MONASH UNIVERSITY

AUSTRALIA

IMPROVED ESTIMATION PROCEDURES FOR NONLINEAR PANEL DATA MODELS

Offer Lieberman and Lászlo Mátyás

Working Paper No. 15/94

July 1994

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0756 6

IMPROVED ESTIMATION PROCEDURES FOR NONLINEAR PANEL DATA MODELS

Offer Lieberman and László Mátyás

Working Paper No. 15/94

July 1994

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

Improved Estimation Procedures for Nonlinear Panel Data Models

Offer Lieberman
University of Bristol
Department of Economics

László Mátyás
Monash University
Department of Econometrics
and
Budapest University of Economics

Abstract: The paper proposes two different estimation procedures for nonlinear panel data models with a general parametric heterogeneity distribution. Using small–sigma and the Laplace approximation, easily computable analytical solutions to the marginal likelihood are presented.

Key words: Panel data, Nonlinear models, Laplace approximation, Solomon–Cox method, Probit model, Duration model.

July 1994

Improved Estimation Procedures for Nonlinear Panel Data Models

Offer Lieberman* and László Mátyás**

* University of Bristol

** Monash University, Australia and Budapest University of Economics

1. Introduction[1]

The use of panel data has become increasingly popular in econometrics over the last two decades. While the early theoretical and applied research focused on linear models, much effort has been made recently to establish appropriate estimation and hypothesis testing procedures capable of dealing with nonlinear panel data models (*Mátyás and Sevestre* [1992]). Unfortunately most results relating to linear models are special and cannot be extended to models having nonlinear specifications.

When the data heterogeneity is considered to be observable, the fixed effects approach is appropriate. In this case, generally, it is not possible to separate the estimation of the structural coefficients from those of the specific effects, and we may face an incidental–parameter (*Hsiao* [1986], [1992]). On the other hand, when the data heterogeneity is believed to be non–observable a random effects approach is suitable. The traditional approach in this case involves two simplifying assumptions. The first sets the conditional mean of the dependent variable to be linear, while the second specifies a restrictive parametric form for the distribution of the specific effects. As the log likelihood of the observations involves integration with respect to the heterogeneity distribution, these two assumptions, and in particular the second one, have been viewed as necessary for the maximization process to be operational. See, for example, *Stiratelli, Laird, and Ware* [1984, p. 964] and *Gourieroux* [1992, pp. 213–222].

Procedures for the estimation of a large class of models involving specific heterogeneity distributions are widely available. *Anderson and Aitkin* [1985], *Im and Gianola* [1988] and others applied Gaussian quadrature to evaluate integrals in panel logit and probit models with normal random effects, whereas *Waldman* [1985] used

[1] Computing assistance by Mark Harris is kindly aknowledged.

this routine for the estimation of duration models. *Kiefer* [1983] developed a series expansion to the same type of integral arising in labour market duration models. *Buttler and Moffit* [1984] reduced a multivariate normal integral into a univariate one for the panel probit model. *Schall* [1991] designed an algorithm for the estimation of generalised linear models with random effects obeying relatively weak assumptions. To date though, there does not seem to exist any analytical solution for the maximum likelihood estimator in the framework of a general nonlinear panel data model with random effects having a general parametric distribution.

Given the restrictions discussed above, we propose two different estimation procedures for panel data models exhibiting, potentially, both nonlinearity in the exogenous variables and/or the parameters of interest, and a general parametric form for the heterogeneity distribution. The first method is a generalisation of *Solomon and Cox's* [1992] technique, giving a small variance approximation to the marginal likelihood. Although this approximation is appealing in its simplicity and computational convenience, a different technique is required if the variance of the specific effects is known to be large. This provides a motivation for the development of a Laplace type approximation to the marginal likelihood. This approach has been successfully applied recently in a variety of statistical problems, notably, *Tierney and Kadane* [1986], *Tierney et al.* [1989 a.b.], *Wolfinger* [1993] and *Lieberman* [1995]. The Laplace approximation has, in general, a relative error of order $O(T^{-1})$.

In Section 2 of the paper we introduce the general problem and present the two approximations leading to computable formulae. The relative advantages in adopting either procedures are discussed. Some typical econometric examples are provided in Section 3. In Section 4 we apply the two methods to a model of youth unemployment in Australia, and analyse their properties through a Monte Carlo simulation. Concluding remarks are given in Section 5 and proofs are provided in the Appendix.

