Limited Dependent Variable Panel Data Models: A Bayesian Approach

Giuseppe Bruno [1]

June 2005

Preliminary Draft

[1]Bank of Italy Research Department.

e-mail: giuseppe.bruno@bancaditalia.it

1 Introduction

Advances in computing power have opened the way for the use of intensive computational techniques to solve and estimate nonlinear models, specifically those arising from nonlinear panel data such as Probit and Tobit. These models are very valuable tools in the applied economic research. In these models, sometimes the dependent variable is dichotomous, reflecting the decision whether to purchase or not a durable good, participate or not into the labour force. In other circumstances the decisions are related to statements such as "whether or not, and if so how much". This case arises in many investment decisions, like the construction of a new plant, and money market decisions, like the borrowing requirements of individual banks from the central bank. For these models, allowing a flexible specification for the correlation induced by firm/individual heterogeneity leads to models involving T-variate multiple integration whose numerical approximation can sometimes be very poor. In these cases, when the value of T (number of periods for each individual) is greater than 4 or 5, maximum-likelihood estimation can be cumbersome if not analytically intractable. Different solutions are offered based variously on integral approximation through simulation, some form of Generalized Method of Moments (GMM), or Markov Chain Monte Carlo (MCMC) methods. This paper compares the outcomes of those methods available in standard econometric packages, providing illustrations that compare prepackaged algorithms and a MCMC Gibbs sampler for nonlinear panel data. Using Chib (1992) and Chib and Carlin (1999), I derive a sampler for Probit/Tobit panel models and provide easy-to-use software for implementing the Gibbs sampler with data augmentation in panel data with discrete/limited dependent variable. We show that, when dealing with a large dataset and in presence of serial correlation, MCMC methods can fill the gap present in the procedures provided by standard econometric packages.

The structure of the paper is as follows. Section 2 presents the panel Tobit model under investigation. In Section 3 it is presented the Gibbs sampling methodology along with the data augmentation process. Section 4 shows an empirical application from the U.S. the labour market. Section 5 and 6 present two different Monte Carlo exercises with independent and serially correlated residuals. Finally section 7 draws some conclusions and lines of evolution.

2 The Panel Tobit Model

Panel datasets provide a very rich source of information for empirical economists, providing the scope to control for individual heterogeneity. While there is a large literature on linear panel data models, less is known about limited dependent variable models. This is especially true for computational comparisons among different methods.

Extending the work of Chib (1992) and Contoyannis et al. (2002) this paper is concerned with the Bayes estimation in the panel Tobit model. The purpose of the paper is three-fold: first to develop an easy-to-use Bayesian estimation approach, second to compare the efficacy of this method relative to others available in some econometric packages and finally to check its numerical features in presence serial correlation in the residuals.

I am concerned with a standard Panel data Tobit model:

$$y_{it}^* = \beta' x_{it} + u_{it}, \qquad i = 1, 2, \ldots, N, \qquad t = 1, 2, \ldots, T_i, \tag{2.1}$$

$$u_{it} = \nu_i + \epsilon_{it}, \qquad (\nu_i \sim NID(0, \sigma_\nu^2)), \qquad (\epsilon_{it} \sim NID(0, \sigma_\epsilon^2)), \tag{2.2}$$

where the observed variables are

$$y_{it} = \begin{cases} y_{it}^* & if \quad y_{it}^* > 0 \\ 0 & otherwise \end{cases} . \tag{2.3}$$

In general the common error term u_{it} in equation (2.2) can be freely correlated over time. Here I consider the error-components model that splits the error u_{it} into a time-invariant individual random effect (RE), ν_i and a time-varying idiosyncratic random error, ϵ_{it}.

