MONASH

WP 4-98

ISSN 1440-771X
ISBN 0 7326 1041 9

MONASH UNIVERSITY

ANCORA · IMPARO

AUSTRALIA

A Comparison of Alternative Estimators

for Binary Panel Probit Models

Mark N. Harris
Lachlan R. Macquarie
Anthony J. Siouclis

Working Paper 4/98
June 1998

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

A Comparison of Alternative Estimators for Binary Panel Probit Models*

Mark N. Harris,[†] Lachlan R. Macquarie[‡] and Anthony J. Siouclis[‡]

[†]The Melbourne Institute of Applied Economics and Social Research,

The University of Melbourne, Parkville, Vic. 3052, AUSTRALIA.

[‡]Department of Econometrics and Business Statistics,

Monash University, Clayton, Vic. 3168, AUSTRALIA.

Mailing Author : Mark.Harris@iaesr.unimelb.edu.au

June, 98

Abstract

Recent advances in computing power have brought the use of computer intensive estimation methods of binary panel data models within the reach of the applied researcher. The aim of this paper is to apply some of these techniques to a marketing data set and compare the results. In addition, their small sample performance is examined via Monte Carlo simulation experiments. The first estimation technique used was maximum likelihood estimation of the cross section probit (ignoring heterogeneity). The remaining.techniques estimated the binary panel probit model using: standard maximum likelihood; the Solomon-Cox approximation to this likelihood and finally; the Gibbs sampler to obtain Bayesian estimates. The results suggested that, in most cases, standard maximum likelihood estimation of the binary panel probit model was the preferred technique primarily because it is readily available to applied practitioners. Although when the variance of the heterogeneity term is small, the computational simplicity of the Solomon-Cox approximation may prove attractive. In large samples, the Gibbs sampler was also found to perform well.

*We are extremely grateful to Tim Fry, Ritchard Longmire and Rosi Doran for helpful guidance and suggestions. We are also greatly indebted to Sid Chib for making the Gibbs sampler Gauss code available to us. All remaining errors are our own.

1. Introduction

Given recent advances in personal computing power, it is now within the reach of the applied researcher to implement computer intensive estimation methods to previously intractable models. This paper compares different estimation techniques of the binary panel probit model via both an application to a marketing data set and a limited set of Monte Carlo experiments.

Four estimation techniques were applied to an empirical and several simulated panel data sets in this study. The first estimation technique used was a cross section binary probit model (ignoring heterogeneity), with coefficients estimated via maximum likelihood (ML). This estimation method is subsequently referred to as the cross section (CS) probit. The remaining three methods all estimated the binary panel probit model (accounting for heterogeneity). The first estimation technique employed was ML estimation of the model (referred to as standard ML), whilst the second approximated this likelihood with the Solomon-Cox (SC) approximation. The estimates derived here maximised the approximated likelihood function. Finally, the Gibbs sampler was used to obtain Bayesian estimates of the binary panel probit model, which does not evaluate the likelihood function directly.

The binary panel probit model has been frequently used in applied research (see for example, Zellner and Rossi [1984], Albert and Chib [1993, 1995] and Harris [1996]). The Gibbs sampler has been applied to binary and multinomial panel probit models (McCulloch and Rossi [1994], Geweke *et al.* [1994b], Albert and Chib [1995] and Chib [1996]) as well as cross-sectional binary and multinomial probit models (Zellner and Rossi [1984], Chib [1993], McCulloch and Rossi [1994] and Geweke *et al.* [1994a]). The Solomon-Cox method is also relatively new to the literature (Lieberman and Mátyás [1994, 1996]). Recently, Solomon-Cox approximations have been used successfully in panel data applications, although there appears to be few applications to the panel probit model (see Lieberman and Mátyás [1994, 1996] for a list of references).

The findings of this paper in general support of use of standard ML estimation of the binary panel probit model. Although when the variance of the heterogeneity term is small, the computational simplicity of the Solomon-Cox approximation may prove an attractive alternative. In large samples, the Gibbs sampler was also found to perform well. The plan of the paper is as follows: Section 2 briefly outlines the binary panel probit model and the various estimators considered; Section 3 describes the data set used; Section 4 provides empirical results; Section 5 contains the Monte Carlo simulation experiment results. Section 6 contains some concluding remarks.

2. The Binary Panel Probit Model

Choice decisions made by individuals are often of economic interest. In general however, only the outcome of the decision process is observed, while the underlying process is unobserved or *latent*. In the simplest case, the individual is faced with a yes/no (binary) decision, which, without loss of generality, can be defined as one (for a yes decision) and zero (for a no).

Let the observed binary variable y_i be the discrete realisation of the (latent) continuous random variable, y_i^*. Assuming a linear functional form for the latent variable gives

$$y_i^* = x_i'\beta + u_i, \quad i = 1, ..., N \tag{2.1}$$

where x_i is a $k \times 1$ vector of explanatory variables,
β is a $k \times 1$ vector of unknown coefficients and,
u_i is the unknown error term.