2. Integrating out the individual effects

We consider an array of panel data specifications in which the observations of the response (dependent) variable y_{it} are independent, each with conditional density

$$f(y_{it} \mid x_{it}, \mu_i; \beta), \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T,$$

where x_{it} and β are both $(k \times 1)$ and fixed. We assume that the heterogeneity factors (individual effects) μ_i, are continuous i.i.d. random varables, each with a density

$$g(\mu_i; \alpha), \quad \mu_i \in \mathbb{R} \qquad i = 1, \ldots, N.$$

Interest lies in estimating the parameter vector β. The conventional approach is to maximize the log of the marginal likelihood

$$\begin{aligned} \ell_M &= \sum_{i=1}^{N} \log \int g(\mu_i; \alpha) \left[\prod_{t=1}^{T} f(y_{it} \mid x_{it}, \mu_i; \beta) \right] d\mu_i \\ &= \sum_{i=1}^{N} \log L_i, \quad \text{say,} \end{aligned} \tag{1}$$

with respect to β. The obvious difficulty in making (1) operational, has so far led to strict parametric settings for $f(\cdot)$ and $g(\cdot)$. To maximize (1) under sufficiently general parametric assumptions, we consider two approaches.

2.1 Small–sigma asymptotics

The method proposed here is a generalisation of the procedure of *Solomon and Cox* [1992], who assumed the μ_i's i.i.d. normal. Put

$$\ell_i(\mu_i, \theta) = \ell_{ci} + \ell_{\mu i}\,, \tag{2}$$

with $\theta = (\beta : \alpha)$,

$$\ell_{ci} = \sum_{t=1}^{T} \log f(y_{it} \mid x_{it}, \mu_i; \beta) \tag{3}$$

and

$$\ell_{\mu i} = \log g(\mu_i; \alpha). \tag{4}$$

Write

$$\ell_i^{(r)} = \ell_{ci}^{(r)} + \ell_{\mu i}^{(r)} = [\partial^r \ell_i / \partial \mu_i^r]_{\mu_i = \xi}\,, \qquad r \geq 1\,, \tag{5}$$

where $\xi = E(\mu_i)\ \forall i$ and the arguments have been suppressed for brevity. We denote the variance of μ_i by σ^2. Given the setup so far, the small variance approximation to ℓ_M is

$$\ell_M = \sum_{i=1}^{N} \log \left\{ \left(\frac{-2\pi}{\ell_i^{(2)}} \right)^{\frac{1}{2}} \left[\prod_{t=1}^{T} f(y_{it} \mid x_{it}, \xi; \beta) \right] g(\xi; \alpha) \exp \left[\frac{-\ell_i^{(1)2}}{2\ell_i^{(2)}} \right] [1 + O(\sigma^4)] \right\}. \tag{6}$$

Proof, regularity conditions and the first correction factor in the expansion are provided in the Appendix. We note that if $\mu_i \sim N(0, \sigma^2)$, then $g(\xi; \sigma^2) = (2\pi\sigma^2)^{-1/2}$, $\ell_{\mu i}^{(1)} = 0$, $\ell_{\mu i}^{(2)} = -\frac{1}{\sigma^2}$ and so (6) collapses to

$$\ell_M = \sum_{i=1}^{N} \log\left\{ \left[\prod_{t=1}^{T} f(y_{it} \mid x_{it}, \xi; \beta) \right] \left[1 - \sigma^2 \ell_{ci}^{(2)} \right]^{-1/2} \times \exp\left[\frac{\sigma^2 \ell_{ci}^{(1)2}}{2(1 - \sigma^2 \ell_{ci}^{(2)})} \right] \left[1 + O(\sigma^4) \right] \right\}, \tag{7}$$

which, apart from the obvious difference in notation, agrees with the leading term of *Solomon and Cox's* [1992, equations (13) and (15)] formula. The error of this approximation is $O(\sigma^4)$. For any given $g(\cdot)$ and $f(\cdot)$, the maximization of (6) with respect to the parameters of interest involves, basically, a standard nonlinear optimization routine.

2.2 Laplace approximation

A second approach exploits the behaviour of the integrand of

$$L_i = \int \exp\{\ell_i(\mu_i, \theta)\}\, d\mu_i\,, \tag{8}$$

as $T \to \infty$. From (2)–(4) we can see that for fixed σ^2, $\ell_i(\mu_i, \theta) = O(T)$. This means that when T is large, $\exp\{\ell_i(\mu_i, \theta)\}$ attains a sharp peak at $\widehat{\mu}_{i_\theta}$, the point which maximizes $\ell_i(\mu_i, \theta)$ for fixed θ. Utilizing this feature, the Laplace approximation to L_i is

$$\left[\frac{-2\pi}{\widehat{\ell}_i^{(2)}} \right]^{\frac{1}{2}} \exp(\widehat{\ell}_i) \left\{ 1 + O(T^{-1}) \right\}, \tag{9}$$

