

# Determinants of social startups in Italy

Lucio Palazzo, Pietro Sabatino, Riccardo Ievoli

## 1. Introduction

The so called Startup Act (Decree Law 179/2012, converted into Law 221/2012), has introduced in Italy the notion of innovative companies with a high technological value, i.e., the innovative start-ups. Among them, the Italian government includes the category of social startups, i.e., “*startup innovative a vocazione sociale*” (hereafter SIAVS), representing a relatively new field of interest in both scientific and normative perspective.

SIAVS must satisfy the same requirements of other innovative startups, but operate in sectors such as social assistance, education, health, social tourism and culture, enjoying also some tax benefits. Furthermore, they have a possible direct (social) impact on the collective well-being, measured through a self-evaluation document named: “Documento di Descrizione dell’Impatto Sociale” published yearly by each SIAVS (Vesperi, Lenzo). Today, social startups are more than doubled with respect to five years ago<sup>1</sup>.

Within Italian academic debate concerning startups and innovative economic enterprises, SIAVS have been considered for their hybrid nature, balancing between profit and non-profit model of business, and for their role of producing value for local communities (Vesperi et al, 2015). Although there are some recent empirical studies on social entrepreneurship intentions (Bacq et al., 2016), little is known about territorial pattern of SIAVS, even if a certain similarity has been observed, at regional scale, with the territorial distribution of overall startups (Maglio, 2019). Italian non-profit organizations present different characteristics compared to innovative companies, notably on gender balance in workforce and territorial diffusion (Istat, 2019; Forum del Terzo Settore, 2017).

The aim of this paper is to investigate the relevant factors influencing the presence of social startups in Italy at the provincial level. The outcome variable is the number of active social startups in Italian provinces while the set of explanatory variables is composed by economic and demographic indicators at the provincial level.

Regarding the explanatory variables, unemployment rate and number of incubators have been used as predictors of the number of startups at regional level in Colombelli (Quartaro), while Hoogerdoorn (2016) considers the GDP per capita. Information regarding registered firms at the provincial level can be found also in the work of Colombelli et al. (2019) to predict the number of new firms at the provincial level (NUTS 3 regions). Furthermore, the effectiveness of incubators for Italian startups is still under debate (Deidda Gagliardo et al., 2017), while Sansone et al. (2020) have introduced a new taxonomy, distinguishing between *business*, *mixed* and *social* incubators. We also consider other variables as *broadband*, which can be viewed as a proxy of the technological level of a province, and the percentage of *NEET* (neither in employment or in education or training between 15 and 29 years) which is a measure of non-attractiveness of a territory for the young people.

Generalized linear models (GLM) for discrete outcomes are applied and compared, even taking into account the zero-inflated issue arising due to the distribution of these particular data.

---

<sup>1</sup>“Relazione Annuale al Parlamento sullo stato di attuazione e l’impatto delle policy a sostegno di startup e PMI innovative” [www.mise.gov.it](http://www.mise.gov.it)

## 2. Materials and Methods

### Data

Information regarding startups and certified incubators are retrieved from the Italian Chambers of Commerce<sup>2</sup>, updated to the third quarter 2020. Other additional variables, at the provincial (NUTS 3) and regional (NUTS 2) level, and the spatial coordinates of these provinces, are obtained through the Italian National Institute of Statistics<sup>3</sup> (ISTAT) and European Statistical Office<sup>4</sup> (EUROSTAT).

A possible drawback is that some variables suffer from timeliness issue. Moreover, for the purpose of this explorative study, this issue seems less severe considering the reasonably not too high variations occurring in the short term period at provincial level. Thus, we retrieved the latest update (i.e., the value for the last available year) for all considered covariates. In some cases we consider the geometric mean to avoid problems related to possible temporal variations.

### Measurement Variables

The dependent variable is the count of SIAVS in Italian provinces. Therefore, the sample size is equal to  $n = 105$ , composed by all Italian provinces except for “Sud Sardegna” and “Andria-Trani-Barletta”, which do not include any kind of startup in their territory.

As mentioned, we identified the following candidates as possible determinants for the presence of SIAVS (the latest update is in brackets):

*Population density*, number of inhabitants divided by the area of province (reference year: 2017). A logarithm transformation is applied;

*GDP per capita*, in thousand of euros (reference year: 2017);

*Incubators*, count of *certified* incubators in each province. These particular companies (Decree Law 179/2012) are registered in the Italian Chambers of Commerce and offer services for the developing of startups (reference year: 2020);

*Unemployment rate*, at the provincial level (reference year: 2019);

*Registration rate*, companies registered in a year divided by the total number of companies registered in the previous year in the Italian “registro imprese” at the provincial level (geometric mean between 2015 and 2018);

*Broadband*, number of ultra-broadband subscriptions as a percentage of inhabitants in each provinces (geometric mean between 2015 and 2017);

*Social employees*, rate of workers in social cooperatives (reference year: 2019)

*NEETs*, percentage of neither in employment or in education or training people between 15 and 29 years (reference year: 2019).