The latent variable y_i^* can be considered as representing the perceived utility of individual i. Assuming the Random Utility Model (RUM), an individual will select an alternative if they perceive that it maximises their utility, such that

$$y_i = \begin{cases} 1 \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \leq 0 \end{cases} . \tag{2.2}$$

Thus, $\Pr(y_i = 1) = \Pr(y_i^* > 0) = \Pr(u_i > -x_i'\beta)$. The evaluation of this probability requires specifying the distribution of u_i. The probit model assumes u_i is normally distributed with a mean of zero, and a variance of one (for identification).

2.1. Binary Probit Estimation

The researcher is therefore interested in modelling the probability of y_i being zero or one. Assuming this probability can be parameterised by the function $F(x_i'\beta)$, the cross section probit results if F is the standard normal cumulative distribution function evaluated at $x_i'\beta$. That is,

$$F(x_i'\beta) = \Phi(x_i'\beta) = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \tag{2.3}$$

where the values of $F(x_i'\beta)$ are bounded between zero and one.

ML estimation of the binary probit (BP) model therefore involves maximising the log-likelihood function

$$l_{BP} = \sum_{y_i=0} \ln F(x_i'\beta) + \sum_{y_i=1} \ln(1 - F(x_i'\beta)), \tag{2.4}$$

with respect to β.

2.2. The Binary Panel Probit Model

The formulation of the binary choice model above assumes that the error term is an independently and identically distributed standard normal variable and hence ignores any heterogeneity between individuals. The binary panel probit model allows us to account for this heterogeneity in the form of individual specific effects, represented by α_i (see Hsiao [1986, 1996]). A higher realisation of α_i increases the probability of $y_i = 1$, for all $t = 1, ..., T$. With data in the form of a panel, the latent model can be reparamaterised to include individual specific effects, given by

$$y_{it}^* = x_{it}'\beta + \alpha_i + \nu_{it} \ , \ i = 1, ..., N, t = 1, ..., T \tag{2.5}$$

where α_i is the individual specific effect and,
ν_{it} is the random error term of the panel probit model.

It is assumed that α_i and ν_{it} are independently and normally distributed with zero mean and variances σ_α^2 and σ_ν^2 respectively.

2.2.1. Classical Maximum Likelihood Estimation

A fixed effect estimation technique applied when the individual specific effects are random will result in, at best, a loss of efficiency in the estimates of β, and at worst, inconsistent estimates attributed to a small T in most panel data sets.[1] Under the assumption of normality of α_i and ν_{it}, the log likelihood function of the panel probit (PP) model with random effects is given by

$$l_{PP} = \sum_{i=1}^{N} \ln\left(P(y_i)\right) \text{ where}$$

$$P(y_i) = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp -\frac{\alpha_i^2}{2} \left\{ \prod_{t=1}^{T} \Phi \left[\left(\frac{x_{it}'\beta}{\sigma_\nu} + \frac{\alpha_i}{\sigma_\nu} \right) (2y_{it} - 1) \right] \right\} d\alpha_i. \tag{2.6}$$

[1] Hsiao (1996).

ML estimation of the log likelihood function provides consistent and efficient estimates for β (under weak regularity conditions), however, it is computationally intensive. ML estimation can be simplified when appropriate assumptions are made upon the composite error term, $\varepsilon_{it} = \alpha_i + v_{it}$. Assuming α_i and ν_{it} are independent, the correlation across i is assumed to be constant such that

$$Corr\left(\varepsilon_{it}, \varepsilon_{is}\right) = \rho = \sigma_{\alpha}^{2} / \left(\sigma_{\alpha}^{2} + \sigma_{v}^{2}\right), t \neq s. \tag{2.7}$$

The estimation problem can now be reduced to a single integral, with its integrand being the product of a single normal density and T differences of normal cumulative distribution functions. Butler and Moffitt [1982], propose using Gaussian quadrature to evaluate the required integral. The formula used is the *Hermite integration formula* of the form

$$\int_{-\infty}^{\infty} e^{-z^2} g(z) = \sum_{j=1}^{J} A_j g(z_j) \tag{2.8}$$

where J is the number of evaluation points, z_j are the nodes at which $g(.)$ is evaluated, with respective weights A_j. In this paper, eight point quadrature was used.

2.3. Solomon-Cox Approximation

The Solomon-Cox (SC) approximation is a general estimation technique proposed to provide an analytical solution for the ML estimates of non-linear panel data models with random effects, which are often very complex, or indeed, intractable. Given the concerns noted above regarding the evaluation of the integral, it appears ideally suited to the panel probit case. The procedure provides a small-variance approximation to the true marginal likelihood function, and is attractive due to its simplicity and computational tractability.

The SC approximation applied to the binary panel probit model with random effects again assumes that α_i are independently normally distributed for all i, with the log likelihood function (Lieberman and Mátyás [1996]) given by

$$\begin{aligned} l_{SC} &= \sum_{i=1}^{N} \ln \left\{ \left[\prod_{t=1}^{T} f(y_{it}|x_{it}, \xi; \beta) \right] \left[1 - \sigma^2 l_{ci}^{(2)} \right]^{-\frac{1}{2}} \times \right. \\ & \left. \exp \left[\frac{\sigma^2 l_{ci}^{(1)2}}{2(1 - \sigma^2 l_{ci}^{(2)})} \right] \left[1 + O(\sigma^4) \right] \right\} \end{aligned} \tag{2.9}$$

