

Continuous versus discrete latent structures in dynamic latent variable models

Strutture latenti continue e discrete nei modelli a variabili latenti dinamici

Silvia Cagnone and Francesco Bartolucci

Abstract The paper proposes a comparison between dynamic models for panel data with continuous and discrete latent variables. We consider Limited Dependent Variable models (LDV) in the first case, and Latent Markov (LM) models in the second case. In both instances we use the maximum likelihood estimation method through the EM algorithm. Since the likelihood of LDV models is not tractable analytically, we implement the Gauss Hermite and the Adaptive Gauss Hermite quadrature methods for approximating the integrals involved. The comparison between the two classes of models is carried out by means of a simulation study.

Abstract *In questo lavoro proponiamo un confronto tra modelli dinamici per dati panel con strutture latenti continue e discrete. In particolare, consideriamo modelli con variabili dipendenti limitate (LDV) nel primo caso e modelli latenti Markoviani (LM) nel secondo caso. Per entrambi i modelli usiamo il metodo di stima della massima verosimiglianza tramite l'algoritmo EM. La verosimiglianza dei modelli LDV non è trattabile analiticamente, pertanto abbiamo implementato la quadratura di Gauss Hermite e la quadratura di Gauss Hermite adattiva per approssimare gli integrali coinvolti in essa. Il confronto tra le due classi di modelli è basato su uno studio di simulazione.*

Key words: Limited Dependent Variable models, Latent Markov models, EM algorithm, Gauss Hermite quadrature, Adaptive Gauss Hermite quadrature

Silvia Cagnone

Department of Statistical Sciences, University of Bologna e-mail: silvia.cagnone@unibo.it

Francesco Bartolucci

Department of Economics, Finance and Statistics, University of Perugia e-mail: bart@stat.unipg.it

1 Introduction

Statistical models for the analysis of panel data aim at accounting for the not observable heterogeneity between individuals, by including time varying latent variables in the model of interest. These latent variables can be treated as continuous or discrete. In the first case the dynamics of the latent variables are typically modeled by assuming that these variables follow an autoregressive process of order 1, denoted as usual by AR(1) [9]. Within the variety of dynamic models discussed in the literature in the case of continuous latent variables we refer, in particular, to Limited Dependent Variable (LDV) models for discrete data, which are very common in social sciences [13].

If the latent variables are considered discrete we can refer to Latent Markov (LM) models that assume that the individual effects follow a first-order Markov chain [2]. Both continuous and discrete processes present advantages and disadvantages. The former allow us to properly capture the variability in the data through an autoregressive structure resulting more parsimonious than the latter. On the other hand, the estimation procedure of models based on a continuous process presents some computational difficulties related to the presence of the time-varying latent variables. These variables have to be integrated out from the likelihood function and an analytical solution for this does not exist. In the literature different solutions to the integration problem of this class of models have been discussed. Among the methods based on numerical integration, [10] proposed a non-linear filter algorithm for the LDV model estimation specified in the state space models framework and approximated the uni-dimensional integrals through the Gauss Hermite (GH) quadrature. However, GH based methods guarantee accurate parameter estimates when several quadrature points are used per each dimension and, for certain target functions, instability problems arise in the phase of maximization when the function is approximated through this quadrature method. To overcome these problems, [7] proposed to use the Adaptive Gaussian Hermite (AGH) quadrature method to approximate the uni-dimensional integrals involved in the likelihood obtained using the non-linear filtering algorithm discussed by [10]. They compared the performance of this approximation with the classical GH and found that AGH outperforms GH in all the cases in which the process is highly persistent.

These computational problems are not present in LM models because of the discrete nature of the latent variables. This class of models can be viewed as semi-parametric models since a parametric form can be assumed for the response variables whereas the initial and transition probabilities of the latent states are estimated from the data. However, these models result less parsimonious and the interpretation of discrete latent variables can result more difficult than the continuous case.

