Marginal Models: recent developments and applications to categorical time series analysis.

Roberto Colombi[1]

[1] Università di Bergamo - Viale Marconi - 24044 Dalmine Italy; colombi@unibg.it

Abstract: Recently general definitions of marginal interactions and marginal models have been introduced by Bergsma, Rudas (2002), Colombi, Forcina (2001) and by Bartolucci, Colombi, Forcina (2004) that considerably improved the flexibility and interpretability of standard hierarchical log-linear models by allowing interactions to be contrasts of four types of Logits defined within different marginal distributions. This paper reviews these recent contributions and shows their relevance in the context of categorical time series analysis.

Keywords: marginal models, categorical time series, non-normal state space models

1 Introduction

In section two of this paper we review the definition of generalized marginal interactions introduced by Bartolucci, Colombi, Forcina (2004) and we show how these interactions are used to build a class of models which generalizes the Hierarchical Marginal Models previously introduced by Bergsma, Rudas (2002). In section three of this paper the proposed marginal models are used to specify a class of dynamic models for multi-categorical time series and in section four some examples are given. The aim of the work is to show that marginal parameterizations can be easily adapted to the context of categorical time series modelling.

2 Marginal interaction parameters and marginal models

Consider the joint probability function of q response variables $A_1, \dots, A_q$, with A_j taking values x_j in $\{1, 2, \dots, a_j\}$. The set of response variables that defines a given marginal distribution will be denoted by the set $\mathcal{M}$ of indices of the corresponding variables and $\mathcal{Q} = \{1, \dots, q\}$ will refer to the joint distribution. The vector of the $\prod_1^q a_j$ joint probabilities will be denoted by $\boldsymbol{\pi}$.

2.1 Generalized Marginal Interactions

We now introduce the Bartolucci, Colombi, Forcina (2004) definition of interaction parameters which includes the four well known types of logits: *local* (l), *global* (g), *continuation* (c) and *reverse continuation* (r) and the sixteen types of log-odds ratios discussed by Douglas *et al.* (1990). Note that it makes sense to use logits of type *local* both with ordinal and non-ordinal variables but that logits of type *global* and *continuation* can be used only with ordinal variables.
For any *category* $x_j < a_j$, define the event $\mathcal{B}(x_j, 0)$ to be equal to $\{x_j\}$ if the logit is of type local or continuation and to $\{1, \ldots, x_j\}$ for global or reverse continuation logits; similarly, the event $\mathcal{B}(x_j, 1)$ is equal to $\{x_j + 1\}$ if the logit is of type local or reverse continuation and to $\{x_j + 1, \ldots, a_j\}$ for global or continuation logits. Finally define the marginal probabilities:

$$p_{\mathcal{M}}(\boldsymbol{x}_{\mathcal{M}}; \boldsymbol{h}_{\mathcal{M}}) = p(A_j \in \mathcal{B}(x_j, h_j),\ \forall j \in \mathcal{M}),$$

where $\boldsymbol{x}_{\mathcal{M}}$ is a row vector of categories x_j, $j \in \mathcal{M}$, and $\boldsymbol{h}_{\mathcal{M}}$ is a row vector whose elements, h_j, $j \in \mathcal{M}$, are equal to zero or to one. These marginal probabilities are probabilities of a table where the variables $A_j, \forall j \in \mathcal{M}$, have been dichotomized according to the categories: $\mathcal{B}(x_j, 0)$, $\mathcal{B}(x_j, 1)$. The marginal generalized interactions are log-linear contrasts of the previous probabilities and are so defined:

$$\eta_{\mathcal{H};\mathcal{M}}(\boldsymbol{x}_{\mathcal{H}} \mid \boldsymbol{x}_{\mathcal{M}\setminus\mathcal{H}}; \boldsymbol{h}_{\mathcal{M}\setminus\mathcal{H}}) = \sum_{\mathcal{K}\subseteq\mathcal{H}} (-1)^{|\mathcal{H}\setminus\mathcal{K}|} \log p_{\mathcal{M}}(\boldsymbol{x}_{\mathcal{M}}; \boldsymbol{h}_{\mathcal{M}\setminus\mathcal{H}}, \boldsymbol{0}_{\mathcal{H}\setminus\mathcal{K}}, \boldsymbol{1}_{\mathcal{K}}). \tag{1}$$