where $\xi = E(\alpha_i)$,
$O(\sigma^4)$ is the order of bias,

$$\prod_{t=1}^{T} f(y_{it}|\xi_{it}, 0, \beta) = \prod_{t=1}^{T} \left[\Phi^{y_{it}} (1-\Phi)^{1-y_{it}}\right],$$

$$l_{ci}^{(1)} = \sum_{t=i}^{T} \tfrac{y_{it}-\Phi}{\Phi(1-\Phi)}\phi,$$

$$l_{ci}^{(2)} = -\sum_{t=1}^{T} \left\{y_{it}\left[\Phi^{-2}\phi^2 + \Phi^{-1}\phi x_{it}'\beta\right]\right\},$$
$$+(1-y_{it})\left[(1-\Phi)^{-2}\phi^2 - (1-\Phi)^{-1}\phi x_{it}'\beta\right]$$

and, Φ and ϕ are the standard normal cumulative distribution and probability density functions respectively, evaluated at $x_{it}'\beta$.

The SC approximation is computationally convenient in that it avoids burdensome numerical integration. Hence, if the variance of the composite error term is relatively small, the SC technique provides a useful method to obtain estimates of the parameters in the binary panel probit model with random effects.

2.4. Gibbs Sampler

The Gibbs sampler is a Monte Carlo Markov Chain method of sampling well suited to Bayesian inference. This method of estimation has simplified the Bayesian analysis of panel data models providing precise finite sample estimates. The Gibbs sampling algorithm simulates the posterior distribution rather than computing the posterior moments, and may be a preferred estimation tool over its SC approximation counterpart.

In a general setting, Bayesian estimation requires the estimation of the posterior distribution of the unknown parameters in the model ($\theta = \beta, \rho$) to be estimated, for example $\pi(\theta|y)$. The Gibbs sampler obtains draws from the posterior distribution by successively drawing from the full conditional distributions of the individual parameters comprising θ. Say there is a data set y, and the parameters of interest are $\theta' = (\beta', \rho)'$. To draw from the joint posterior distribution $\pi(\beta, \rho|y)$, start with any value of ρ that supports the joint distribution of β and ρ, and draw from the *full conditional distribution*, $p(\beta|\rho, y)$. Using this draw of β as the new value, a drawing is made from the full conditional distribution $p(\rho|\beta, y)$, providing an updated value of ρ, which is in turn used to obtain an updated value for β. This process is continued for a large number of iterations, with the sequence being a stationary Markov chain. The equilibrium distribution of this Markov chain can be shown to be the posterior joint distribution $\pi(\theta|y)$.

Thus, by recursively sampling from the full conditional distributions, draws from the posterior distribution can be obtained.

The distinguishing feature of the binary panel probit model is the existence of the unobservable latent data. The Gibbs sampler overcomes this via data augmentation. This was proposed by Albert and Chib [1993] in a cross-sectional context, but can be easily applied to the panel probit (Albert and Chib [1995]). Given the assumptions underlying the model, the distribution of the latent variable is known. Thus, *conditional* on the parameter vector β, $p(y_{it}^*|\beta) \sim N(x_{it}'\beta, 1)$. Draws for y_{it}^* are relatively easy to obtain from truncated normal distributions, dependant on the observed value of y_{it}. Specifically

$$y_{it}^* \begin{cases} TN_{(0,\infty)}(x_{it}'\beta,\ 1) \text{ if } y_{it} = 1 \\ TN_{(-\infty,0)}(x_{it}'\beta,\ 1) \text{ if } y_{it} = 0. \end{cases} \tag{2.10}$$

The full conditional distributions used in the Gibbs sampler are thus $p(\beta\,|y_{it}^*)$ and $p(y_{it}^*\,|y_{it},\beta)$. By sampling from these full conditional distributions a set of simulated latent variables is obtained. Treating the simulated values of y_i^* as observed data, the model can then estimated as a standard Bayesian regression.

The Albert and Chib algorithm can easily be applied to a panel probit model with random effects. This procedure avoids the problem of evaluating the integral of the likelihood function in equation (2.6).

The Gibbs sampler can be implemented by first drawing values of the latent variable y_{it}^*. The full conditional distribution of the latent variable in this case is given by the following truncated normal distributions

$$p(y_{it}^*\,|\beta,\alpha_i,D,y_{it}) = \begin{cases} TN_{(0,\infty)}\left(x_{it}'\beta+\alpha_i,\ \sigma_\alpha^2+\sigma_\nu^2\right) \text{ if } y_{it} = 1 \\ TN_{(-\infty,0)}\left(x_{it}'\beta+\alpha_i,\ \sigma_\alpha^2+\sigma_\nu^2\right) \text{ if } y_{it} = 0. \end{cases} \tag{2.11}$$

By obtaining draws for the latent variable, y_{it}^*, the problem is reduced to one of estimating a standard Bayesian panel regression, with the observed binary data rendered redundant in the subsequent steps of the Gibbs sampling routine.

