure* Proportion of women aged 15-49 years (married or in union) who make their own decision on all three selected areas.

*Data source* National health survey.

*Indicator 5.6.2* “Number of countries with laws and regulations that guarantee full and equal access to women and men aged 15 years and older to sexual and reproductive health care, information and education”.

*Unit measure* Proportion of Countries.

*Data source* Official government responses collected through the United Nations Inquiry among Governments on Population and Development. Data collection is scheduled every 4 years.

**Target 5.a Undertake reforms to give women equal rights to economic resources, as well as access to ownership and control over land and other forms of property, financial services, inheritance and natural resources, in accordance with national laws**

*Indicator 5.a.1* (a) Proportion of total agricultural population with ownership or secure rights<sup>1</sup> over agricultural land, by sex; and (b) share of women among owners or rights-bearers of agricultural land, by type of tenure.

*Unit measure* % of women on total agricultural population.

*Data source* Agriculture Census, Agricultural Administrative Registers.

**Target 5.b Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women**

*Indicator 5.b.1* Proportion of individuals who own a mobile telephone. The mobile phone is a personal device that, if owned and not just shared, provides women with a degree of independence and autonomy, including for professional purposes.

*Unit measure* % of females have a mobile telephone.

*Data source* National household surveys.

**Target 5.c Adopt and strengthen sound policies and enforceable legislation for the promotion of gender equality and the empowerment of all women and girls at all levels**

*Indicator 5.c.1* Systems to track and make public allocations for gender equality and women’s empowerment, to measure government efforts to track budget allocations for gender equality throughout the public finance management cycle.

This indicator aims to encourage national governments to develop appropriate budget tracking and monitoring systems and commit to making information about allocations for gender equality readily available to the public.

---

<sup>1</sup> Secure rights” in the context of indicator 5.a.1 is defined as secure tenure rights, i.e., rights to use, manage and control land, fisheries and forests

The above indicators are classified as:

**Tier 1:** Indicator is conceptually clear, has an internationally established methodology and standards are available, data are regularly produced by countries for at least 50 per cent of countries and of the population in every region where the indicator is relevant.

**Tier 2:** Indicator is conceptually clear, has an internationally established methodology and standards are available, but data are not regularly produced by countries.

**Tier 3:** No internationally established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested.

According to the update of 28 December 2020, 130 indicators belong to Tier I, 97 indicators to Tier II, and four indicators belong to several Tiers (different components of the indicator are classified into different levels), while none indicator is in Tier III.

With regard to SGD 5 the indicators 5.2.1, 5.3.1, 5.3.2 e 5.6.1 are Tier 1, the others Tier 2.

The global indicator framework set was approved during the 48th Conference of UN Statistical Commission. Through the activities of the High-Level Political Forum on Sustainable Development (HLPF) (central element of the United Nations), each year the progress and results of the political actions of all members of the United Nations are evaluated (ONU, <https://unstats.un.org/sdgs/>).

The initial set of indicators to be refined annually and reviewed comprehensively by the Commission at its fifty-first session, held in 2020, and its fifty-sixth session, to be held in 2025, and will be complemented by indicators at the regional and national levels, which will be developed by Member States.

### 3. Indicators for monitoring SDG 5 in Italy

Data production is essential to guide, inform and empower governance and decision-making. For this reason and to make up for the non-constant availability and reliability of up-to-date information, the United Nations, in addition to the various specialised agencies, trust in the responsibility of the individual States to submit regularly on a voluntary basis (Voluntary National Review, VNR), accessible, rigorous data. and transparent, disaggregated by sex, age, income and any other relevant characteristic to assess the progress of the SDGs at national and regional level.

In Italy, the official body in charge for SDGs metrics is the National Institute of Statistics. Istat has the task of coordinating the institutions belonging to the National Statistical System (Sistan) in the statistical production of data, but, as a matter of fact, other bodies are also involved in sub-regional monitoring action (e.g., PoliS-Lombardia).

The Italian Alliance for Sustainable Development (ASviS)<sup>2</sup> produces an annual report entitled “Rapporto ASviS. L’Italia e gli obiettivi di Sviluppo Sostenibile”, in which he analyses the achievement of the SDGs at the national level and presents policy proposals. ASviS established an interactive online database (available in the page “[The numbers of sustainability](#)”), this allows stakeholders, the media and the public to verify the Italy’s progress with respect to the SDGs, using a wide range of statistical indicators, among those selected by the UN for the 2030 Agenda, released by the Istat, as well as the composite indicators relative to each SDG calculated by ASviS for Italy and the Italian regions (cfr. [ASSET \(futurast.it\)](#)).

In this paragraph, we analyse the Italian status of SDG 5 monitoring, based on recent official reports, highlighting critical issues in the availability of statistical data. For each Italian indicator we assessed the timeliness and reliability of the statistical data required for the development internationally harmonised indicators.

In case of lack of data, we evaluated other indicators produced in our country able to signal the phenomenon to be monitored, as indicated below.

---

<sup>2</sup> ASviS mission is to raise the awareness of the Italian society, economic stakeholders and institutions on the importance of the Global Agenda for sustainable development, bringing together actors who already deal with specific aspects related to the Sustainable Development Goals.

The **indicator 5.1.1** “Whether or not legal frameworks are in place to promote, enforce and monitor equality and non-discrimination on the basis of sex” is very difficult to calculate, since it is a qualitative rather than quantitative assessment of the legislation in force in our Country. This indicator only plays a part in the international comparison, counting Countries that have legal frameworks versus those without.

With regards to **indicator 5.2.1** “Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age” and **indicator 5.2.2** “Proportion of women and girls aged 15 years and older subjected to sexual violence by persons other than an intimate partner in the previous 12 months, by age and place of occurrence”, a significant proportion of data are obtained from Italian Crime Victimization Surveys, that Istat carried out every 5 years.

To permit ongoing monitoring, it is necessary to use administrative data of the Ministry of the Interior on complaints of violence and murders, and administrative data from health, justice and social public services (e.g., the number of calls to anti-violence 1522, number of women who were welcomed into shelters, etc.). However, it is not yet possible to establish the reliability of this administrative information.

Regarding the **indicator 5.3.1** “Proportion of women aged 20–24 years who were married or in an informal union before age 15 and before age 18”, it is to be emphasised that the measure of child marriage is retrospective in nature by design, capturing age at first marriage among a population that has completed the risk period (i.e., adult women). While it is also possible to measure the current marital status of girls under age 18, such measures would provide an underestimate of the level of child marriage, as girls who are not currently married may still do so before they turn 18. The problem is that early marriage is in large part a submerged and hard to detect reality.

**Indicator 5.3.2** “Proportion of girls and women aged 15–49 years who have undergone female genital mutilation/cutting, by age”. These data must be analysed in light of the extremely delicate and often sensitive nature of the topic. Self-reported data on FGM need to be treated with caution for several reasons. Women may be unwilling to disclose having undergone the procedure because of the sensitivity of the issue or the illegal status of the practice in their country. We have to remember the retrospective nature of these data, which results in this indicator not being sensitive to recent change.

As of 2018, UNICEF launched a new country consultation process with National Statistical Offices (or other national authorities) on selected child-related global SDG indicators.

**Indicator 5.4.1** “Proportion of time spent on unpaid domestic and care work, by sex, age and location” provides an assessment of gender equality, by highlighting discrepancies between how much time women and men spend on unpaid work, like cooking, cleaning or taking care of children. The main data source is the time Istat survey carried every 5 years.

Consequently, an indirect indicator of women's involvement in care has been established given by the ratio between the employment rate of aged women 20-49 years with preschool children and the employment rate of women 20-49 without children (Labour force survey), published yearly by Istat<sup>3</sup>.

With reference to the **Indicator 5.5.1** “Proportion of seats held by women in (a) national parliaments and (b) local governments” there are updated basis on the election results at national and territorial levels. Instead, for the **Indicator 5.5.2** “Proportion of women in managerial positions”, only the percentage of women on the boards of the publicly listed companies is detected, according the Golfo-Mosca Law (L.120/2011).

**Indicator 5.6.1** “Proportion of women aged 15–49 years who make their own informed decisions regarding sexual relations, contraceptive use and reproductive health care” it is based on data collected as part of the five-years Health Survey. The sensitivity of the topics addressed in health surveys, in particular, those examining women's health, making them a feasible instrument for

---

<sup>3</sup> The national plan for gender equality, in contrast, chose the indicator of the difference between the two female employment rates, also this kind of information is not yet published by Istat.



incorporating questions on women's experience of decision making in sex relations, use of contraceptive, and health care for themselves. There is no other national information.

**Indicator 5.a.1** consists of two sub-indicators: a) Proportion of total agricultural population with ownership or secure rights over agricultural land, by sex; and (b) share of women among owners or rights-bearers of agricultural land, by type of tenure. The first one focuses on gender parity, measuring the extent to which women are disadvantaged in ownership or secure rights over agricultural land. Agricultural Censuses can be used for collecting data on SDG 5.a.1, however, the Census is usually conducted every 10 years, therefore, it cannot provide data to closely monitor the progress on indicator.

**Indicator 5.b.1** "Proportion of individuals who own a mobile telephone". Mobile phone networks have spread rapidly over the last decade, however, not every person uses or owns a mobile-cellular telephone. Mobile phone, if owned and not just shared, provides women with a degree of independence and autonomy, including for professional purposes. Currently available data from household survey are referred for cellular use.

To conclude with **indicator 5.c.1** "Proportion of countries with systems to track and make public allocations for gender equality and women's empowerment", we underline that it is aimed only at international comparison. At Country level, it is necessary to know how many Public Administrations have drawn up the gender budgeting. Gender budgeting is an application of gender mainstreaming in the budgetary process. It means a gender-based assessment of budgets, incorporating a gender perspective at all levels of the budgetary process and restructuring revenues and expenditures in order to promote gender equality. As 2018 Italian Ministry of Economy and Finance (State General Accounting Department) publishes gender budgeting (see [Ragioneria Generale dello Stato - Ministero dell'Economia e delle Finanze - Bilancio di genere 2020 \(mef.gov.it\)](https://www.mef.gov.it/ragioneria-generale-dello-stato-ministero-dell-economia-e-delle-finanze-bilancio-di-genere-2020)).

#### 4. Conclusions

The Agenda 2030 and ambitious scope of the Sustainable Development Goals (SDGs) has resulted in a long list of indicators that will need to be monitored at national, regional, and global levels. Many of these indicators are 'aspirational' and will take time and significant resources to produce.

There is a clear lack of detailed and up-to-date information to construct monitoring indicators.

Finally, a further problem arises from the need to monitor SDG5 at the regional level, often due to the non-reliability of data derived from sample surveys.

On the other hand, it should be noted that Italy has not established specific quantitative targets to be achieved by 2030 for topics relating to SDG 5

The United Nations Body for Gender Equality and Women's Empowerment (UN Women) highlights the need to improve the statistical production of data and to identify targeted analysis and monitoring procedures to assess the progress achieved in gender equality. In particular, it suggests strengthening the capacity of national statistical systems and increasing the quantity and quality of data through the use of innovative technologies and methods (Data Revolution).

#### References

- Cavalli, L., Lizzi, G., Toraldo, S. (2020). L'Agenda 2030 in Italia a cinque anni dalla sua adozione: una review quantitativa, *Fondazione Eni Enrico Mattei Report N.12*
- Gennari, F., (2015). *L'uguaglianza di genere negli organi di corporate governance*. Franco Angeli, Milano
- Gennari, P., D'Orazio, M.(2020). A statistical approach for assessing progress towards the SDG targets, *Statistical Journal of the IAOS*, vol. 36, no. 4, pp. 1129-1142.
- Richiedei, A., Pezzagno, M. (2022). Territorializing and Monitoring of Sustainable Development Goals in Italy: An Overview. *Sustainability* **2022**, *14*(5), 3056
- United Nations (2017). *Resolution adopted by the General Assembly on 6 July 2017. A/RES/71/313*

# An experimental annotation task to investigate annotators' subjectivity in a misogyny dataset

Alice Tontodimamma, Stefano Anzani, Marco Antonio Stranisci, Valerio Basile,  
Elisa Ignazzi, Lara Fontanella

## 1. Introduction

In recent years, hatred directed against women has spread exponentially, especially in online social media, where the detachment resulting from being enabled to write without any obligation to reveal oneself directly allows people to feel greater freedom in the way they express themselves, and even to attack a chosen target without risk of being recognised or traced. Although this alarming phenomenon has given rise to many studies both from the viewpoint of computational linguistics and from that of machine learning, less effort has been devoted to analysing whether models for the detection of misogyny are affected by bias (Nozza et al., 2019).

During the last years, the problem of social bias in the field of Natural Language Processing (NLP) has been increasingly considered. Obtaining multiple annotator judgements on the same data instances is a common practice in NLP in order to improve the quality of final labels.

However, the fact that annotators are individuals obviously means that they have their own biases and values, and therefore are often likely to disagree with each other, especially when they are working on subjective tasks which involve detecting offensive language, misogynistic language, and hate speech. These disagreements can have a positive value, since they isolate subtleties in tasks of this kind that are obscured when annotations are combined to create a single ground truth (Davani et al., 2022).

In this work, we present two corpora: a corpus of messages posted on Twitter after the liberation of Silvia Romano on the 9th of May, 2020 and corpus of comments constructed starting from posts on Facebook that contained misogyny, developed through an experimental annotation task, to explore annotators' disagreement. In particular, we propose a qualitative-quantitative analysis of the resulting corpora.

## 2. Related work

The notion of a 'single correct answer' fails to take into account the subjectivity and complexity of many tasks. A task can be defined as 'subjective' when the human judgement is inherently influenced by factors pertaining to the judges themselves, rather than by the linguistic phenomenon, whereas human judgement applied to an 'objective' task depends solely on the object that is being judged. Different people, while annotating a highly subjective task such as offensive language, can differ greatly in how offensive they find various expressions to be: in such cases, the opinions of all the annotators could be seen as valid. In the subjective task scenario, the one-truth assumption is no longer valid (Basile, 2020).

In recent years, proposals have been made to consider disagreement as an information content that can be exploited to improve the performance of tasks (Basile et al., 2021). Uma et al (2020)

Alice Tontodimamma, University of Chieti-Pescara G. D'Annunzio, Italy, [alice.tontodimamma@unich.it](mailto:alice.tontodimamma@unich.it)  
Stefano Anzani, University of Chieti-Pescara G. D'Annunzio, Italy, [s.anzani92@gmail.com](mailto:s.anzani92@gmail.com), 0009-0000-5408-0104  
Marco Antonio Stranisci, University of Turin, Italy, [marcoantonio.stranisci@unito.it](mailto:marcoantonio.stranisci@unito.it), 0000-0001-9337-7250  
Valerio Basile, University of Turin, Italy, [valerio.basile@unito.it](mailto:valerio.basile@unito.it), 0000-0001-8110-6832  
Elisa Ignazzi, University of Chieti-Pescara G. D'Annunzio, Italy, [elisa.ignazzi@studenti.unich.it](mailto:elisa.ignazzi@studenti.unich.it)  
Lara Fontanella, University of Chieti-Pescara G. D'Annunzio, Italy, [lara.fontanella@unich.it](mailto:lara.fontanella@unich.it), 0000-0002-5441-0035

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Alice Tontodimamma, Stefano Anzani, Marco Antonio Stranisci, Valerio Basile, Elisa Ignazzi, Lara Fontanella, *An experimental annotation task to investigate annotators' subjectivity in a misogyny dataset*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.49, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 281-286, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

and Basile (2020) studied the impact of disagreement-informed data on the quality of NLP evaluation, and found it to be beneficial and providing complementary information for the quality of classification tasks. There are also authors in contrast with this approach: Bowman and Dahl (2021) recently proposed to study biases and artifacts in data to eliminate them; Beigman Klebanov et al. (2009) adopted a slightly softer stance, proposing to only evaluating on ‘easy’ instances. Basile et al. (2021) argue against this approach, based on the evidence about the prevalence of disagreement in NLP judgments. Removing the disagreement could lead to better evaluation scores, but fundamentally it hides the true nature of tasks. Furthermore, the reduction of noise in the data leads to a loss of information.

Our work contributes to the topic of investigating the impact of disagreement on computational resources by presenting an experimental annotation pipeline aimed at enhancing the subjectivity of annotators. Rather than being bound to a rigid set of labels, annotators were asked to label texts with an open-ended annotation, highlighting the portion of text that they considered to be misogynistic. This type of task had already been proposed, for example in Toxic Spans Detection, which is a task at SemEval 2021 (Pavlopoulos et al., 2021). In fact, in Toxic Span Detection participants were asked to identify toxic spans, i.e., proportion of text that were responsible for the toxicity of the posts, when identifying such spans was possible.

### **3. Dataset creation and description**

The dataset creation process involved trainees engaged in an internship program, who participated in two annotation tasks. They first annotated a corpus of 760 messages posted on Twitter after the liberation of Silvia Romano on the 9th of May, 2020. Tweets were obtained through the official Twitter API and filtered by keywords: only messages published from the 9th to the 16th of May and containing the mention of Silvia Romano were collected and sampled.

For the second task, trainees labelled 784 Facebook comments. We started from a total of 57826 Facebook comments to post directed to women and selected by the trainees themselves. These comments were scraped using exportcomments.com. For the annotation task, we extracted a sample from this corpus using the revised HurtLex dictionary (Tontodimamma et al., 2022), an Italian lexicon of offensive, aggressive, and hateful words divided into 21 categories. Specifically, we used three categories: derogatory words, words related to prostitution, and words used to offend, insult, or denigrate women, which we consider could be used to create a subset. Using this filter, we retained only comments containing words that belong to these three categories and that occur at least 8 times. The final dataset for the annotation task comprises 784 comments.