Note that any interaction is defined by the *interaction set* $\mathcal{H}$ of the variables involved, by the marginal distribution $\mathcal{M}$ where it is defined and by the logit type assigned to each variable of $\mathcal{M}$. According to this definition the kind of dichotomy implied by the type of logit adopted for each variable should carry over when defining higher order interactions within the same marginal distribution. As an example consider the bivariate case, $q = 2$, where the continuation logit type is assigned to each variable and the marginals of interest are: $\mathcal{M}_1 = \{1\}$, $\mathcal{M}_2 = \{2\}$ and $\mathcal{M}_3 = \{1, 2\}$. Let π_{ij}, $\pi_{i\cdot}$ and $\pi_{\cdot j}$ denote the joint and marginal probabilities, then

$$\boldsymbol{\eta}_{\{1\};\{1\}}(i) = \ln \frac{p(A_1 \in \mathcal{B}(i,1))}{p(A_1 \in \mathcal{B}(i,0))} = \ln \frac{\sum_{n=i+1}^{a_1} \pi_{n\cdot}}{\pi_{i\cdot}},$$

$$\boldsymbol{\eta}_{\{2\};\{2\}}(j) = \ln \frac{p(A_2 \in \mathcal{B}(j,1))}{p(A_2 \in \mathcal{B}(j,0)} = \ln \frac{\sum_{n=j+1}^{a_2} \pi_{\cdot n}}{\pi_{\cdot j}},$$

and

$$\boldsymbol{\eta}_{\{1,2\};\{1,2\}}(ij) =$$

$$= \ln \frac{p(A_1 \in \mathcal{B}(i,1), A_2 \in \mathcal{B}(j,1))}{p(A_1 \in \mathcal{B}(i,0), A_2 \in \mathcal{B}(j,1))} - \ln \frac{p(A_1 \in \mathcal{B}(i,1), A_2 \in \mathcal{B}(j,0))}{p(A_1 \in \mathcal{B}(i,0), A_2 \in \mathcal{B}(j,0)} =$$

$$= \ln \frac{\sum_{m=j+1}^{a_2} \sum_{n=i+1}^{a_1} \pi_{nm}}{\sum_{m=j+1}^{a_2} \pi_{im}} - \ln \frac{\sum_{n=i+1}^{a_1} \pi_{nj}}{\pi_{ij}},$$

are continuation log-odds ratios.

2.2 Complete and Hierarchical Families of Interaction Sets

We now examine the problem of allocating the *interaction sets* among the marginals within which they may be defined.
Denote by $\mathcal{F}_m$ the family of interaction sets defined within the marginal distribution $\mathcal{M}_m$. Let also $\mathcal{P}(\mathcal{J})$ be the family of all non empty subsets of $\mathcal{J}$ and $\mathcal{P}_m$ be a short-hand notation for $\mathcal{P}(\mathcal{M}_m)$.
Given a non-decreasing sequence of marginals $\mathcal{M}_1, \ldots, \mathcal{M}_s$, a family of interactions sets is called complete and hierarchical if (i) any interaction set is defined in one marginal distribution $\mathcal{M}_m$, (ii) $\mathcal{F}_1 = \mathcal{P}_1$ and $\mathcal{F}_m = \mathcal{P}_m \setminus \bigcup_{h<m} \mathcal{F}_h$.
The previous definition implies that $\mathcal{M}_s = \mathcal{Q}$, that $\mathcal{M}_m \in \mathcal{F}_m$, for every m, that every family $\mathcal{F}_m$ is a non-empty ascending class of subsets of $\mathcal{M}_m$ and that every interaction is defined within only one marginal distribution.
In the following, for every interaction set $\mathcal{I} \in \mathcal{M}_m$ of a complete hierarchical family of interactions sets, we will consider only the interactions:

$$\eta_{\mathcal{I};\mathcal{M}_m}(\boldsymbol{x}_\mathcal{I}) = \eta_{\mathcal{I};\mathcal{M}_m}(\boldsymbol{x}_\mathcal{I} \mid \mathbf{1}_{\mathcal{M}_m \setminus \mathcal{I}}; \mathbf{0}_{\mathcal{M}_m \setminus \mathcal{I}})$$

where the conditioning variables of $\mathcal{M}_m \setminus \mathcal{I}$ are fixed to their first category. When all the conditioning variables in $\mathcal{M}_m \setminus \mathcal{I}$ have assigned logits of type local Bartolucci, Colombi, Forcina (2004) showed that the interactions $\eta_{\mathcal{I};\mathcal{M}_m}(\boldsymbol{x}_\mathcal{I} \mid \boldsymbol{x}_{M_m \setminus \mathcal{I}}; \boldsymbol{h}_{M_m \setminus \mathcal{I}})$ are linear functions of the interactions $\eta_{\mathcal{H};\mathcal{M}_m}(\boldsymbol{x}_\mathcal{H})$, $\mathcal{H} \supseteq \mathcal{I}$, so that at least in this case there is no restriction in limiting the attention to these parameters.