In the literature there is a wide debate on the comparison between considering discrete and continuous latent variables. For the item response models [6] and [1] have shown that the shape of the distribution does not greatly affect the parameter estimates. Because of the arbitrariness about the assumption of distribution of the latent variables, [6] suggested to use a semi-parametric ML estimation where they estimated the probabilities of the discrete distribution defined on a fixed grid of

support points. [18] extended this approach by considering a fully semi-parametric method where both support points and probabilities are estimated from the data. A nonparametric estimation of item response models was examined by [12] and in the more general framework of generalized linear latent variable models [11] proposed a semi-parametric specification of the density of the latent variable. More recently, [4] combined continuous and discrete processes by proposing a more flexible model based on a mixture of AR(1) for the latent process.

In this paper we compare the performance of LDV and LM models by means of a simulation study. LM models are estimated using the maximum likelihood estimation via the EM algorithm [5, 3]. We propose the same estimation method for LDV models and we implement both GH and AGH quadrature methods for approximating the nested uni-dimensional integrals involved in the likelihood function of these models. The version of AGH that allows us to easily approximate nested integrals is the one proposed by [16] in the Bayesian context and by [17] for multilevel models. We also compare the results with the rectangular quadrature method proposed by [4].

The paper is organized as follows. In Section 2 we review the specification of the dynamic latent variable models in the case of continuous and discrete latent variables. In particular we focus on the case of ordinal observed variables. Section 3 discusses model estimation. For LDV models GH and AGH approximation methods are implemented. Section 4 shows some results of the simulation study and Section 5 gives some remarks.

2 Dynamic latent variable models for panel data

Panel data consist of repeated observations on the same individuals over time. They can be analyzed by means of LDV models that assume continuous latent variables α_{it} , interpreted as time and unit dependent random intercepts that account for the unobserved heterogeneity between subjects. They are assumed to follow a first order stationary autoregressive process. We denote the response variable for unit i at occasion t by y_{it} , $i = 1, \dots, n$, $t = 1, \dots, T$, and, since covariates are also typically observed, we denote by $\mathbf{x}_{it}$ the vector of the covariates corresponding to y_{it} . The proposed formulation is based on the following equations for $i = 1, \dots, n$ and $t = 1, \dots, T$:

$$\begin{aligned} y_{it} &= G(y_{it}^*), \\ y_{it}^* &= \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_{it} + \varepsilon_{it}, \\ \alpha_{it} &= \alpha_{i,t-1}\rho + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_\eta^2), \end{aligned}$$

where y_{it}^* is a continuous unobservable variable underlying y_{it} and $G(\cdot)$ is a parametric function, the specification of which depends on the nature of the observed variable. Moreover, ε_{it} and η_{it} are error terms assumed to be mutually independent. Different distributions may be assumed for the error terms ε_{it} . In LDV models for

panel data the response variable y_{it} is typically discrete so that $G(\cdot)$ is not the identity function. Here we consider the case of ordinal response variables with J categories and we define a set of thresholds $\tau_1 \leq \dots \leq \tau_{J-1}$ with $\tau_0 = -\infty$ and $\tau_J = +\infty$ such that

$$G(y_{it}^*) = j \Leftrightarrow \tau_{j-1} < y_{it}^* \leq \tau_j \quad j = 1, \dots, J.$$

Different parameterizations can be considered for ordinal observed variables. A very common one is given by the following Proportional Odds model for cumulative probability functions [14]

$$\log \left(\frac{p(y_{it} \leq j | \alpha_{it}, \mathbf{x}_{it})}{p(y_{it} > j | \alpha_{it}, \mathbf{x}_{it})} \right) = \tau_j - \alpha_{it} - \mathbf{x}_{it}' \boldsymbol{\beta} \quad j = 1, \dots, J-1, i = 1, \dots, n, t = 1, \dots, T. \quad (1)$$

The discrete latent process formulation assumes that α_i follows a first-order homogenous Markov chain with s states denoted by $u_1, \dots, u_s$. For this model we define the initial probabilities as

$$\pi_h = p(\alpha_{i1} = u_h), \quad h = 1, \dots, s$$

and the transition probabilities as

$$\pi_{h_2|h_1} = p(\alpha_{it} = u_{h_2} | \alpha_{i,t-1} = u_{h_1}), \quad h_1, h_2 = 1, \dots, s, \quad t = 2, \dots, T.$$

In LM models the initial ad transition probabilities are directly estimated from the data. The great flexibility of this class of models can lead to a better goodness of fit of the analyzed data but to a greater number of parameters to be estimated. In this case in the Markov chain they are $s^2 - 1$, since we assume that the initial probabilities are equal to those of the stationary distribution of the chain. On the other hand, LDV models assume a parametric distribution for the latent variables completely defined by two parameters, ρ and σ .