### **4. Annotation task**

For a given comment, the annotation procedure consists in selecting one or more chunk from each text that is regarded as misogynistic and establishing whether a gender stereotype is present. Each comment is annotated by at least three annotators in order to better analyse their subjectivity. The annotation process was carried by 13 trainees (2 males, 11 females, students on the Sociology degree course) who were engaged in an internship program in the Computational Social Research Lab<sup>1</sup>.

### **5. Quantitative-qualitative analysis of disagreement**

As a result of the annotation task, 2,207 annotations of tweets about Silvia Romano and 4,942 annotations of Facebook posts were collected. Each Facebook message obtained 3 annotations, while 4 annotations were provided for each Tweet.

---

<sup>1</sup> <http://csrlab.unich.it/>.

Since annotation tasks about abusive language are highly prone to subjectivity (Basile *et al.*, 2021) and chunk selection tasks often result in significant disagreement, in this section a quantitative and qualitative analysis of disagreement is provided. The computation of the Inter Annotation Agreement (IAA) relied on Cohen’s Kappa (Fleiss, 1969) for labels, and F1-measure (Lehnert, 1992) for spans.

Specifically, Cohen’s kappa is designed for measuring the agreement between two raters and it is defined in the following way:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}.$$

Here  $p_0 = \sum_{i=0}^1 p_{ii}$  denotes the proportion of observed agreement in the labels between two annotators, and  $p_e = \sum_{i=0}^1 p_{i.} p_{.i}$  the proportion of chance agreement.

When multiple raters are considered, the kappa statistics computed from each possible pair of raters are averaged. Kappa has value 1 if there is perfect agreement between the raters, and value 0 if the observed agreement is equal to agreement expected by chance. Several authors have suggested interpretation or benchmark guidelines for values between 0 and 1. Landis and Koch (1977) proposed the following guidelines: 0.00 - 0.20 indicates slight agreement, 0.21-0.40 fair agreements, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 indicates almost perfect agreement.

The IAA on chunk selection was computed only on messages annotated with the same label and was computed through averaged pairwise F1-measure, which is the harmonic of precision and recall. In this setting, the annotations of one annotator are used as the reference against which the annotations of the other annotator are compared. The average F1-measure among all pairs of raters can be used to quantify the agreement among the raters. The higher the average F1-measure, the more the raters agree in the span selection.

Table 1 shows the IAA agreement for both labelling and span detection activities. Values are the average of Cohen’s Kappa scores and F1-measures obtained by each annotator against the others who annotated the same part of the dataset. In order to account the differences between single annotators we also computed the standard deviation for all tasks and activities.

		Twitter’s Corpus	Facebook’s Corpus
labels Cohen’s Kappa	Mean	0.228	0.210
	<i>Std</i>	<i>0.120</i>	<i>0.090</i>
spans F1-measure	Mean	0.232	0.299
	<i>Std</i>	<i>0.070</i>	<i>0.190</i>

Table 1: Mean and standard deviation of Cohen’s Kappa coefficients scored by each annotator and F1-measure.

From a general overview of Cohen’s Kappa scores first emerges a low agreement in both tasks. Annotators averaged an agreement of 0.228 on the Silvia Romano’s task, and of 0.210 on the Facebook posts task. It is worth mentioning the high standard deviation between annotators, which is 0.12 for the former task and 0.09 for the latter.

For the F1-measures results show that annotators obtained a higher agreement selecting span from Facebook posts than from tweets about Silvia Romano. However, the standard deviation is significantly higher: 0.19 for Facebook posts against 0.07 for Silvia Romano tweets.

The qualitative analysis was carried out by manually inspecting the highlighted chunks from couples of annotations that scored particularly high or particularly low on the measure of similarity. From the quantitative analysis, it emerges that annotators obtained a higher agreement selecting span from Facebook posts than from tweets about Silvia Romano: such a result could be explained by the different domains of the Silvia Romano dataset. In fact, even though the tweets mention Silvia Romano, this dataset also contains many offensive comments and words on Islamophobia

and choices made by the Italian government, and not always as offensive comments against Silvia Romano.

Looking at annotations from this last dataset, the comments with more overlap are often those in which the highlighted spans coincide with the entire text. Moreover, it is possible to observe that some of these comments are directed to Silvia Romano, specifically on her body (traces of body-shaming are evident), others show scepticism about Stockholm syndrome, and some are explicit death threats (see table 2 Silvia Romano Id 1040). On the other hand, the comments with less overlap are often those pertaining different domains, such as the government, or religion, which were not the main target of the annotation task (see table 2 Silvia Romano Id 395).

Source	Text	Chunk 1	Chunk 2
Silvia Romano Id 1040	Silvia Romano stai attenta che se si dovesse manifestarsi qualche attentato da parte del gruppo in cui ti sei convertita,ti troveremo e ti faremo a pezzi,altro che sciabole...	<i>Silvia Romano stai attenta che se si dovesse manifestarsi qualche attentato da parte del gruppo in cui ti sei convertita,ti troveremo e ti faremo a pezzi,altro che sciabole...</i>	<i>Silvia Romano stai attenta che se si dovesse manifestarsi qualche attentato da parte del gruppo in cui ti sei convertita,ti troveremo e ti faremo a pezzi,altro che sciabole</i>
Silvia Romano Id 395	Ha chiesto il corano. Si è convertita all'Islam. Torna in Italia con gli stessi abiti che indossano le donne islamiche. Abbiamo regalato milioni di euro a terroristi. E Conte e Di Maio l'hanno pure accolta a braccia aperte. Schifo. #SilviaRomano #LiveNoneLadUrso	<i>Conte e Di Maio l'hanno pure accolta a braccia aperte</i>	<i>Schifo.</i>

Table 2: Example of comments with more and less agreement for Silvia Romano dataset.

Regarding Facebook dataset, the comments with more agreement are generally shorter, so again the annotators selected chunks corresponding to the full phrases, it is also noteworthy that almost all of the comments with a very high degree of similarity refer to physical aspects (see table 3 Facebook Id 299). While the comments with less overlap seem to be longer and generally with more offensive terms (see table 3 Facebook Id 77).

Source	Text	Chunk 1	Chunk 2
Facebook Id 299	Bruttissima fa schifo il suo viso sembra plastica 🐼	<i>Bruttissima fa schifo il suo viso sembra plastica</i>	<i>Bruttissima fa schifo il suo viso sembra plastica</i>
Facebook Id 77	Capra,capra,capra!!! NN TOCCARE LA SICILIA!!! Soprattutto noi siciliani!!! Cn moltissimi valori!!!Quelli che nn tieni tu'!!! GALLINA SPENNATA!!	<i>GALLINA SPENNATA</i>	<i>Capra,capra,capra!!</i>

Table 3: Example of comments with more and less agreement for Facebook dataset.

## 6. Conclusion and future work

In this work we present two corpora developed through an experimental annotation task designed to explore disagreement among annotators. For a given comment, the annotation procedure consisted in selecting one or more chunks from each text that is regarded as misogynistic and establishing whether a gender stereotype is present. As a result of the annotation task, 2,207 annotations of tweets about Silvia Romano and 4,942 annotations of Facebook posts were collected.

The analysis of annotations showed a high level of disagreement in both tasks. From the quantitative analysis it emerged that annotators obtained a higher agreement when selecting span from Facebook posts than from tweets about Silvia Romano: such a result could be explained by the different domains of the Silvia Romano dataset. In fact, even though the tweets mention Silvia Romano, this dataset also contains many offensive comments and words on Islamophobia and choices made by the Italian government, and not always as offensive comments against Silvia Romano. In general, the comments with more overlap are often those in which the highlighted spans coincide with the entire text, while the comments with less overlap tend to be longer and generally contain more offensive terms.

Future work will focus on expanding this work into different domains, in order to better analyse how disagreement impacts on computational resources and try to integrate disagreement into modelling and evaluation.

## References

- Basile, V. (2020). It's the end of the gold standard as we know it. on the impact of pre-aggregation on the evaluation of highly subjective tasks. In *2020 AIXIA Discussion Papers Workshop, AIXIA 2020 DP* (Vol. 2776, pp. 31-40). CEUR-WS.
- Basile, V., Fell, M., Fornaciari, T., Hovy, D., Paun, S., Plank, B., ... & Uma, A. (2021). We Need to consider disagreement in evaluation. In *1st Workshop on Benchmarking: Past, Present and Future* (pp. 15-21). Association for Computational Linguistics.
- Beigman Klebanov B., Beigman E., and Diermeier D. 2008. Analyzing disagreements. In *Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics*, pages 2-7, Manchester, UK. Coling 2008 Organizing Committee.
- Bowman, S. R., & Dahl, G. E. (2021). What Will it Take to Fix Benchmarking in Natural Language Understanding?. arXiv preprint arXiv:2104.02145.
- Davani, A. M., Diaz, M., & Prabhakaran, V. (2022). Dealing with disagreements: Looking beyond the majority vote in subjective annotations. *Transactions of the Association for Computational Linguistics*, 10, 92-110.
- Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological bulletin*, 72(5), 323.
- Landis JR., Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics*. 1977 Mar;33(1):159-74. PMID: 843571.
- Lehnert, W., Cardie, C., Fisher, D., McCarthy, J., Riloff, E., & Soderland, S. (1992). University of Massachusetts: MUC-4 test results and analysis. In *Fourth Message Understanding Conference (MUC-4): Proceedings of a Conference Held in McLean, Virginia, June 16-18, 1992*.
- Nozza, D., Volpetti, C., & Fersini, E. (2019, October). Unintended bias in misogyny detection. In *Ieee/wic/acm international conference on web intelligence* (pp. 149-155).
- Pavlopoulos, J., Sorensen, J., Laugier, L., & Androutsopoulos, I. (2021, August). Semeval-2021 task 5: Toxic spans detection. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)* (pp. 59-69).
- Uma, A., Fornaciari, T., Dumitrache, A., Miller, T., Chamberlain, J., Plank, B., ... & Poesio, M. (2021). Semeval-2021 task 12: Learning with disagreements. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)* (pp. 338-347). Association for Computational Linguistics.

Tontodimamma A., Fontanella L., Anzani S., Basile V. (2022). An Italian lexical resource for incivility detection in online discourses. *Quality & Quantity*. <https://doi.org/10.1007/s11135-022-01494-7>.

# Potential risk of gambling products and online gambling among European adolescents

Elisa Benedetti, Gabriele Lombardi, Rodolfo Cotichini, Sonia Cerrai,  
Marco Scalese, Sabrina Molinaro

## 1. Introduction

Gambling addiction is a widespread research topic, suggesting that pathological gambling has characteristics that are similar to those of substance abuse (Blanco et al., 2001), and that a relevant part of increasing social costs associated to gambling are more likely to be paid by the less-well off, and potentially most vulnerable members of the society (Resce et al., 2019). Nowadays, a greater focus is devoted to adolescent gambling behavior, which is caused by the greater availability and accessibility of gambling activities, at the same time generating personal, social and economic costs for the new generations (Hardoon and Derevensky, 2002). Furthermore, it is well-recognized how certain categories of people are more at-risk of becoming problematic gamblers: among them, who experienced difficulties at school, drug users, children of gamblers and, in general, males (Winters et al., 1993), whose participation seem to be favoured by the current gaming culture (Lopez-Fernandez et al., 2019). On the other side, more recent findings about problematic adolescent gamblers suggest how having high support both by families (e.g. parental monitoring) and institutions (in terms of benefits, financial support and inequality reduction) reduces the risks of problematic behaviors (Colasante et al., 2022).

This situation appears to be exacerbated by the venue of online gambling, which makes even more accessible these kind of games for adolescents. The undoubted proficiency of young people in using social media and online tools increases their chances of being exposed to online gambling, especially casino and poker (Griffiths and Parke, 2010; Molinaro et al., 2020). Accordingly, Chóliz (2016) highlights how the characteristics of online gaming make them way more addictive, and their usage (jointly with the number of young pathological gamblers) is increased with their growth and promotion.

The paper is organized as follows: in the second section data are presented, from the 2019 ESPAD cross-sectional survey on European adolescents. Jointly, the estimation strategy based on a probit model with sample selection (Van de Ven and Van Praag, 1981) will be briefly described. In the third section results of the main model will be shown and discussed. Moreover, predicted probabilities will be plotted for subsamples based on four different types of games (lotteries, cards, betting and slot-machines) in order to explore how different games influence the probability of problematic gambling, conditioned on online gaming. Finally, some conclusions are drawn from the obtained results.

The analysis will show how factors important in increasing the chance of playing, are not necessarily important for generating a problematic gambler, who seems to be triggered by a lacking of family support, high money availability, and a social context with many slot and betting gamblers. Indeed, slot-machines emerge as the main game able to induce problematic behaviors also in other games, while young people are less sensible to lotteries, among others. Online gaming always increases the chances of becoming a problematic gambler.

Elisa Benedetti, CNR, National Research Council of Italy, Italy, elisa.benedetti@ifc.cnr.it, 0000-0002-7345-3962  
Gabriele Lombardi, University of Florence, Italy, gabriele.lombardi@unifi.it, 0000-0003-4337-569X  
Rodolfo Cotichini, CNR, National Research Council of Italy, Italy, rodolfo.cotichini@ifc.cnr.it, 0000-0002-3586-1033  
Sonia Cerrai, CNR, National Research Council of Italy, Italy, sonia.cerrai@ifc.cnr.it, 0000-0002-0774-8586  
Marco Scalese, CNR, National Research Council of Italy, Italy, marco.scalese@ifc.cnr.it, 0000-0002-7470-2422  
Sabrina Molinaro, CNR, National Research Council of Italy, Italy, sabrina.molinaro@ifc.cnr.it, 0000-0001-7221-0873

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Elisa Benedetti, Gabriele Lombardi, Rodolfo Cotichini, Sonia Cerrai, Marco Scalese, Sabrina Molinaro, *Potential risk of gambling products and online gambling among European adolescents*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.50, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 287-292, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



## 2. Data and Methods

Data were drawn from the ESPAD cross-sectional survey that collected comparable data on risk-behaviours among students in several European and neighbouring countries, every four years since 1995. The sample (n= 85,420) comes from 33 countries that participated in the 2019 data collection. The data collection was conducted through the self-administration of questionnaires to students in the classroom setting. The study methodology used nationally representative samples of randomly selected classes/schools in which the cohort of students turning 16 years in the survey year completed the standardized ESPAD questionnaire.

The dependent variable of gambling was based on the question asking students about both the frequency of their gambling activity in general and the types of games played (slot machines, cards or dice, lotteries or betting on sports/animals) in the last 12 months. Gamblers were identified as those who had gambled for money on at least one of the four games of chance (slot machines, cards or dice, lottery, betting on sports or animal races) in the last 12 months.

The dependent variable of problem gambling was based on the Consumption Screen for Problem Gambling (CSPG) (Rockloff, 2012). The CSPG consists of three questions measuring: (1) gambling frequency; (2) time spent on gambling; and (3) gambling intensity. Summing up scores, those scoring 4+ points were considered at high risk of problem gambling based on the cut-off indicated in Rockloff (2012). For the purposes of this paper, the terms "gamblers at high risk of problem gambling" and "problem gamblers" are used interchangeably.

The following independent variables - summarized in Table 1 - entered the analysis: Gender; Perceived family support; Perceived friend support; Days of school missed; Highest parental education; Self-reported family well-off w.r.t. other families in the country; Parental monitoring indicator; Indicator of how often parents give money to their children.

**Table 1:** Means and standard deviations (between brackets) for the covariates in the total sample and divided by the four examined outcomes.

	Total		Not Player		Player		Not at-risk		At-risk	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Female	0.529	(0.355)	0.579	(0.494)	0.350	(0.477)	0.385	(0.487)	0.143	(0.350)
Family Support Index	5.747	(1.632)	5.767	(1.624)	5.677	(1.659)	5.703	(1.621)	5.526	(1.854)
Friend Support Index	5.544	(1.647)	5.556	(1.647)	5.502	(1.646)	5.525	(1.615)	5.368	(1.809)
<i>School Missed (Ref: 0 days)</i>										
1-2 days	0.326	(0.469)	0.332	(0.471)	0.305	(0.460)	0.314	(0.464)	0.250	(0.433)
3-5 days	0.201	(0.404)	0.201	(0.401)	0.224	(0.417)	0.225	(0.417)	0.222	(0.416)
5 days or more	0.233	(0.423)	0.217	(0.412)	0.291	(0.454)	0.277	(0.447)	0.375	(0.484)
<i>Highest Parental Education (Ref: Up to non-completed Secondary School)</i>										
Non completed university	0.398	(0.489)	0.391	(0.488)	0.421	(0.494)	0.420	(0.494)	0.432	(0.495)
Completed university	0.408	(0.491)	0.409	(0.492)	0.402	(0.490)	0.403	(0.491)	0.393	(0.488)
<i>Family Well off (Ref: Less off)</i>										
About the same	0.462	(0.499)	0.471	(0.499)	0.428	(0.495)	0.436	(0.496)	0.375	(0.484)
Well off	0.453	(0.498)	0.444	(0.497)	0.487	(0.500)	0.339	(0.473)	0.530	(0.499)
<i>Parental Monitoring (Ref: About Always)</i>										
Sometimes	0.283	(0.450)	0.268	(0.443)	0.335	(0.472)	0.339	(0.473)	0.315	(0.465)
About Never	0.139	(0.346)	0.123	(0.328)	0.198	(0.399)	0.180	(0.384)	0.305	(0.461)
<i>Parents give money (Ref: Seldom/Never)</i>										
Often/Sometimes	0.479	(0.450)	0.480	(0.450)	0.475	(0.499)	0.484	(0.500)	0.418	(0.493)
Almost Always	0.305	(0.460)	0.300	(0.458)	0.325	(0.468)	0.317	(0.465)	0.372	(0.483)
<i>GPI<sub>slot</sub></i>	-0.011	(1.006)	-0.053	(0.966)	0.141	(1.125)	0.061	(1.044)	0.600	(1.426)
<i>GPI<sub>cards</sub></i>	-0.005	(1.006)	-0.050	(0.987)	0.148	(1.009)	0.112	(0.991)	0.354	(1.084)
<i>GPI<sub>lotteries</sub></i>	0.005	(1.006)	-0.055	(0.964)	0.218	(1.118)	0.149	(1.037)	0.620	(1.435)
<i>GPI<sub>betting</sub></i>	-0.007	(0.995)	-0.052	(0.974)	0.166	(1.095)	0.089	(1.022)	0.612	(1.360)
Online Gaming	0.066	(0.249)	0.007	(0.084)	0.280	(0.449)	0.219	(0.413)	0.634	(0.482)
Number of observations	85,420		66,843		18,577		15,837		2,740	

To calculate the Gambling Product Index (GPI) a question was asked for each type of gambling product: Slot machines (fruit machine, new slot, etc.); Cards or dice (poker, bridge, dice,

etc.); Lotteries (scratchcards, bingo, etc.); Betting on sports or animals (horses, dogs etc.). The Gambling Product Index (GPI) is obtained as the standardization of the following formula:

$$GPI_{c,g} = \frac{0 \times N_{c,g,ans1} + 1 \times N_{c,g,ans2} + 24 \times N_{c,g,ans3} + 104 \times N_{c,g,ans4}}{N_c},$$

where  $c$  represents the country,  $g$  the type of game,  $ansx$  indicates the answer that each subject provided to the questions about gambling frequency,  $N$  represents the number of individuals in our sample. Each  $N_{c,g,ansx}$  is multiplied by the yearly frequency of gambling declared in the answers. In order to be as conservative as possible, when the answers indicate an interval, the lower bound of the interval is chosen (e.g. 2-4 times a month corresponds to at least 24 times in a year). Thus, the  $GPI$  can be interpreted as an indicator of the average frequency of gambling for a particular game in a specific country.