2.3 Complete and Hierarchical Marginal Parametrizations

The interactions $\eta_{\mathcal{I};\mathcal{M}_m}(\boldsymbol{x}_\mathcal{I})$ associated to a complete hierarchical family of interactions may be arranged into the vector $\boldsymbol{\eta}$ which may be explicitly written in matrix form as

$$\boldsymbol{\eta} = \boldsymbol{C} \log(\boldsymbol{M}\boldsymbol{\pi}), \qquad (2)$$

where the rows of $\boldsymbol{C}$ are contrasts and $\boldsymbol{M}$ is a matrix of zeros and ones which sums the probabilities of appropriate cells to obtain the necessary marginal probabilities of the type described by (2.1). A detailed description of these matrices is given by Colombi, Forcina (2001). Bartolucci, Colombi, Forcina (2004) showed that (2) is invertible. The result extends the Bergsma, Rudas (2002) important contribution on marginal models and earlier works of Lang, Agresti (1994), Glonek, McCullagh (1995) and Glonek (1996). Parameters defined by a function of the joint probabilities of the type (2) have a long history starting from the seminal works of Grizzle *et al.* (1969) and of Forthofer, Koch (1973) and here we stress the fact that the representation of the link function (2) is important, in the context of maximum likelihood estimation, both from the theoretical point of view and from the computational point of view. The importance of the representation will carry over also to the context of categorical time series as it will be shown in the next section.

A parameterization of the joint probabilities in term of the generalized marginal interactions $\eta_{\mathcal{I};\mathcal{M}_m}(\boldsymbol{x}_{\mathcal{I}})$ defined as above will be called *complete hierarchical marginal parameterization.*

The advantages of a marginal parameterization with respect to the log-linear one come from the flexibility in the choice of the interactions and from the interpretability of the parameters. Marginal parametrizations allow a direct and straightforward parameterization of the marginal probabilities of interest and in the framework of a marginal parameterization it is easier to state that a given marginal distribution is stochastically larger than another or that the strength of the dependence between two variables increase with a third variable or that two variables are marginally independent or positively associated. In fact these hypotheses can be defined by linear inequality and equality constraints on generalized marginal interactions as shown in Dardanoni, Forcina (1998), Bartolucci, Forcina, Dardanoni (2001), Colombi, Forcina (2001) and Bartolucci, Colombi, Forcina (2004). Moreover complete hierarchical marginal parameterizations are very useful in parametrizing block recursive models as shown by Bartolucci, Colombi, Forcina (2004).

As an example consider the seemeengly unrelated logit regressions represented by the dashed edges graph of figure 5.3(a) of Cox, Wermuth (1996); under this model the variables A_3 and A_4 are explanatory for the variables A_1 and A_2, A_2 is independent from A_3 given A_4 and A_1 is independent from A_4 given A_3. The model can be parametrized choosing the complete hierarchical parameterization defined by the marginals $\mathcal{M}_1 = \{3,4\}$, $\mathcal{M}_2 = \{1,3,4\}$, $\mathcal{M}_3 = \{2,3,4\}$, $\mathcal{M}_4 = \{1,2,3,4\}$, and the constraints:

$$\boldsymbol{\eta}_{\{2,3\};\{2,3,4\}}(\boldsymbol{i}_{\{2,3\}}) = 0, \quad \boldsymbol{\eta}_{\{2,3,4\};\{2,3,4\}}(\boldsymbol{i}_{\{2,3,4\}}) = 0,$$

$$\boldsymbol{\eta}_{\{1,4\};\{1,3,4\}}(\boldsymbol{i}_{\{1,4\}}) = 0, \quad \boldsymbol{\eta}_{\{1,3,4\};\{1,3,4\}}(\boldsymbol{i}_{\{1,3,4\}}) = 0.$$

If the four categorical variables are ordinal it is sensible to choose logits of type global for A_3 and A_4 within $\mathcal{M}_1$ and for A_1 and A_2 within $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$. As explained in Bartolucci, Colombi, Forcina (2004), who gave a general description of block recursive models of this type, it is convenient to use logits of type local for the explanatory variables A_3 and A_4 within $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$.