3 Model estimation

Estimation of the models illustrated above may be carried out by the maximum likelihood method, which is based on the maximization of

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{y}_i | \mathbf{x}_i), \quad (2)$$

where $\boldsymbol{\theta}$ is the vector of all model parameters which affects the manifest joint distribution of the observed variables. For LDV models, the latter can be expressed as

$$\begin{aligned} f(\mathbf{y}_i | \mathbf{x}_i) &= \int \dots \int \prod_{t=1}^T p(y_{it} | \alpha_{it}, \mathbf{x}_{it}) f_\alpha(\alpha_{it} | \alpha_{i,t-1}) d\alpha_{it} \dots d\alpha_{i1} = \\ &= \int p(y_{i1} | \alpha_{i1}, \mathbf{x}_{i1}) f_\alpha(\alpha_{i1}) \int p(y_{i2} | \alpha_{i2}, \mathbf{x}_{i2}) f_\alpha(\alpha_{i2} | \alpha_{i1}) \dots \end{aligned} \quad (3)$$

$$\dots \int p(y_{iT} | \alpha_{iT}, \mathbf{x}_{iT}) f_{\alpha}(\alpha_{iT} | \alpha_{i,T-1}) d\alpha_{iT} \dots d\alpha_{i1}.$$

The above expression is based on the conditional independence assumption, according to which $p(y_{it} | y_{i1}, \dots, y_{i,t-1}, \alpha_{i1}, \dots, \alpha_{iT}) = p(y_{it} | \alpha_{it})$ and on the first order Markov assumption on the latent variables, that is $f_{\alpha}(\alpha_{it} | y_{i1}, \dots, y_{i,t-1}, \alpha_{i1}, \dots, \alpha_{i,t-1}) = f_{\alpha}(\alpha_{it} | \alpha_{i,t-1})$. Notice that $f_{\alpha}(\alpha_{i1} | \alpha_{i0}) = f_{\alpha}(\alpha_{i1})$.

If a latent Markov model is assumed, expression (3) becomes

$$f(\mathbf{y}_i | \mathbf{x}_i) = \sum_{h_1} p(y_{i1} | u_{h_1}, \mathbf{x}_{i1}) \pi_{h_1} \sum_{h_2} p(y_{i2} | u_{h_2}, \mathbf{x}_{i2}) \pi_{h_2 | h_1} \dots \sum_{h_T} p(y_{iT} | u_{h_T}, \mathbf{x}_{iT}) \pi_{h_T | h_{T-1}}.$$

The likelihood in (2) can be maximized by using the EM algorithm [5]. The algorithm is based on the computation of the complete log-likelihood that for LDV model is given by

$$LC(\boldsymbol{\theta}) = \sum_i \left[\sum_t \log p(y_{it} | \alpha_{it}, \mathbf{x}_{it}) + \log f_{\alpha}(\alpha_{i1}) + \sum_{t>1} \log f_{\alpha}(\alpha_{it} | \alpha_{i,t-1}) \right],$$

and for LM model can be expressed as

$$LC(\boldsymbol{\theta}) = \sum_i \left[\sum_t \sum_{h_t} \log p(y_{it} | u_{h_t}, \mathbf{x}_{it}) + \sum_{h_1} \log \pi_{h_1} + \sum_{t>1} \sum_{h_t} \sum_{h_{t-1}} \log \pi_{h_t | h_{t-1}} \right].$$

The algorithm is based on two steps, the E-step and the M-step. The E-step consists of computing the expected value of $LC(\boldsymbol{\theta})$ with respect to the posterior distribution of the latent variables given the observed variables at the current values of the parameter estimates. The M-step consists of maximizing the expected value $E(LC(\boldsymbol{\theta}) | \mathbf{y})$ with respect to the vector of parameters $\boldsymbol{\theta}$, obtaining the current estimates of the model parameters. For LM models the algorithm is described in details in [3, 2].

