Stochastic models for ordinal panel data with individual and time-varying latent effects

FULVIA PENNONI
University of Milano-Bicocca
Department of Statistics and Quantitative Methods
Via Bicocca degli Arcimboldi 8, 20126 Milan
ITALY
fulvia.pennoni@unimib.it

GIORGIO VITTADINI
University of Milano-Bicocca
Department of Statistics and Quantitative Methods
Via Bicocca degli Arcimboldi 8, 20126 Milan
ITALY
giorgio.vittadini@unimib.it

Abstract: We illustrate a model based approach for clustering when we deal with longitudinal data. The model formulation is specifically tailored when we deal with a continuous variable which is properly discretized so that a univariate ordinal response variable results. We show how to estimates the effects of interest by means of a suitable parameterization based on global logits as well as to group units by considering the unobserved heterogeneity among them. According to the chosen distribution for the latter an hidden Markov model or a mixture of auto-regressive process AR(1) results. The model estimation is carried out by means of the expectation-maximization algorithm and the Newton-Raphson algorithm. Standard errors are obtained by using the observed information matrix. A way to obtain reliable individual prediction is illustrated. We also show an application to real data.

Key–Words: EM algorithm, Latent Markov model, Mixture latent auto-regressive model, Path prediction

1 Introduction

We show a flexible model formulation to deal with panel data which may be seen a promising tool to be used in many practical applications as it is a model-based approach for clustering. In fact, it can be applied when the interest lies in an ordinal response variable which may results from a properly discretized continuous variable having a limited number of categories and both when few or many time occasions are available.

The proposed model is based on two different formulations of the distribution of the latent process in order to properly model the unobserved heterogeneity. When the model is formulated according to a discrete distributive assumption on the unobserved heterogeneity a latent Markov model results (for a review see Bartolucci, Farcomeni, Pennoni, 2013 and Bartolucci, Farcomeni, Pennoni, 2014). Instead, when it is formulated according to continuous distributive assumption on the unobserved heterogeneity a mixture of auto-regressive processes results (Bartolucci, Bacci, Pennoni, 2014). The adopted parameterization accounts for a model which is parsimonious, easy to interpret and with hypotheses which may be interesting to test.

We illustrate the model formulation by focusing on a derived ordinal variable. We show the maximum likelihood estimation procedures for each type of model formulation. We also illustrate the specific model selection procedure required by the two types of model.

To show an application of the models we relay on data on some hospitals in Lombardy to examine if they have efficiency gains during the period 2008-2011. In such a context, the latent variables have the role to account for hospital unobserved heterogeneity by introducing a unit specific random intercept. This aspect is in connection with the inclusion of observed covariates which may also be time-varying and do not fully explain the heterogeneity between the level of efficiency gained by different hospitals.

2 Notation and specification of the proposed models

With reference to a sample of \( n \) units observed at \( T \) time occasions, let \( y_{it} \) be the ordinal response variable for unit \( i \) at occasion \( t \). Let the number of categories denoted by \( J \), and let \( x_{it} \) be a corresponding column vector of covariates, with \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \). We also denote by \( y_i = (y_{i1}, \ldots, y_{iT}) \) the vector of response variables and by \( X_i = (x_{i1} \ldots, x_{iT}) \) the matrix of time-varying and time-constant covariates for unit \( i \).

The model we formulate is based on the assumption that \( y_{it} = G(y_{it}^*) \), where \( y_{it}^* \) follows the model

\[
y_{it}^* = \alpha_{it} + x_{it}' \beta + \eta_{it}, \quad t = 1, \ldots, T,
\]
for $i = 1, \ldots, n$, with $\eta_{it}$ being independent error terms with a standard logistic distribution, and $G(\cdot)$ is a link function which models the relationship between each response variable $y_{it}$ and the corresponding latent variable $\alpha_{it}$ and the vector of covariates $x_{it}$. In such a case it is a function of cutpoints $\mu_1 \geq \cdots \geq \mu_{J-1}$ and it can be formulated as

$$G(y^*) = \begin{cases} 1 & y^* \leq -\mu_1, \\ 2 & -\mu_1 < y^* \leq -\mu_2, \\ \vdots & \vdots \\ J & y^* > -\mu_{J-1}. \end{cases}$$

The basic assumptions of the model are that for every sample unit $i$, $y_{it}^*$, $\ldots$, $y_{iT}^*$ are conditionally independent given $(\alpha_{i1}, \ldots, \alpha_{iT})$ and $X_i$.