The analysis is conducted through a probit model with sample selection correction (Heckman, 1979) as proposed by Van de Ven and Van Praag (1981). In particular, in the selection equation all the individual variables are included in order to determine how they influence the probability of being or not a player, and the environmental variables (i.e.  $GPI_s$ ). In the second step, in order to determine gambler at risk of problematic behavior, the four  $GPIs$  are removed, but the dummy variable indicating the usage of online gaming, which would have shown a perfectly predicted outcome in the first stage, is included. As a robustness checks, estimation are presented also for two separate probit models, not commented for the sake of brevity, since the correlation between the equation implies inconsistent estimations in these cases (Miranda and Rabe-Hesketh, 2006). Finally, through the estimation of separate models on the subsamples of players and problematic gamblers for each of the four type of games, we are able to predicted probabilities (Williams, 2012) for the effect of each game on the others, conditioned to the usage of online gaming.

### 3. Estimation and Results

Starting from the selection equation (Model 3.1 in Table 2), we can notice that females exhibit less chances of playing, a well-known result in the field. As the support of the family decreases the probability of becoming a player, the opposite happens for the friend support, probably due to a peer effect that makes adolescents gamble when a close friend plays, too. The school experience matters a lot: as many days of school are missed, as greater are the chances of playing. The parental education is weakly significant: surprisingly, parents without a secondary education degree have less chances to have gambling children. As it will be explained, this result could be associated to the lower money availability. Economic conditions do not seem to affect the probability in terms of *social comparisons*: indeed, as adolescents who perceive of being poorer than or in line with the average of families in the country do not differ in terms of probability, so those who think of being richer have greater chances of playing. Also in this case, it seems that money availability is very important in generating a player. This is confirmed by the covariate called *Parents give money*: individuals who claim their parents give them money often or sometimes play more than those who receive money never or seldom, and those who obtain money almost always play more than anyone else. Finally, parents who less control where and with whom children are during evening outings have higher chances to have gambling kids. Looking at the  $GPIs$  for the four types of games, it is possible to observe that only lotteries have a positive effect on generating players, while all the other indexes are not significant. Indeed, as betting and slot-machines will emerge as triggers for problematic behaviors, and cards are associated to a playful environment, so lotteries are more enslaving for

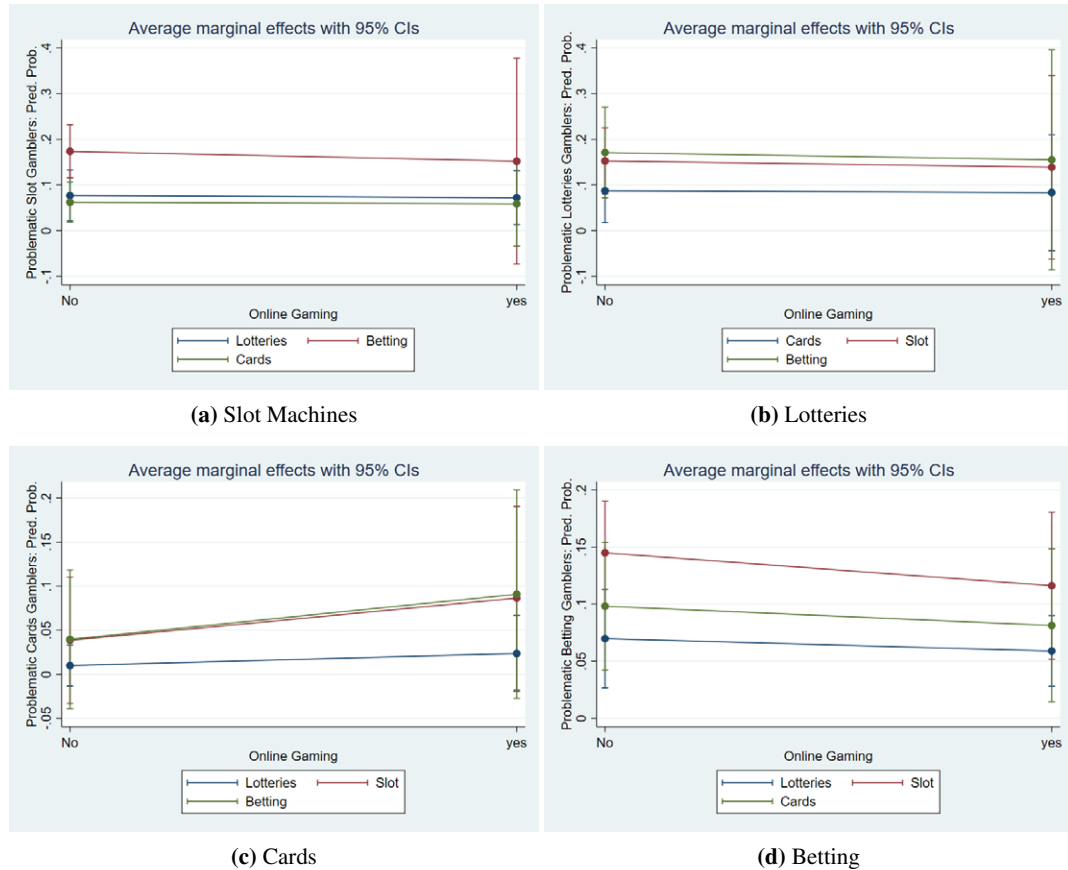
older people, still a benchmark for adolescents (Welte et al., 2007).

**Table 2:** Estimations for two separated Probit models and the joint two-equations Heckprobit model. All observations are weighted.

	Probit Outcome: Not Player/Player (1)		Probit Outcome: Not at-risk/At risk (2)		Heckman Probit First Step: Not Player/Player; Second Step: Not at-risk/At risk; (3.1) (3.2)			
	$\hat{\beta}$	s.d.	$\hat{\beta}$	s.d.	$\hat{\beta}$	s.d.	$\hat{\beta}$	s.d.
Female	-0.540***	(0.035)	-0.521***	(0.056)	-0.539***	(0.034)	0.069	(0.201)
Family Support Index	-0.025***	(0.006)	-0.018*	(0.010)	-0.026***	(0.006)	0.001	(0.009)
Friend Support Index	0.022***	(0.006)	0.004	(0.009)	0.022***	(0.006)	-0.012*	(0.007)
<i>School Missed</i> (Ref: 0 days)								
1-2 days	0.158***	(0.017)	-0.001	(0.036)	0.158***	(0.017)	-0.123***	(0.030)
3-5 days	0.268***	(0.022)	0.087**	(0.035)	0.266***	(0.022)	-0.149**	(0.060)
5 days or more	0.361***	(0.027)	0.255***	(0.041)	0.360***	(0.027)	-0.107	(0.106)
<i>Highest Parental Education</i> (Ref: Up to non-completed Secondary School)								
Non completed university	0.111***	(0.030)	0.081	(0.069)	0.106***	(0.029)	-0.063**	(0.032)
Completed university	0.048*	(0.027)	-0.017	(0.068)	0.049*	(0.027)	-0.082***	(0.030)
<i>Family Well off</i> (Ref: Less off)								
About the same	-0.019	(0.024)	-0.086*	(0.048)	-0.016	(0.024)	-0.026	(0.028)
Well off	0.042*	(0.022)	-0.037	(0.073)	0.047**	(0.021)	-0.020	(0.032)
<i>Parental Monitoring</i> (Ref: About Always)								
Sometimes	0.228***	(0.018)	-0.007	(0.029)	0.228***	(0.018)	0.146***	(0.043)
About Never	0.345***	(0.029)	0.265***	(0.047)	0.344***	(0.029)	0.054	(0.120)
<i>Parents give money</i> (Ref: Seldom/Never)								
Often/Sometimes	0.075***	(0.013)	-0.079**	(0.034)	0.073***	(0.013)	-0.111***	(0.022)
Almost Always	0.145***	(0.018)	0.123**	(0.052)	0.138***	(0.019)	-0.061	(0.046)
Online Gaming	-	-	0.883***	(0.032)	-	-	0.553***	(0.157)
<i>GPI<sub>slot</sub></i>	-0.040	(0.079)	-	-	-0.010	(0.064)	-	-
<i>GPI<sub>cards</sub></i>	0.035	(0.037)	-	-	0.026	(0.029)	-	-
<i>GPI<sub>lotteries</sub></i>	0.174**	(0.072)	-	-	0.147*	(0.076)	-	-
<i>GPI<sub>betting</sub></i>	-0.012	(0.084)	-	-	0.001	(0.067)	-	-
<i>Constant</i>	-0.985***	(0.040)	-1.309***	(0.080)	-1.252***	(0.485)	0.467	(0.417)
Number of observations	85,420		18,577		85,420			
Log-Likelihood	-39950.323		-40471.453		-46829.15			
Pseudo R <sup>2</sup>	0.0681		0.1520		-			
Error Correlations	-	-	-	-	-1.252**		(0.485)	

The model 3.2 in Table 2 analyze the probability of becoming a problematic gambler, conditioned to the fact of having played in the last year, as the error correlation term is significant. Surprisingly, controlling for the selection bias we discover that the gender differences in the probability of becoming a problematic player - conditioned to the fact of having experienced gambling, yet - disappear. Also the support of the family has no effect in conditioning the chances of being at risk, even if as much educated the parents are, as little the probability of being problematic is. Nonetheless, the support of friends is weakly significant and negatively correlated with problematic behaviors: apparently, as friends can stimulate playing, so they can be able to save from problematic gambling. Regarding days of school missed, those who miss the few and the most experience the higher risk. Even if perceived economic conditions are not significant in this context, money availability remains an important factor not only for playing, but also for developing gambling problems. In fact, no difference appears among young people who receive money by parents seldom, never, or almost always. Namely, in this case adolescents who receive less money have the same chances of developing problematic behaviors of those who obtain more. Probably, after having become a player, a social comparison effect can more easily arise, which foster the will of improving their own economic condition, as well as a gambling problem. Playing online is positively significant.

**Figure 1:** Marginal effects for the probability of becoming an at-risk gambler in a specific game by type of game, conditioned to online gaming (95% CIs).



In Figure 1 predicted probabilities are plotted for four separated models in which players are restricted to those who play a specific game and three regressors are added for the other types of game. Thus, it is possible to observe how each game affects the probability of being an at-risk player of another gambling activity. As expected, having been a player of slot-machines and betting increases the probability of being a problematic player on the other games. Accordingly, cards and even more lotteries are the games less effective in causing problematic players in other games. Online gaming increases the chances of problematic behaviors especially in playing cards, while it has no effect with regard to lotteries and slot-machines, and a negative effect looking at betting. Probably, the addiction developed in playing lotteries and slots reflects itself in the high accessibility of online cards game (e.g. poker online).

## 4. Conclusions

This article, based on 2019 ESPAD cross-sectional survey, explores the determinant of gambling and problematic gambling among European adolescents. As a general conclusion, it seems that starting to gamble attains more to what can be called a “social dimension”, while problematic behaviors to the “individual behavior”. Namely, playing in the first step is favored

by factors such as friend support, and parental education, which are components involved in the social context lived by kids. On the other side, these factors lose effectiveness for problematic gambling, much more favored by individual characteristics as the perception of their own economic availability. Indeed, both a very high and very low money availability are always important in strengthening both gambling and problematic gambling. At-risk players are also fostered by those countries with higher shares of lotteries gamers. It is confirmed that on-line gaming, with its high accessibility and availability, is an important trigger for problematic gambling behaviors. Regarding types of games, slot machines and betting emerge as the most addictive.

## References

- Blanco, C., Moreyra, P., Nunes, E.V., Saiz-Ruiz, J., Ibanez, A. (2001) Pathological gambling: addiction or compulsion?, in *Seminars in clinical neuropsychiatry*, **6**(3), pp. 167-176.
- Chóliz, M. (2016). The challenge of online gambling: the effect of legalization on the increase in online gambling addiction *Journal of Gambling Studies*, **32**(2), pp. 749-756.
- Colasante, E., Pivetta, E., Canale, N., Vieno, A., Marino, C., Lenzi, M., Benedetti, E., King, D.L., Molinaro, S. (2022). Problematic gaming risk among European adolescents: a cross-national evaluation of individual and socio-economic factors. *Addiction*, pp. 1-10.
- Griffiths, M. D., Parke, J. (2010). Adolescent gambling on the Internet: A review. *International journal of adolescent medicine and health*, **22**(1), pp. 59-75.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pp. 153-161.
- Hardoon, K. K., Derevensky, J. L. (2002). Child and adolescent gambling behavior: Current knowledge. *Clinical Child Psychology and Psychiatry*, **7**(2), pp. 263-281.
- Lopez-Fernandez, O., Williams, A. J., Griffiths, M. D., Kuss, D. J. (2019). Female gaming, gaming addiction, and the role of women within gaming culture: A narrative literature review. *Frontiers in Psychiatry*, **10**, pp. 454.
- Miranda, A., Rabe-Hesketh, S. (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *The Stata Journal*, **6**3, pp. 285-308.
- Molinaro, S., Vicente, J., Benedetti, E. (2020). ESPAD Report 2019 Results from the European School Survey Project on Alcohol and Other Drugs. *Luxembourg: Publications Office of the European Union*, isbn: 978-92-9497-546-1. doi:10.2810/877033.
- Resce, G., Lagravinese, R., Benedetti, E., Molinaro, S. (2019). Income-related inequality in gambling: evidence from Italy. *Review of Economics of the Household*, **17**(4), pp. 1107-1131.
- Rockloff, M. J. (2012). Validation of the consumption screen for problem gambling (CSPG). *Journal of Gambling Studies*, **28**(2), pp. 207-216.
- Van de Ven, W. P., Van Praag, B. M. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics*, **17**(2), pp. 229-252.
- Welte, J. W., Barnes, G. M., Wieczorek, W. F., Tidwell, M. C. O., Hoffman, J. H. (2007). Type of gambling and availability as risk factors for problem gambling: A Tobit regression analysis by age and gender. *International Gambling Studies*, **7**2, pp. 183-198.
- Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, **12**2, pp. 308-331.
- Winters, K. C., Stinchfield, R., Fulkerson, J. (1993). Patterns and characteristics of adolescent gambling. *Journal of Gambling Studies*, **9**(4), pp. 371-386.

# An Open Data platform for decision making in local public administration

Giuseppe Sindoni, Matteo Massenzio

## 1. Introduction

This paper presents the Milan Open Data (OD) platform as a means of providing statistics and data in the framework of “Data-Driven Milan”, a city where policy decisions are taken in an “informed and aware” way using data. Such a strategic approach is enabled by the enormous amount of data available to public administrations. This includes not just the well-known big data, but also all the data automatically produced by digitalized systems, such as citizen relations systems or systems issuing permits for occupation of public areas. The former can be analysed in real time to understand citizens’ needs and adjust service development policies accordingly, while the latter, integrated with maps of the city, enable every event to be kept under control and any clashes between events of a different nature to be managed more efficiently.

Data-Driven Milan has been implemented since 2016 through a data exploitation strategy aimed at developing a digital platform system to collect and safely integrate data for use in analysis reports, dashboards and geographic intelligence applications and to publish easily accessible open data to share knowledge with citizens.

OD are ever more important in providing citizen communities with useful information. Over the last 10 years, the municipality of Milan has developed its OD platform from an experimental portal to a fully-fledged portal with more than 1,600 datasets, implemented a Linked Open Data (LOD) system and 8 advanced data visualization projects, and produced OD policies and operating guidelines.