For LDV models the E-step requires the computation of the posterior distributions $f(\alpha_{it} | \mathbf{y}_i, \mathbf{x}_i)$ $t = 1, \dots, T$ and $f(\alpha_i, \alpha_{i,t-1} | \mathbf{y}_i, \mathbf{x}_i)$ $t = 2, \dots, T$ for every individual i . They can be obtained by applying to expression (3) suitable recursions used by [5, 3] for LM models. But the integrals involved in formula (3) cannot be computed analytically and hence have to be approximated. Numerical quadratures can be used in order to solve this problem.

3.1 Gauss-Hermite quadrature

A widely used method is represented by the Gauss-Hermite (GH) quadrature that allows to evaluate numerically all the integrals of the form

$$\int e^{-z^2} f(z) dz \simeq \sum_{k=1}^q w_k f(z_k), \quad (4)$$

where z_k are the zeros of the Hermite orthogonal polynomial H_k , w_k are the correspondent weights, and q is number of quadrature points [8]. The approximation is exact if $f(z)$ is a polynomial of degree equal to $2q - 1$.

In order to apply GH to expression (3) we rewrite it as follows

$$\begin{aligned} f(\mathbf{y}_i | \mathbf{x}_i) &= \int p(y_{i1} | \alpha_{i1}, \mathbf{x}_{i1}) f_\alpha(\alpha_{i1}) \int p(y_{i2} | \alpha_{i2}, \mathbf{x}_{i2}) \frac{f_\alpha(\alpha_{i2} | \alpha_{i1})}{f_\alpha(\alpha_{i2})} f_\alpha(\alpha_{i2}) \dots \\ &\dots \int p(y_{iT} | \alpha_{iT}, \mathbf{x}_{iT}) \frac{f_\alpha(\alpha_{iT} | \alpha_{i,T-1})}{f_\alpha(\alpha_{iT})} f_\alpha(\alpha_{iT}) d\alpha_{iT} \dots d\alpha_{i1}, \end{aligned} \quad (5)$$

where $f_\alpha(\cdot)$ is the marginal distribution of α_{it} and it is a normal density distribution with mean 0 and variance $\sigma^2 = \sigma_\eta^2 / (1 - \rho^2)$. Denoting with $g(\alpha_{it} | \alpha_{i,t-1}) = f_\alpha(\alpha_{it} | \alpha_{i,t-1}) / f_\alpha(\alpha_{it})$ and integrating over the standardized $\tilde{\alpha} = \frac{\alpha}{\sigma}$, we obtain

$$\begin{aligned} f(\mathbf{y}_i | \mathbf{x}_i) &= \int p(y_{i1} | \sigma \tilde{\alpha}_{i1}, \mathbf{x}_{i1}) f(\sigma \tilde{\alpha}_{i1}) \int p(y_{i2} | \sigma \tilde{\alpha}_{i2}, \mathbf{x}_{i2}) g(\sigma \tilde{\alpha}_{i2} | \sigma \tilde{\alpha}_{i1}) f(\sigma \tilde{\alpha}_{i2}) \dots \\ &\dots \int p(y_{iT} | \sigma \tilde{\alpha}_{iT}, \mathbf{x}_{iT}) g(\sigma \tilde{\alpha}_{iT} | \sigma \tilde{\alpha}_{i,T-1}) f(\sigma \tilde{\alpha}_{iT}) d\sigma \tilde{\alpha}_{iT} \dots d\sigma \tilde{\alpha}_{i1} \simeq \\ &\simeq \sum_{k_1=1}^q p(y_{i1} | \sigma z_{k_1}^*, \mathbf{x}_{i1}) w_{k_1}^* \sum_{k_2=1}^q p(y_{i2} | \sigma z_{k_2}^*, \mathbf{x}_{i2}) g(\sigma z_{k_2}^* | \sigma z_{k_1}^*) w_{k_2}^* \dots \\ &\dots \sum_{k_T=1}^q p(y_{iT} | \sigma z_{k_T}^*, \mathbf{x}_{iT}) g(\sigma z_{k_T}^* | \sigma z_{k_{T-1}}^*) w_{k_T}^*, \end{aligned} \quad (6)$$