In this case, assuming independence between the ν's and the ϵ's, letting $d_{it} = 1$ for uncensored observations and $d_{it} = 0$ for censored observations, the likelihood contribution for each individual, marginalized with respect to the random effect ν_i, is

$$l_{it} = \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_\epsilon} \phi \left(\frac{y_{it} - \beta' x_{it} - \nu_i}{\sigma_\epsilon} \right) \right]^{d_{it}} \cdot \left[\Phi \left(\frac{-\beta' x_{it} - \nu_i}{\sigma_\epsilon} \right) \right]^{(1-d_{it})} f(\nu_i, \sigma_i) d\nu_i \quad , \tag{2.4}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the probability density function (pdf) and the cumulative distribution function (cdf) of the standard normal distribution, $f(\nu_i, \sigma_i)$ is the normal pdf with mean ν_i and standard deviation σ_i.

In general, for T_i observations belonging to individual i, one has the following likelihood contribution

$$L_i = \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{t=T_i} \left[\frac{1}{\sigma_\epsilon} \phi \left(\frac{y_{it} - \beta' x_{it} - \nu_i}{\sigma_\epsilon} \right) \right]^{d_{it}} \left[\Phi \left(\frac{-\beta' x_{it} - \nu_i}{\sigma_\epsilon} \right) \right]^{(1-d_{it})} \right\} f(\nu_i, \sigma_i) d\nu_i. \tag{2.5}$$

The likelihood function for the whole sample is the product of the contribution L_i over the N individuals, and the log-likelihood is

$$\mathcal{L} = \sum_{i=1}^{N} ln(L_i), \tag{2.6}$$

from this expression one sees that the log likelihood in equation (2.6) does not collapse in a sum, as it would in the case of a time-series or a simple cross-sectional Tobit model, because the likelihood function for individual i is an integral of a product instead of just a product and the log operator cannot be carried through the integral sign.

The situation gets even more complex in presence of serial correlation in the disturbance ϵ_{it} for each individual. In this case, the lack of independence among the observations prevents the possibility of factoring out the likelihood contribution of the T_i periods for individual i and we end up with a T-dimensional integral that makes classical estimation methods numerically infeasible when the number of time periods is more than three or four.

3 Gibbs sampling with data augmentation

The Gibbs sampler is a Monte Carlo Markov Chain method for sampling from probability densities that are analytically intractable (see for example (see for example Chib (2001) or Casella and George (1992) for an introductory presentation).

This method, also called *alternating conditional sampling*, has made possible the Bayesian approach to the estimation of nonlinear panel-data models providing accurate finite-sample estimates.

Gibbs sampling is based on a preliminary splitting of the parameter vector into s groups, $\theta = (\theta_1, \ldots, \theta_s)$. In our panel Tobit model the parameter vector is already subdivided according to $\theta = (\beta, \sigma_\epsilon, \sigma_\nu)$. Then one can obtain draws from the posterior distribution of the parameter vector conditional on the data by means of an iterative sampling scheme. Each one of the draw is built up by drawing each group of parameters from its own probability distribution conditional on the data and the rest of the other parameters. In our case we have the following procedure:

1) pick arbitrary initial values for $\Theta^0 = \beta^0, \sigma_\epsilon^0, \sigma_\nu^0$ (one possibility could be the GLS random effects even if biased and inconsistent),

2) draw β^k from the distribution $\pi(\beta|y, \mathbf{x}, \sigma_\epsilon^{\mathbf{k-1}}, \sigma_\nu^{\mathbf{k-1}})$,

3) draw σ_ϵ^k from the distribution $\pi(\sigma_\epsilon|y, \mathbf{x}, \beta^{\mathbf{k}}, \sigma_\nu^{\mathbf{k-1}})$,

4) draw σ_ν^k from the distribution $\pi(\sigma_\nu|y, \mathbf{x}, \beta^{\mathbf{k}}, \sigma_\epsilon^{\mathbf{k}})$.

Under general assumptions, after a certain number of iterations, this process produces samples from the desired posterior distribution. Then point estimates and confidence intervals are computed as averages from the generated sample.