The parameters of the model can then be drawn via the following full conditional distributions:[2]

$$p(\alpha_i|D,\beta,y_{it}^*) \sim N(\hat{\alpha}_i, V_i); \tag{2.12}$$
$$p(\beta|D,y_{it}^*,\alpha_i,x_{it}) \sim N(\hat{\beta},\Sigma_\beta); \tag{2.13}$$

[2]A full exposition of these distributions can be found in Chib [1996].

$$p(D^{-1}|\beta, y_{it}^*, \alpha_i, x_{it}) \sim W\left(\rho_0 + N, \left(R_0^{-1} + \sum_{i=1}^{n} \alpha_i \alpha_i\right)^{-1}\right); \quad (2.14)$$

where α_i are independent of x_{it}, $V_i = (D^{-1} + 1)^{-1}$, $\hat{\alpha}_i = V_i(y_i^* - x_i\beta)$, $\hat{\beta}$ and Σ_β are the usual panel estimates for the mean and variance of β and, ρ_0 and R_0 are parameters of the prior distribution of D^{-1}. Successive draws from the full conditional distributions above are obtained and repeated a large number of times. A salient feature of the Gibbs sampling procedure is that it does not require the evaluation of the likelihood function of the panel probit model at any stage. In this application, non-informative priors for all of the parameters were used. The Gibbs sampler was run for 6100 iterations, with the first 100 draws omitted to account for "start-up" noise.[3] The Bayesian estimates reported are the means of the full conditional distributions.

3. An Empirical Example: The Data

The data used in this application was extracted from the Roy Morgan Research Consumer Panel of Australia, over the financial year July 1992 - June 1993, for the Melbourne metropolitan area. The data was diary based and collected monthly for a variety of product fields. Daily purchases of products and a range demographic variables were recorded for each household. The subset comprised 284 respondents who purchased laundry detergent during the year.

In order to fit into the binary panel probit framework, the outcome of interest was coded into a binary decision, representing whether the household purchased any laundry detergent in a given month or not (one and zero respectively).

Not all of the respondents in the panel returned a full set of diaries for the year. Out of the 284 individuals that purchased laundry detergent, 216 returned all twelve diaries. This however may not be due to attrition in the panel. No information was available as to which diaries were not returned. It was therefore impossible to discern whether incomplete diary returns were due to attrition of panel members. An incomplete set of diaries may have also been due to such things as forgetfulness, or because the panel member was on vacation, and so on. Empirical work was therefore based on the 216 members who returned all twelve diaries.

[3] Estimates obtained from the Gibbs sampler could be made more precise by increasing the number of iterations. The obvious trade-off is the associated increase in running time.

The only choice specific variable considered was the average price of all laundry detergents in all stores during each month. Other variables initially analysed were individual specific demographics, such as sex, type of work, marital status, household size and household income. Preliminary analysis suggested that only household size was statistically significant.. However, price was also included in further modelling as it was deemed to be an important economic variable.

The expected sign of the household size coefficient is not obvious. One might expect that the larger the household the more detergent is required. However, modelling the purchase decision renders the volume of purchases irrelevant. It could also be argued that larger households would buy laundry detergent in bulk, and hence fewer purchases would occur throughout the year than for a single person household for example. Nevertheless, one would expect on average *a priori*, that household size would have a positive coefficient.

Microeconomic demand theory would lead one to expect that price would exert a negative effect on the probability of purchase. However, as the price variable used was aggregated across all Melbourne metropolitan stores for the entire month, some of the price effect may have been smoothed out. Prices of laundry detergent can potentially vary a great deal from day to day both between and within stores. Furthermore, the nature of the product means that panel members can postpone (although not indefinitely) purchases according to price changes. The price variable used can therefore not be expected to account for such behaviour.

4. An Empirical Example: Results and Evaluation

Ordinary least squares estimates appeared to provide suitable starting values for each procedure as convergence was achieved rapidly in each estimation technique. The results are presented in table 4.1.[4] A preliminary examination of table 4.1 indicates that the four procedures yielded similar parameter estimates. In each method, the household size coefficient was found to be statistically significant and had the expected sign. The coefficients of the average price variable, although not statistical significant, did have the expected negative sign, complying with *a priori* expectations.

Tests for heterogeneity can be based upon the significance of the estimate of ρ or upon maximised likelihood functions in the form of a likelihood ratio test.

[4] Gauss code for the empirical example and subsequent simulation (see section 5) is available on request from the authors.

Table 4.1: Laundry Detergent Purchases Parameter Estimates (N=216, T=12[a])

Explantory Variables	Probit (using ML)	Panel Probit	Solomon-Cox approximation	Bayesian (Gibbs Sampling)
Constant	1.0682 (0.9508)	1.4533 (1.5021)	1.2831 (1.4686)	1.240 (1.307)[b]
Household Size	0.1697 (0.0176)	0.2138 (0.0427)	0.2205 (0.0347)	0.207 (0.041)
Average Price	-0.0049 (0.0028)	-0.0064 (0.0044)	-0.0060 (0.0043)	-0.006 (0.004)
ρ	-	0.3539 (0.0312)	0.3487 (0.0470)	0.366 (0.080)
Max. log-likelihood	-1744.37	-1569.56	-1601.99	-

[a] Asymptotic standard errors in parentheses.

[b] Standard deviation of the posterior distribution.

Table 4.2: Hit-Miss Table for the Cross Section Probit

	Predicted		
Actual	0	1	**Total**
0	36.7	16.2	**52.9**
1	23.9	23.2	**47.1**
Total	**60.6**	**39.6**	**100**

Under both methods heterogeneity was detected and therefore there seemed to be strong evidence that the cross section probit model was seriously misspecified.