where $\widehat{\ell}_i^{(r)}$ is as defined in (5), except that it is evaluated at $(\widehat{\mu}_{i_\theta}, \theta)$. The existence and uniqueness of $\widehat{\mu}_{i_\theta}$ are both implicitely assumed. For a readable account on the method of Laplace, the reader is referred to *Barndorff-Nielsen and Cox* [1989, Ch.3]. To apply the approximation, we fix θ at some initial value and solve

$$\widehat{\ell}_i^{(1)}(\widehat{\mu}_{i_\theta}, \theta) = 0, \qquad i = 1, \ldots N. \tag{10}$$

or in full,

$$L_i = \left[1 - \sigma^2 \widehat{\ell}_{ci}^{(2)}\right]^{-\frac{1}{2}} \exp\left[-\frac{\widehat{\mu}_{i_\theta}}{2\sigma^2} + \widehat{\ell}_{ci}\right] \left[1 + O(\sigma^2)\right] .$$

In comparison, the approximation (6) has a superior error of order $O(\sigma^4)$, but it embodies in it no direct mechanism for which its accuracy is improved as $T \to \infty$.

From a computational view point, the Laplace approximation has two chief drawbacks. The first is the requirement of evaluation of third order derivatives in (12), which for certain problems, may be a cumbersome task. The application of the method effectively necessitates iteration between (10) and a numerical maximization of (11). Evidently, the computation will be slow when N is large. This constitutes the second difficulty. However, given *Solomon and Cox's* [1992, Sec.4] findings, the Laplace approximation should be more reliable when σ is large. As analytical alternatives to the methods suggested exist only in some special cases, the Laplace approximation remains an important and valuable tool.

3. Examples

3.1 Count data

The traditional specification for the number of events of a given type occurring at a given time period is

$$f(y_{it} \mid x_{it}, \mu_i; \beta) = (y_{it}!)^{-1} \left[\exp(x'_{it}\beta + \mu_i)\right]^{y_{it}} \times \exp\left[-\exp(x'_{it}\beta + \mu_i)\right] .$$

For illustration purposes only, we assume that

$$g(\mu_i; \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-\frac{\mu_i^2}{2\sigma^2}),$$

although it should be clear from the foregoing discussion that neither methods rely on normality. The log marginal likelihood is

$$\ell_M = \sum_{i=1}^{N} \log \int (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\frac{\mu_i^2}{2\sigma^2}) \left\{ \prod_{t=1}^{T} (y_{it}!)^{-1} \left[\exp(x'_{it}\beta + \mu_i)\right]^{y_{it}} \times \exp\left[-\exp(x'_{it}\beta + \mu_i)\right] \right\} d\mu_i. \quad (13)$$

In view of (1), (8) and (9), the Laplace approximation to ℓ_M is

$$\ell_M = \sum_{i=1}^{N} \log\left\{ \left[\frac{-2\pi}{\widehat{\ell}_i^{(2)}}\right]^{\frac{1}{2}} \exp(\widehat{\ell}_i)\,[1+O(T^{-1})]\right\}. \tag{11}$$

Differentiating (10) with respect to θ, we see that

$$\frac{\partial \widehat{\mu}_{i_\theta}}{\partial \theta} = -(\widehat{\ell}_i^{(2)})^{-1}\frac{\partial \widehat{\ell}_i^{(1)}}{\partial \theta}.$$

Using the last expression, the MLE of θ is defined by

$$\frac{\partial \ell_M}{\partial \theta} = \sum_{i=1}^{N}\left\{\frac{1}{2}(\widehat{\widehat{\ell}}_i^{(2)})^{-1}\left[(\widehat{\widehat{\ell}}_i^{(2)})^{-1}\widehat{\widehat{\ell}}_i^{(3)}\frac{\partial \widehat{\widehat{\ell}}_i^{(1)}}{\partial \theta} - \frac{\partial \widehat{\widehat{\ell}}_i^{(2)}}{\partial \theta}\right] - \widehat{\widehat{\ell}}_i^{(1)}(\widehat{\widehat{\ell}}_i^{(2)})^{-1}\frac{\partial \widehat{\widehat{\ell}}_i^{(1)}}{\partial \theta} + \frac{\partial \widehat{\widehat{\ell}}_i}{\partial \theta}\right\} = 0, \tag{12}$$

where the "double hat" indicates that the relevant functions are evaluated at $(\widehat{\mu}_{i_{\widehat{\theta}}}, \widehat{\theta})$. The current estimate, $\widehat{\theta}$, found in (12), is substituted into (10) for the updating of $\widehat{\mu}_{i_\theta}$. The iteration between (10) and (12) proceeds until convergence. A similar approach was proposed by *Wolfinger* [1993] who assumed both $f(\cdot)$ and $g(\cdot)$ to be normal and β to have a flat prior.