To implement the sampler one needs to specify the different conditional pdf's. For each one of the three groups of parameters I use non-informative conjugate priors for simplifying all the computations. That is, I adopt the following distributions:

$$\beta \sim \mathcal{N}\left(\beta_0, \Omega_0\right), \tag{3.7}$$

$$\sigma_\nu^2 \sim \mathcal{IG}(\eta_0, \gamma_0), \tag{3.8}$$

$$\sigma_\epsilon^2 \sim \mathcal{IG}(\nu_0, \delta_0), \tag{3.9}$$

here all the variables indexed by 0 are the hyperparametes of our distributions, $\mathcal{N}(\mu, \Sigma)$ is the multivariate Normal distribution with mean vector μ and variance-covariance matrix Σ, and $\mathcal{IG}(\nu, \delta)$ is the inverse Gamma distribution with shape ν and scale δ.

A peculiar feature of nonlinear panel-data models such as the Tobit, is the presence of unobservable latent data that would make the previous sampling loop very complex. Thus, following the suggestion of Tanner and Wong (1987), adopted by Chib (1992) in a cross-sectional context, I enrich the Gibbs sampler by means of the data-augmentation strategy. Given the assumptions underlying the model, the distributions of the latent variables are truncated normals. So I can augment the dataset with an estimate for the censored variables. Using the augmented dataset renders this problem a classic linear panel-data model.

The Gibbs sampler as previously described has been modified by adding at the beginning a step for sampling the censored variables. For example, in case (2.3), I have to simulate a random sample from a truncated normal distribution with support $(-\infty, 0)$ and pdf given by

$$y_{it}^* \sim \frac{\mathcal{N}(x_{it}'\beta + \nu_i, \sigma_\nu^2 + \sigma_\epsilon^2)}{1 - \Phi\left(\frac{x_{it}'\beta}{\sqrt{\sigma_\nu^2 + \sigma_\epsilon^2}}\right)}. \tag{3.10}$$

To sample from this truncated normal, I use the one-for-one draw technique described in Hajivassiliou and McFadden (1990) that is much more cost effective than the acceptance-rejection method.

Here, then, is a complete description of the algorithm for estimating the random-effects Tobit model:

1) run a GLS estimation with the original truncated data to fix the initial values for $\beta^0,\ \sigma_\epsilon^0,\ \sigma_\nu^0$;

2) sample the censored variables from the pdf (3.10) to build the augmented dataset;

3) run a GLS estimation on the panel with the augmented dataset for computing new mean values for $\beta,\ \sigma_\epsilon,\ \sigma_\nu$;

4) draw $\beta^{\mathbf{k}}$ from the distribution $\pi(\beta|\mathbf{y}, \mathbf{x}, \sigma_\epsilon^{\mathbf{k-1}}, \sigma_\nu^{\mathbf{k-1}}) = \mathcal{N}(\beta, \mathbf{\Omega})$;

5) estimate the individual effects using the residuals from the previous step ;

6) draw σ_ν^k from the distribution $\pi(\sigma_\nu|y, \mathbf{x}, \beta^{\mathbf{k-1}}, \sigma_\epsilon^{\mathbf{k-1}}) = \mathcal{IG}\left(\nu_{\mathbf{0}}, \delta_{\mathbf{0}}\right)$;

7) draw σ_ϵ^k from the distribution $\pi(\sigma_\epsilon|y, \mathbf{x}, \beta^{\mathbf{k-1}}, \sigma_\nu^{\mathbf{k-1}}) = \mathcal{IG}\left(\eta_{\mathbf{0}}, \gamma_{\mathbf{0}}\right)$.

The implementation of this sampler is written in the language and computing framework provided by the Modeleasy+ software.

4 The empirical application

To illustrate the behaviour of this sampler, I use a dataset available from the web site of the STATA package [1]. This is an unbalanced panel dataset taken from the National Longitudinal Survey on economic and demographic variables. It includes 4140 women of age comprised between 14 and 46 years leaving 19151 observations once those with at least one missing values have been removed. Each individual is observed on a time interval spanning from 1 to 12 time periods starting from 1968 (the average is 4.6 periods).