For many researchers, the choice of the most appropriate estimation technique may be a function of the required computational complexity. The least computer and time intensive procedure was the binary probit model. However, diagnostic tests suggested that, in this instance, it was misspecified. The three panel estimation procedures provide similar results in terms of expected signs and parameter significance, and subsequently, the SC approximation was preferred, as it is less burdensome than its counterparts.

An additional tool used to help differentiate between the estimation techniques was to examine their within sample prediction accuracy. A common measure of predictive accuracy in discrete choice models is the hit-miss table. The respective hit-miss tables for the four estimators are provided in tables 4.2, 4.3, 4.4 and 4.5. Tabulated values are in percentages relative to the total number of observations in the sample.

From these results, it was apparent that standard ML estimation and the Solomon-Cox approximation (Tables 4.3 and 4.4 respectively) correctly predict

Table 4.3: Hit-Miss Table for Standard ML Estimation

	Predicted		
Actual	0	1	**Total**
0	35.8	17.1	**52.9**
1	22.5	24.6	**47.1**
Total	**58.3**	**41.7**	**100**

Table 4.4: Hit-Miss Table for the Solomon-Cox Approximation

	Predicted		
Actual	0	1	**Total**
0	35.8	17.1	**52.9**
1	22.5	24.6	**47.1**
Total	**58.3**	**41.7**	**100**

an identical number of purchases of laundry detergent. Furthermore, these two procedures provided the closest predictions to the actual results. Both methods under-predicted the number of purchases by 5.4 percentage points. In contrast, the CS probit and Bayesian estimation under-predicted the number of purchases by 7.7 and 10.1 percentage points respectively. These tables seem to support standard ML estimation and the SC approximation as the most appropriate estimation techniques of the binary panel probit using this data set.

In summary, despite similar parameter estimates for each methodology employed in this paper, standard ML and the SC approximation of the binary panel probit model are the favoured procedures as they achieved a prediction accuracy of 60.4% compared with 59.9% and 59.1% for the cross section probit and Bayesian estimator respectively. Coupled with the computational complexity of each estimation technique, it seems that maximum likelihood estimation of the SC procedure would appear to be the most appropriate for the applied researcher. However, further examination of these results was conducted using a set of Monte

Table 4.5: Hit-Miss Table for Bayesian Estimation

	Predicted		
Actual	0	1	**Total**
0	37.5	15.4	**52.9**
1	25.5	21.6	**47.1**
Total	**63.0**	**37.0**	**100**

Carlo experiments.

5. Monte Carlo Simulation

The data used in the simulation experiments was generated according to the following process

$$y_{it}^{*} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2i} + \alpha_i + \nu_{it} \tag{5.1}$$

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

$$y_{it} = \begin{cases} 1 \text{ if } y_{it}^{*} > 0 \\ 0 \text{ otherwise.} \end{cases}$$

The values for the three β coefficients were 0.5, -1, and 1 respectively. These values were chosen for their simplicity and because they gave an appropriate split of zeros and ones. The initial values of the time and individual varying variable were generated according to $x_{1it} \sim U(-2, 2)$. Subsequent values of x_{1it} followed an autoregressive process given by

$$\begin{aligned} x_{1it} &= 0.1 \times (trend) + 0.5\,(x_{1it-1}) + u_{it}, \\ u_{it} &\sim U(-\frac{1}{2}, \frac{1}{2}). \end{aligned}$$

The time invariant variable of the model, x_{2i} was generated as $x_{2i}^{*} \sim U(0, 1)$, where

$$x_{2i} = \begin{cases} 0 \text{ if } 0 \leq x_{2i}^{*} < 0.5 \\ 1 \text{ if } 0.5 \leq x_{2i}^{*} \leq 1. \end{cases}$$

To complete the simulation, the generation process of the composite error term was specified. The individual specific effects parameter (α_i) was generated according to $\alpha_i \sim N(0, \sigma_\alpha^2)$, where σ_α^2 was specified as 0.5, 1 and 2. The choice of σ_α^2 corresponded to values for ρ equal to $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{3}$ respectively, implying weak, mild and strong correlation over time for each individual. Finally, the random error term (ν_{it}) was independently drawn from a standard normal distribution. The simulations attempted to replicate commonly used data sets. For example, the regressor x_{1it} attempted to proxy time and individual varying variables, such as age, experience, income etc., whilst the time invariant regressor, x_{2i} was used as a proxy for such variables as gender, race etc. (see Harris [1996]).

Although ρ is directly estimated by standard ML, the Gibbs sampler and the SC approximation estimate the variance of the individual effects. Therefore a calculation based on (2.7) was required to find the corresponding value of ρ. To complete the simulation process, the dimensions of the panel needed to be specified. To replicate the empirical example, a panel of dimension $N = 200$ by $T = 12$ was simulated, corresponding to 200 observations over 12 months. Following this, experiments of smaller dimension were simulated, consisting of a T dimension of 3, 6 and 12, with $N = 100$.