Furthermore together with the previous equality constraints the following inequality constraints:

$$\boldsymbol{\eta}_{\{2,4\};\{2,3,4\}}(\boldsymbol{i}_{\{2,4\}}) \geq 0, \quad \boldsymbol{\eta}_{\{1,3\};\{1,3,4\}}(\boldsymbol{i}_{\{1,3\}}) \geq 0,$$

state that the distributions of A_2 conditioned by the explanatory variables are stochastically increasing with the categories of A_4 and that the conditional distributions of A_1 are stochastically increasing with the categories of A_3. The problem of testing linear inequality constraints on marginal parameters has been discussed by Dardanoni, Forcina (1998), Colombi, Forcina (2001) and by Bartolucci, Colombi, Forcina (2004).

3 Multinomial State Space Models

In this section marginal models are used to introduce a class of dynamic models for multicategorical time series. For a survey of the state of art on categorical time series analysis see Fahrmeir, Tutz (1994), MacDonald, Zucchini (1997), Davis, Wang (1999) and Kedem, Fokianos (2002). Let $\boldsymbol{\pi}_t$ be the vector of the joint probabilities of the categories of q categorical variables given the information set $\mathcal{F}_{t-1}$ available at time t. We parametrize the joint probabilities $\boldsymbol{\pi}_t$ by inverting at time t the link function:

$$\boldsymbol{\eta}_t = \boldsymbol{C} \ln \boldsymbol{M}\boldsymbol{\pi}_t, \tag{3}$$

where the vector of marginal parameters is a linear function of time varying regressors: $\boldsymbol{\eta}_t = \boldsymbol{X}_t\boldsymbol{\beta}_t$ and where $\boldsymbol{\beta}_t$ changes according to a standard normal transition model:

$$\boldsymbol{\beta}_t = \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{H}\boldsymbol{\varepsilon}_t. \tag{4}$$

Here $\boldsymbol{\varepsilon}_t$ are independent multinormal random variables with null expected value and unknown diagonal variance matrix $\boldsymbol{Q}$. For a discussion of state space models for categorical data and count data see Kedem, Fokianos (2002), Durbin, Koopmann (1997) and Fahrmeir, Tutz (1996). Special cases of the previous general model (for example $\boldsymbol{X}_t = \boldsymbol{I}$, $\boldsymbol{H} = \boldsymbol{I}$ and $\boldsymbol{F} = \boldsymbol{I}$) are easily obtained and the advantage of defining the transition model in function of the marginal parameters rather than the log-linear ones come from the fact that the normal transition model applied to log-linear parameters is often difficult to interpret. On the contrary the transition model applied to marginal interactions and in first place to marginal Logits is very easy to interpret and a more natural and direct modelling strategy. Moreover in the context of categorical time series many important non-Granger causality type hypotheses, which state that a set of categorical variables doesn't depend on the past of another set of variables, given $\mathcal{F}_t$, are equivalent to linear hypotheses on marginal interactions and this fact enhances the importance of marginal models in this context. Finally in the context of marginal models it is easier to distinguish between hypotheses of simultaneous independence between categorical variables and hypotheses of independence of a categorical variable from the past of the others. These advantages of marginal modelling have been firstly pinpointed by Giordano (2003) in the context of models for the joint transition probabilities of multivariate Markov Chains and the problem of testing Granger non-causality under Markov assumptions was firstly considered by Bouissou *et al* (1986). The important topic of modelling multivariate Markov Chain was started by the works of Fahrmeir, Kaufmann (1987) and Kaufmann (1987) and generalized to a less stringent assumption than the one of Markovianity by Fokianos, Kedem (1998). Hidden Markov models (MacDonald, Zucchini, 1997) can also be considered in this context by substituting the normal transition model (4) with the following one:

$$\boldsymbol{\beta}_t = S_t \boldsymbol{\delta}_1 + (1 - S_t)\boldsymbol{\delta}_2$$

where the binary variable S_t indicates the state at time t of a two state markov Chain.
In this last case the maximum likelihood estimates are easily computed (MacDonald, Zucchini 1997, Krolzig 1997) and in the case of a normal transition model maximum likelihood estimation of the unknown parameters of the multivariate normal distribution of $\boldsymbol{\varepsilon}_t$ can be performed by the Montecarlo likelihood method of Durbin, Koopman (1997, 2001) or by the Montecarlo EM algorithm of Chan, Ledolter (1995). Less computationally demanding methods are the EM-type algorithm of Fahrmeir, Wagenpfeil (1997) and the method based on the maximization of an approximation of the log-likelihood of Durbin, Koopman (1997, 2001). Note that in the case of marginal models all the previous methods are more computationally demanding, than in the cases previously considered, because at every iteration the relation $\boldsymbol{\eta}_t = \boldsymbol{C} \ln \boldsymbol{M}\boldsymbol{\pi_t}$ must be inverted for every t.
The asymptotic properties of the M.L. estimator of the unknown parameters in the case of a latent Markov Chain with time homogeneous transition probabilities follow from the results of Bickel, Ritov, Ryden (1998) on Hidden Markov Models. The asymptotic normality of the M.L. estimators for non-normal state-space model is discussed in Jensen, Petersen (1999).