The application studies the determinants of wages. Using the example provided in the STATA manual, I fit a random-effects Tobit model on the log of wages against a set comprising the following explanatory variables:

1) **union**, dummy variable equal to 1 if the individual belongs to a workers' union;
2) **age**, the individual's age ;
3) **grade**, the years of schooling completed;
4) **not_smsa**, dummy variable equal to 1 if the individual does not live in a standard metropolitan statistical area (smsa);
5) **south**, dummy variable equal to 1 if the individual lives in the south;
6) **southXt**, interaction-term variable indicating how long the individual has lived in the south;
7) **occ_cod**, a categorical variable indicating the occupational code of the individual (larger numbers mean a lower rank);

[1] The dataset nlswork.dta is downloadable from the URL http://www.stata-press.com/data/r8

This model has been estimated using the random-effects methods in LIMDEP and STATA. Even if both package use the Maximum Likelihood procedure with the Hermite quadrature formulae, I got results that are numerically quite different. The Gibbs sampler was run for 500, 1000, 1500 iterations, the first 10% iterations are discarded as a burn-in phase. I chose to present the outcome of the 1000 iteration sampler. The results are summarized in the following table:

Table 1: Parameter Estimates for the nls dataset (19151 observations and 4140 individuals)

		LIMDEP	STATA	Gibbs Sampler
const	β_0	.75297 (.02649)	.56572 (.03308)	.63202(.02953) [-.28331]
union	β_1	.15946 (.00533)	.15449 (.00698)	.14632 (.00684) [.36156]
age	β_2	.00785 (.00038)	.00871(.00054)	.00788 (.00005) [.82761]
grade	β_3	.06653 (.00179)	.07803 (.00216)	.07332 (.00191) [.08662]
nots_msa	β_4	-.13871 (.00609)	-.12669 (.00898)	-.12408 (.00843) [-.43407]
south	β_5	-.12874 (.00887)	-.11686 (.01224)	-.11764 (.01133) [-.32051
southxt	β_6	.00263 (.00060)	.00309 (.00084)	.00358 (.00008) [.74192]
occ_code	β_7	-.01952 (.00078)	-.01829 (.00111)	-.01749 (.00106) [-1.4917]
	$\boldsymbol{\sigma_\epsilon}$	.2542 (.0010)	.2483 (.0018)	.2378 (.00010)
	$\boldsymbol{\sigma_\nu}$	.3341 (.0039)	.2911 (.0048)	.2582 (.00219)

Notes: Standard errors in parentheses, Convegence Diagnostic in brackets

The similarity between the estimates provided by the Gibbs Sampler and those of LIMDEP and STATA is remarkable. Here it is important to highlight the simplicity of the sampler with respect to the classical algorithms, including quadrature approximations and maximization.

5 Monte Carlo Simulation

Starting from the previous results, I wish to establish a kind of a benchmark. I generate a dataset according to the following process (see Harris et al (2000)) :

$$y^*_{it} = \beta_0 + \beta_1 x1_{it} + \beta_2 x2_i + \nu_i + \epsilon_{it} \quad \nu_i \sim NID(0, 1.0), \qquad \epsilon_{it} \sim NID(0, 1.0), \tag{5.11}$$

where the mapping from the latent variable to the observed variable is

$$y_{it} = \begin{cases} y^*_{it} & if \quad y^*_{it} > 0 \\ 0 & otherwise \end{cases}. \tag{5.12}$$

The values for the three β's are 0.5, -1, and 1 respectively. These values result in a roughly 50% split between censored and non-censored variables.

Values of x_{it} follow an auto-regressive process given by

$$x1_{it} = 0.1 \cdot trend + 0.5 \cdot x1_{i,t-1} + u_{it}, \tag{5.13}$$

where $u_{it} \sim U(-.5, .5)$.

The time-invariant variable x_{2i} is generated according to

$$x_{2i} = \begin{cases} 0 & if \quad 0 \leq x^*_{2i} < 0.5 \\ 1 & if \quad 0.5 \leq x^*_{2i} \leq 1 \end{cases}, \tag{5.14}$$

where the latent variable is generated according to $x^*_{2i} \sim U(0, 1)$.