Due to the ordinal nature of the response variable we have

$$\log \frac{p(y_{it} \geq j|\alpha_{it}, x_{it})}{p(y_{it} < j|\alpha_{it}, x_{it})} = \mu_j + \alpha_{it} + x_{it}'\beta,$$  \quad (1)

with $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 2, \ldots, J$. These parameterization is based on global logits for the conditional distribution of each response variable and it is particularly suitable as we deal with an underlying continuous outcome which is suitable discretized (McCullagh, 1980).

In the following we illustrate the use two different types of distribution of the latent variable. The discrete latent process formulation assumes that, for all $i$, $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{iT})$ follows a first-order homogeneous Markov chain with $k$ states denoted by $\xi_1, \ldots, \xi_k$. This chain has initial probabilities $\pi_h$ and transition probabilities $\pi_{h_1h_2}$, with

$$\pi_h = p(\alpha_{i1} = \xi_h), \quad h = 1, \ldots, k,$$

$$\pi_{h_1h_2} = p(\alpha_{it-1} = \xi_{h_1}, \alpha_{it} = \xi_{h_2}),$$

where $h_1, h_2 = 1, \ldots, k$, $t = 2, \ldots, T$. It is assumed that every $\alpha_{it}$ is conditionally independent of $\alpha_{i1}, \ldots, \alpha_{i,t-2}$ given $\alpha_{i,t-1}$, but apart from this assumption, the distribution of $\alpha_i$ is unconstrained. To ensure identifiability we require that $\sum_h \pi_h = 1$ and $\sum_{h_2} \pi_{h_1h_2} = 1$ and one component of the support point is constrained to be zero. In such case a latent Markov (LM) model with covariates results where the covariates affect the measurement model.

The continuous latent process formulation assumes that the hidden response variables in $y_{i1}^*, \ldots, y_{iT}^*$ are conditionally independent given $X_i$ and the latent process $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{iT})$. Another hypothesis is that every hidden variable and then every response variable, only depends on $\alpha_{it}$ and $x_{it}$ and that the latent process $\alpha_i$ has distribution given by a mixture of $k$ AR(1) processes with common variance $\sigma^2$. According to the latter we assume the existence of a discrete latent variable $u_i$, for $i = 1, \ldots, n$, having a distribution with $k$ support points and mass probabilities $\pi_1, \ldots, \pi_k$ such that when $u_i = h$ we assume that

$$\alpha_{it} = \xi_h + \eta_{it}, \quad i = 1, \ldots, n,$$

and that

$$\alpha_{it} = \xi_h + (\alpha_{i,t-1} - \xi_h)\rho_h + \eta_{it}\sqrt{1 - \rho_h^2},$$

where $i = 1, \ldots, n$, $t = 2, \ldots, T$, and $\eta_{it} \sim N(0, \sigma^2)$ for all $i$ and $t$ and $(\xi_h, \rho_h)$ are parameters which for $h = 1, \ldots, k$ are estimated jointly with the common variance. To ensure identifiability of the model, we require that $\xi_1 = 0$ or, $\sum_h \xi_h \pi_h = 0$.

## 2.1 Estimation details

Given a sample of $n$ independent units, the model log-likelihood is

$$\ell(\theta) = \sum_{i=1}^{n} \log p(y_i|X_i)$$

where $\theta$ is the vector of all free parameters affecting $p(y_i|X_i)$. The latter is the manifest distribution of the response vector $y_i$ given all the observable covariates $X_i$. The model estimation is performed by the Expectation-Maximisation (EM) algorithm which is based on the complete data log-likelihood.