3.1 Bivariate Markov Driven Marginal Models

Often multi-categorical time series exhibit two different regimes. The starting time and the length of the spells in the regimes are random. To model the different behavior of the time series under the two regimes the parameters of a Marginal Model can be let to depend on the state of an unobservable Markov Chain which models the transitions between the regimes. A latent variable problem arises because the regime is not an observable variable. More precisely the model must consist of two parts:

I) a *Marginal Model* which specifies the joint probabilities of the categories of the variables at time t given the categories of the variables at the previous lag times $t-1, t-2, ..., t-lag$, given the values (at time $t-1$) of a vector of regressors $\mathbf{x}_{t-1}$ and given the regime S_t at time t ($S_t = 1$ or $S_t = 0$ in the case of two regimes).
II) a *two states Markov Chain* that models the history of the unobservable regimes S_t.
According to this model the observed multi-categorical time series is not Markovian, however conditionally on the series $\{S_t\}$ of the regimes it is a Markov Chain of order lag.
Here we examine the case of a bivariate categorical time series $\{A_{1,t}, A_{2,t}\}$. The joint probability function of $A_{1,t}$ and $A_{2,t}$ conditionally on the past can be specified by a log-linear model. Let $\boldsymbol{Z}_t$ be the vector of predetermined variables at time t and of the unobservable regime S_t. Then, the log-linear model:

$$\ln \pi_{ij,t} = \lambda_t + \lambda_{i,t}^{A_1} + \lambda_{j,t}^{A_2} + \lambda_{ij,t}^{A_1 A_2}, \qquad i = 1, 2, ..., a_1, j = 1, 2, ..., a_2,$$

could be introduced by allowing the interaction parameters *lambda* to depend on the vector $\boldsymbol{Z}_t$ of predetermined variables. This approach doesn't allow a direct parameterization of the marginal probabilities $\pi_{i.,t}$, $\pi_{.j,t}$. For this reason we prefer to parametrize the marginal probabilities directly with univariate logit Models. For example the Continuation logit Parameterization (Colombi, Forcina 1999) for the marginal probabilities is given by the following formulae:

$$\pi_{i.,t} = \frac{\exp\{-\eta_{1,t}(i)\}}{\prod_{m=1}^{i}\left[1 + \exp\{-\eta_{1,t}(m)\}\right]}, i = 1, 2, ..., a_1 - 1,$$

$$\pi_{.j,t} = \frac{\exp\{-\eta_{2,t}(j)\}}{\prod_{m=1}^{j}\left[1 + \exp\{-\eta_{2,t}(m)\}\right]}, j = 1, 2, ..., a_2 - 1.$$

Here we have slightly simplified the notation of interactions given in section two by omitting curly brackets and the indication of the marginal within which the interaction is defined. The Continuation Logits $\eta_{1,t}(i)$ and $\eta_{2,t}(j)$ depend on the vector of predetermined variables $\boldsymbol{Z}_t$ according to linear predictors of the type commonly used in the context of logit regression (see section 4 for an example). Note that the Continuation logit of a categorical variable may depend also on the past of the other categorical variable. The joint probabilities $\pi_{ij,t}$ are specified by the marginal continuation logits and by the logarithms of the Continuation Odds Ratios (Colombi, Forcina 1999):

$$\eta_{12,t}(ij) = \ln \frac{\pi_{ij,t} \cdot \sum_{m=i+1}^{a_1} \sum_{n=j+1}^{a_2} \pi_{mn,t}}{\sum_{m=i+1}^{a_1} \pi_{mj,t} \cdot \sum_{n=j+1}^{a_2} \pi_{in,t}},$$
$$i = 1, 2, ..., a_1 - 1, \quad j = 1, 2, ..., a_2 - 1.$$

The following hypotheses on the Continuation Odds Ratios are relevant:

$$\begin{aligned}\eta_{12,t}(ij) &= \eta_{12}(ij),\\ \eta_{12,t}(ij) &= \eta_{12}(ij)+\rho S_t.\end{aligned}$$