## 2. The open data portal for Open Government Data

According to the Organisation for Economic Co-operation and Development, “Open Government Data (OGD) is a philosophy - and increasingly a set of policies - that promotes transparency, accountability, and value creation by making government data available to all.” [OECD, 2020]. OGD is about using public data to enforce the transparency of public administration, which generates trust and in turn improves citizen participation and collaboration between public and private organizations.

Citizen participation is about getting feedback, suggestions, ideas and help through public debates on the development of public policies. Collaboration must be implemented by tearing down watertight compartments and hierarchical structures inside and between organizations, by working “horizontally” and locally between organizations with service design tools and flexible methods, and through the involvement of citizens and promotion of cooperation.

In this context, data can help to enforce transparency through the monitoring of public policies, for example through data-based communication strategies and impact indicators, and through citizen education, using advanced data visualization and explaining the governance process with data and infographics.

Figure 2.1 shows how Milan’s OD strategy has developed from the launch of the portal 10 years ago to the publication of the first Report on the Council’s results, as well as its constantly

increasing number of datasets.

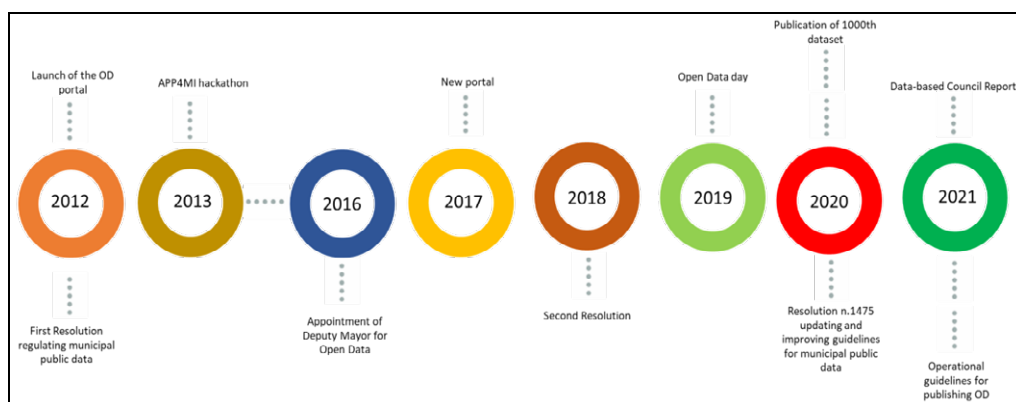


Fig. 2.1 – The history of the Milan Open Data strategy.

The constant rise in the number of published datasets<sup>1</sup> is depicted in Figure 2.2. This growth is due to the publishing strategy, by which all datasets representing evolving phenomena are updated whenever a new version is available, and new datasets are created from internal sources or collected from external sources.

The push for the qualitative and quantitative improvement of the municipality of Milan’s public information assets in open format arises from the provisions of European legislation and the digital administration code (so-called CAD [CAD, 2005]) as well as, since 2012, from a series of municipal council resolutions regulating the open data sector.

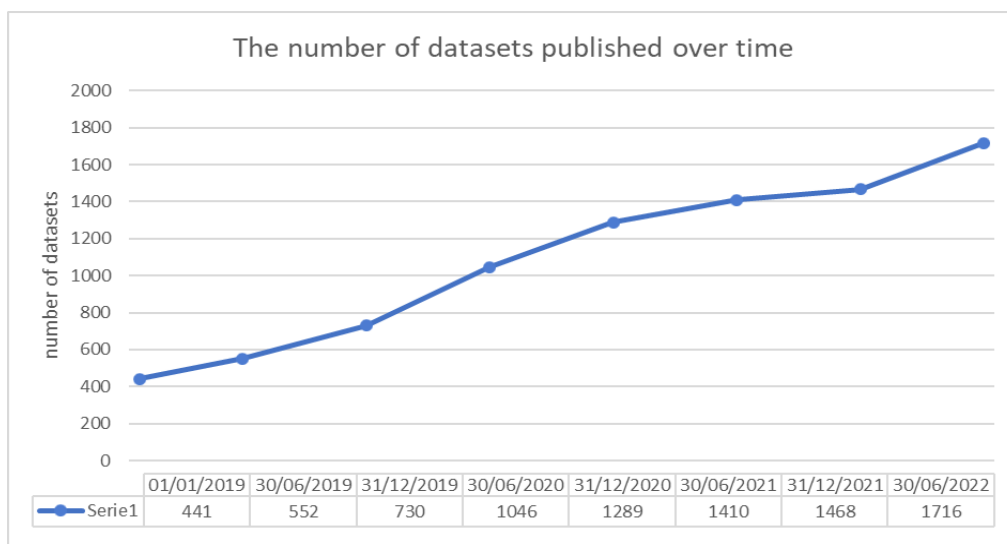


Fig. 2.2 – Increase over time in published datasets.

The increasing number of datasets produced and maintained by Milan has seen it become a national leader in Open Data (winner of the ICity rank editions 2020 and 2021 [ICity rank 2020, 2021]) and place it on a par with major international cities such as London (data.london.gov.uk): 1817 published datasets; Paris (opendata.paris.fr): 335; and New York (data.cityofnewyork.us): 3589.

<sup>1</sup> The Open Data portal is available at: [dati.comune.milano.it](http://dati.comune.milano.it)

The datasets cover the themes of the DCAT-AP-IT (DCAT, 2022) metadata profile. Figure 2.3 shows how they are distributed across the themes.

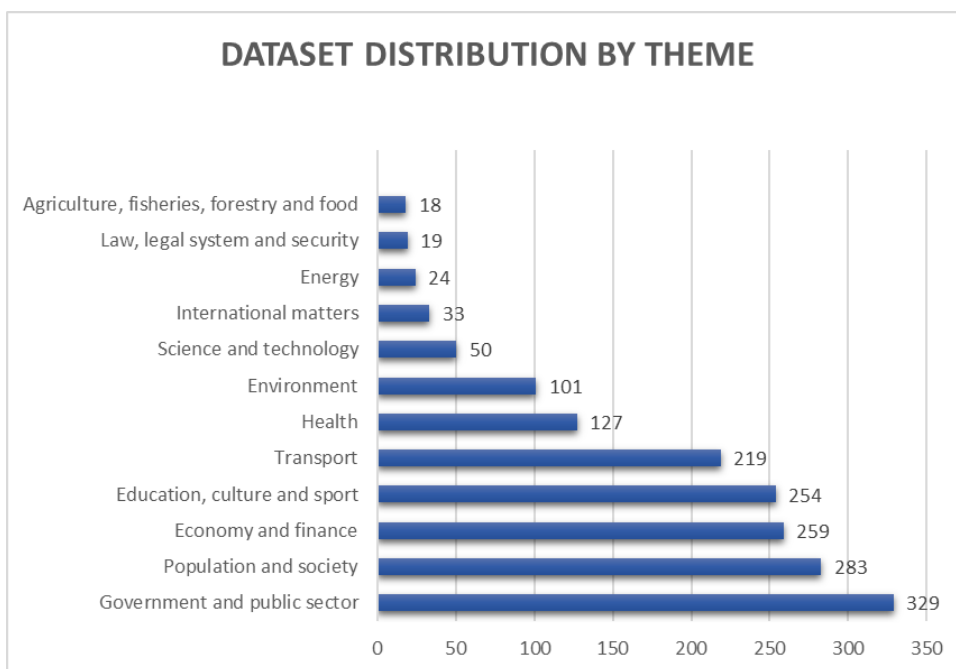


Fig. 2.3 – Distribution of the datasets by theme.

The highly biased distribution is due to the fact that most datasets come from internal sources - for the main part, the digital systems supporting the administration's processes and services. The distribution hence reflects the core themes of the various services.

The portal is based on CKAN technology, which makes datasets available via both download and Application Programming Interfaces. It currently has about 9,500 visits per month. Data are also published as tables on the statistical portal and maps on the geo-portal.

### 3. Linked Open Data

Linked Open Data are semantically enriched machine-readable data that help data interoperability between distributed systems. The international community classifies OD on a 5-star scale based on 3 characteristics: information, access and services. The stars represent an increasing level of usability and accessibility, with 5 stars awarded to the most valuable data: Linked Open Data, which enables both human and automated access to data.



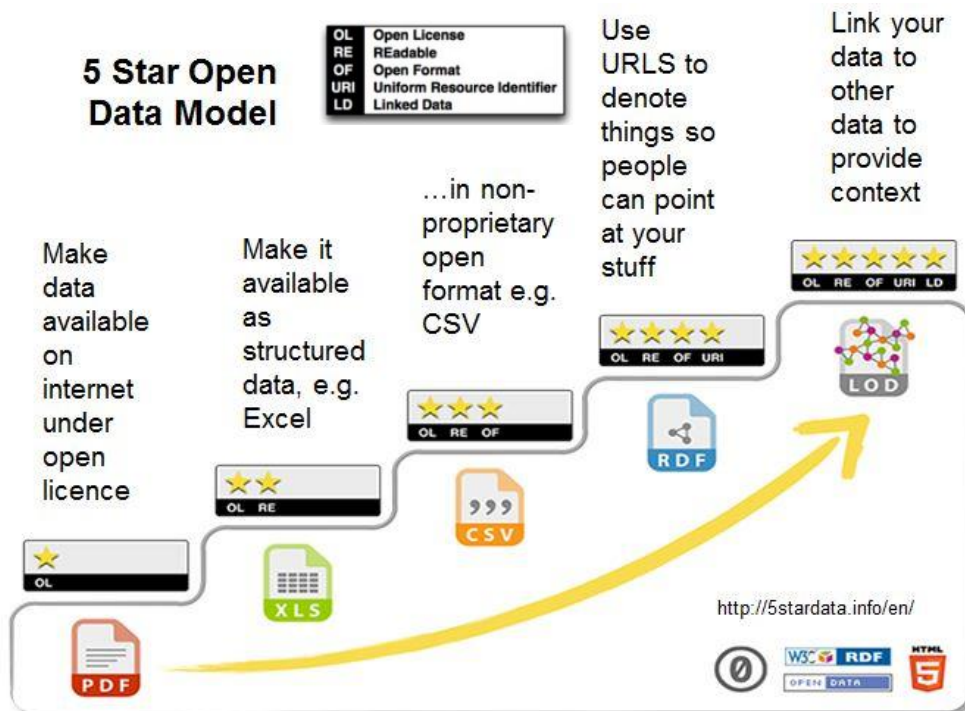


Fig.3.1 – The five stars of Open Data.

LOD are semantically enriched and interlinked, so they enable the development of very efficient data services based on data mashups, where datasets can be used machine-to-machine with automatic integration made possible by their semantic representation through ontologies. The Milan LOD platform<sup>2</sup> is based on 6 ontologies allowing semantic access to the datasets available for the topics covered by each ontology: libraries, public acts, schools, consumer prices, limited traffic zones and sports facilities.

The working model for the design of the ontologies and the implementation of LOD is based on three phases: ontology design, data census and preparation, data loading and graph generation. An ongoing LOD automation project aims to improve the current system by minimizing manual operations in the dataset lifecycle.

#### 4. Data visualization projects

Data visualization projects are part of Milan’s strategy for “data democratization”, i.e. making data accessible and usable to the greatest possible audience. This includes people without specific data manipulation skills who just need objective, easy-to-understand information about the council’s activities and performance. In addition to LOD, seven more special projects have been carried out in the last 4 years to better exploit the open data assets for the benefit of citizens.

<sup>2</sup> The LOD platform is accessible from the OD portal or directly at: [dati.comune.milano.it/sparql/home.html](http://dati.comune.milano.it/sparql/home.html)

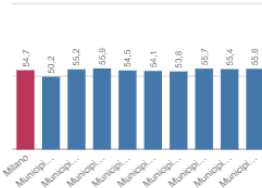


Fig. 4.1 –Openbilancio - current expenses

## Politica e Istituzioni

Partecipazione elettorale (primo turno elezioni comunali 2016)

per 100 aventi diritto



Percentuale di persone che hanno votato al primo turno alle elezioni comunali sul totale degli aventi diritto.

% Donne negli organi decisionali

41,7<sup>2016</sup>  
Anno

Percentuale di donne assessori comunali sul totale degli assessori.

% Donne e rappresentanza pol...

37,5<sup>2016</sup>  
Anno

Percentuale di donne elette nei Consigli Comunali sul totale degli eletti.

Donne e rappresentanza politica a livello locale (consiglio comunale 2016)

per 100 eletti



Percentuale di donne elette nei Consigli Comunali sul totale degli eletti.

Età media negli organi decisionali

45,2<sup>2016</sup>  
Anno

Età media degli assessori comunali alla data delle elezioni

Età media consiglieri comunali

46,9<sup>2016</sup>  
Anno

Età media dei consiglieri comunali alla data delle elezioni

Elaborazione dell'Unità Statistica su fonte dati: Comune di Milano

Fig.4.2 Local analysis of the BES index (sustainable equitable well-being)

The most important data visualization projects are Open Budget and the Council's Mandate Report. Open Budget is an advanced project for the publication of both the final balance and anticipatory municipal budget data. It provides a very advanced user experience and, from this perspective, can be seen as a true data democratization effort. The site was made public in 2018 and is based on data from the Management Executive Plan, published as open datasets since 2013. It has been continuously improved ever since, to provide better usability and more data

transparency.

The Council's Mandate Report is another data democratization project aimed at reporting, through data, the results achieved by the Council during its 2016-2020 term of office (a 2021 update is ongoing). The web site complements the traditional document-based report and offers readers a quantitative view of the Council's performance.

Milan's open data is also widely used by various socioeconomic operators to create their own applications. From this point of view, the Municipality tries to anticipate the needs of stakeholders right from the "Demand" stage, by carrying out various thematic meetings. These meetings have shown that the major users are universities, companies and citizens, who use the data to better direct their choices.

Since 2018, the municipality of Milan has constantly monitored and published information on individual accesses to each dataset:

[https://dati.comune.milano.it/dataset/ds916\\_accessi\\_unici\\_ai\\_dataset](https://dati.comune.milano.it/dataset/ds916_accessi_unici_ai_dataset)

Political decision-makers also make extensive use of open data.

## 5. Next Steps

As part of the broader project to create a data-driven administration, the Municipality of Milan intends to continue strengthening the Opendata Portal. The cornerstone of this approach will be the creation of datasets based on ontologies and glossaries in order to develop an increasing number of high-quality datasets that can be easily made available as Linked Open Data.

## References

- OECD (2020). Open Government Data <https://www.oecd.org/gov/digital-government/open-government-data.htm> Web page
- AGID (2022). *DCAT-AP\_IT v1.1 – Profilo italiano di DCAT-AP*. <https://dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0>. Web page
- ICity rank (2021). <https://www.forumpa.it/citta-territori/icity-rank-2021-firenze-milano-e-bologna-le-citta-piu-digitali-ditalia/> Web page
- ICity rank (2022). <https://www.forumpa.it/citta-territori/icity-rank-2020-firenze-bologna-e-milano-sono-le-citta-piu-digitali-ditalia/> Web page
- CAD (2005) Law decree 7 March 2005, n. 82

# Educational mismatch and productivity: evidence from LEED data on Italian firms

Laura Bisio, Matteo Lucchese

## 1. Introduction<sup>1</sup>

Over the past years, the role of the potential mismatch between the demand and supply of skills and qualifications has received considerable attention in Italy. However, the empirical evidence about the impact of this mismatch onto firms' productivity has not been fully documented so far. In the present paper, we investigate this issue empirically, exploiting the information available from the System of Statistical Registers built within the Italian National Institute of Statistics (Istat).

In particular, we focus on the "educational mismatch", defined as the difference between the educational attainment of workers (the highest level of education the worker has completed) and that "needed" for their job. In this way, over (under) education refers to situations where the individual's educational attainment is higher (lower) than the "required" level, thereby producing a surplus (deficit) of education. Indeed, this mismatch is the result of several overlapping factors, ranging from the adequacy of training to the (in)efficiency of the labour market or the ability of the economic system to absorb skilled labour. The latter issue increasingly depends on the speed at which technological change, and in particular the digitalization process, has changed the demand for skills in the last decades, especially for high-tech and knowledge-intensive industries.

The role of human capital as a key factor in improving firm's competitiveness has been already highlighted by Istat (Istat, 2018); investments in this area have been also recently found to be associated with an increase in firms resilience during the pandemic crisis (Istat, 2021). An analysis of the skill and qualification mismatch for the Italian economy is proposed by OECD (2016) and Monti e Pellizzari (2016), which aimed to provide statistical evidence of the roots of skill mismatch, based on the PIAAC survey results. More recently, the correlation between the ability to match the skills need and labour productivity has been pointed out by Fanti et al. (2021) for a representative sample of Italian firms based on the INAPP PEC survey.

In this paper we explore the effect of over/under education of employed workers on firms' productivity for the Italian economy on the basis of the work of Kampelmann and Rycx (2012), which provides evidence about the direct impact of educational mismatch on productivity using linked employer-employee data for a panel of Belgian firms covering the period 1999-2006.<sup>2</sup> By means of the Istat System of Statistical Registers, we are able to adapt the same analytical framework to Italian data, to contribute filling the gap in the literature about the link between human capital and firms' competitiveness in our economy. The results suggest that over/under education affects productivity growth in both manufacturing and services firms: in particular, over-education rises firm's productivity in medium and high-tech manufacturing firms as well as in less knowledge-intensive services, whereas under-education hampers productivity in manufacturing and services industries with a higher intensity of technology and knowledge.

This paper is organized as follows: section 2 presents the dataset and the empirical methodology; section 3 offers some preliminary descriptive statistics, section 4 shows the results and section 5 draws conclusions.

---

<sup>1</sup> An earlier version of this analysis appeared in the 2022 edition of "Istat Report on Competitiveness" (Istat, 2022).