When we are assuming the discrete latent formulation the complete data log-likelihood has expression

$$\ell^*(\theta) = \sum_{i=1}^{n} \left\{ \sum_{t=1}^{T} \sum_{h=1}^{k} \sum_{y=0}^{J-1} \alpha_{ih}^{(t)} \log p(y_{it}|h, x_{it}) ight. \\
left. + \sum_{h_{1}}^{k} \sum_{h_{2}}^{k} \sum_{t=2}^{T} \sum_{h_{1}=1}^{k} \sum_{h_{2}=1}^{k} b_{ih}^{(t)} \pi_{ih} \pi_{ih} \right\},$$

where $b_{ih}^{(t)}$ is a dummy variable for unit $i$ in component $h$ at occasion $t$, with reference to the same occasion and the same unit, $b_{ih}^{(t)}$ is a dummy variable equal to 1 if this unit moves from state $h_{1}$ to state $h_{2}$, whereas $\alpha_{ih}^{(t)}$ is equal to 1 if the unit is in state $h$ and provide response $y$ and covariate configuration $x$. The conditional response probabilities $p(y_{it}|h, x_{it})$ are computed efficiently by using some recursions known in the hidden Markov model literature (Zucchini and MacDonald, 2009).

Whereas when we are assuming the continuous latent formulation the manifest distribution of the response vector $y_i$ given all the observable covariates
\( X_i \) is expressed through a \( T \)-dimensional integral which is approximately computed by a quadrature method based on a series of \( q \) nodes properly chosen. The expression for \( p(y_i | X_i) \) based on the quadrature is an approximation which depends on the number of integration points used. The nodes are taken on an equispaced grid of points, to which we refer as \( \nu_m, m = 1, \ldots, q \). On the basis of this choice the complete data log-likelihood may be expressed as

\[
\ell^c(\theta) = \sum_{i=1}^{n} \sum_{h=1}^{k} w_{ih} \left\{ \sum_{m=1}^{q} \sum_{t=1}^{T} z_{imt} \log p(y_{it} | \nu_m, x_{it}) + \log \pi_h, + \sum_{m_1=1}^{q} \sum_{m_2=1}^{q} \sum_{t=2}^{T} z_{im_1m_2t}^{*} \log \omega_{m_1m_2}^{(h)} \right\},
\]

where \( w_{ih} = I \{ u_i = h \} \) is a dummy variable for unit \( i \) in component \( h \), \( z_{imt} = I \{ \alpha_{it} = \nu_m \} \) is a dummy variable for unit \( i \) given the \( m \)-th quadrature point for the integral with respect to \( \alpha_{ij} \) \( z_{im_1m_2t}^{*} = z_{im_1t-1} z_{im_2t} \) is a dummy variable for unit \( i \) given the \( m_2 \)-th quadrature point for the integral with respect to \( \alpha_{it} \) given the \( m_1 \)-th quadrature point used for the integral with respect to \( \alpha_{it-1} \) and \( \omega_{m_1m_2}^{(h)} \) denotes the \( m_2 \)-th weight for the integral with respect to \( \alpha_{it} \) given the \( m_1 \)-th quadrature point for the integral with respect to \( \alpha_{it-1} \).

The EM algorithm alternates the following steps until convergence: the E-step of the algorithm computes the conditional expected values of dummy variables given the observed data and the current parameter vector \( \theta \); the M-step of the algorithm updates the model parameters by maximising the posterior probabilities. Since the EM algorithm is rather slow to converge, after a certain number of EM steps we switch to a full Newton-Raphson algorithm to maximise the model log-likelihood \( \ell(\theta) \).

From the EM algorithm we obtain the score vector as

\[
s(\theta) = E_{\hat{\theta} \sim \theta} [s^*(\theta)|obs.data],
\]

where the expected value is at the parameter value \( \hat{\theta} \) and \( s^*(\theta) = \partial \ell^c(\theta)/\partial \theta \) is the score vector of the complete-data log-likelihood whose expected value is computed at the beginning of each M-step. We then compute the observed information matrix \( J(\theta) \) as the minus the numerical derivative of \( s(\theta) \) obtained as above (Pennoni, 2014). The standard errors for the parameter estimates are obtained from \( J(\theta)^{-1} \).