Both are hypotheses of constant association in the sense that the Continuation Odds Ratios do not depend on the past of A_1 and A_2 and on a vector of regressors $\mathbf{x}_{t-1}$. In the first case, the Odds Ratios are also regime independent, whereas in the second case the Continuation Odds Ratios depend on the latent regime but the effect of the regime is the same for all i and j ($i = 1, 2, ..., a_1 - 1; j = 1, 2, ..., a_2 - 1$). A more parsimonious model is given by the following hypotheses of Uniform Constant association:

$$\begin{aligned}\eta_{12,t}(ij) &= \eta_{12},\\ \eta_{12,t}(ij) &= \eta_{12}+\rho S_t.\end{aligned}$$

Finally the transition probabilities of the Hidden Markov Chain $p_{00t} = p(S_{t+1} = 0|S_t = 0)$ and $p_{11t} = p(S_{t+1} = 1|S_t = 1)$ can assumed to be function of a vector of regressors $\mathbf{x}_{t-1}$ according to the logit Models:

$$\ln\frac{p_{iit}}{1-p_{iit}} = \alpha_{0i} + \boldsymbol{\alpha}'_{1i}\mathbf{x}_{t-1}, i = 0, 1. \tag{5}$$

The case of a time homogeneous transition matrix is obtained by putting $\boldsymbol{\alpha}_{1i} = \mathbf{0}$, $i = 0, 1$.

Given the marginal continuation logits and the Continuation Odds Ratios the joint probabilities $\pi_{ij,t}$ can be computed with the iterative algorithm introduced by Colombi, Forcina (1999) and described in Colombi, Zanarotti (2002).

Let $\boldsymbol{\vartheta}'{=}\left[\alpha_{00}, \boldsymbol{\alpha}_{10}, \alpha_{01}, \boldsymbol{\alpha}_{11}, \boldsymbol{\theta}'\right]$ be the vector of the parameters to be estimated where $\boldsymbol{\theta}$ is the vector of the parameters of the bivariate marginal model. Given the parameters, the BLHK filter and smoother (Krolzig, 1997) can be used to marginalize with respect the unobservable Markov Chain and to compute the log-likelihood at every iteration of the Fisher Scoring algorithm.

3.2 State Space Trend Models for categorical data

Marginal State Space Models for categorical data can be specified in many ways thanks to the flexibility of the definition of $\boldsymbol{\eta}_t$ and of the transition model: $\boldsymbol{\eta}_t = \boldsymbol{X}_t\boldsymbol{\beta}_t, \quad \boldsymbol{\beta}_t = \boldsymbol{F}\boldsymbol{\beta}_{t-1} + \boldsymbol{H}\boldsymbol{\varepsilon}_t$. A first important and useful case is given by the $(k-1)$- polynomial stochastic trend where some components $\eta_{i,t}$ change according to the transition model:

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{F}\boldsymbol{\beta}_{i,t-1} + \boldsymbol{\varepsilon}_t \quad \eta_{i,t} = \beta_{1,it}$$

and $\boldsymbol{F}$ is a $k \cdot k$ upper triangular matrix of ones. A second important example is the case of k order random walk where some components $\eta_{i,t}$ of $\boldsymbol{\eta}_t$ change according to the transition model:

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{F}\boldsymbol{\beta}_{i,t-1} + \boldsymbol{h}\boldsymbol{\varepsilon}_{i,t}$$
$$\eta_{i,t} = \beta_{1,it}.$$

here $\boldsymbol{h}$ is the first column of a $k \cdot k$ identity matrix and $\boldsymbol{F}$ is a $k \cdot k$ identity matrix with the first row replaced by the row vector $\boldsymbol{c}$, the $i-th$ element of which is $c_i = (-1)^{i-1}\begin{pmatrix} k \\ i \end{pmatrix}$. These models are useful to model local-trends for logits defined within different marginals. For example in the Bivariate Case introduced in the previous section a local level model (k=1) can be applied to the two marginal continuation logits:

$$\eta_{1,t}(i) = \eta_{1,t-1}(i) + \varepsilon_{1,t},$$
$$\eta_{2,t}(i) = \eta_{2,t-1}(i) + \varepsilon_{2,t}.$$