<sup>2</sup> Mahy et al. (2015) extend the period of analysis to 2010 and highlight, among other results, that the effect of over-education on productivity is stronger in firms belonging to high-tech/knowledge-intensive industries – but with no distinction between manufacturing and services firms.

## 2. Data and empirical analysis

Our analysis is based on the integration of two different Statistical Registers (*Asia-Employment Register* and *Frame-SBS Register*), covering almost the totality of Italian firms. The *Asia-Employment Register* (*Asia Occupazione*) is a LEED-type (Linked Employer Employee Database) one, which allows to obtain information related to firms, the workers employed therein and the main aspects of the work contracts; this dataset also provides information on the level of educational attainment achieved by each worker – via matching to the 2011 edition of the Population Census, updated through the “Information Base on education and qualifications” (*Base Informativa su istruzione e titoli di studio*, BIT). The “*Frame-SBS Register*”, instead, provides data on firms’ main economic and structural characteristics, including labour productivity.

The empirical analysis covers a large set of Italian firms with at least 20 workers over the period 2014–2019. Both labour productivity and mismatch variables are evaluated with respect to employees – i.e. self-employed are excluded from the analysis –, while employment is measured in terms of annual average job positions, based on the worker’s weekly work attendance.<sup>3</sup>

As already mentioned, the empirical analysis is an application to the Italian case of the ORU (Over, Required and Under Education) model performed by Kampelmann and Rycx (2012), based on a longitudinal LEED data structure. The ORU model consists in a two-step procedure. The first step is aimed at computing the aggregate measures of over/under-education of workers at the firm level. The latter are calculated on the basis of the years of education “required” for a given type of “occupation”, that – in our case – is identified by the combination of three elements: the economic sector in which the firm operates (2-digit economic sector according to Nace Rev.2), the workers’ qualification (blue-collar, white-collar, apprentice, middle manager, manager/supervisor, other type) and their age class (15-29, 30-49, 50 and more). The “required” years of education correspond to the modal years of education of the workers employed within each type of occupation<sup>4</sup>. A worker is defined as over (under) educated if his/her years of education are higher (lower) than those required by the type of occupation in which he/she is employed. Once the years of over/under-education are calculated at the worker-level, three distinct measures are derived at the firm level by averaging the number of years of, respectively, “required” (REQ), over- (OVER) and under-education (UNDER) of the workers within each firm. As in Kampelmann and Rycx (2012), the following equations describe the firm-level “mismatch” variables:

- $REQ_j = \frac{1}{n_j} \sum_{i=1}^{n_j} REQ_{i,j}$  where  $REQ_{i,j}$  is the “required” years of education for the type of occupation of the worker  $i$  employed in firm  $j$ , while  $n_j$  it is the number of employees in firm  $j$ ;
- $OVER_j = \frac{1}{n_j} \sum_{i=1}^{n_j} OVER_{i,j}$  where  $OVER_{i,j}$  is the difference between the number of years of education attained by worker  $i$  employed in firm  $j$  and the “required” number of years for his/her occupation, when the worker is over-educated; it is 0 otherwise;

<sup>3</sup> We are only able to investigate a specific type of mismatch occurring in the labour market, i.e. the poor matching in terms of education required/attained at the firm/worker. We cannot study e.g. the lack of matching between workers’ skills or professional status and those needed by the firms. In addition, though the imbalances – of either qualification or skills – that can occur at the aggregate level are found to be related to mismatches at the individual level (Montt, 2015), they fall out of the scope of our analysis.

<sup>4</sup> The Asia-Employment Register reports the following 7 levels of educational attainment (i.e. 7 degrees) to which specific amounts of educational years are associated (in parenthesis): no education or primary education (5); lower secondary education (8); technical and professional upper secondary education (11); upper secondary education (13); tertiary education, 1st level degree (16); tertiary education, 2<sup>nd</sup> level degree (18); Ph.D. (21). It should be noticed that the educational attainment level does not have full coverage in the Asia-Employment Register (see below).

- $UNDER_j = \frac{1}{n_j} \sum_{i=1}^{n_j} UNDER_{i,j}$  where  $UNDER_{i,j}$  is the difference between the number of years of education attained by the worker  $i$  employed in firm  $j$  and the “required” number of years for his/her occupation, when the worker is under-educated; it is 0 otherwise.

Thus, the sum of the three measures ( $REQ_j$ ,  $OVER_j$  and  $UNDER_j$ ) is equal to the average years of education of the employees employed in firm  $j$ .

The second step is the estimate of a labour productivity function at the firm level, where the dependent variable is defined as value added per worker and the measures of educational mismatch are the key explanatory variables:

$$\ln PROD_{j,t} = \beta_0 + \beta_1 \ln PROD_{j,t-1} + \beta_2 REQ_{j,t-1} + \beta_3 OVER_{j,t-1} + \beta_4 UNDER_{j,t-1} + \beta_5 \bar{F}_{j,t} + \beta_6 \bar{W}_{j,t} + \delta_t + \varepsilon_{j,t}$$

The regression also includes two vectors of control variables,  $\bar{F}_{j,t}$  and  $\bar{W}_{j,t}$ , respectively related to firm’s (2-digit economic sector according to Nace Rev.2, firm age, firm size, unit labour costs) and labour force characteristics (firm’s average age of workers, the share of workers under 29 and over 50 years old, the share of female workers, the share of workers by professional status, the share of temporary and part-time workers). In addition, the lagged dependent variable controls for the potential persistency of labour productivity, while business cycle-related effects are taken into account by year dummies ( $\delta_t$ ).

The aim of the analysis is to verify how over/under-education can affect productivity (value added per worker) at the firm level, conditional to the average years of education required in each firm. The productivity equation can be consistently estimated by pooled ordinary least square (OLS), but the existence of firm-specific time-invariant factors influencing both labour productivity and the explanatory educational variables can make the estimated coefficients by OLS biased. The so-called “heterogeneity bias” can be properly tackled by a fixed-effects (FE) estimator. However, a second source of bias may also arise due to time-varying unobserved factors making educational mismatch being determined by the dynamics of firms’ productivity (and *viceversa*)<sup>5</sup>. Such endogeneity issue undermines the unbiasedness of the FE estimator. Thus, to take into account of both the heterogeneity and the simultaneity issues –as properly proposed by Kampelmann and Rycx (2012) – we adopt the dynamic “System-GMM” (Generalized Method of Moments) estimator by Arellano and Bover (1995) and Blundell and Bond (1998).

Finally, we apply this analysis to a balanced panel of over 36,500 manufacturing and services firms with at least 20 workers, operating during the whole period 2014-2019<sup>6</sup>. For the sake of robustness and adapting the work of Kampelmann and Rycx (2012) to our dataset, the original microdata underwent a few cleaning steps. In particular, we exclude firms with a share of missing values concerning workers’ educational degree above 20%<sup>7</sup>, type of “occupations” with less than 30 observations (workers) and firms for which labour productivity value lies below/above the 1<sup>st</sup>/99<sup>th</sup> percentile.

### 3. Descriptive statistics

Figure 1 shows the evolution of the number of required years of education, over-education and

<sup>5</sup> The interested reader may refer to Kampelmann and Rycx (2012) for a more thorough review of studies addressing this issue in the educational mismatch literature.

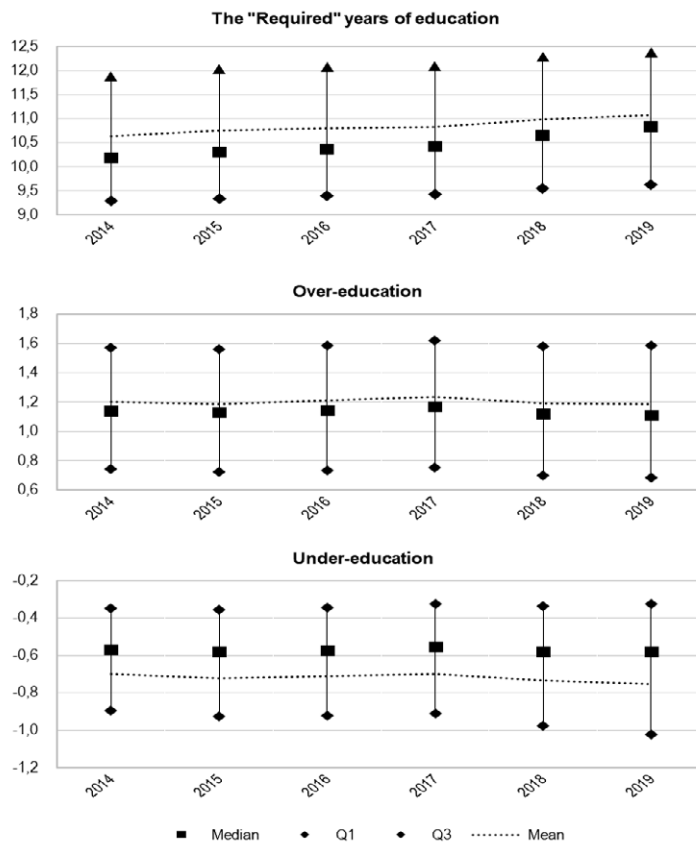
<sup>6</sup> We consider the following sections: C, G, H, I, J, L, M and N according to the Nace Rev.2 classification.

<sup>7</sup> The remaining missing values have been replaced with the required years of education in the relative type of occupation (we recode about 4% of total workers each year). It is worth noting that the share of missing values is rather constant across years, thus the cleaning procedure – either in the form of replaced or deleted observations – has been applied uniformly across time.

under-education at the firm level between 2014 and 2019, for the whole set of manufacturing and services firms in each year, according to the quartiles of their annual distribution and the mean values. The average number of required years grew from 10.65 to 11.07, with a slow but steady upward shift of the distribution, stronger in 2018 and 2019. In addition, the inter-quartile range increased from 2.59 in 2014 to 2.77 in 2019, revealing a widening of the dispersion of required years of education.

In the same period, over-education remained almost steady (around 1.2 years), while under-education increased from -0.70 years in 2014 to -0.75 in 2019 – indeed, a shrinking of years of under-education in absolute terms corresponds to an increase of the phenomenon. Both over and under-education exhibit standard deviation and interquartile range increasing over time, pointing to a growing divergence among firms in terms of their educational mismatch. At the sectoral level – not shown in Figure 1 –, the required years of education in the manufacturing sector slightly grew from 10.12 in 2014 to 10.36 in 2019, while the increase has been stronger in the service sectors (from 11.15 to 11.67). In 2019, over-education is more pronounced in the manufacturing sector (1.29 years and 1.10 years respectively), while under-education is higher in services (-0.61 and -0.87 years).

Figure 1. The required years of education, over-education and under-education - 2014-2019 (annual average by firm for the whole set of firms each year)



## 4. Results

The results of our estimates by GMM-SYS are reported in Table 1.<sup>8</sup> They show the effects of the educational mismatch on firm labour productivity, according to the different technological and knowledge intensity of sectors.<sup>9</sup> In each specification, the absence of second-order autocorrelation of the residuals to the differences has been verified using the Arellano-Bond test (Arellano and Bond, 1991), while the set of instruments is valid according to the Hansen test (Hansen, 1982). Results from POLS and FE estimators are not shown for the sake of brevity, but are available upon requests from the authors.

A one-unit (year) increase in the mean required years of education leads to an increase in firm productivity in both manufacturing and services sectors, but with greater intensity in high-tech industry; in addition, over-educated workers appear to be more productive and to bring a productivity premium to the firms in which they work, while under-educated workers hamper the productivity of the firms where they are employed<sup>10</sup>. Among manufacturing firms, the influence of over-education raises with the technological intensity, while it acts as a competitive factor especially for less knowledge-intensive services. Interestingly, our estimates highlight a (negative) impact of under-education for firms in high and medium-high technological industries and in knowledge-intensive services, where the relatively higher degree of complexity of production processes probably entails higher costs of using less educated human capital.

Table 1. *The impact of the educational mismatch on firm productivity in Italian firms*

	Dep: Labour productivity			
	GMM-SYS			
	High and Medium-High Tech Manufacturing	Low and Medium-Low Tech Manufacturing	Knowledge Intensive Services	Less Knowledge Intensive Services
<i>Lagged Labour Productivity</i>	0.067 ** (0.026)	0.375 *** (0.065)	0.466 *** (0.069)	0.065 *** (0.018)
<i>Required Education</i>	0.115 *** (0.036)	0.046 * (0.027)	0.038 * (0.022)	0.083 *** (0.021)
<i>Over-Education</i>	0.099 *** (0.037)	0.031 ** (0.015)	0.056 * (0.029)	0.119 *** (0.020)
<i>Under-Education</i>	0.081 ** (0.036)	0.020 (0.026)	0.050 * (0.027)	0.053 (0.039)
<i>Workers and Firm variables</i>	Yes	Yes	Yes	Yes
<i>Year dummies</i>	Yes	Yes	Yes	Yes
Arellano-Bond test (AR2) <sup>a</sup>	0.135	0.252	0.271	0.298
Hansen-J test <sup>b</sup>	0.228	0.645	0.215	0.176
Observations	32,064	65,429	16,087	66,583
Firms	6,616	13,386	3,348	13,637

Standard errors in parentheses. Significance levels: \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

a) P-value associated to the Arellano-Bond statistics testing null of absence of serial correlation of differentiated errors at the second lag.

b) P-value associated to the Hansen-J statistics testing the null of exogeneity of instruments.

<sup>8</sup> Table 1 only shows the estimated coefficients related to the mismatch variables, while those related to the control variables are not reported. Anyway, the results are in line with our expectations and available for the interested reader.

<sup>9</sup> We use Eurostat “High-tech aggregation by NACE Rev. 2” (3-digit for manufacturing, 2-digit for services), available at: [https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec\\_esms\\_an3.pdf](https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf).

<sup>10</sup> Because mean years of under-education take negative values by construction, a positive regression coefficient indicates a negative correlation between under-education and productivity – i.e. productivity rises when mean years of over-education increase or under-education decreases.



## 5. Conclusions

Providing strong empirical evidence of the relationship between human capital and firm productivity at the different levels of the technology ladder, our results offer some relevant implications that may steer the policy action towards an increase of the education levels achieved by the working population and a reduction of the mismatch between the demand and supply of skills and qualifications. The availability of longitudinal microdata at the firm level is indeed the main strength of this analysis, which applies and adapts to the Italian case the ORU framework proposed by Kampelmann and Rycx (2012) for a panel of Belgian firms.

There are, of course, several enhancements of our empirical analysis – e.g. improving the identification of specific types of occupations, controlling for potential “birth cohort” effects, exploring the potential mismatch among types of occupations and workers’ relative fields of study – that have to be tackled by future work. And it would be also important to try disentangling the channels through which the productivity premium is achieved – e.g. those linked to the complementarities with digital technologies (see OECD, 2022). However, as the empirical evidence on this phenomenon is relatively scarce, we think that this analysis offers a useful, though preliminary, contribution to the ongoing debate on this crucial issue for the development of the Italian economy.

## References

- Arellano, M., and Bond, O. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**(2), pp. 277–297.
- Arellano, M., and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68**(1), pp. 29–51.
- Blundell, R., and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87**(1), pp. 115–143.
- Fanti, L., Guarascio, D., Tubiana, M. (2021). Skill mismatch and the dynamics of Italian companies’ productivity. *Applied Economics*, **53**(59), pp. 6790–6803.
- Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), pp. 1029–1054.
- Istat (2018). *Rapporto sulla Competitività dei settori produttivi*. Istat, Roma.
- Istat (2022). *Rapporto sulla Competitività dei settori produttivi*. Istat, Roma.
- Istat (2021). *Rapporto Annuale sulla situazione del Paese*. Istat, Roma.
- Kampelmann S., and Rycx F. (2012). The Impact of Educational Mismatch on Labour Productivity. Evidence from Linked Panel Data. *Economics of Education Review*, **31**(6), pp. 918–931.
- Mahy, B., Rycx, F., Vermeylen, G. (2015). Educational Mismatch and Firm Productivity: Do Skills, Technology and Uncertainty Matter?. *De Economist*, **163**, pp. 233–262.
- Monti, P., and Pellizzari M. (2016). Skill Mismatch and Labour Shortages in the Italian Labour Market. *Innocenzo Gasparini Institute for Economic Research Policy Brief*, **2**.
- Montt, G. (2015). The causes and consequences of field-of-study mismatch: An analysis using PIAAC. *OECD Social, Employment and Migration Working Papers*, **167**, OECD Publishing, Paris.
- OECD (2016). *Skills Matter. Further Results from the Survey of Adult Skills*. OECD Publishing, Paris.
- OECD (2022). *Closing the Italian digital gap: The role of skills, intangibles and policies*. OECD Science, Technology and Industry Policy Papers, **126**, OECD Publishing, Paris.

# A paradata-driven statistical approach to improve fieldwork monitoring: the case of the Non-Profit Institutions census

Gabriella Fazzi, Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti, Valentino Parisi, Roberta Piergiovanni, Luigi Arlotta, Maura Giacummo

## 1. Introduction

A complex process requires relevant information on the crucial nodes of the process itself to make more effective decisions. This is the case for large complex surveys where, among the several causes of wrong or inappropriate interviewers' behaviours, only the crucial ones have to be identified and corrected to avoid a knock-on effect. An example of such a survey is the Non-Profit Institutions (NPIs) census, for which fieldwork monitoring is improved by using a paradata-driven approach based on quality control tools (Jans *et al.*, 2013).

The complexity of the NPIs census is due to the variety of unit-typologies: from large and structured institutions to very small associations. The complexity depends also on the different data collection modes and on the several communication channels. Besides, two questionnaires with different research aims – to assess the quality of statistical registers (short form) and to collect information (long form) – contribute to boosting the complexity.