2.3 Discussion

The principal difference between the two techniques is in the point about which $\ell_i(\mu_i, \theta)$ is expanded. The first approach utilizes an expansion about $\mu_i = \xi$, whereas the second about $\mu_i = \widehat{\mu}_{i_\theta}$. While the approximation (6) is only expected to behave well for small values of σ, the Laplace approximation should be reliable when either σ is small, T large, or both. To illustrate this point analytically, suppose, without loss of generality, that

$$\ell_{\mu i} = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{\mu_i^2}{2\sigma^2}.$$

Then for fixed T, as σ tends to zero, $\ell_i = O(\sigma^{-2})$ and the Laplace approximation to L_i is

$$L_i = \left[\frac{-2\pi}{\widehat{\ell}_i^{(2)}}\right]^{\frac{1}{2}} \exp(\widehat{\ell}_i)[1+O(\sigma^2)],$$

A closed form solution to the integral in (13) is not available. To apply (7) to (13), we first see that

$$\ell_{ci}^{(1)} = \frac{\partial \sum_{t=1}^{T} \log f(y_{it} \mid x_{it}, \mu_i; \beta)}{\partial \mu_i}|_{\mu_i=0} = \sum_{t=1}^{T} [y_{it} - \exp(x_{it}'\beta)] \tag{14}$$

and

$$\ell_{ci}^{(2)} = -\sum_{t=1}^{T} \exp(x_{it}'\beta)\,. \tag{15}$$

Substituting (14) and (15) into (7), we obtain

$$\ell_M = \sum_{i=1}^{N} \log\Bigg\{ \Bigg\{ \prod_{t=1}^{T} (y_{it}!)^{-1} \left[\exp(x_{it}'\beta)\right]^{y_{it}} \exp\left[-\exp(x_{it}'\beta)\right] \Bigg\} \times$$

$$\left[1 + \sigma^2 \sum_{t=1}^{T} \exp(x_{it}'\beta)\right]^{-1/2} \exp\left\{ \frac{\sigma^2 \left[\sum_{t=1}^{T} (y_{it} - \exp(x_{it}'\beta))\right]^2}{2(1 + \sigma^2 \sum_{t=1}^{T} \exp(x_{it}'\beta))} \right\} \{1 + O(\sigma^4)\} \Bigg\}\,.$$

For the Laplace approximation, we set $\alpha = \sigma^2$ and

$$\ell_i(\mu_i, \theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{\mu_i^2}{2\sigma^2}$$
$$+ \sum_{t=1}^{T} \Big\{ -\log(y_{it}!) + y_{it}(x_{it}'\beta + \mu_i) - \exp(x_{it}'\beta + \mu_i) \Big\}\,.$$

The $\widehat{\mu}_{i_\theta}$ defining equation is

$$\widehat{\ell}_i^{(1)}(\widehat{\mu}_{i_\theta}, \theta) = -\frac{\widehat{\mu}_{i_\theta}}{\sigma^2} + \sum_{t=1}^{T} [y_{it} - \exp(x_{it}'\beta + \widehat{\mu}_{i_\theta})] = 0. \tag{16}$$

Further,

$$\widehat{\ell}_i^{(2)} = -\frac{1}{\sigma^2} - \sum_{t=1}^{T} \exp(x_{it}'\beta + \widehat{\mu}_{i_\theta}).$$

The Laplace approximation to ℓ_M is reduced to

$$\ell_M = \sum_{t=1}^{T} \log\Bigg\{ \left[1 + \sigma^2 \sum_{t=1}^{T} \exp(x_{it}'\beta + \widehat{\mu}_{i_\theta})\right]^{-1/2} \exp\Big[-\frac{\widehat{\mu}_{i_\theta}^2}{2\sigma^2} + \sum_{t=1}^{T}[-\log(y_{it}!)$$
$$+ y_{it}(x_{it}'\beta + \widehat{\mu}_{i_\theta}) - \exp(x_{it}'\beta + \widehat{\mu}_{i_\theta})]\Big] \left[1 + O(T^{-1}\right] \Bigg\}\,. \tag{17}$$

The execution of the method requires an initial solution to (16), followed by a maximization of (17). The current estimates of $\widehat{\beta}$ and $\widehat{\sigma}^2$ are substituted into (16) to update $\widehat{\mu}_{i_\theta}$. We iterate between (16) and (17) until convergence.