where $z_k^* = \sqrt{2} z_k$ and $w_k^* = (1/\sqrt{\pi}) w_k$.

3.2 Adaptive Gauss-Hermite quadrature

An improved version of the GH approximation is given by the Adaptive Gauss Hermite (AGH) quadrature rule introduced in the Bayesian context with the aim of efficiently computing posterior densities if they are approximately normal [15, 17]. It essentially consists of adjusting the GH quadrature locations with the mean and the variance of the posterior density so that the nodes are more concentrated around the peak of the integrand, and a better approximation of the function to be integrated is obtained compared to the classical GH. The application of the AGH quadrature requires to multiply and divide each integrand in (5) by the normal density $\phi(\tilde{\alpha}_{it}, \tilde{\mu}_{it}, \tilde{\sigma}_{it})$ as follows

$$\begin{aligned} f(\mathbf{y}_i | \mathbf{x}_i) &= \int \frac{p(y_{i1} | \sigma \tilde{\alpha}_{i1}, \mathbf{x}_{i1}) g(\sigma \tilde{\alpha}_{i1} | \sigma \tilde{\alpha}_{i1}) f(\sigma \tilde{\alpha}_{i1})}{\phi(\tilde{\alpha}_{i1}, \tilde{\mu}_{i1}, \tilde{\sigma}_{i1})} \phi(\tilde{\alpha}_{i1}, \tilde{\mu}_{i1}, \tilde{\sigma}_{i1}) \dots \\ &\dots \int \frac{p(y_{iT} | \sigma \tilde{\alpha}_{iT}, \mathbf{x}_{iT}) g(\sigma \tilde{\alpha}_{iT} | \sigma \tilde{\alpha}_{i,T-1}) f(\sigma \tilde{\alpha}_{iT})}{\phi(\tilde{\alpha}_{iT}, \tilde{\mu}_{iT}, \tilde{\sigma}_{iT})} \phi(\tilde{\alpha}_{iT}, \tilde{\mu}_{iT}, \tilde{\sigma}_{iT}) d\sigma \tilde{\alpha}_{iT} \dots d\sigma \tilde{\alpha}_{i1}. \end{aligned} \quad (7)$$

Since we have to approximate nested sets of univariate integrals we cannot simply compute posterior mean and variances of each random effect separately but we have to take into account the non-null posterior correlations between subsequent random effects. [16, 17] proposed to orthogonalize, center and scale the random effects by means of a sequential procedure. This procedure simplify noticeably the application of AGH. The sequential procedure consists of the following steps:

1. $w_1 = \tilde{\alpha}_1$
2. $w_t = \tilde{\alpha}_t + \psi_{t,t-1} w_{t-1}$ $\psi_{t,t-1} = -Cov(\tilde{\alpha}_t, w_{t-1})/var(w_{t-1})$ $t = 2, \dots, T$
3. $z_t = (w_t - \tilde{\mu}_t)/\tilde{\sigma}_t$ $t = 1, \dots, T$

where $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ are the mean and variance of w_t . Expression (7) is then approximated as follows

$$p(\mathbf{y}_i | \mathbf{x}_i) \simeq \sum_{k_1=1}^q p(y_{i1} | \sigma v_{i1k_1}, \mathbf{x}_{i1}) w_{i1k_1}^{**} \sum_{k_2=1}^q p(y_{i2} | \sigma v_{i2k_2}, \mathbf{x}_{i2}) g(\sigma v_{i2k_2} | \sigma v_{i1k_1}) w_{i2k_2}^{**} \dots \\ \dots \sum_{k_T=1}^q p(y_{iT} | \sigma v_{iT k_T}, \mathbf{x}_{iT}) g(\sigma v_{iT k_T} | \sigma v_{i,T-1,k_{T-1}}) w_{iT k_T}^{**}$$