Examining the results in table 6.1, one can observe that, when σ_α^2 and N were kept constant, standard ML provided parameter estimates closer to the true values. Furthermore, the mean squared errors (MSE) for each technique were typically relatively small and consistent between procedures. As N increased, the parameter estimates converged to their true values, and their respective MSE functions decreased monotonically. When σ_α^2 was varied, a number of interesting results appeared. In particular, the SC approximation, being a small-variance approximation, was more accurate when $\sigma_\alpha^2 = \frac{1}{2}$. As the variance increased, the parameter estimates diverged from their true values, and correspondingly, the MSE functions increased rapidly. Despite being a small-variance approximation, the SC was outperformed by standard ML for all of the three chosen values of σ_α^2. As ML and the Gibbs sampler are not a small sigma approximation, they did not appear to be adversely affected by changes in σ_α^2. An overall examination of the MSE's in table 6.1 suggested that the preferred estimation procedure was standard ML.

When T was adjusted, keeping all other factors constant, more accuracy and precision in the estimates was observed, accompanied by universally lower MSE functions. The results of table 6.2 concur with those of table 6.1 in that standard ML estimation seemed to outperform the other techniques, particularly for small variance simulations when $N = 200$. However, when $N = 100$, more precise parameter estimates were observed as the variance was increased. On the other hand, the MSE functions were smaller for the small variance simulations. For all values of σ_α^2, standard ML estimation outperformed the SC approximation.

The final simulation involved further increasing T to 12. Once again, standard ML estimation performed well. Precision of the estimates increased with N, which is consistent with the previous simulations, and for the SC approximation, increased precision was observed when σ_α^2 decreased. For $N = 200$, estimates from the Gibbs sampler performed better than the ML estimates. This seems to suggest that the Bayesian technique may be preferred when T is large. The CS probit model did not perform well in any setting. This seemed to suggest that not accounting for heterogeneity adversely effects parameter estimates.

In summary, the simulation results suggested that standard ML estimation was the preferred technique. The SC approximation provided accurate estimates in small variance simulations, but was outperformed by ML estimation for all levels of σ_{α}^{2}. Although the likelihood function of the SC is simpler and provides a closed analytical form, standard ML is available in econometrics packages such as Limdep^{TM}. For a given level of T, increased precision was observed in the estimates as N increased. Finally, the Gibbs sampler outperformed ML when $T = 12$. While the Gibbs sampler did not perform as well in the other experiments, a distinct advantage it has is that it provides a *distribution* of the parameters of interest, rather than a point estimate. Thus, the increased programming and computational burden may be warranted. Although the increased programming and computational burden of the SC approximation and the Gibbs sampler are not prohibitive, standard ML is recommended for most applied situations.

6. Conclusion

Four estimation techniques using consumer panel data were applied and compared in this paper. First, the CS probit model (ignoring heterogeneity) was estimated using maximum likelihood. The three remaining methods estimated the binary panel probit model (accounting for heterogeneity) using: standard ML estimation of the likelihood function; the Solomon-Cox approximation to the likelihood function and; a Bayesian approach utilising the Gibbs sampler, using data augmentation to avoid direct evaluation of the likelihood function. The starting values used for each technique were least squares estimates.

Each technique was applied to a consumer panel data set of laundry detergent purchases. The parameter estimates of the four estimators all had the expected signs, while individual effects were found to have a pronounced influence on purchases from diagnostic tests. Using within sample predictive accuracy to differentiate between procedures, the SC approximation and standard ML outperformed the Bayesian and CS probit estimators, with slightly superior predictive accuracy.

Finally, the results from some Monte Carlo experiments were presented. The results confirmed that the SC approximation would perform well when σ_{α}^{2} is small. From the empirical example, the estimated value of ρ (normalising on σ_{ν}^{2}) implied $\sigma_{\alpha}^{2} \approx 0.5$, in which setting the SC approximation performed well in the Monte Carlo experiments.. However, for all samples and parameter settings, standard ML estimation emerged as the best performed estimation procedure. While results for each of the binary panel probit estimators were quite close, standard ML estimation was considered to be the preferred technique. It is readily available to practitioners in many statistical packages, avoiding programming and alleviating some of the computational burden.

References

[1] **Albert, J.H. and Chib, S. (1993),** Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, pp. 669-679.

[2] **Albert, J.H. and Chib, S. (1995),** Bayesian Residual Analysis for Binary Response Regression Models, *Biometrika*, 82, pp. 747-759.

[3] **Butler, J.S. and Moffitt, R. (1982),** A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model, *Econometrica*, 50, pp. 761-764.

[4] **Chib, S. (1996),** Inference in Panel Data Models Via Gibbs Sampling in: *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Second Revised Edition, L. Mátyás and P. Sevestre (eds.), Ch. 24, Kluwer Academic Publishers, Dordrecht.

[5] **Geweke, J.F., Keane, M.P. and Runkle, D.E (1994a),** Alternative Computational Approaches to Inference in the Multinomial Probit Model, *The Review of Economics and Statistics*, 76, pp. 609-632.

[6] **Geweke, J.F., Keane, M.P. and Runkle, D.E. (1994b),** Statistical Inference in the Multinomial Multiperiod Probit Model, Federal Reserve Bank of Minneapolis, Research Department Staff Report #177.

[7] **Harris, M.N. (1996),** Modelling the Probability of Youth Unemployment in Australia, *The Economic Record*, 72, pp. 118-129.