3.2 Duration models

Duration models with heterogeneity have been considered by, among others, *Gourieroux* [1992], *Keifer* [1983], *Lancaster* [1979]. In most cases, the heterogeneity distribution is chosen on pure mathematical grounds, rather than economic considerations. The endogenous variables in this context is positive. One formulation, considered by *Keifer* [1983], is based on the exponential model

$$f(y_{it} \mid x_{it}, \mu_i; \beta) = \begin{cases} \exp(x'_{it}\beta + \mu_i) \exp\{-\exp(x'_{it}\beta + \mu_i) y_{it}\} & \text{if } y_{it} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

with

$$g(\mu_i; \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(\frac{-\mu_i^2}{2\sigma^2}).$$

Under normality, Gaussian Quadrature or Kiefer's method can be used, but not so under a more general specification of the mixing distribution.

The approximation (7) is easily verified to yield

$$\ell_M = \sum_{i=1}^{N} \log\Bigg\{ \left[\prod_{t=1}^{T} \exp(x'_{it}\beta) \exp[-y_{it} \exp(x'_{it}\beta)] \right] \left[1 + \sigma^2 \sum_{t=1}^{T} y_{it} \exp(x'_{it}\beta) \right]^{-\frac{1}{2}} \times$$
$$\exp\left[\frac{\sigma^2 [\sum_{t=1}^{T} [1 - y_{it} \exp(x'_{it}\beta)]]^2}{2[1 + \sigma^2 \sum_{t=1}^{T} y_{it} \exp(x'_{it}\beta)]} \right] [1 + O(\sigma^4)] \Bigg\}.$$

The Laplace approximation results in

$$\ell_M = \sum_{i=1}^{N} \log\Bigg\{ \left[1 + \sigma^2 \sum_{t=1}^{T} [y_{it} \exp(x'_{it}\beta + \widehat{\mu}_{i_\theta})] \right]^{-\frac{1}{2}} \times$$
$$\exp\left[-\frac{\widehat{\mu}_{i_\theta}^2}{2\sigma^2} + \sum_{t=1}^{T} [x_{it}\beta + \widehat{\mu}_{i_\theta} - y_{it} \exp(x'_{it}\beta + \widehat{\mu}_{i_\theta})] \right] \times \quad (18)$$
$$[1 + O(T^{-1})] \Bigg\},$$

where $\widehat{\mu}_{i_\theta}$ satisfies

$$\sum_{t=1}^{T}[1 - y_{it}\exp(x'_{it}\beta + \widehat{\mu}_{i_\theta})] - \frac{\widehat{\mu}_{i_\theta}}{2\sigma^2} = 0\,. \qquad (19)$$

As in the previous example, the MLE's of β and σ^2 are achieved by the iterative solution to (19) and the maximization of (18).

3.3 The Probit model

The Probit Model is one of the most popular modelling approaches in discrete choice analysis. The difficulties in the maximization of ℓ_M even under normality of the random effects are well known. It is assumed that

$$y_{it} = \begin{cases} 1 & \text{if } x'_{it}\beta + \mu_i + u_{it} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

with $\mu_i \sim N(0, \sigma^2)$, $u_{it} \sim N(0,1)$, μ_i and u_{it} are independent $\forall i$ and $\forall t$, $i = 1, \ldots, N$ $t = 1, \ldots, T$. The log marginal likelihood is

$$\ell_M = \sum_{i=1}^{N} \log \int (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{\mu_i^2}{2\sigma^2}\right) \left[\prod_{t=1}^{T} [\Phi(x'_{it}\beta + \mu_i)]^{y_{it}} \times \right.$$
$$\left. [1 - \Phi(x'_{it}\beta + \mu i)]^{1-y_{it}}\right] d\mu_i\,.$$

To apply (7), we put

$$\prod_{t=1}^{T} f(y_{it} \mid x_{it}, 0, \beta) = \prod_{t=1}^{T} \left[\Phi^{y_{it}}(1-\Phi)^{1-y_{it}}\right] \qquad (20)$$

$$\ell_{ci}^{(1)} = \sum_{t=1}^{T} \frac{y_{it} - \Phi}{\Phi(1-\Phi)}\phi \qquad (21)$$

$$\ell_{ci}^{(2)} = -\sum_{t=1}^{T}\left\{ y_{it}\left[\Phi^{-2}\phi^2 + \Phi^{-1}\phi x'_{it}\beta\right] + (1-y_{it})\left[(1-\Phi)^{-2}\phi^2 - (1-\Phi)^{-1}\phi x'_{it}\beta\right]\right\}, \qquad (22)$$

where Φ and ϕ are the standard normal cdf and pdf evaluated at $x'_{it}\beta$. Equations (20)- (22) are the relevant components in (7) for this problem.