The use of computer-assisted survey instruments offers the opportunity to automatically record paradata, making it possible to apply statistical procedures that allow for near real-time monitoring. To this end, a set of performance indicators is defined to assess the adequacy and observance of survey protocols and to uncover any problematic situations that need to be addressed quickly. Once indicators are defined, control charts can be used to display them (Reed and Reed, 1997).

This work focuses on the system of indicators and control charts developed for the 2022 NPIs census carried out in Italy by the National Statistical Institute (Istat). The paper is organized as follows. Section 2 provides a brief introduction to the survey. Section 3 describes the data collection system. Then, the procedure specifically developed to monitor the interviewers' work is presented, focusing on indicators (section 4), control charts (sub-section 5.1), and possible interventions for the main types of out-of-control events (sub-section 5.2). Finally, some conclusions are drawn (section 6).

## 2. The Non-Profit Institutions census

The NPIs census aims to expand the extent of information available on the non-profit sector by investigating specific issues and by verifying and supplementing the data from the Statistical Register of NPIs, which is based on various administrative sources.

The survey runs from March 10 to November 23, 2022, and involves a sample of approximately 110,000 NPIs. A letter, signed by the president of Istat, is sent to all the sample units to inform them about the purpose of the census, the modes of participation, the deadline, the obligation to participate, and the penalty in case of no participation in the survey.

The survey sample is drawn from the Statistical Register of NPIs and is divided in two sub-samples that differ in terms of units' characteristics, data collection mode, and questionnaire. Besides, each sample is associated to different aims.

The first sub-sample includes about 11,000 NPIs, selected among those units with “weak”

Gabriella Fazzi, ISTAT, Italian National Institute of Statistics, Italy, [fazzi@istat.it](mailto:fazzi@istat.it), 0000-0001-5661-3963  
Manuela Murgia, ISTAT, Italian National Institute of Statistics, Italy, [murgia@istat.it](mailto:murgia@istat.it), 0000-0003-4154-3784  
Alessandra Nuccitelli, ISTAT, Italian National Institute of Statistics, Italy, [nuccitel@istat.it](mailto:nuccitel@istat.it)  
Francesca Rossetti, ISTAT, Italian National Institute of Statistics, Italy, [frosset@istat.it](mailto:frosset@istat.it), 0000-0002-5974-0826  
Valentino Parisi, ISTAT, Italian National Institute of Statistics, Italy, [parisi@istat.it](mailto:parisi@istat.it), 0000-0001-8429-4250  
Roberta Piergiovanni, ISTAT, Italian National Institute of Statistics, Italy, [piergiiov@istat.it](mailto:piergiiov@istat.it), 0000-0003-2489-8786  
Luigi Arlotta, ISTAT, Italian National Institute of Statistics, Italy, [arlotta@istat.it](mailto:arlotta@istat.it)  
Maura Giacummo, ISTAT, Italian National Institute of Statistics, Italy, [magiacum@istat.it](mailto:magiacum@istat.it)

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Gabriella Fazzi, Manuela Murgia, Alessandra Nuccitelli, Francesca Rossetti, Valentino Parisi, Roberta Piergiovanni, Luigi Arlotta, Maura Giacummo, *A paradata-driven statistical approach to improve fieldwork monitoring: the case of the Non-Profit Institutions census*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.53, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 305-310, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

administrative signals in the Register: these are mainly small units that are assigned to the CAPI<sup>1</sup> mode for the administration of a short questionnaire (for this reason, such a sub-sample is called “short”). The main aim is to assess the quality of the Statistical Register of NPIs.

The second sub-sample includes about 99,000 NPIs selected among those units with “strong” administrative signals in the Register. The aim is to collect new information or consolidate the existing one: all these NPIs are initially assigned to the CAWI<sup>2</sup> mode to complete the full version of the questionnaire (for this reason, such a sub-sample is referred to as “long”).

To boost cooperation, CAWI non-respondents are sent a maximum number of four reminders. The reminder letter restates the purpose of the census, the modes of participation, the deadline, and the regulatory framework. In case the NPIs prefer to change the survey technique, they can request the support of a CAPI interviewer by calling the contact center or by accessing a dedicated survey page. To reach all the sample units, the CAPI mode is used also for those NPIs that do not receive the information letter.

The CAPI mode is implemented by an external company on behalf of Istat. Each interviewer from the external company is instructed to find the NPIs and to conduct a targeted survey in a given geographical area. Specifically, the interviewer has to find the NPIs by following online traces (such as website, pages on social media, *etc.*) and by visiting them at their postal addresses. If no signs of “activity” emerge, then the interviewer is required to make an in-person visit at least three times, trying to obtain useful contact information and to administer the interview itself. The units without digital and physical signs of activity are registered as “untraceable”; the units with signs of activity, but untraceable after three visits, are coded as “impossible to be interviewed”.

### 3. The data collection system

The NPIs census is a complex survey. From a technical point of view, the complexity is related to the presence of various actors (respondents, interviewers, fieldwork supervisors, survey managers) with different views on data, the management of several communication channels, and the use of mutually exclusive techniques (CAPI or CAWI). Besides, each unit is assigned a data collection mode, but the unit can ask to change it during the survey.

An integrated web-based information system supports all the different stages of the survey process. The system consists of two web applications that can be customized for any type of survey (they were already used for the Agriculture and Population censuses):

- i)* the data acquisition application, *i.e.*, the online questionnaire used by both respondents (CAWI) and interviewers (CAPI);
- ii)* the management and monitoring application (SGI), accessible to all census operators.

The two applications interact in such a way that they look like a single one to the end user.

SGI is designed to support the various activities of the data collection process. Each actor has a specific profile associated to an appropriate view of data, functions, and outcomes. In this way, each actor can only process data or enter information for which he/she is responsible or authorized. In particular, each authorized actor can enter and manage his/her own data collection network and assign units to the interviewers. This makes it possible to intervene at any time to avoid, for instance, work overloads that might compromise the data quality. In addition to the profile, a key element of SGI is the user-entered outcome, which allows actions to be activated or deactivated via a previously configured workflow. This also enables a unit to be assigned to a different technique.

The information system automatically collects a variety of paradata. As regards the accesses to the data acquisition application (*i*), the number of work sessions and the timestamps of the first and last visit to the online questionnaire are stored for each user.

As for SGI (*ii*), the application records and historicizes each transaction, collecting paradata at the unit level. They are stored in tables that are updated weekly and include the survey technique,

---

<sup>1</sup> Computer-Assisted Personal Interviewing

<sup>2</sup> Computer-Assisted Web Interviewing

the delivery status of the information letter, the address changes, the date and the author of each contact attempt, the latest – temporary or final – outcome of the various contact attempts (e.g., completed interview, refusal, break-off, eligibility status). This information can be used both during the survey to monitor the fieldwork, intervening promptly if necessary, and at the end of the data collection process to understand what needs to be improved.

#### 4. The monitoring indicators

A set of indicators can be adopted to monitor the work of each interviewer involved in the NPIs census. This set is defined taking into account the constraints dictated by both the available information and the interview protocol (which was agreed with the external company).

The monitoring indicators are defined as outcome rates based on the main survey disposition codes, namely the set of codes that SGI uses to record the outcomes of the various contact attempts (section 3). Of the several indicators that can be derived from the available paradata<sup>3</sup>, the following are considered the most effective in highlighting any anomalies in fieldwork:

- a) *response rate*, the no. of completed interviews divided by the no. of eligible units in the sample;
- b) *activity rate*, the no. of units with a final outcome divided by the no. of contacted units;
- c) *eligibility rate*, the no. of eligible units divided by the no. of total (eligible plus ineligible) units in the sample;
- d) *non-interview rate*, the no. of units for which the interview could not be carried out divided by the no. of units with a final outcome.

Rates (a) and (b) are sufficient to monitor the scheduling and carrying out of the interviews, while indicator (c) makes it possible to verify that when those rates are high, it is because the interviewer is working well, in the sense that he/she is not making up ineligible units (these are given a short form that is paid as a completed interview). Moreover, some problems in contacting the NPIs may be detected by an excessive proportion of non-interviews (d).

All the above rates are produced at regular time intervals (weekly) during the fieldwork period, only for those interviewers who have been working in the last four weeks. In fact, it may happen that, also due to the difficulties experienced in the data collection, some interviewers stop carrying out the field activity. Besides, the indicators are calculated by province to understand whether problems arise in specific areas of the country – and are therefore common to all the operators working in those areas – or whether the problems concern certain interviewers only.

Finally, given the relevant impact that both the type of administrative signal and the questionnaire length have on the fieldwork (section 2), the set of rates is produced separately for:

- the NPIs in the short sub-sample;
- the NPIs in the long sub-sample that did not receive the information letter<sup>4</sup>.

In this way, any anomalies more directly attributable to the interviewer's behaviour are better highlighted.

#### 5. The monitoring procedure

##### 5.1 Control charts

The monitoring procedure for the NPIs census is mainly aimed at understanding whether the CAPI operators are working in compliance with the interviewing protocol or, if not, what actions must be taken to improve their work. Besides, it tries to simplify the monitoring activities so that

---

<sup>3</sup> It is worth noting that the time interval between the first and last access to the online questionnaire is a too rough estimate of the interview duration and, therefore, is of little help in monitoring the interviewers' work.

<sup>4</sup> The indicators are not calculated for the NPIs (long sub-sample) that ask for a change of technique (from CAWI to CAPI), as for these units both the contact phase and the interview are less troublesome (response rate and eligibility rate very close to 1).

costs and efforts of this phase of the data collection process are reduced: thanks to this procedure, survey and fieldwork managers can immediately detect any potential problem interviewers might encounter and take the proper actions to solve it in due time.

The procedure is designed as an alternative monitoring tool to the traditional contingency tables that report the values of performance indicators by interviewer, week, geographical area, *etc.* Contingency tables are extremely useful in monitoring data collection, but they might become hard to read when the number of variables and cases to be monitored increases. Displaying the values of each indicator on a control chart, instead, makes it much easier to find critical situations, as out-of-control cases are highlighted by statistical evidence. Moreover, in this way, contingency tables can only be produced for a restricted number of variables and cases.

Each indicator introduced in section 4 is displayed using a Shewhart  $p$ -chart, where the central line represents the mean, and the upper and lower limits – respectively, UCL and LCL – bound the range of variation of the mean when the process is in statistical control (Montgomery, 2009).

The control charts are implemented with SAS/QC software (SAS Institute Inc., 2018) and are produced weekly in two steps:

1. a first set of charts (called *screening charts*) is produced for the provinces and for the interviewers who have been working in the last four weeks;
2. for each interviewer with at least an out-of-control event from the first step, a second set of charts (called *in-depth charts*) is produced to monitor each indicator over the whole period the interviewer has been working.

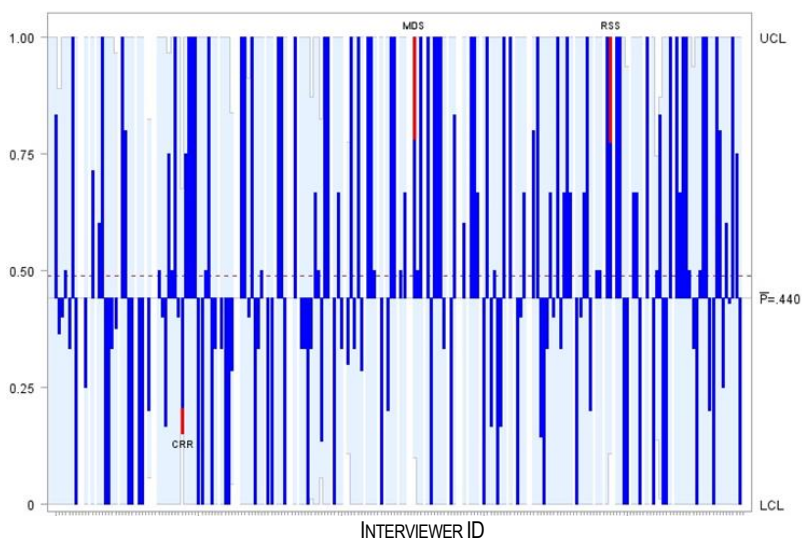
What differs in the two types of charts are the sub-groups of elements for which the mean is calculated: in the screening charts, the sub-groups are the provinces or the interviewers, while for the in-depth charts they are the fieldwork weeks.

The in-depth charts are fundamental to understand whether an out-of-control event that has occurred in the last four weeks is occasional or systematic. In the latter case, the survey manager can decide whether and how to intervene on each interviewer.

Some examples of charts are reported below to better explain how they work.

Figure 1 provides the screening chart of the eligibility rate for the interviewers who have been working in the four weeks preceding June 13, 2022. Three interviewers have out-of-control rates: for CRR the value falls below the LCL, while for both interviewers MDS and RSS the value is 1.

**Figure 1.** Screening control chart of the *Eligibility rate* for all interviewers (up to June 13)



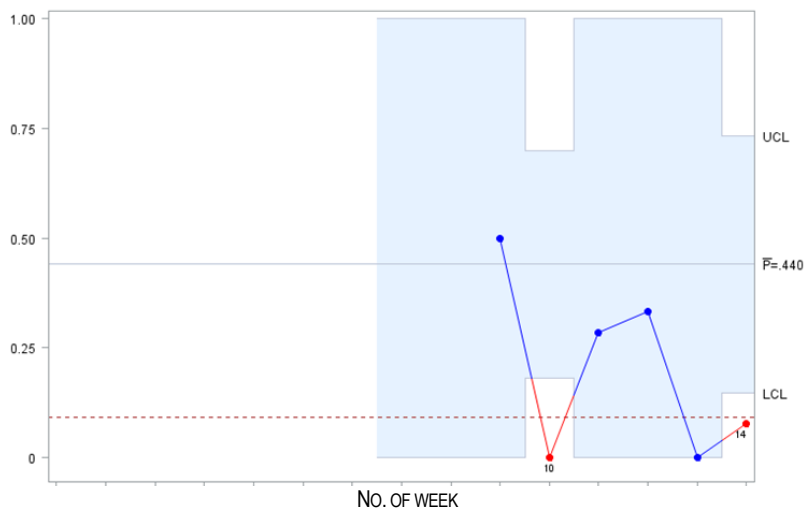
Source: NPIs census data, Short sub-sample, 2022

The control limits<sup>5</sup> are calculated with respect to the mean value  $\bar{P}=0.44$ , which is referred to all the interviewers who have been active for at least one fieldwork week (from March 10 to June 13). The average rate in the last four weeks is plotted as a dashed red line.

To understand whether the out-of-limits values are occasional or systematic, an in-depth chart is produced for each of the three interviewers. For the sake of brevity, only the in-depth chart for CRR – who started working from the 9<sup>th</sup> week of fieldwork – is shown (Figure 2). In this chart, the average eligibility rate for the interviewer (dashed red line) is below the mean value ( $\bar{P}=0.44$ ), suggesting that the NPIs surveyed by him/her are mostly ineligible (especially in the last two weeks). It is important to analyse the charts for the other indicators before taking a proper decision.

In the case of interviewer CRR, if the activity rate is very high and, at the same time, the non-interview rate falls below the LCL, further investigation is required to exclude that he/she is making up interviews. Instead, for MDS and RSS, if the response rate turns out to be excessively low, it is quite likely that they need to be trained again on the contact strategy with the respondents.

**Figure 2.** In-depth control chart of the *Eligibility rate* for interviewer CRR (up to June 13)



Source: NPIs census data, Short sub-sample, 2022

## 5.2 Out-of-control events and types of intervention

In addition to the above-mentioned indicators (section 4) and charts (sub-section 5.1), the monitoring procedure automatically produces two tabular reports listing, respectively, the provinces and the interviewers with at least an out-of-control event, along with the limit values at which each out-of-control event occurs. The absolute values of the variables used to build the indicators are also reported to take in due account those “signals” based on many units.

The information from the two reports helps to understand whether the out-of-control events subtend a structural issue affecting the entire province (when no interviewer is flagged within a flagged province) or an interviewer-specific problem (when the interviewer is flagged regardless of whether the province in which he/she operates is flagged or not). In the latter case, targeted actions, such as de-briefing or additional training sessions, might be undertaken. Some of the interventions suggested by the output of the procedure are summarized in Table 1.

Any doubts about the actions to be undertaken are removed by analysing all other available information – traditional reports and questionnaires – and/or by randomly contacting some NPIs for feedback on whether the interview was actually conducted and/or whether some of the data reported in the questionnaire are accurate.

<sup>5</sup> The control limits are 3 times the standard error, above and below the central line, and depend on the sub-group size.

**Table 1.** Possible interventions by the main results of the monitoring procedure

TYPES OF OUT-OF-CONTROL EVENT				POSSIBLE ACTIONS
RESPONSE RATE	ACTIVITY RATE	ELIGIBILITY RATE	NON-INTERVIEW RATE	
Below the LCL or in control	Above the UCL	Below the LCL (too many ineligible units)	Below the LCL or in control	<i>Interviewer to be checked:</i> he/she might cheat in surveying ineligible units (they are given a short questionnaire that is paid as a completed interview)
Below the LCL	Above the UCL	--	Above the LCL or in control	<i>Interviewer to be re-trained or generalized problem:</i> he/she shows a high activity rate (i.e., he/she assigns a lot of final outcomes), but this is due to a high non-interview rate. It is important to understand if he/she faces any difficulties in the contact phase or if the contact information is not updated. If the entire province is flagged, then the problem of outdated contact information is generalized
Below the LCL	--	Above the UCL	--	<i>Interviewer to be re-trained:</i> units surveyed by him/her are mostly eligible, but he/she has problems in completing the interview
Below the LCL	Below the LCL	--	--	<i>Interviewer to be checked:</i> he/she is slow and possibly lazy and should be invited to put more effort in his/her work

## 6. Conclusions

The monitoring procedure for the NPIs census is developed to understand whether the operators are working in compliance with the interviewing protocol or, if not, what actions must be taken to improve their work. Specific indicators are defined using recorded paradata to support the survey-specific monitoring goals and then assist in finding inefficiencies in the data collection.