where $v_{itk} = (\tilde{\sigma}_{it} z_k^* + \tilde{\mu}_{it}) - \psi_{t,t-1}(\tilde{\sigma}_{i,t-1} z_k^* + \tilde{\mu}_{i,t-1})$ and $w_{itk}^{**} = \tilde{\sigma}_{it} w_k^* \exp(z_k^{*2}/2) \exp(-v_{itk}^2/2)$. The covariances $\psi_{t,t-1}$ as well as $\tilde{\mu}_{it}$ and $\tilde{\sigma}_{it}$ are obtained from the posterior means, variances and covariances of the random effects.

3.3 EM algorithm

The steps of the EM algorithm for LDV models are defined as follows:

1. Choose initial estimates for the model parameters $\boldsymbol{\theta} = (\tau_1, \dots, \tau_{J-1}, \rho, \sigma)'$
2. E-step: compute approximated univariate and bivariate posterior distributions by means of backward and forward recursions. If AGH is used, the posterior moments are computed by means of an iterative routine.
3. M-step: obtain improved estimates for the parameters by solving the non-linear maximum likelihood equations for the parameters $\tau_1, \dots, \tau_{J-1}, \rho$ by means of a Newton Raphson iterative scheme and explicit solutions for the parameters σ .
4. Return to step 2 and continue until convergence is attained.

4 Simulation study

LDV and LM models considered in this simulation study refer to observed ordinal data. We generated data from both models. In the first case we selected $T = 8$ time points, sample size $n = 500$, values of the correlation parameter equal to $\rho = (0.50, 0.95)$ and $\sigma_\eta = 1$. We chose five categories for the ordinal responses with thresholds fixed equal to $\tau_1 = -1.65, \tau_2 = -0.5, \tau_3 = 0.5, \tau_4 = 1.65$. In the

second case we chose $T = 8$ time points, sample size $n = 500$, number of latent states $s = 3$ and transition matrix equal to

$$\boldsymbol{\Pi} = \begin{pmatrix} 0.80 & 0.15 & 0.05 \\ 0.10 & 0.80 & 0.10 \\ 0.05 & 0.15 & 0.80 \end{pmatrix}$$

We considered three categories for the ordinal responses with thresholds fixed equal to $\tau_1 = -1.65$, $\tau_2 = 1.65$. In all the designs we chose two slopes equal to $\beta_1 = 1.12$ and $\beta_2 = 2.34$. They correspond to a time-constant covariate generated from a Bernoulli distribution with parameter 0.7 and to a time-varying covariate generated from a stationary AR(1) with autocorrelation parameter equal to 0.5. We considered 100 replications for each condition of the study. As for LDV, we used $q = 51$ points with GH, $q = 51$ points with the rectangular quadrature (RQ) and $q = 21$ with AGH. The choice of the number of quadrature points derived from the findings obtained in previous studies where the same approximation methods were used [10, 4, 7]. As for LM, for each condition of the study we considered number of latent states varying $s = 1, \dots, 5$ and we used the BIC information criterion to choose it. The comparison between LDV with different numerical approximations and LM is assessed in terms of accuracy of the estimates and goodness of fit of the models.

In Table 1 the results for LDV-generated data for $\rho = 0.5$ are shown. As for LM, only the results for $s = 2$ are reported, since BIC preferred it in all the samples. LDV with GH seems to perform better than the other methods in terms of both bias and standard deviations (s.d.) for ρ , σ_η and β 's estimates. All the other quadrature approximations as well as LM give similar results. In terms of goodness of fit BIC preferred 24% of time GH, 33% of time AGH, 39% of time RQ, and only 4% of times LM.