### Statistical Models

The number of SIAVS in Italian provinces can be modelled applying GLM family (see e.g. Nelder, Wedderburn; McCullagh, Nelder, among others). The general formulation of GLM (Agresti, 2003) is carried out through a link function  $g(\cdot)$ , which transforms the expectation of the response variable, i.e.  $\mu_i = E(Y_i)$ , to the linear predictor:

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n. \quad (1)$$

where  $p = 8$  is the number of variables previously discussed.

<sup>2</sup><https://www.registroimprese.it/>

<sup>3</sup><https://www.istat.it/>

<sup>4</sup><https://ec.europa.eu/eurostat>

In this context, two main competing models can be considered: Poisson (POI) and Negative Binomial (NB) regression. In the former case,  $Y_i \sim Poi(\lambda_i)$  and the corresponding log-link function is  $g(\mu_i) = \log \lambda_i$ , while in the latter case  $Y_i \sim NegBin(\mu_i, \omega)$ . In the POI model, the observed counts are equidispersed, i.e.  $E(Y_i) = Var(Y_i) = \mu_i$ . Moreover, the scale parameter  $\omega$  in NB model takes into account for the presence of overdispersion i.e.  $Var(Y_i) = \mu_i + \mu_i^2/\omega$ .

A possible issue related to the count of SIAVS (and startups in general) is the possible presence of excess of zeros in the data, i.e. provinces without any registered SIAVS. Thus, previously introduced models may be modified to take into account the zero inflation. The zero inflated Poisson (ZIP) model is derived as a mixture of a binary logistic and POI (Lambert, 1992). The responses  $Y_i$  are independent and  $Y_i \sim 0$  with probability  $\pi_i$  and  $Y_i \sim Poi(\lambda_i)$  with probability  $1 - \pi_i$ . The resulting link function can be written as follows:

$$g(\mu_i) = \begin{cases} \log \frac{\pi_i}{1-\pi_i}, & \text{if } y_i = 0 \\ \log \lambda_i, & \text{if } y_i > 0 \end{cases} \quad (2)$$

The zero inflated NB (ZINB) model, introduced in Greene (1994), is derived by substituting the POI link function with the NB (when responses are not equal to zero). We remark that ZIP and ZINB assume that the zero inflation effect is generated by a separate process apart from the count values.

### 3. Results and Discussion

The number of SIAVS is equal to 240 and the 87.5% of them are classified in the service sector. The remaining 12.5% is divided in industry and/or craft sector (7.9%) and sectors such as agriculture, tourism and commerce (4.6%). Registered SIAVS present almost 40 activity codes. The main activities of SIAVS can be divided in: *a*) software production and IT consultancy (17.5%), *b*) scientific research and development (12.9%), *c*) education (10.4%), *d*) information and other services (9.6%), *e*) non-residential social assistance (8.8%), *f*) activities related to libraries, archives and museums (3.8%) *g*) art and entertainment (2.9%). The remaining 34.2% of SIAVS are classified in 33 different activity codes.

Almost a quarter of SIAVS (24.2%) is located in the province of Milan (58), while provinces of Rome and Turin include respectively 27 (11.2%) and 13 (5.4%) SIAVS. In general, 65 provinces (62%) contain almost a SIAVS but only 20 (19%) of them registered more than 2 social startups. SIAVS also present a higher frequency of female prevalence (measured in terms of at least 50% of women in the company) compared to other startups, exceeding them by the 10%. Moreover, differences can not be found in practice (with respect to other startups) regarding the proportion of young people (under 35) and foreigners.

In Figure 1 we can observe the distribution of SIAVS (left panel), the distribution of startups (center panel) at the provincial level and the distribution of non-profit institutions at the regional level (right panel). Main differences between startups and SIAVS can be viewed in the provinces of Centre Italy. Nonetheless, startups and SIAVS are concentrated in the metropolitan areas (especially in the provinces of Milan and Rome) and also the non-profit subjects can be found especially in the North-East (Lombardia Region). In addition, the provinces of Sardinia present the lower counts of startups and SIAVS, even if the number of non-profit institutions appears comparable with respect to the other regions.