[8] **Hsiao, C. (1986),** *Analysis of Panel Data*, Cambridge University Press, Cambridge.

[9] **Hsiao, C. (1996),** Logit and Probit Models in: *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Second Revised Edition, L. Mátyás and P. Sevestre (eds.), Ch. 16, Kluwer Academic Publishers, Dordrecht.

[10] **Lieberman, O. and Mátyás, L. (1994),** Improved Estimation Procedures for Nonlinear Panel Data Models, Monash University, Department of Econometrics, Working Paper 15/94.

[11] **Lieberman, O. and Mátyás, L. (1996),** Improved Estimation Procedures in: *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Second Revised Edition, L. Mátyás and P. Sevestre (eds.), Ch. 21, Kluwer Academic Publishers, Dordrecht.

[12] **McCulloch, R. and Rossi, P. (1994),** An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics*, 64, pp. 207-240.

[13] **Zellner, A. and Rossi, P. (1984),** Bayesian Analysis of Dichotomous Quantal Response Models, *Journal of Econometrics*, 25, pp. 365-393.

Table 6.1: Monte Carlo Parameter Estimates for T = 3[1]

		N=100			N=200		
	$\sigma_\alpha^2 =$	$\frac{1}{2}$	1	2	$\frac{1}{2}$	1	2
Probit	$\beta_0 = 0.5$	0.417 (0.026)	0.356 (0.039)	0.296 (0.063)	0.409 (0.016)	0.358 (0.030)	0.290 (0.055)
	$\beta_1 = -1$	−0.836 (0.046)	−0.726 (0.093)	−0.594 (0.182)	−0.831 (0.038)	−0.715 (0.090)	−0.579 (0.186)
	$\beta_2 = 1$	0.831 (0.076)	0.732 (0.119)	0.596 (0.216)	0.828 (0.051)	0.715 (0.103)	0.581 (0.199)
Solomon-Cox	$\beta_0 = 0.5$	0.469 (0.026)	0.431 (0.031)	0.389 (0.050)	0.462 (0.012)	0.436 (0.018)	0.385 (0.032)
	$\beta_1 = -1$	−0.939 (0.028)	−0.874 (0.040)	−0.772 (0.077)	−0.934 (0.016)	−0.864 (0.031)	−0.756 (0.072)
	$\beta_2 = 1$	0.937 (0.066)	0.887 (0.082)	0.787 (0.137)	0.935 (0.032)	0.870 (0.049)	0.772 (0.093)
	ρ	0.229 (0.020)	0.360 (0.028)	0.505 (0.034)	0.232 (0.015)	0.370 (0.022)	0.519 (0.026)
Standard ML	$\beta_0 = 0.5_0$	0.511 (0.0315)	0.503 (0.038)	0.515 (0.080)	0.500 (0.013)	0.508 (0.020)	0.503 (0.035)
	$\beta_1 = -1$	−1.025 (0.036)	−1.028 (0.042)	−1.031 (0.060)	−1.016 (0.017)	−1.015 (0.022)	−1.007 (0.026)
	$\beta_2 = 1$	1.019 (0.081)	1.034 (0.103)	1.042 (0.195)	1.013 (0.036)	1.011 (0.048)	1.007 (0.076)
	ρ	0.319 (0.0154)	0.484 (0.092)	0.651 (0.009)	0.323 (0.008)	0.492 (0.006)	0.660 (0.004)
Gibbs Sampler	$\beta_0 = 0.5_0$	0.542 (0.036)	0.528 (0.041)	0.540 (0.075)	0.514 (0.014)	0.519 (0.021)	0.516 (0.036)
	$\beta_1 = -1$	−1.082 (0.045)	−1.069 (0.049)	−1.068 (0.066)	−1.043 (0.020)	−1.034 (0.024)	−1.028 (0.028)
	$\beta_2 = 1$	1.072 (0.092)	1.071 (0.113)	1.069 (0.191)	1.039 (0.039)	1.030 (0.050)	1.027 (0.077)
	ρ	0.400 (0.090)	0.535 (0.160)	0.688 (0.321)	0.363 (0.041)	0.517 (0.095)	0.680 (0.214)

[1] Average parameter estimates over 1000 Monte Carlo repetitions with Mean Squared Errors in parentheses.

Table 6.2: Monte Carlo Parameter Estimates for T = 6[1]