4 Ground O3 and CO data analysis

The Hidden Markov models described in section 3.1 are used to analyze daily levels of ground *O3* (variable A_{1t}) and *CO concentration* (variable A_{2t}) both with three categories (*low (1), normal(2)* and *high(3))*. Data are taken by San Giorgio (Bergamo-Italy) measurement unit from 1997 to 1999. In this application the covariates that affect the continuation Logits are: *temperature* and *solar radiation*.
The general effects of the linear predictors are assumed to change according to the hidden regime and the other parameters (additive effects, interactions, regression coefficients) are regime independent. More precisely the most general linear predictor used for the $\eta_{1,t}(i), i = 1, 2, ..., a_1 - 1$ is:

$$\eta_{1,t}(i) = \left(\mu_j^{(0)} + \delta_j S_t\right) +$$
$$+\left(\sum_{l=1}^{lag}\sum_{m=1}^{2}\theta_{ml}^{A_1} I_{\{A_{1,t-l}=m\}} + \sum_{l=1}^{lag}\sum_{m=1}^{2}\theta_{ml}^{A_2} I_{\{A_{2,t-l}=m\}}\right) +$$
$$+\left(\sum_{l=2}^{lag}\sum_{m=1}^{2}\delta_{ml}^{A_1}\prod_{k=t-l}^{t-1} I_{\{A_{1,k}=m\}} + \sum_{l=2}^{lag}\sum_{m=1}^{2}\delta_{ml}^{A_2}\prod_{k=t-l}^{t-1} I_{\{A_{2,k}=m\}}\right) +$$
$$+\beta_1 x_{1t} + \beta_2 x_{2t}.$$

A similar predictor is used for the $\eta_{2,t}(j), j = 1, 2, ..., a_2 - 1$. In the first column of Table 1 it is given the number LAG of past pollutant levels that

TABLE 1. Switching Bivariate Marginal Models (O3 and CO)

lag	link	association	log-lik.	n. par.
1	add.	$\eta_{12,t}(ij) = 0$	-919.08	18
2	add.	$\eta_{12,t}(ij) = 0$	-894.08	26
3	add.	$\eta_{12,t}(ij) = 0$	-873.49	34
4	add.	$\eta_{12,t}(ij) = 0$	-858.88	42
4	add.+int.	$\eta_{12,t}(ij) = 0$	-846.27	78
4	add.+int.	$\eta_{12,t}(ij) = \eta_{12}$	-846.26	79
4	add.+int.	$\eta_{12,t}(ij) = \eta_{12}(ij)$	-845.22	82
4	add.+int.+reg.	$\eta_{12,t}(ij) = \eta_{12}(ij)$	-842.52	86

TABLE 2. One step forecasts-O3

predicted→ / *observed↓*	**low**	**normal**	**high**.	tot.
low	**704**	61	0	765
normal	86	**189**	3	278
high.	0	12	**5**	17
tot.	790	262	8	1060

TABLE 3. One step forecasts-CO

predicted→ / *observed↓*	**low**	**normal**	**high**	tot.
low	**152**	102	0	254
normal	56	**709**	5	770
high	0	23	**13**	36
tot.	208	834	18	1060

affects the current one. In the second column the linear predictor used is described (*add.* means that the effect of the LAG previous levels is additive and *add.+int.* means that interactions between time adjacent past levels of the same pollutant are also allowed and *add.+int.+reg.* is the general case where also the effects of the covariates *temperature* and *solar radiation* are introduced). In the third column the type of association between CO and O3, given the past levels and the hidden regime, is described. In the fourth column the value of the log-likelihood is reported and in the last column the number of parameters is given. For all the models considered the transition probabilities of the Hidden Markov Chain are time invariant. In the last two tables the one-step predicted levels are crossed with the actual ones, using the model in the last row of Table 1.

In Table 4 the results obtained by using some State Space Trend Models

TABLE 4. Bivariate State Space Models (O3 and CO)

model	number of states	log-lik.
M_1	8	-117.558
M_2	5	-118.444
M_3	12	-115.358
M_4	9	-116.771

introduced in section 3.2 are reported. In this case only the first 100 observations were used, covariates effects were not included and local logits and local odds-ratios were used instead of the continuation ones. In the case of the first model M_1 the four local logits and the four local odds-ratios that parametrize the joint distribution at time t changes according to a random walk. In model M_2 the four odds ratios are assumed to be equal and the five parameters still changes according to a random walk. According to model M_3 the four odds ratios changes according to a random walk and the four logits changes according to a local level local trend model (local polynomial of order one). In model M_4 the transition equation for the logits is as in model M_3 and the four odds-ratios are equal and change according to a random walk. Initial states have been treated as unknown parameters so that the number of parameters to be estimated is twice the number of states. The method based on the maximization of the approximate log-likelihood of Durbin, Koopman (2001) were used but after convergence the log-likelihood was computed with the importance-sampling method of Durbin, Koopman (2001).

Acknowledgments: This work has been supported by the COFIN 2002 project, reference 2002133957

References

Bartolucci, F., Colombi, R., Forcina, A. (2004). An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints, submitted to *Annals of Statistics* .