Table 1 summarizes the main results of statistical models discussed in Section 2. First of all, we check the usefulness of the whole set of regressors in all models, by observing the decreasing of the Bayesian Information Criterion (BIC) between the null models ( $BIC_0$ ), including only the intercept, and the models with all considered covariates. For each model, the BIC is function

of a different likelihood, and the decreasing is more (numerically) evident in the POI and ZIP models than in the NB and ZINB. Another similar check can be also carried out (only for the first two models) through the McFadden’s *Pseudo R*<sup>2</sup>. We also report, for each model specification, the likelihood ratio test statistic to formally test for the departure from the “null” model (which only includes the intercept) and its associated p-value. This check also confirms the usefulness of proposed regressors. We have to remark that it is not possible to make a proper comparison between the four models in terms of likelihood-based statistics. Therefore, we use a leave-one-out cross-validation (CV) approach to compare the prediction of four models, estimating  $R = 105$  times the model and then computing  $MSE(CV) = n^{-1} \sum_r (\hat{y}_r - y_r)^2$ . Regarding this performance indicator, the conventional POI exhibits the lower  $MSE(CV)$ , followed by the ZIP and ZINB. Finally, the (here not reported) results of two Vuong tests (Vuong, 1989) suggest the rejection of null hypothesis of POI and NB in favour of ZIP and ZINB.

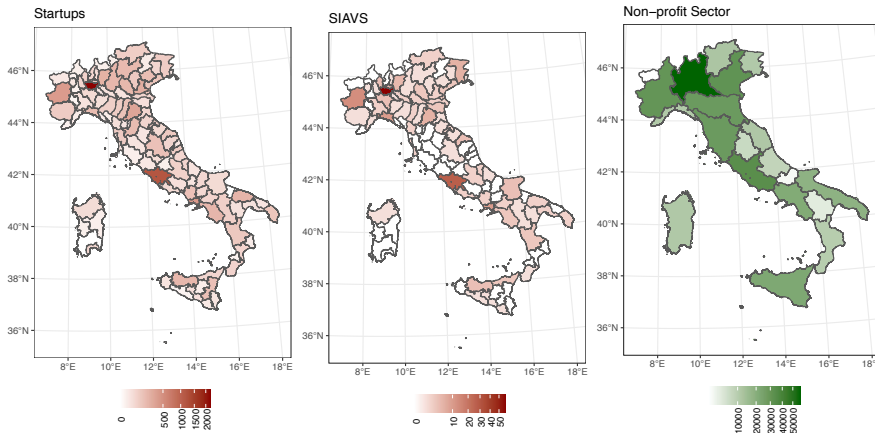


Figure 1: Geographical distribution of number of startups, number of SIAVS (provincial level) and non-profit sector (regional level).

Conventional GLM models help to identify log population density, (certified) incubators and broadband as positive determinants of the counts of SIAVS at the provincial level considering a nominal error rate of the 1%. Conversely, in more robust zero-inflated regressions, the coefficient of population density is no longer statistically significant. Therefore, in ZIP and ZINB, unemployment rate is identified as a possible positive driver for the arise of SIAVS, while the percentage of young people neither in employment or in education or training can be considered as a negative indicator for the arise of SIAVS. Surprisingly, GDP per capita and social employees are not statistically significant in any considered model.

Certified incubators appears fundamental for the presence of SIAVS. At a descriptive level, 64% of SIAVS (153) is located in provinces including almost a certified incubator. This percentage is slightly lower considering all innovative startups (56%).

To conclude, SIAVS arise in provinces with higher technological levels, including ecosystem to develop and assist startups. Basing on our results, also population density and unemployment may have an influence on the presence of SIAVS, but further investigation will be

Table 1: Input variables setting scheme used in each model.

	POI	NB	ZIP	ZINB
<i>Intercept</i>	0.7569	-0.3718	0.7480	0.7441
<i>log(Population density)</i>	0.3491 ***	0.3949 ***	0.1636	0.1637
<i>Incubators</i>	0.2373 **	0.2326 *	0.2269 **	0.2269 **
<i>Registration rate</i>	-0.2995 .	-0.2228	-0.1976	-0.1972
<i>Unemployment rate</i>	0.0468	0.0409	0.1512 ***	0.1512 ***
<i>Broadband</i>	0.1969 ***	0.1810 ***	0.1987 ***	0.1987 ***
<i>GDP per capita</i>	-0.0304	-0.0184	-0.0146	-0.0146
<i>Social employees</i>	-0.0746	-0.0591	-0.0500	-0.0499
<i>NEETs</i>	-0.0312	-0.0247	-0.0764 **	-0.0764 **
BIC <sub>0</sub>	803.8809	415.8049	714.4403	420.4589
BIC	359.6001	362.5449	357.1889	361.8049
McFadden's $R^2$	0.6025	0.2226	-	-
LR Test	481.5100 ***	90.4920 ***	431.7100 ***	133.1200 ***
MSE(CV)	6.1835	25.5918	10.7048	10.7284