For the Laplace approximation, we set again $\alpha = \sigma^2$ and

$$\widehat{\ell}_i = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{\widehat{\mu}_{i_\theta}^2}{2\sigma^2} + \sum_{t=1}^{T}\left[y_{it}\log\widehat{\Phi} + (1-y_{it})\log(1-\widehat{\Phi})\right], \tag{23}$$

where $\widehat{\mu}_{i_\theta}$ is defined by

$$-\frac{\widehat{\mu}_{i_\theta}}{\sigma^2} + \sum_{t=1}^{T}\frac{y_{it}-\widehat{\Phi}}{\widehat{\Phi}(1-\widehat{\Phi})}\widehat{\phi} = 0,$$

and $\widehat{\Phi}, \widehat{\phi}$ are evaluated at $x'_{it}\beta + \widehat{\mu}_{i_\theta}$. The second derivative of $\ell_i(\mu_i, \theta)$ in this setting is

$$\begin{aligned}\widehat{\ell}_i^{(2)} = -\frac{1}{\sigma^2} - \sum_{t=1}^{T}\Big\{ & y_{it}\left[\widehat{\Phi}^{-2}\widehat{\phi}^2 + (x'_{it}\beta + \widehat{\mu}_{i_\theta})\widehat{\Phi}^{-1}\widehat{\phi}\right] \\ & + (1-y_{it})\left[(1-\widehat{\Phi})^{-2}\widehat{\phi}^2 - (1-\widehat{\Phi})^{-1}(x'_{it}\beta + \widehat{\mu}_{i_\theta}\widehat{\phi}\right]\Big\}.\end{aligned} \tag{24}$$

The Laplace approximation to ℓ_M is obtained upon substitution of (23) and (24) into (11).

4. Application and numerical evaluation

Harris [1994] analysed youth unemployment in Autralia for the period 1985–88 using the Australian Labour Force Survey. Given the heterogeneity of the data he estimated a Probit model on four sub-groups of the original data set (male and female, high and low education groups). The dependent variable of the model was set to zero if the individual for given period was employed and to unity if she/he was unemployed. The model was estimated by Limdep (see *Greene* [1992]) which uses the Gaussian quadrature proposed by *Buttler and Moffit* [1982] to evaluate the necessary integral.

Table 1 contains the original results produced by Limdep for the female low education group and our parameter estimates for the same model and data using the small–sigma and Laplace approximations.[2]

The fact that the parameter estimates in Table 1 are similar but not identical, provides a motivation for further investigation of the accuracy of the analytical

[2] Similar results were obtained for the remaining three groups.

Table 1:
Parameter Estimates for Unemployment Model
($N = 566$, $T = 4$)

Explanatory variable	Limdep	Small–sigma	Laplace
Constant	1.46	1.93	1.52
Age ($\exp[-0.1 \times \text{age}]$)	-6.05	-7.14	-7.35
Married/not married	-0.63	-1.06	-1.27
Education $\leq$ 10 years	0.31	0.39	0.36
Education 11 years	0.62	0.82	0.51
Working partner	0.87	1.45	1.64
Reservation wage	-0.40	-0.63	-0.58
Disabled / not	-0.31	-0.52	-0.65
Number of chidren	-0.20	-0.30	-0.40
Rho ($\sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_u^2)$)	0.78	0.59	0.44
Max log–likelihood	-594.3	-620.4	-477.0

approximations, using Monte Carlo simulation. For the artificial data generation we used the following specification:

$$
\begin{aligned}
y_{it}^* &= \beta_0 + \beta_1 x_{it}^{(1)} + \beta x_{it}^{(2)} + \alpha_i + \varepsilon_{it} \\
x_{it}^{(j)} &= x_{i,t-1}^{(j)} + v_{it}^{(j)} \qquad j = 1, 2 \\
v_{it} &\sim \text{Uniform}[-0.5,\ 0.5] \qquad \varepsilon_{it} \sim N(0,\ 1) \\
y_{it} &= \begin{cases} 1 & \text{if } y_{it}^* > \text{its mean} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

and $\beta_0 = \beta_1 = \beta_2 = 0.5$, $N = 100$, $T = 5$ and we performed 1000 Monte Carlo replications for the small-sigma and 500 for the Laplace approximation.[3] In the Laplace iteration procedure we used the OLS estimates as starting values for β_i and 1 for σ. Results are summarized in Table 2.

The small–sigma approximation works generally well. It appears to be reliable and accurate even for moderate to large values of sigma. The Laplace procedure works, in general, nicely as well, but for large N it can be slow and, as an additional

[3] The Laplace approximation is much more time consuming, so it was not possible to run the Monte Carlo experiment over 1000 replications.

nuisance, it is very sensitive to the starting values of the iteration procedure. Our results suggest that it is advantageous to use the Laplace approximation only when sigma is relatively large.