The individual specific effects are generated according to $\nu_i \sim NID(0, \sigma^2_\nu)$, where σ_ν is specified as 1 and 2 to provide two values of correlation over time. For the idiosyncratic random error term I choose $\epsilon_{it} \sim NID(0, 1)$. I carry out the simulation using a panel of 100 and 200 individuals with 3, 6, and 12 time periods.

From the following tables one can observe that, for small T the Maximum Likelihood method outperforms the Gibbs sampler. When T is increased, the Gibbs sampler produces estimates with standard errors smaller than those achieved by Maximum Likelihood. LIMDEP is the package that gives the smallest bias in this experiment. Further investigation should be done to pin down the reason of that.

These results seem to indicate that the MCMC method may be preferred when T is large.

Table 2: Monte Carlo Parameter Estimates for T= 3

	true parameters	*N = 100*		*N = 200*	
	σ_ν^2	1	2	1	2
LIMDEP	$\beta_0 = .5$	.49251	.50126	.49963	.50492
		(.17447)	(.40942)	(.13044)	(.31964)
	$\beta_1 = -1$	-1.00833	-1.00189	-1.0005	-.99568
		(.15673)	(.21836)	(.12094)	(.16022)
	$\beta_2 = 1$	1.00614	1.02014	.99512	.99185
		(.23524)	(.57546)	(.17795)	(.43363)
STATA	$\beta_0 = .5$	.50665	.58218	.48599	.53866
		(.18361)	(.43688)	(.12084)	(.30257)
	$\beta_1 = -1$	-.99535	-.99622	-1.008711	-1.00074
		(.17466)	(.22397)	(.12614)	(.15656)
	$\beta_2 = 1$	.98447	.90551	.99246	.972846
		(.23816)	(.54675)	(.16304)	(.41163)
Gibbs Sampler	$\beta_0 = .5$	.34488	.25174	.40712	.27154
		(.19324)	(.3223)	(.12252)	(.23569)
	$\beta_1 = -1$	-1.4212	-1.5717	-1.0413	-1.1164
		(.17854)	(.26103)	(.10936)	(.1666)
	$\beta_2 = 1$	1.4782	1.3316	1.2843	1.3979
		(.25057)	(.42141)	(.1541)	(.29773)

Notes: Average parameter estimates over 1000 Monte Carlo replications with Mean Squared errors in parentheses

Table 3: Monte Carlo Parameter Estimates for T= 6

	true parameters	*N = 100*		*N = 200*	
	σ_ν^2	1	2	1	2
LIMDEP	$\beta_0 = .5$	.50077	.60865	.50191	.56321
		(.15350)	(.45406)	(.11994)	(.34987)
	$\beta_1 = -1$	-1.00293	-1.0009	-1.00338	-.99789
		(.10386)	(.12187)	(.07276)	(.08684)
	$\beta_2 = 1$	1.00381	.95696	1.0005	.98633
		(.23508)	(.61510)	(.16837)	(.49703)
STATA	$\beta_0 = .5$	.48925	.73507	.50136	.71314
		(.17993)	(.51041)	(.12651)	(.43339)
	$\beta_1 = -1$	-.99317	-.99398	-.98986	-.99998
		(.10384)	(.12044)	(.07018)	(.09130)
	$\beta_2 = 1$	1.0046	.76978	.98526	.82756
		(.23267)	(.65634)	(.16087)	(.53801)
Gibbs Sampler	$\beta_0 = .5$	.44098	.22519	.46997	.31492
		(.16824)	(.3081)	(.11271)	(.21301)
	$\beta_1 = -1$	-1.0914	-1.0801	-.94969	-.94218
		(.10101)	(.14672)	(.07052)	(.09632)
	$\beta_2 = 1$	1.2695	1.1548	1.1752	1.2777
		(.22202)	(.39303)	(.14297)	(.28197)