Table 2 shows the results for LDV-generated data in the case of $\rho = 0.95$. Also in this case we considered LM with $s = 2$, since BIC preferred it in all the samples. GH does not produce convergent solutions in all the generated samples because of instability problems. This behaviour of the GH approximation was found also by [7] for the same model estimated using a direct maximum likelihood estimation through a non-linear filter algorithm. This is due to the fact that for $\rho = 0.95$, the integrand has a very sharp peak and GH does not properly capture it. Also in this case all the other methods perform similarly but we have to notice that AGH uses only 21 quadrature points. In terms of goodness of fit BIC preferred 58% LDV with AGH and 42% LDV with RQ.

In Table 3 the results for LM-generated data are shown. For LDV we reported only the results for GH and AGH and for LM we reported the results for $s = 3$ that was chosen by BIC in 91% of the generated samples. We can observe that GH shows a better behaviour in terms of bias for threshold and slope estimates. In terms of s.d. LDV with the two approximation methods and LM perform similarly. In this case BIC chose GH in only 4% of cases, AGH in 60% of cases and LM in 36% of cases.

Table 1 Estimated bias and s.d. for the LDV model parameters under GH , AGH , RQ and for the LM model parameters, LDV-generated data, $T = 8$, $\rho = 0.50$, $n = 500$, 100 replications

True value	GH_{51}		AGH_{21}		RQ_{51}		LM	
	Bias	s.d.	Bias	s.d.	Bias	s.d.	Bias	s.d.
$\tau_1 = -1.65$	0.032	0.092	0.034	0.107	0.046	0.168	-0.050	0.096
$\tau_2 = -0.50$	0.022	0.082	0.025	0.083	0.012	0.099	-0.013	0.088
$\tau_3 = 0.50$	0.010	0.085	0.004	0.089	-0.011	0.093	0.022	0.085
$\tau_4 = 1.65$	-0.004	0.086	-0.009	0.117	-0.038	0.161	0.064	0.105
$\beta_1 = 1.12$	-0.003	0.092	-0.012	0.108	-0.019	0.140	-0.046	0.101
$\beta_2 = 2.34$	-0.021	0.055	-0.033	0.128	-0.061	0.199	-0.083	0.092
$\sigma_\eta = 1$	-0.046	0.047	0.081	0.192	0.066	0.282	-	-
$\rho = 0.50$	0.034	0.061	0.064	0.091	-0.007	0.120	-	-
Av loglik	-4093.05		-4096.94		-4.095.07		-4.095.75	

Table 2 Estimated bias and s.d. for the LDV model parameters under GH , AGH , RQ and for the LM model parameters, LDV-generated data, $T = 8$, $\rho = 0.95$, $n = 500$, 100 replications

True value	GH_{51}		AGH_{21}		RQ_{51}		LM	
	Bias	s.d.	Bias	s.d.	Bias	s.d.	Bias	s.d.
$\tau_1 = -1.65$	-	-	-0.025	0.116	-0.033	0.111	-0.037	0.117
$\tau_2 = -0.50$	-	-	-0.004	0.107	-0.029	0.118	-0.009	0.122
$\tau_3 = 0.50$	-	-	0.016	0.099	-0.016	0.111	0.017	0.112
$\tau_4 = 1.65$	-	-	0.042	0.107	-0.014	0.108	0.048	0.112
$\beta_1 = 1.12$	-	-	0.039	0.113	0.010	0.128	-0.040	0.137
$\beta_2 = 2.34$	-	-	0.046	0.072	0.016	0.058	-0.066	0.055
$\sigma_\eta = 1$	-	-	0.098	0.118	0.010	0.083	-	-
$\rho = 0.95$	-	-	-0.012	0.048	0.008	0.034	-	-
Av loglik	-	-	-4011.70		-4032.85		-4041.69	

Table 3 Estimated bias and s.d. for the LDV model parameters under GH , AGH and for the LM model parameters, LM-generated data, $T = 8$, $n = 500$, 100 replications