5. Conclusion

The conventional use of nonlinear econometric panel data models with random effects requires generally a restrictive set of assumptions to handle the mathematical complexity of the problem. In this paper we have suggested two tractable procedures for the estimation of such models under general parametric and functional settings. The small–sigma approximation is particularly convenient and reliable when the variance of the heterogeneity distribution is not too large. The Laplace approximation can be regarded as either a large T or a small–sigma approximation and is a viable alternative to the small–sigma approximation. Given the non-existence of alternative analytical solutions to the marginal likelihood when the heterogeneity distribution is nonnormal, both techniques enable theorists and practitioners to explore these models under much more realistic parametric assumptions.

Table 2:

Monte Carlo Simulation Results

($N = 100$, $T = 5$)

	parameter	average $\widehat{\beta}$	mean bias	std. dev. of bias
Small–sigma				
$\sigma_\alpha^2 = 0.01$	β_1	0.510	0.099	0.073
$\sigma_\alpha^2 = 0.01$	β_2	0.508	0.099	0.075
$\sigma_\alpha^2 = 0.1$	β_1	0.509	0.109	0.081
$\sigma_\alpha^2 = 0.1$	β_2	0.500	0.105	0.080
$\sigma_\alpha^2 = 0.2$	β_1	0.492	0.113	0.090
$\sigma_\alpha^2 = 0.2$	β_2	0.500	0.112	0.087
$\sigma_\alpha^2 = 0.3$	β_1	0.492	0.123	0.089
$\sigma_\alpha^2 = 0.3$	β_2	0.493	0.111	0.084
$\sigma_\alpha^2 = 1.0$	β_1	0.447	0.152	0.116
$\sigma_\alpha^2 = 1.0$	β_2	0.450	0.150	0.117
Laplace				
$\sigma_\alpha^2 = 0.01$	β_1	0.439	0.171	0.130
$\sigma_\alpha^2 = 0.01$	β_2	0.466	0.176	0.136
$\sigma_\alpha^2 = 0.1$	β_1	0.490	0.232	0.189
$\sigma_\alpha^2 = 0.1$	β_2	0.487	0.206	0.171
$\sigma_\alpha^2 = 1.0$	β_1	0.529	0.279	0.234
$\sigma_\alpha^2 = 1.0$	β_2	0.554	0.327	0.258

Due to the setup of the experiment the regression constant β_0 is not identified.

Appendix

Here we prove (5). The integral in (1) may be written as

$$L_i = \int \exp\left[\ell_i(\mu_i, \theta)\right] d\mu_i \, .$$

Expanding $\ell_i(\mu_i, \theta)$ about $\mu_i = \xi$ and suppressing the arguments and the index i temporarily, we obtain

$$L = \sigma e^{\ell} \int \exp\left[\sigma \ell^{(1)} \mu^* + \frac{1}{2}\sigma^2 \ell^{(2)} \mu^{*2}\right] \left[1 + \frac{1}{6}\sigma^3 \ell^{(3)} \mu^{*3} + \frac{1}{24}\sigma^4 \ell^{(4)} \mu^{*4} + \ldots\right] d\mu^* ,$$

where $\mu^* = (\mu - \xi)/\sigma$. Transforming $-\sigma^2 \ell^{(2)} \mu^{*2} = Z^2$, the last integral becomes

$$L = \left[-\ell^{(2)}\right]^{-1/2} \exp\left[\ell - \frac{\ell^{(1)2}}{2\ell^{(2)}}\right] \int \exp\left[-\frac{1}{2}\left[Z - \frac{\ell^{(1)2}}{\sqrt{-\ell^{(2)}}}\right]^2\right] \times$$

$$\left[1 + \frac{\ell^{(3)}}{6[-\ell^{(2)}]^{3/2}} Z^3 + \frac{\ell^{(4)}}{24\ell^{(2)2}} Z^4 + \ldots\right] dZ \, .$$

This yields the expansion

$$L = \left[-\frac{2\pi}{\ell^{(2)}}\right]^{\frac{1}{2}} \exp\left\{\ell - \frac{\ell^{(1)2}}{2\ell^{(2)}}\right\} \exp\left\{1 + \frac{\ell^{(3)}}{6[-\ell^{(2)}]^{3/2}} \left[\frac{3\ell^{(1)}}{\sqrt{-\ell^{(2)}}} + \frac{\ell^{(1)3}}{[-\ell^{(2)}]^{3/2}}\right]\right.$$

$$\left. + \frac{\ell^{(4)}}{24\ell^{(2)2}} \left[3 - \frac{6\ell^{(1)2}}{\ell^{(2)}} + \frac{\ell^{(1)4}}{\ell^{(2)2}}\right] + \ldots\right\} .$$

To establish the order of the error of the approximation we impose the following conditions:

condition 1; $\sigma^2 \ell^{(2)}_{\mu_i} = O(1)$ as $\sigma^2 \to 0$, $\forall i$, $i = 1, \ldots, N$.

condition 2; $\ell^{(r)}_{\mu_i} = O(1)$ as $\sigma^2 \to 0$, $\forall i$, $i = 1, \ldots, N$, $r = 1, 3, 4, \ldots$.