The system of control charts, which is used to display the proposed indicators, helps balance cost and thoroughness of monitoring activities by using statistical principles to differentiate potentially problematic cases from those that vary naturally around a process average. In this way, fieldwork supervisors and survey managers are guided in making targeted interventions, without spending time exploring false alarms.

The procedure is used next to the traditional reports and under a close cooperation among methodologists, fieldwork supervisors, and survey managers. This allows the latter – fieldwork supervisors and survey managers – to get acquainted with the new instrument and the former to understand whether any improvement in terms of usability or efficacy is required.

Finally, this experience will be extremely important to understand whether this approach is suitable for other censuses or any other survey that needs to monitor the fieldwork.

## References

- Jans, M., Sirkis, R., Morgan, D. (2013). Managing data quality indicators with paradata based statistical quality control tools: the keys to survey performance, in *Improving surveys with paradata: analytic uses of process information*, ed. F. Kreuter, John Wiley & Sons Inc., Hoboken, (NJ).
- Montgomery, D. C. (2009). *Introduction to statistical quality control, Sixth edition*, John Wiley & Sons Inc., Hoboken, (NJ).
- Reed, S. J, Reed, J. H. (1997). The use of statistical quality control charts in monitoring interviewers, in *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section*, American Statistical Association, Alexandria, (VA).
- SAS Institute Inc. (2018). *SAS/QC® 15.1 User's guide*. SAS Institute Inc., Cary, (NC).

# Gender INequality Indicator for Academia (GINIA)

Margherita Silan, Giovanna Boccuzzo

## 1. Introduction

Gender equality is a fundamental right, a common value of Europe, and a necessary condition for the achievement of the EU objectives of growth, employment and social cohesion (European Commission, 2019). Over the last few decades, women in all countries in Europe have caught up with or even surpassed men in terms of their level of education, but they are still facing segregation in different forms. Indeed, the career of women remains markedly characterized by strong vertical segregation throughout the Europe. The term vertical segregation refers to the under-representation of a clearly identifiable group of workers (in this case women) in top levels of occupations or sectors.

Another problem is that Science and Technology have historically been and still are male dominated areas. In this case, there is a problem of horizontal segregation, which shows that there is an unequal distribution of women and men in different scientific fields.

To strengthen the role of women in scientific research, the European Commission funded the Gender Time Project (Gender Time, 2012), from which this work originated.

The main aim of this work consists of a methodological proposal for a composite indicator that, together with a system of indicators, represents and measures gender inequality in academia. In this paper, the indicator is shaped in order to represent gender inequality in the staff of University of Padova (Unipd), however, the proposal is extremely flexible with the purpose to fit also different academic environments. We called the composite indicator GINIA (Gender INequality Indicator for Academia) and, for the sake of brevity, the acronym will be used in the following.

## 2. Measuring Gender Equality

In recent decades, several indices have been proposed in the literature in order to measure gender equality in different contexts and areas. In order to properly define the aspects and dimensions to be considered in the theoretical definition of GINIA we carefully considered them and converted their specification into an academic environment.

Among others, the proposal made by the European Institute for Gender Equality (EIGE), the Gender Equality Index (EIGE-GEI), represents a solid methodology for measuring gender disparity among European countries. Its value has been continuously updated since 2005, both for Europe and for the Member States (Barbieri et al., 2021). The entire system of the EIGE's Gender Equality Index is based on an interesting framework of collecting data divided into six core domains and two satellite domains (violence and intersecting inequalities).

In the existing literature on the systems for measuring gender equality in Academic and Research Institutions, a good solution may come from the GenisLab project (Genis Lab, 2010), funded by the European Commission in 2010. Three elements were highlighted as fundamental dimensions in gender budgeting: the allocation of funds and the management of time and space.

Margherita Silan, University of Padua, Italy, silan@stat.unipd.it, 0000-0001-5541-0603

Giovanna Boccuzzo, University of Padua, Italy, giovanna.boccuzzo@unipd.it, 0000-0003-2143-7730

Referee List (DOI 10.36253/fup\_referee\_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Margherita Silan, Giovanna Boccuzzo, *Gender INequality Indicator for Academia (GINIA)*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.54, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 311-316, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3



### 3. Theoretical framework

The first step to define the structure of GINIA consists in the definition of the theoretical framework that supports it. Starting from existing indexes described in Section 2, in our approach the gender gap is detected in seven domains (Figure 1): work, money, knowledge, time, power, health and space (Boccuzzo et al., 2016). These seven domains are better specified and declined through twelve sub-domains that are measured by seventeen variables. The composite indicator is the result of a three-step aggregation of variables, sub-domains and domains and provides a synthetic measure of gender inequality in the University of Padova.

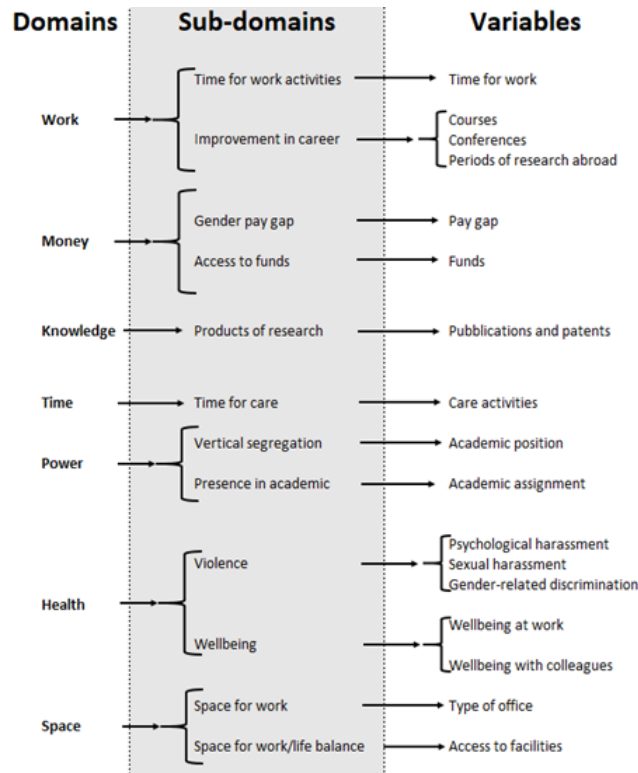


Figure 1: The theoretical framework to measure gender equality in the University of Padova.

### 4. Data and Population

Data used to build and compute the gender equality index in the University of Padova come both from administrative official datasets (numbers of people per role, action plans, code of conducts, expertise, etc.) and from an ad-hoc survey that was carried out in September/October 2015 by Unipd research group as part of the GenderTime Project. The questionnaire was distributed to all academic staff of the University of Padova. The target population of the questionnaire is Unipd academic staff members at 31<sup>st</sup> December 2014, including Full and Associated Professors, Assistant Researchers, Research Fellows (fixed-term) and Post-Doc Fellows. All members of the target population were asked to be part of the survey; however, only the 31% replied to the questionnaire. This response rate is in line with the expected response rate for a web survey, especially with respect to such a delicate topic. There are some differences between

the target population and the respondents. It is really important to evaluate those differences in order to evaluate the representativeness of the respondents' population. For instance, there is an over representation of women and young academics at the beginning of their carrier. This result is probably due to a stronger involvement in the survey contents.

The following analysis are based on respondents of the survey; but, since they do not reflect the distribution (for gender, age, academic position and school) of the whole population, it will be necessary to weight answers. Thus, we compute post-stratification weights for each intersection of gender, academic position and school.

## 5. Methods

### 5.1 Normalization and age standardization

All indicators need to have the same direction defined in the theoretical framework. In GINIA's system, the direction is given by "higher is better", which means that all indicators have higher values for better situations. When this is not the case, the indicator has to be reversed.

Having different data sources and several measurement scales, the need to make all the variables of the system of indicators vary between in a common interval has to be addressed in order to compare them. We chose the Min-Max method for normalization, which makes the variables vary in a range between 0 and 1. So, the normalised variable  $I_{ji}$  related to the person  $i$ , who has gender  $j$ , is:

$$I_{ji} = \frac{\text{Observed Value}_{ji} - \text{Theoretical Minimum}}{\text{Theoretical Maximum} - \text{Theoretical Minimum}} \quad (1)$$

Since there are differences in the age distribution among the male and female population employed in the University of Padova at 31<sup>st</sup> December 2014, the comparison between male and female could be biased by the different age structure of the two populations. Indeed, even the academic position could depend on the age structure. To take into account the different age structures in the calculation of the indicator, we calculated crude and also standardized indicators considering three main age classes, applying direct standardization and using as a reference the whole academic staff of the University of Padova.

### 5.2 Weighting

After the definition of the theoretical framework, the data selection and normalization and imputation of missing data, weighting and aggregation techniques should be taken into account. Their choice should be done along the lines of the underlying theoretical framework.

The assigning weights to single indicators is necessary when not all of them contribute to the formation of the composite indicator in the same measure.

In this work, we will consider two weighting methods: equal weights and preference matrix weights (based on the importance respondents have given to each dimension in the final question of the web survey). Indeed, most composite indicators rely on equal weighting (EW), i.e. all variables are given the same weight. This essentially implies that all variables have the same relevance in the composite (or there is insufficient knowledge of causal relationships or a lack of consensus on the alternative).

Respondents are asked to order items that represent domains according to their importance. The answers are used to compute a weighting system based on preference analysis which is used to aggregate the domains into the final composite indicator. The main advantage of this weighting method is that it takes into account the ranking made by the respondents. In addition,

Table 1: Alternative weighting and aggregation methods used for the computation of the composite indicator. The combination of weighting and aggregation techniques chosen for GINIA is underlined.

	Variables	Sub-domains	Domains
Aggregation	<u>Arithmetic Mean</u>	<u>Arithmetic Mean</u> Geometric Mean	Arithmetic Mean <u>Geometric Mean</u>
Weighting	<u>Equal Weighting</u>	<u>Equal Weighting</u>	Equal Weighting <u>Preference Matrix Weights</u>

it can be used both for qualitative and quantitative data, and it increases the transparency of the composite. The main disadvantages are that it requires a high number of pairwise comparisons, and thus it can be computationally costly; furthermore, the results depend on the set of respondents.

### 5.3 Aggregation

The literature on composite indicators offers several examples of aggregation techniques. In this work we use two common aggregation methods: arithmetic and geometric mean aggregation.

The arithmetic mean is a complete compensatory method, which means that poor performance in some indicators can be compensated for by sufficiently high values for others. Although widely used, this aggregation entails restrictions on the nature of indicators and the interpretation of weights. Furthermore, it requires that the indicators have to be preferentially independent, which is a very strong condition, especially in this application.

If we want some degree of non-compensability, geometric aggregation is better suited. It is a less compensatory approach, indeed, while in a linear aggregation, the compensability degree is constant, in a geometric aggregation, the compensability is lower for composite indicators with low values (a low score on one indicator will need a much higher score on the others to improve the situation). It is very sensitive to data far from the central value, and it will be nullified if there is an indicator equal to zero.

### 5.4 GINIA composition

Every indicator of the GINIA system of indicators for the University of Padova is the result of the comparison between the elementary indicators corresponding for men and women. The comparison is carried out by the following formula (Boccuzzo et al., 2016):

$$\text{Inequality Indicator} = \frac{\text{Indicator for women}}{\text{Indicator for men}} \quad (2)$$

Thus, the indicator is close to 1 in the most equalitarian scenario, when indicators for men and women are more similar. Moreover, when the value of the indicator is below one, there is a situation in which women are penalized compared to men; whereas, when it is above 1, women are privileged with respect to men.

In order to compute GINIA, we are dealing with three levels of aggregation: one for variables (arithmetic mean), one for sub-domains (arithmetic mean) and the final step that puts together domains in order to obtain the final composite indicator with geometric mean (Table 1). Indeed, according to our theoretical framework, variables related to the same domain can compensate each other, while this consideration is not plausible for the domains.

The computation of confidence intervals of the GINIA is not trivial, especially due to the correlation between indicators. Thus, we computed confidence intervals using bootstrap (10000

iterations). Bootstrap samples are extracted with replacement assigning to each unit a probability to be selected proportional to post-stratification weights, then in each sample the GINIA is computed.

The choice of weighting and aggregation methods is a fundamental step because the indicator may substantially change modifying the weighting and aggregation methods. This is why we performed a sensitivity analysis considering different combinations of weighting and aggregation techniques (shown in Table 1) to assess the robustness of the composite indicator as a final step in the analysis.

## 6. Results

Looking at the indicators computed for the seven domains disaggregated, it is possible to detect which are the aspects where women are more disadvantaged with respect to men. Since in Table 2 only standardized indicators are reported, the observed differences do not depend on a different age structure.

Table 2: The standardized indicators for each domain by gender and their rate.

	Work	Time	Power	Knowledge	Space	Health	Money
Women	0.447	0.815	0.254	0.142	0.643	0.642	0.404
Men	0.461	0.815	0.304	0.195	0.687	0.752	0.530
W/M	0.970	1.000	0.836	0.728	0.936	0.853	0.761
CI 95% lower	0.943	0.981	0.771	0.643	0.908	0.825	0.682
CI 95% upper	0.997	1.019	0.903	0.826	0.964	0.882	0.844

In domain *time*, we find perfect equality between men and women with respect to satisfaction in work-life balance. This does not mean that men and women working at the University of Padova have the same time allocation in terms of family care and work, but it means that they are equally satisfied with respect to their desired time allocation. On the other hand, in all the other domains we find a significant disadvantage for women, with a more serious situation for domains *knowledge* and *money*. The domain *knowledge* is based on the number of publications in the last two years, and the fact that women have more difficulties to get published is an important limitation that needs to be acknowledged. Indeed, having a low number of publications also affects other aspects of academic life, such as career possibilities and access to funds. This second aspect is a part of the *money* domain (the other with a mostly low value). Since in Italy academic salaries are fixed and linked to the covered position, this domain is composed by access to research funds and additional activities that yield an extra return.

In Table 3, the values of GINIA are reported both crude and standardized. Both show a marked disadvantage for women. The crude indicator is lower than the standardized one; this is probably because a part of the disadvantage detected by the crude indicator is actually due to the different age structure between man and women in academia.

Table 3: The crude and standardized composite indicators (arithmetic mean and then weighted geometric mean) by gender, with bootstrap confidence intervals.

	Crude indicators	Standardized indicators
Women	0.395 (0.382-0.408)	0.405 (0.392-0.418)
Men	0.477 (0.466-0.488)	0.473 (0.463-0.483)
W/M (GINIA)	0.829 (0.796-0.862)	0.856 (0.824-0.888)

From the sensitivity analysis, whose results are shown in Figure 2, the use of geometric mean as aggregation methods results in indicators with lower values due to the lack of compensability, especially when it is used at the domains' level. The use of weights computed by preferences analysis results in values slightly lower than the equal weights solution, probably because the domains stated as more important are also those in which women are more disadvantaged.

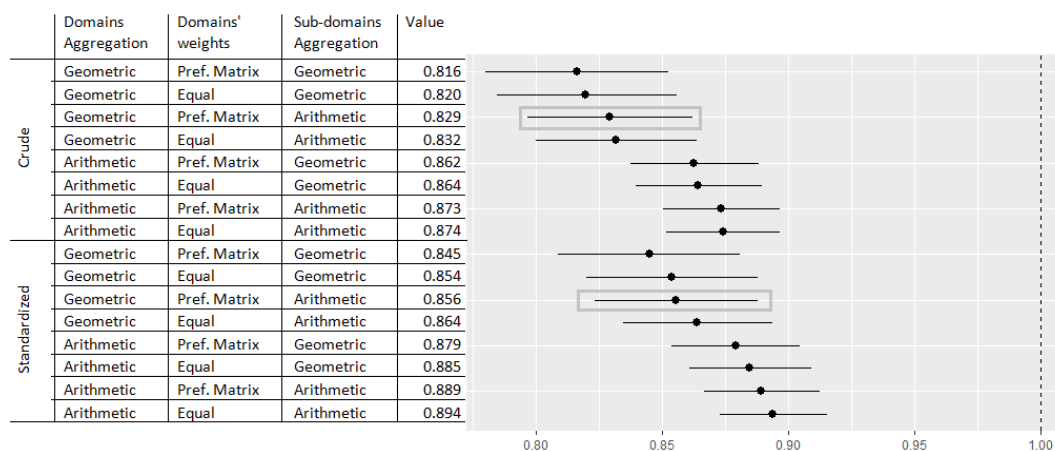


Figure 2: GINIA values and respective bootstrap confidence intervals in all cases considered by the sensitivity analysis.

## 7. Conclusions

As concluding remarks, we may say that the GINIA indicator seems useful for measuring and monitoring gender equality in academia. The situation at the date of the questionnaire seems improvable; therefore, it would be interesting and useful to repeat the experience in order to evaluate changes. The observation of the disaggregated domains' indicators shows a critical aspect referred to publications that could be a good starting point to meditate on effective policies to reduce the gap.