Bartolucci, F., Forcina, A., Dardanoni, V. (2001). Positive quadrant dependence and marginal modelling in two-way tables with ordered margins, *Journal of the American Statistical Association*, 96, pp. 1497-1505.

Bergsma, W. P., Rudas, T. (2002). Marginal models for categorical data, *Annals of Statistics*, 30, pp. 140-159.

Bickel, P.J., Ritov, Y., Ryden, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics*. 26, pp. 1614-1635.

Bouissou, B., Laffont, J., Vuong, H. (1986) Tests of noncausality under Markov assumptions for qualitative panel data. *Econometrica*, 54, pp. 395-414.

Chan, S., Ledolter, J. (1995). Monte Carlo EM Estimation for Time Series Models Involving Counts, *Journal of The American Statistical Association*, 90, pp. 242-252.

Colombi, R., Forcina, A. (1999). An instance of generalized log-linear models with inequality constraints: the continuation logit parameterization, *Proceedings of the 14th International Workshop on Statistical Modelling*, edited by H. Friedl., Gratz.

Colombi, R., Forcina, A. (2001). Marginal regression models for the analysis of positive association of ordinal response variables, *Biometrika*, 88, pp. 1007-1019.

Colombi, R., Zanarotti, C. (2002). A markov driven bivariate logit model. *Studi in onore di Angelo Zanella*, edited by Frosini B.V., Magagnoli U., Boari G., pp 125-135, Vita e Pensiero, Milano.

Cox, R., Wermuth, N. (1996). *Multivariate dependencies, Models analysis and interpretation*, Chapman Hall, London.

Dardanoni, V., Forcina, A. (1998). A Unified approach to likelihood inference on stochastic orderings in a nonparametric context, *Journal of the American Statistical Association*, 93, pp. 1112-1123.

Davis, R., Wang, Y. (1999). Modelling Time series of Count Data, *Asymptotics, Nonparametrics and Time Series* edited by Ghosh S., Marcel Dekker, New York.

Douglas, R., Fienberg, S. E., Lee, M. T., Sampson, A. R., Whitaker, L. R. (1990). Positive dependence concepts for ordinal contingency tables, in *Topics in statistical dependence*, edited by Block H. W., Sampson A. R. and Sanits T. H., Institute of Mathematical Statistics, Lecture Notes, Monograph series, Haywar, California.

Durbin J., Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika*, 84, pp. 669-684.

Durbin, J., Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press, New York.

Fahrmeir, L., Kaufmann, H. (1987). Regression models for nonstationary categorical time series, *Journal of Time Series Analysis.* 8, pp. 147-160.

Fahrmeir, L., Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models* Springer, Berlin.

Fahrmeir, L., Wagenpfeil, W. (1997). Penalized Likelihood estimation and iterative Kalman smoothing for non-Gaussian dynamic regression models, *Computational Statistics and data analysis*, 24, pp.295-320

Fokianos, K., Kedem, B. (1998). Prediction and classification of non-stationary categorical time series. *Journal of Multivariate Analysis*, 67, pp. 277-296.

Forthofer, R., Koch, G. (1973). An analysis of compounded functions of categorical data, *Biometrics*, 29, pp. 143-157.

Giordano, S. (2003). *Modelli Parametrici per catene di Markov bivariate*, Tesi di Dottorato, XV ciclo, Università di Milano-Bicocca, Milano.

Glonek, G. (1996). A class of regression models for multivariate categorical responses, *Biometrika*, 83, pp. 15-28.

Glonek, G., McCullagh, P. (1995). Multivariate Logistic Models, *Journal of the Royal Statistical Society B*, 57, pp. 533-546.

Grizzle, J., Starmer, F., Koch, G. (1969). Analysis of categorical data by linear models, *Biometrics*, 25, pp. 489-505.

Jensen, J., Petersen, N. (1999). Asymptotic normality of the maximum likelihood estimator in State Space models. *Annals of Statistics*, 27, pp. 514-535.

Kaufmann, H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory, *The Annals of Statistics*, 15, pp. 79-98.

Kedem, B., Fokianos, K. (2002). *Regression Models for Time Series Analysis*, Wiley, New York.

Krolzig, M. (1997). *Markov Switching Vector Autoregression*, Springer Berlin.

Lang, J., Agresti, A. (1994). Simultaneously modelling the joint and marginal distributions of multivariate categorical responses, *Journal of the American Statistical Association*, 89, pp. 626-632.

MacDonald, I., Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete valued Time Series*, Chapman Hall, London.