We note that conditions 1 and 2 are both satisfied when, for example, $\mu_i \sim N(0, \sigma^2)$. Under these conditions, the desired expansion is

$$L = \left[-\frac{2\pi}{\ell^{(2)}}\right]^{\frac{1}{2}} \exp\left[\ell - \frac{\ell^{(1)2}}{2\ell^{(2)}}\right] \left[1 + \frac{4\ell^{(3)}\ell^{(1)} + \ell^{(4)}}{8\ell^{(2)2}} + o(\sigma^4)\right] .$$

The correction factor appearing in the expansion is of order $O(\sigma^4)$ under conditions 1 and 2. The error of neglecting further terms is $o(\sigma^4)$. On substitution of the leading term of the last expression into (1) and rearranging, the result (6) follows.

References

Anderson, D. A., and Aitkin, M. [1985]: Variance components models with binary response: interviewer variability; *Journal of the Royal Statistical Society B*, 47, pp. 203–210.

Barndorff-Nielsen, O. E., and Cox, D. R. [1989]: *Asymptotic techniques for use in statistics*; Chapman and Hall, London.

Buttler, J. S., and Moffitt, R. [1982]: A computationally efficient quadrature procedure for the one-factor multinomial probit model; *Econometrica*, 50, pp. 761–764.

Gourieroux, C. [1992]: *Introduction to nonlinear models*; in Matyas and Sevestre (eds.): The Econometrics of Panel Data, Kluwer Academic Publishers, Dordrecht, 1992.

Greene, W. [1992]: *Limdep, Version 6.0: User's manual*; Bellport, N.Y. Econometric Software, 1992.

Harris M. [1994]: *Modelling the probability of youth unemployment in Australia: 1985–1988*; Monash University, Dept. of Econometrics, Working Paper 12/94.

Hsiao C. [1992]: *Logit and probit models*; in Matyas and Sevestre (eds.): The Econometrics of Panel Data, Kluwer Academic Publishers, Dordrecht, 1992.

Hsiao C. [1986]: *Analysis of panel data*; Cambridge University Press, Cambridge, 1986.

Im, S., and Gianola, D. [1988]: Mixed models for binomial data with an application to lamb mortality; *Applied Statistics*, 37, pp. 196–204.

Lancaster, T. [1979]: Econometric methods for the duration of unemployment; *Econometrica*, 47, pp. 939–956.

Lieberman, O. [1995]: A Laplace approximation to the moments of a ratio of quadratic forms; *Biometrika*, (forthcoming).

Kiefer, N. M. [1983]: An integral occurring in duration models with heterogeneity; *Economics Letters*, 11, pp. 251–256.

Mátyás L. and Sevestre, P. [1992]: *The econometrics of panel data*; Kluwer Academic Publishers, Dordrecht, 1992.

Schall, R. [1991]: Estimation in generalized linear models with random effects; *Biometrika*, 78, pp. 719–727.

Solomon, P. J., and Cox, D. R. [1992]: Nonlinear component of variance models; *Biometrika*, 79, pp. 1–11.

Stiratelli, R., Laird, N. and Ware, J. H. [1984]: Random effects models for serial observations with binary response; *Biometrika*, 40, pp. 961–971.

Tierney, L., and Kadane, J. B. [1986]: Accurate approximations for posterior moments and marginal densities; *Journal of the American Statistical Association*, 81, pp. 82–86.

Tierney, L., Kass, R. E., and Kadane, J. B. [1989]: Fully exponential Laplace approximations to expectations and variances of nonpositive functions; *Journal of the American Statistical Association*, 84, pp. 710–716.

Tierney, L., Kass, R. E., and Kadane, J. B. [1989]: Approximate marginal densities of nonlinear functions; *Biometrika*, 76, pp. 425–433.

Waldman, D. M. [1985]: Computation in duration models with heterogeneity; *Journal of Econometrics*, 28, pp. 127–134.

Wolfinger, R. [1993]: Laplace's approximation for nonlinear mixed models; *Biometrika*, 80, pp. 791–795.