The selection of the appropriate number of latent states in the latent Markov model formulation is made relying on the Bayesian Information Criterion (Schwarz, 1978). For the mixture latent autoregressive model the selection of the number components is made first selecting the number of quadrature points and then selecting the number of mixtures first trying to increase values of \( q \) until the maximum of \( \ell(\theta) \) does not significantly change with respect to the previous value of \( q \). Then we select the number of states relying on the BIC criterion in which takes into account the goodness-of-fit and the parsimony of the model.

It is important to mention that for both models we adopted a proper estimation procedure to prevent problems due to the multimodality of the likelihood function by applying a multi-start strategy combining a deterministic rule with a random starting rule.

Moreover on the basis of the final parameters estimates \( \hat{\theta} \) we can compute the predictions of the entire sequence of latent states \( \alpha_{it} \) for unit \( i \) which corresponds to the maximum with respect to \( h_1, \ldots, h_k \) of the posterior probabilities.

### 3 Some main findings from an application

We applied the proposed model data deriving from a large administrative database provided by the health care department of Lombardy region regarding hospital’s features. The data cover the full population of patients for the general medicine ward. They are related to 120 hospitals and cover the years 2008, 2009, 2010 and 2011. The variables of interest are the yearly revenues from discharges and the number of outpatient discharges in the ward. We suggest to consider the ratio between them as it accounts for the more complex case mix of the patients and it can be interpreted as an efficiency monetary measure of the hospital. For every hospital the following time-varying covariates are also available: the total number of beds, the yearly hours of activity of physicians, nurses and other employees of the hospital and the hours of activity of the surgery rooms.

We estimate the models proposed in Section 2 to the data. First we employ the model selection procedures as illustrated in Section 2.1 for the latent Markov model and on the basis of the BIC index we select \( k = 4 \) latent states. We select the number of quadrature points for the mixture latent autoregressive model. The results are \( q = 91 \) for \( k = 1 \), 81 for \( k = 2 \) and 111 for \( k = 3 \). Then on the basis of the BIC value we select one mixture component.

In Table 1 we show the estimates of the parameters referred to the cutpoints and the regression coefficients in equation (1), together with the corresponding standard errors for both types of models to which we refer as LM and MLAR. We use the translog function for the covariates. We have implicitly changed
the scale of the latent response variable (Bauer, 2009) and therefore we get that the estimates of the thresholds and of the regression coefficients of the mixed latent auto-regressive model are on another scale with respect to that of the other model.

### 4 Conclusion

The proposed model accounts for two types of formulation of the latent effect. The first one assuming a discrete distribution gives rise to a latent Markov model which is not very complex to fit. It is more natural in many contexts and very suitable for classification even if the number of parameters increases with the number of latent states. The second model formulation relies on a continuous distribution for the unobserved heterogeneity. It gives rise to a mixture latent auto-regressive model which is more complex to fit as the distribution of the observed data given the covariates is obtained by solving a $T$ dimensional integral. Maximum likelihood estimation of the model parameters is performed by a joint use of the Expectation-Maximization algorithm and of the Newton-Raphson algorithm. Standard errors for the parameter estimates are obtained by exploiting the observed information matrix. The number of latent states are selected by considering the BIC index for the latent Markov model and by an appropriate strategy for the mixture components. We can also obtain the prediction of the individual effect for every unit at each time occasion on the basis of the parameter estimates.

To illustrate the proposal we applied the models to data related to the hospitals by considering a derived ordinal variable which accounts for the efficiency of the hospital spending policy observed over a four year period. When we estimate the model by relying on a discrete distribution of the unobserved heterogeneity we select a latent Markov model with four latent states. The latter are clusters of hospitals sharing the same propensity towards efficiency gains. When we estimate the model by relying on a continuous distribution of the unobserved heterogeneity we select a mixture latent auto-regressive model with one component and none of the covariates are significant.

**Acknowledgements:** We acknowledge “Finite mixture and latent variable models for causal inference and analysis of socio-economic data” (FIRB - Futuro in ricerca) funded by the Italian Government (RBFR12SHVV). F. Pennoni also thanks the financial support of the STAR project “Statistical models for human perception and evaluation”, University of Naples Federico II.

**References:**