Notes: Average parameter estimates over 1000 Monte Carlo replications with Mean Squared errors in parentheses

Table 4: Monte Carlo Parameter Estimates for T= 12

	true parameters	*N = 100*		*N = 200*	
	σ_ν^2	1	2	1	2
LIMDEP	$\beta_0 = .5$	.49836	.51373	.50552	.55766
		(.18473)	(.42772)	(.13403)	(.33558)
	$\beta_1 = -1$	-1.00137	-1.0001	-1.00004	-.99499
		(.05735)	(.06967)	(.0403)	(.04589)
	$\beta_2 = 1$	.99662	.88625	.98522	.86962
		(.26741)	(.55242)	(.20670)	(.47691)
STATA	$\beta_0 = .5$	.51550	.84474	.50769	.74664
		(.21648)	(.49741)	(.20163)	(.47334)
	$\beta_1 = -1$	-.99929	-.99316	-.99649	-.99346
		(.051797)	(.06029)	(.04159)	(.04356)
	$\beta_2 = 1$	.97826	.65932	1.02023	.78933
		(.27569)	(.63651)	(.22279)	(.58367)
Gibbs Sampler	$\beta_0 = .5$	.50481	.24922	.51361	.37178
		(.1593)	(.31227)	(.10384)	(.20215)
	$\beta_1 = -1$	-1.1088	-1.0583	-.99628	-.97195
		(.06301)	(.092724)	(.04319)	(.05606)
	$\beta_2 = 1$	1.1271	1.0106	1.0897	1.1451
		(.1995)	(.39059)	(.13083)	(.26029)

Notes: Average parameter estimates over 1000 Monte Carlo replications with Mean Squared errors in parentheses

6 Monte Carlo Simulation with serial correlation

I extended the framework of the previous Monte Carlo set-up by introducing an autoregressive term in the error structure of the disturbance.

This time in the equation (2.2) the error term will be characterized by the following structure:

$$u_{it} = \nu_i + \epsilon_{it}, \qquad (\nu_i \sim NID(0, \sigma_\nu^2), \quad \epsilon_{it} = \rho\epsilon_{i,t-1} + \delta_{it}, \quad \delta_{it} \sim NID(0, \sigma_\delta^2)), \tag{6.15}$$

With this error structure the variance-covariance matrix of the disturbances for will imply the need of a T-variate integral for evaluating the likelihood of the individual i, thus rendering infeasible classical estimation methods.

Therefore, I have modified my original Gibbs sampler for taking into account the presence of autocorrelation in the disturbance term. This problem has been considered in the framework of panel probit model in Contoyannis et al. 2002). In this case it is not straighforward to draw sample from the distribution of the autocorrelation coefficients ρ conditional to the other parameters of the model.

Sampling for ρ has been carried out by inserting a Metropolis-Hastings step within the Gibbs sampler (see Chib and Greenberg (1995), Chib (1993)) . In the panel tobit model with serial correlation the censored observations will be transformed according to the estimated variance-covariance matrix. After the introduction of this feature this sampler has been used for a Monte Carlo experiment by introducing a serial correlatin of order one in the same model described in 5.12 in which the idiosyncratic shock is described by equation (6.15). Four examples have been considered by using two different values of first order autocorrelation $\rho = .3$ and $\rho = .5$ and two different level of censoring 20% and 50%.

For this Monte Carlo experiment, because of the lack of a benchmark provided by the numerical results of a commercial package, we need to find some accuracy measures and convergence checks. The Gibbs sampler, and in general, all the MCMC methods produce sequences that are neither independent nor identically distributed. Therefore particular care has to be exerted in deriving asymptotic results. In this work we have adopted the approach suggested in Geweke (1992) where the *Numerical Standard Error* (henceforth *NSE*) and the *Convergence Diagnostic* (henceforth *CD*) have been put forward as a computationally efficient mean for dealing with numerical accuracy and convergence checks in the MCMC framework.