## References

- European Commission, Directorate-General for Employment, Social Affairs and Inclusion (2019). Manual for gender mainstreaming : employment, social inclusion and social protection policies, Publications Office, <https://data.europa.eu/doi/10.2767/19842>
- Barbieri, D., Cazorla, A. G., THIL, L., Mollard, B., Ochmann, J., Peciukonis, V. (2021). Gender Equality Index 2021: Health (doi: 10.2839/035225).
- Boccuzzo, G., Rocco, I., Silan, M., Badaloni, S. (2016). Methodology of the System of Indicators, in *A model for building a Gender Equality Index for academic institutions*, eds. S. Badaloni and L. Perini, Padova University Press, pp. 59–79.
- GENIS LAB (2011). The Gender in Science and Technology LAB ? GENIS LAB. <https://cordis.europa.eu/project/id/266636>, last access: 20<sup>th</sup> November, 2022.
- Gender Time (2012). Transferring Implementing Monitoring Equality Gender Time. <https://gendertime.org/>, last access: 20<sup>th</sup> November, 2022.

# Given $N$ Forecasting Models, What To Do?

Fabrizio Culotta

## 1. Introduction

It is well known that the future is uncertain. Against this uncertainty, economic agents plan their economic activity accordingly. In this planning, producing forecasts of the quantity of interest is the traditional way of uncovering possible not-yet-realized trajectories. Feedback from estimated future dynamics will then influence actual planning and business activities. This is true also for private decision-makers, like firms and other types of organizations, but especially for public policy-makers since their activities produce effects at the whole country level.

The increasing availability of data, together with progress in computational techniques, have incentivized researchers to construct more sophisticated forecasting models and to increase the accuracy of their performances. Nowadays, available forecasting models range from classical econometric models, e.g. ARIMA, to non-parametric models, e.g. exponential smoothing, to machine-learning, e.g. trees and neural networks. It results in a plethora of single forecasting models available to both private and public decision-makers. Since the late '70s, a group of academic researchers proposed the idea of competition among different forecasting models (Makridakis et al., 1982). It emerged that statistically sophisticated models do not necessarily produce more accurate forecasts, whereas combinations of them outperform vis-à-vis single models. Moreover, the ranking of forecasting models depends on the accuracy measure being as well as on the adopted forecast horizon. The success of the first so-called *M-competition* (M stands to Makridakis) allowed us to carry on the tradition of forecasting competitions (Hyndman, 2020) until today with the recent M4 and M5 competitions (Petropoulos and Makridakis, 2020; Makridakis et al., 2021). Given a set of time series at different frequencies, several models compete to produce the best forecast. Models' performances are then ranked based on some accuracy measures. Based on the idea of competition among different forecasting methods, this work compares their forecasting performances on a given time horizon. Unlike the tradition of Ms competitions, which are based on thousands of time series at different time frequencies, a single univariate time series is selected at the monthly frequency.

The motivation of this choice is to show that, in the simplest exercise of forecasting a single time series, the ex-ante choice of the model is likely to be misleading because a model ranking exists and it is specific to time (hence, frequency) and of measurement object of the single series. Indeed, when a set of forecasting models is available, a semi-automatic algorithm of model selection based on some performance measures would be a superior choice for the various decision-makers. In the case at hand, the choice of the monthly unemployment rate is dictated by the fact that it is the most common measure of the (mis-)functioning of the labour market and, as such, is of utmost importance for policymakers.

Forecasting models are finally ranked based on some accuracy measures. The main findings confirm that, given  $N$  forecasting models, combination techniques outperform single uncombined models in terms of accuracy and reduce the risk of adopting a single forecasting model.

## 2. Forecasting Models

The comparative forecasting exercise presented in this work comprises a set of 23 different uncombined and combined models. The selected time series on which all models are trained is the deseasoned dynamics of the Italian unemployment rate over the years 2004 – 2019 at the monthly frequency freely available from the ISTAT data warehouse (<http://dati.istat.it/>). The observational period is split between the training set, from January 2004 to June 2019, and the test set, from July to December 2019. The set of selected forecasting models contains some ARIMA-like models, some Exponential Smoothing models, to machine learning models. It also contains combinations of them based on some model averaging techniques. For sake of brevity, the succinct list is reported in table 1. All the computations are carried out with the statistical software  $\mathcal{R}$  by using the most recent packages. Model specifications and other details can be provided upon request.

FAMILY	Label	Model	Reference	R package
ARIMA	ARIMA	ARIMA	Hyndman and Khandakar (2008)	forecast
	ARFIMA	Fractionally-differenced ARIMA	Peiris and Perera (1988)	forecast
	GARMA	Gegenbauer-ARIMA	Dissanayake et al. (2016)	garma
	SSARIMA	State-space ARIMA	Svetunkov and Boylan (2020)	smooth
Exponential Smoothing	ES	Exponential Smoothing	Brown (1956)	ets
	HOLT	Linear Exponential Smoothing	Holt and Modigliani (1960)	forecast
	THETA	Exponential Smoothing with drift	Assimakopoulos and Nikolopoulos (2000)	forecast
	CES	Complex Exponential Smoothing	Svetunkov and Kourentzes (2018)	smooth
	GUM	State-space Exponential Smoothing	Svetunkov and Kourentzes (2018)	smooth
Machine Learning	ARML	Bagged AR		caretForecast
	BAG	Bagged Exponential Smoothing	Bergmeir et al. (2016)	forecast
	NN	Fast-forward Neural Network		forecast
Hybrid	ADAM	Augmented Dynamic Adaptive Model	Hyndman and Khandakar (2008)	smooth
	BATS	GUM with ARMA errors	De Livera et al. (2011)	forecast
	ATA	Combination of ES and ARIMA	Yapar et al. (2017)	ATAforecasting
	SPL	Cubic Spline	Chambers and Hastie (2017)	forecast
Combinations	COMB1	Combination of ETS,SSARIMA,GUM and CES		smooth
	COMB2	Combination of ARIMA,ETS,THETA,NN and BATS		forecastHybrid
	COMB3	Combination of ARML and SPL with simple weights		ForecastCombinations
	COMB4_BG	COMB3 with Bates-Granger weights		ForecastComb
	COMB4_InW	COMB3 with Inverse Rank approach		ForecastComb
	COMB4_Me	COMB3 with Dynamic weighting scheme		ForecastComb
COMB5	Combination of all models except COMBs		ForecastCombinations	

Table 1: *Selection of forecasting models.*

Once all forecasting models have been estimated, it is interesting to compare statistics of model fitting in terms of moments of the corresponding error distribution. At this aim, table 2 below provides rank values (column RANK) for each forecasting model based on a total score (SCORE). The latter statistics is computed as the sum of the single scores reported in terms of mean (RANK\_MEAN), standard deviation (RANK\_SD), skewness (RANK\_SKEWNESS), and kurtosis (RANK\_KURTOSIS).

FAMILY	MODEL	RANK_MEAN	RANK_SD	RANK_SKEWNESS	RANK_KURTOSIS	SCORE	RANK
ARIMA	ARIMA	2	14	23	10	49	13
	ARFIMA	20	15	20	7	62	19
	GARMA	15	9	11	6	41	11
	SSARIMA	21	18	19	18	76	23
Exponential Smoothing	ES	14	8	9	3	34	5
	HOLT	12	7	10	4	33	4
	THETA	22	22	15	16	75	21
	CES	5	20	17	19	61	18
	GUM	23	21	16	15	75	21
Machine Learning	ARML	18	11	1	23	53	14
	BAG	19	12	13	9	53	14
	NN	1	23	5	8	37	8
Hybrid	ADAM	17	17	22	14	70	20
	BATS	13	10	12	5	40	10
	ATA	3	19	14	12	48	12
	SPL	6	6	8	1	21	1
Combinations	COMB1	4	16	21	13	54	16
	COMB2	16	13	18	11	58	17
	COMB3	9	3	2	21	35	7
	COMB4_BG	7	1	7	17	32	3
	COMB4_InvW	8	2	4	20	34	5
	COMB4_MED	10	4	3	22	39	9
	COMB5	11	5	6	2	24	2

Table 2: *Ranking of forecasting models in terms of model fitting.*

What emerges from table 2 is that, in terms of model fitting, the best-performing forecasting model is SPL followed by COMB5, COMB4\_BG, COMB4\_InvW, and so on. In detail, the error distribution of the NN model is associated with the lowest mean error, COMB4\_BG with the lowest dispersion. Whereas ARML and SPL are characterized by the lowest skewness and kurtosis, respectively. Despite model fitting being an important quality feature of forecasting models, it is not the definitive dimension to consider when a decision-maker needs to adopt a single forecasting model. As shown in the next section, the accuracy of forecasting performances may deliver different conclusions.

### 3. Results

Figure 1 shows the forecasts produced by each model on the test set over a time horizon of six months. It is possible to observe that ARML model fails in capturing the dynamics of actual data despite its model fitting performances being characterized by the lowest skewness. On the contrary, the COMB2 forecasts closely mimic the dynamics of the Italian unemployment rate despite its model fitting performance are not the best in any moments of the error distribution.



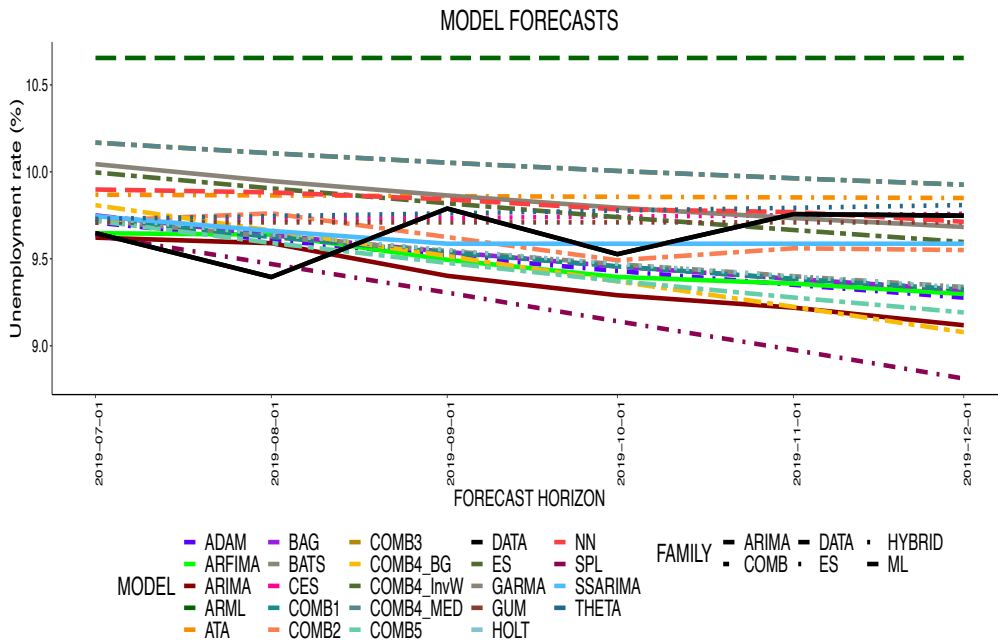


Figure 1: *Forecasts of Italian unemployment rate. ARIMA models (solid line): ARFIMA, ARIMA, GARMA, SSARIMA. Combinations (COMB, two-dashed line): COMB1, COMB2, COMB3, COMB4\_BG, COMB4\_InvW, COMB4\_MED. Exponential Smoothing (ES, dotted line): CES, ES, GUM, HOLT, THETA. Hybrid models (dot-dashed line): ADAM, ATA, BATS, SPL. Machine Learning models (ML, long-dashedline): ARML, BAG, NN.*

These considerations confirm that model fitting, despite being an important aspect to consider for the selection of forecasting models, does not necessarily ensure that forecast performances are aligned with model fitting performances. Instead, the use of various ensembling techniques delivers satisfactory results compared to those of single uncombined models. On this point, note also from figure 1 that the actual dynamics of the unemployment rate is contained within the full set of forecasts. This means that a suitable model combination can be obtained by ensembling appropriately some of the models under scrutiny.

Finally, table 3 provides the values of various accuracy measures used in the various forecasting competitions: ME (mean error), MAE (mean absolute error), MPE (mean percentage error), MSE (mean squared error), MAPE (mean absolute percentage error), RMSSE (root mean squared scaled error), RAME (relative absolute mean error), RMAE (root mean absolute error) and RRMSE (relative root mean squared error).

FAMILY	MODEL	ME	MAE	MPE	MSE	MAPE	RMSSE	RAME	RMAE	RRMSE	SCORE	RANK
ARIMA	ARIMA	19	18	19	18	18	18	19	18	18	165	18
	ARFIMA	14	15	13	15	15	15	14	15	15	131	16
	GARMA	15	10	15	16	10	16	15	10	16	123	14
	SSARIMA	1	4	1	3	4	3	1	4	3	24	3
Exponential Smoothing	ES	9	12	8	11	12	11	9	12	11	95	11
	HOLT	6	11	6	10	11	10	6	11	10	81	9
	THETA	7	3	10	4	3	4	7	3	4	45	5
	CES	4	1	4	2	1	2	4	1	2	21	2
	GUM	3	2	3	1	2	1	3	2	1	18	1
Machine Learning	ARML	23	23	23	23	23	23	23	23	23	207	23
	BAG	10	14	9	13	14	13	10	14	13	110	13
	NN	13	6	14	6	6	6	13	6	6	76	7
Hybrid	ADAM	12	16	12	14	16	14	12	16	14	126	15
	BATS	8	9	7	9	9	9	8	9	9	77	8
	ATA	17	7	17	7	7	7	17	7	7	93	10
	SPL	22	22	22	22	22	22	22	22	22	198	22
Combinations	COMB1	11	13	11	12	13	12	11	13	12	108	12
	COMB2	2	5	2	5	5	5	2	5	5	36	4
	COMB3	20	19	20	19	19	19	20	19	19	178	19
	COMB4_BG	18	21	18	21	21	21	18	21	21	180	21
	COMB4_InvW	5	8	5	8	8	8	5	8	8	63	6
	COMB4_MED	20	19	20	19	19	19	20	19	19	178	19
COMB5	16	17	16	17	17	17	16	17	17	150	17	

Table 3: *Ranking of forecasting models in terms of accuracy measures.*

As expected, the overall rank of forecasting models in terms of accuracy measures differs from the ranking in terms of model fitting presented in table 2. Now, the best-performing forecasting model is GUM, followed by CES and SSARIMA. Among all model combinations, only COMB2 and COMB4\_InvW lie in a good position, being the fourth and the sixth best performing models respectively. Forecasting models SPL and ARML occupy the next-to-last and last positions, respectively.


## 4. Conclusions

Results confirm that it does not exist yet a single superior universal model. On the contrary, the ranking of different forecasting models is specific to the adopted training set. For example, when the time series of interest switches to the employment rates instead of unemployment rates, the rank of model performances changes. Secondly, results confirm that performances of machine learning and neural network models offer satisfactory alternatives to the traditional econometric models like ARIMA or the non-parametric Exponential Smoothing. Finally, the results stress the importance of model ensemble techniques as a solution to model uncertainty as well as a tool to improve forecast accuracy (Shaub, 2020).

Overall, the flexibility provided by a rich set of forecasting models, and the possibility to combine them, together represent an advantage for decision-makers often constrained to adopt solely pure, uncombined, forecasting models.

## References

- Assimakopoulos, V. and Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16(4):521–530.
- Bergmeir, C., Hyndman, R. J., and Benítez, J. M. (2016). Bagging exponential smoothing methods using stl decomposition and box-cox transformation. *International Journal of Forecasting*, 32(2):303–312.
- Brown, R. G. (1956). Exponential smoothing for predicting demand. *NBER*.
- Chambers, J. M. and Hastie, T. J. (2017). Statistical models. In *Statistical models in S*, pages 13–44. Routledge.
- De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527.
- Dissanayake, G., Peiris, M. S., and Proietti, T. (2016). State space modeling of gegenbauer processes with long memory. *Computational Statistics & Data Analysis*, 100:115–130.
- Holt, M. and Modigliani, J. (1960). Planning production, inventories, and workforce.
- Hyndman, R. J. (2020). A brief history of forecasting competitions. *International Journal of Forecasting*, 36(1):7–14.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of Statistical Software*, 27:1–22.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, 1(2):111–153.
- Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., and Winkler, R. L. (2021). The m5 uncertainty competition: results, findings and conclusions. *International Journal of Forecasting*.
- Peiris, M. and Perera, B. (1988). On prediction with fractionally differenced arima models. *Journal of Time Series Analysis*, 9(3):215–220.
- Petropoulos, F. and Makridakis, S. (2020). The m4 competition: bigger. stronger. better. *International Journal of Forecasting*, 36(1):3–6.
- Shaub, D. (2020). Fast and accurate yearly time series forecasting with forecast combinations. *International Journal of Forecasting*, 36(1):116–120.
- Svetunkov, I. and Boylan, J. E. (2020). State-space arima for supply-chain forecasting. *International Journal of Production Research*, 58(3):818–827.
- Svetunkov, I. and Kourentzes, N. (2018). Complex exponential smoothing for seasonal time series.
- Yapar, G., Yavuz, İ., and Selamlar, H. T. (2017). Why and how does exponential smoothing fail? an in depth comparison of ata-simple and simple exponential smoothing. *Turkish Journal of Forecasting*, 1(1):30–39.



This volume collects the contributions presented at the conference “Data-driven Decision Making” organized by the Italian Association for Applied Statistics, held in Genoa from 12 to 14 September 2022. The papers cover a broad range of topics, with a common thread: the use of statistical methods to support decision-making both in public administrations and in private companies.

Enrico di Bella is Associate Professor of Social Statistics at the Department of Political and International Sciences of the University of Genoa.

Luigi Fabbris is President of the Italian Association for Applied Statistics and former Professor of Social Statistics at the University of Padua.

Corrado Lagazio is Professor of Statistics at the Department of Economics of the University of Genoa.

ISSN 2704-601X (print)  
ISSN 2704-5846 (online)  
ISBN 979-12-215-0106-3 (PDF)  
ISBN 979-12-215-0107-0 (XML)  
DOI 10.36253/979-12-215-0106-3

[www.fupress.com](http://www.fupress.com)