True value	GH_{51}		AGH_{21}		LM	
	Bias	s.d.	Bias	s.d.	Bias	s.d.
$\tau_1 = -1.65$	0.047	0.135	-0.077	0.142	0.093	0.120
$\tau_2 = 1.65$	-0.007	0.137	-0.037	0.135	0.038	0.141
$\beta_1 = 1.12$	-0.007	0.145	-0.045	0.120	-0.029	0.122
$\beta_2 = 2.34$	-0.022	0.109	-0.065	0.141	-0.051	0.079
σ_η	-	0.139	-	0.187	-	-
ρ	-	0.036	-	0.082	-	-
Av loglik	-2853.01		-2849.13		-2853.37	

5 Remarks

The comparison between LDV and LM highlighted that in general the accuracy of the estimates is quite similar in all the conditions of the study considered. In more detail, for some parameters LDV with GH seems to give better results than both the

other quadrature methods and LM when $\rho = 0.5$ but presents instability problems when $\rho = 0.95$. AGH, RQ and LM perform similarly but AGH uses less quadrature points than RQ. In terms of goodness of fit in most cases the model with continuous latent variables is preferred to the model with discrete latent variables.

References

1. Bartholomew, D. J.: The sensitivity of latent trait analysis to choice of prior distribution, *Brit J Math Stat Psy*, **41**, 101–107 (1988)
2. Bartolucci, F., Farcomeni, A., Pennoni, F.: *Latent markov models for longitudinal data*, Chapman & Hall/CRC Press (2013)
3. Bartolucci, F., Farcomeni, A.: A multivariate extension of the dynamic logit model for longitudinal data based on latent Markov heterogeneity structure. *J Am Stat Assoc*, **104**, 816–831 (2009)
4. Bartolucci, F., Bacci, S., Pennoni, F.: Longitudinal analysis of the self-reported health status by mixture latent autoregressive models, *J Roy Stat Soc C-App*, **63**, 267–288 (2014)
5. Baum, L., Petrie, T., Soules, G., Weiss, N. (1970): A maximization technique occurring in the statistical analysis of propbabilistic functions of markov chains. *Ann Math Stat*, **41**, 164–171 (1970)
6. Bock, R.D., Aitkin, M.: Marginal maximum likelihood estimation of item parameters application of an EM algorithm, *Psychometrika*, **46**, 443–459 (1981)
7. Cagnone, S., Bartolucci F.: Adaptive quadrature for maximum likelihood estimation of dynamic latent variable models, submitted (2014)
8. Davis, P. J., Rabinowitz P.: *Methods of numerical integration*. Academic Press, (1975)
9. Frees, E.W.: *Longitudinal and Panel Data: Analysis and Application in Social Sciences*. Cambridge University Press, Cambridge UK (2004)
10. Heiss, F.: Sequential numerical Integration in nonlinear state space models for microeconomic panel. *J Appl Econom*, **23**, 373–389 (2008)
11. Irincheeva I., Cantoni E., Genton M.G.: Generalized linear latent variable models with flexible distribution of latent variables. *Scad J Statist*, **39**, pp 664–680 (2012)
12. Lindsay, B.G., Clogg, C. C., Grego, J.: Semiparametric estimation in the Rash model and related exponential response models, including a simple latent class model for item analysis, *J Am Stat Assoc*, **86**, 96–107 (1991)
13. Maddala, G.S: *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press (1983)
14. McCullagh, P: Regression models for ordinal data (with discussion), *J Am Stat Assoc*, **42**, 109–142 (1980)
15. Naylor, J. C., Smith, A.F.M.: Applications of a method for the efficient computation of posterior distributions, *Appl. Statist*, **51**, 214–225 (1982)
16. Naylor, J. C., Smith, A.F.M.: Economeitrc illustration of novel numerical integration strategies for Bayesian inference , *J Econometrics*, **38**, 103–125 (1988)
17. Rabe-Hesketh, S., Skrondal A., Pickels A.: Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects, *J Econometrics*, **128**, 301–323 (2005)
18. Tzamourani, P., Knott, M.: Fully semiparametric estimation of the two-parameter latent trait model for binary data. In: Marcoulides G.A., Moustaki I. (Eds) *Latent variable and latent structure models*, pp. 63–84, Mahwah, NJ:Erlbaum (2002)