Significance codes:  $0 \leq '***' < 0.001 \leq '**' < 0.01 \leq '*' < 0.05 \leq '.' < 0.1 \leq ' ' < 1$

conducted at the territorial level.

Future interesting analysis will concern the trend of new SIAVS in time (using quarterly data), even considering autoregressive models for integer data (see e.g Palazzo, 2019).

## References

- Agresti, A. (2003). Categorical data analysis. Second Edition (Vol. 482), John Wiley & Sons, New York (US).
- Bacq, S., Hartog, C., Hoogendoorn, B. (2016). Beyond the moral portrayal of social entrepreneurs: An empirical approach to who they are and what drives them. *Journal of Business Ethics*, **133**(4), pp. 703–718.
- Colombelli, A., Quatraro, F. (2019). Green start-ups and local knowledge spillovers from clean and dirty technologies. *Small Business Economics*, **52**(4), pp. 773–792.
- Colombelli, A., Orsatti, G., Quatraro, F. (2019). Local knowledge composition and the emergence of entrepreneurial activities across industries: evidence from Italian NUTS-3 regions. *Small Business Economics*, **56**, pp. 613–635.
- Deidda Gagliardo, E., Gobbo., G., Papi, L., Bigoni, M. (2017). L'efficacia dei programmi di incubazione nel successo e nel fallimento delle startup. *RIREA*. (5-8), pp. 225–239.
- Forum del Terzo Settore (2017). Le reti del Terzo Settore. Terzo Rapporto. Eds. Peruzzi, G. Forum Nazionale del Terzo Settore, Rome (IT), pp. 1-72.
- Greene, W.H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working paper No. 94-10 Stern School of Business, New York University, New York (US), pp.1–36.
- Hoogendoorn, B. (2016). The prevalence and determinants of social entrepreneurship at the macro level. *Journal of Small Business Management*, **54** (sup1), pp. 278–296.
- ISTAT (2019). Struttura e profili del settore non profit. Report censimento permanente su istituzioni non profit anno 2017. ISTAT, Rome (IT), pp. 1–15.

- Maglio, M. (2019). Il framework territoriale nel processo di diffusione del fenomeno delle startup a vocazione sociale. *XIII Colloquio Scientifico sull'impresa sociale*, 24-25 maggio 2019, Università degli Studi di Roma Tor Vergata, Rome (IT), pp. 1–14.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1), pp. 1–14.
- McCullagh, P., Nelder, J. A. (1989). Generalized linear models. Second Edition, *Chapman and Hall*, New York (US).
- Nelder, J. A., Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, **135**(3), pp. 370–384.
- Palazzo, L. (2019) Integer Autoregressive modeling: a new gretl routine, in *Proceedings of the International Conference on the GNU Regression, Econometrics and Time-series Library*, eds. F. Di Iorio, R. Lucchetti, Università degli studi di Napoli Federico II, Naples (IT), pp. 95–103.
- Ventura, M., Vesperi, W., Reina, R. (2018). Innovazione sociale nelle startup innovative: un approccio narrativo. *Iris Network*, pp. 1–18.
- Sansone, G., Andreotti, P., Colombelli, A., Landoni, P. (2020). Are social incubators different from other incubators? Evidence from Italy. *Technological Forecasting and Social Change*, **158**, pp. 120–132.
- Vesperi, W., Lenzo, P. (2016). La start-up innovativa a vocazione sociale e il documento di descrizione dell'impatto sociale: una comparazione di casi. *Colloquio scientifico sull'impresa sociale*, 10-11 giugno 2016, Università degli studi di Napoli Federico II, Naples (IT), pp. 1–18.
- Vesperi, W., D'Aleo, V., Morabito, G., Lo Bue, S. (2015). Imprenditorialità vs. responsabilità sociale: La startup innovativa a vocazione sociale. *Colloquio scientifico sull'impresa sociale*, 22-23 maggio 2015, Università degli studi Mediterranea di Reggio Calabria, Reggio Calabria (IT), pp. 1–14.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, **57**(2) pp. 307–333.