For our purposes it is enough to say that the *NSE* is given by the square root of the spectral density of the sequence of draws, while assuming θ is a sequence of draws for a given parameter $CD = (\bar{\theta}_a - \bar{\theta}_b)/\sqrt{[NSE_a^2 + NSE_b^2]}$ where $\bar{\theta}_a$ and $\bar{\theta}_b$ are, respectively, the average of the first 10% and the last 50% of the whole sequence of draws, the denominator of *CD* is given by the total *NSE* of the two subsequences. The *CD* statistics is asymptotically distributed as a standard normal distribution and therefore absolute values greater than 3 are a clear symptom of lack of convergence.

The preliminary results are shown in the following tables.

Table 5: Monte Carlo Parameter Estimates for a censoring level = 20%

	true parameters	*N = 100*	*N = 200*
Gibbs Sampler	$\beta_0 = .5$	.48963	.49961
		(.06802)	(.04673)
	NSE	(.00214)	(.00147)
	CD	(.07846)	(.66355)
	$\beta_1 = -1$	-.96141	-.97946
		(.04061)	(.02899)
	NSE	(.00128)	(.00085)
	CD	(.57035)	(-2.2845)
	$\beta_2 = 1$	.71648	.98482
		(.08107)	(.05734)
	NSE	(.00257)	(.00178)
	CD	(-.04088)	(.80533)
	$\rho = .2$	.1043	.10816
		(.03429)	(.02548)
	NSE	(.00110)	(.00069)

Table 6: Monte Carlo Parameter Estimates for a censoring level = 50%

	true parameters	*N = 100*	*N = 200*
Gibbs Sampler	$\beta_0 = .5$	.51462	.58084
		(.07841)	(.05645)
	NSE	(.00246)	(.00171)
	CD	(-2.0979)	(-.24265)
	$\beta_1 = -1$	-.91923	-.93922
		(.05012)	(.03516)
	NSE	(.00151)	(.00107)
	CD	(.60973)	(-.19055)
	$\beta_2 = 1$	.63481	.87037
		(.09252)	(.06445)
	NSE	(.00293)	(.00208)
	CD	(.93937)	(-.26784)
	$\rho = .2$	.07142	.08461
		(.03529)	(.016339)
	NSE	(.00093)	(.00075)

Notes: Averages over 1000 Monte Carlo replications with Mean Squared errors in parentheses, NSE is the Numerical Standard Error, CD is the Convergence Diagnostic

Table 7: Monte Carlo Parameter Estimates for a censoring level = 20%

	true parameters	*N = 100*	*N = 200*
	$\beta_0 = .5$	.47997	.49182
		(.07818)	(.05622)
	NSE	(.00256)	(.00179)
	CD	(1.0794)	(.11975)
Gibbs Sampler	$\beta_1 = -1$	-.97775	-.99698
		(.04302)	(.03201)
	NSE	(.00137)	(.00102)
	CD	(.15844)	(-.91087)
	$\beta_2 = 1$	.71063	1.0114
		(.09228)	(.06963)
	NSE	(.00293)	(.00209)
	CD	(-.98468)	(.21645)
	$\rho = .5$	.33332	.33404
		(.03094)	(.014627)
	NSE	(.00106)	(.00066)

Table 8: Monte Carlo Parameter Estimates for a censoring level = 50%

	true parameters	*N = 100*	*N = 200*
	$\beta_0 = .5$	.47526	.54075
		(.08448)	(.05894)
	NSE	(.00264)	(.00186)
	CD	(-1.9756)	(-1.427)
Gibbs Sampler	$\beta_1 = -1$	-.92801	-.95228
		(.05616)	(.03629)
	NSE	(.00167)	(.00114)
	CD	(.53347)	(-1.2541)
	$\beta_2 = 1$	.66679	.92862
		(.09747)	(.06867)
	NSE	(.00324)	(.00216)
	CD	(1.7469)	(1.1878)
	$\rho = .5$	.22888	.18282
		(.03349)	(.00673)
	NSE	(.00099)	(.00023)