		N=100			N=200		
	$\sigma_\alpha^2 =$	$\frac{1}{2}$	1	2	$\frac{1}{2}$	1	2
Probit	$\beta_0 = 0.5$	0.421 (0.021)	0.364 (0.035)	0.295 (0.062)	0.413 (0.015)	0.360 (0.028)	0.293 (0.052)
	$\beta_1 = -1$	−0.833 (0.040)	−0.721 (0.090)	−0.591 (0.179)	−0.818 (0.039)	−0.712 (0.089)	−0.582 (0.180)
	$\beta_2 = 1$	0.826 (0.057)	0.709 (0.118)	0.589 (0.210)	0.819 (0.046)	0.710 (0.100)	0.572 (0.201)
Solomon-Cox	$\beta_0 = 0.5$	0.484 (0.020)	0.451 (0.030)	0.398 (0.052)	0.476 (0.011)	0.448 (0.017)	0.400 (0.030)
	$\beta_1 = -1$	−0.962 (0.016)	−0.900 (0.024)	−0.804 (0.052)	−0.946 (0.010)	−0.891 (0.019)	−0.793 (0.049)
	$\beta_2 = 1$	0.981 (0.037)	0.944 (0.065)	0.887 (0.107)	0.974 (0.020)	0.946 (0.031)	0.862 (0.061)
	ρ	0.288 (0.006)	0.452 (0.008)	0.638 (0.006)	0.290 (0.004)	0.460 (0.005)	0.638 (0.004)
Standard ML	$\beta_0 = 0.5$	0.514 (0.022)	0.510 (0.033)	0.509 (0.075)	0.504 (0.011)	0.506 (0.017)	0.498 (0.032)
	$\beta_1 = -1$	−1.016 (0.017)	−1.012 (0.019)	−1.008 (0.022)	−0.999 (0.008)	−1.004 (0.010)	−0.994 (0.011)
	$\beta_2 = 1$	1.007 (0.039)	1.000 (0.072)	1.008 (0.162)	1.000 (0.020)	1.000 (0.032)	0.969 (0.070)
	ρ	0.324 (0.005)	0.488 (0.088)	0.652 (0.004)	0.327 (0.002)	0.494 (0.002)	0.646 (0.002)
Gibbs Sampler	$\beta_0 = 0.5$	0.524 (0.023)	0.520 (0.033)	0.517 (0.058)	0.509 (0.011)	0.511 (0.017)	0.511 (0.026)
	$\beta_1 = -1$	−1.033 (0.018)	−1.025 (0.020)	−1.026 (0.023)	−1.007 (0.009)	−1.011 (0.010)	−1.009 (0.011)
	$\beta_2 = 1$	1.020 (0.040)	1.008 (0.070)	1.021 (0.121)	1.007 (0.021)	1.005 (0.032)	0.989 (0.056)
	ρ	0.356 (0.027)	0.512 (0.070)	0.677 (0.187)	0.342 (0.012)	0.507 (0.039)	0.669 (0.111)

[1] Average parameter estimates over 1000 Monte Carlo repetitions with Mean Squared Errors in parentheses.

Table 6.3: Monte Carlo Parameter Estimates for T = 12[1]

		N=100			N=200		
	$\sigma_\alpha^2 =$	$\frac{1}{2}$	1	2	$\frac{1}{2}$	1	2
Probit	$\beta_0 = 0.5$	0.416 (0.018)	0.358 (0.035)	0.299 (0.056)	0.410 (0.015)	0.360 (0.027)	0.292 (0.053)
	$\beta_1 = -1$	−0.823 (0.035)	−0.713 (0.086)	−0.586 (0.175)	−0.819 (0.034)	−0.711 (0.085)	−0.582 (0.176)
	$\beta_2 = 1$	0.816 (0.053)	0.707 (0.113)	0.577 (0.212)	0.816 (0.044)	0.707 (0.098)	0.579 (0.195)
Solomon-Cox	$\beta_0 = 0.5$	0.438 (0.020)	0.366 (0.048)	0.305 (0.081)	0.431 (0.014)	0.371 (0.031)	0.295 (0.069)
	$\beta_1 = -1$	−0.962 (0.006)	−0.905 (0.013)	−0.814 (0.038)	−0.958 (0.004)	−0.905 (0.011)	−0.811 (0.037)
	$\beta_2 = 1$	1.026 (0.031)	1.028 (0.060)	0.967 (0.096)	1.027 (0.017)	1.032 (0.028)	0.975 (0.053)
	ρ	0.327 (0.003)	0.512 (0.005)	0.716 (0.007)	0.328 (0.002)	0.516 (0.002)	0.719 (0.005)
Standard ML	$\beta_0 = 0.5$	0.508 (0.017)	0.500 (0.033)	0.525 (0.088)	0.501 (0.010)	0.505 (0.016)	0.514 (0.052)
	$\beta_1 = -1$	−1.005 (0.005)	−0.998 (0.005)	−0.991 (0.006)	−1.001 (0.002)	−0.998 (0.003)	−0.986 (0.003)
	$\beta_2 = 1$	0.995 (0.029)	0.988 (0.067)	0.974 (0.194)	0.997 (0.015)	0.993 (0.003)	0.974 (0.109)
	ρ	0.328 (0.003)	0.484 (0.003)	0.626 (0.004)	0.329 (0.001)	0.484 (0.001)	0.619 (0.003)
Gibbs Sampler	$\beta_0 = 0.5$	0.511 (0.017)	0.502 (0.028)	0.513 (0.043)	0.503 (0.009)	0.508 (0.014)	0.504 (0.027)
	$\beta_1 = -1$	−1.011 (0.005)	−1.004 (0.005)	−1.010 (0.007)	−1.004 (0.003)	−1.005 (0.003)	−1.007 (0.003)
	$\beta_2 = 1$	1.003 (0.029)	0.997 (0.054)	0.994 (0.096)	1.000 (0.015)	0.999 (0.024)	1.000 (0.052)
	ρ	0.347 (0.015)	0.505 (0.041)	0.671 (0.129)	0.339 (0.007)	0.503 (0.021)	0.669 (0.077)

[1] Average parameter estimates over 1000 Monte Carlo repetitions with Mean Squared Errors in parentheses.