Notes: Averages over 1000 Monte Carlo replications with Mean Squared errors in parentheses, NSE is the Numerical Standard Error, CD is the Convergence diagnostic

By looking at these tables we see that the coefficient estimates are very close to the correct values while the autocorrelation coefficient shows a downward bias. The bias is increasing with the censoring level. The numerical standard errors (NSE) seem to indicate a good accuracy of the estimates. The Convergence Diagnostic statistics (CD) is always less than 3 in absolute value for all the parameters. These CD values indicate that convergence of the MCMC algorithm has been achieved. It seems that further modifications on the Metropolis-Hastings sampling procedure for the autocorrelation coefficient could provide some improvements.

7 Concluding remarks and further research

In the paper we have compared two methods for the estimation of nonlinear panel data models. The first method is the Classical Maximum Likelihood (ML) with quadrature for the computation of the likelihood function. Both the LIMDEP and STATA canned procedures have been used. Moreover a Bayesian approach based on the Gibbs sampling has been developed in the Modeleasy+ computing environment. We have taken advantage of the data augmentation technique proposed in Tanner and Wong (1987) for simplifying the analytics involved in the computation of the conditional posterior pdf's. The three procedures have been applied to a segment of the national longitudinal survey on labour statistics. Although the parameters estimates have always the right signs there are remarkable numerical difference among to two procedures implementing the ML with the same quadrature formulae. Estimates from the Gibbs sampler are close to the range defined by the two ML estimates. The Gibbs sampler is much easier than ML from the computational standpoint but so far computing time is still something that needs some improvements.

Finally to validate the methods some Monte Carlo experiments were presented. For a given level of T, increasing N has produced an increased precision.

It should be remarked that when there is a more complex correlation structure in the disturbance, such as a first order autocorrelation, quadrature formula become cumbersome and inaccurate making the MCMC methods simple practical solutions to the estimation problem.

In presence of serial correlation in the residuals the proposed method produced accurate estimates for the covariate coefficients while a certain bias seems affecting the estimation of the autocorrelation coefficient. This bias increases with the level of censoring and the value of the autocorrelation coefficient. This result is coherent with Zangari and Tsurumi (1996). A thorough set of convergence diagnostics for the posterior simulator will be inserted in a next release of the code. A future line of study should explore the behaviour of the method against actual empirical applications.

References

[1] G. Casella and George E. Explaining the gibbs sampler. *The American Statistician*, 46:167–174, 1992.

[2] S. Chib. Bayes inference in the tobit censored regression model, 1992.

[3] S. Chib. Bayes regression with autoregressive errors: A gibbs sampling approach. *The Journal of Econometrics*, 58 n.3:275–294, 1993.

[4] S. Chib. *Handbook of Econometrics 5*. North-Holland, Amsterdam, 2001.

[5] S. Chib and Carlin B.P. On mcmc sampling in hierarchical longitudinal models. *Statistics and Computing*, 1999.

[6] S. Chib and Greenberg E. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49 n.4:327–335, 1995.

[7] P. Contoyannis, Jones M.A., and Leon-Gonzalez R. Using simulation based inference with panel data in health economics. Working paper series 13, McMaster University, Hamilton, Ontario Canada, 2002.

[8] J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, 4:169–193, 1992.

[9] V.A. Hajivassiliou and McFadden D.L. The method of simulated scores for the estimation of ldv models with an application to external debt crises. Discussion Paper 967, Cowles Commission, Chicago, 1990.

[10] M.N. Harris, Macquarie L.R., and Siouclis A.J. A comparison of alternative estimators for binary panel probit models. Working Paper 3/00, Melbourne Institute, Melbourne Australia, 2000.

[11] M. Tanner and Wong W. The calculation of posterior density by data augmentation. *Journal of the American Statistical Association*, 1987.

[12] P.J. Zangari and Tsurumi H. A bayesian analysis of censored autocorrelated data on exports of japanese passenger cars to the united states. *Advances in Econometrics*, 11,A:111–143, 1996.