Two-Step Inference in Dynamic Non-Linear Panel Data Models

Chris D. Orme*
School of Economic Studies
University of Manchester
U.K.

June 2001

Abstract

This paper develops a simple two-step test procedure for assessing the so-called initial conditions problem in parametric non-linear dynamic panel data models. Asymptotic local theory suggests that it can also provide useful inferences for certain parameters of interest.

1 Introduction

This paper is concerned with asymptotic inference in a variety of (now common) parametric non-linear dynamic panel data models. Such models includes random effects probit, tobit, multiple spell duration models or count data models, where the number of cross-section (individual) units, N, is large but the number observations over time, T, is small. It is well known that the standard random effects Maximum Likelihood Estimator (MLE) will be inconsistent if the inherent initial condition problem is ignored; see, for example, Heckman (1981a,b). Even when the initial condition problem is 'weak' (i.e., the correlation between the initial observation and the random effect is low), consistent estimation routines are not generally available in existing econometrics/statistics packages. Therefore it may be useful, for those interested in pursuing parametric inference in such models, if a quick check were available for the presence, and possible impact on estimates, of the initial conditions problem.

*Tel: +44 161 275 4856. Fax: +44 161 275 4928. email: chris.orme@man.ac.uk

In view of this, and following recent work on inference using local approximations (see, for example, Chesher, 1991, 1997, 1998; Chesher and Santos-Silva, 1995; and, Gourieroux and Visser,1997), an artificial model is obtained under the assumption of a weak initial condition problem. This model can be estimated using procedures currently available in standard packages to provide: (i) a simple test for the presence of an endogenous initial condition; and, (ii) two-step 'bias-corrected' MLE's. At the very least, this procedure could provide a useful inferential tool *en route* to developing more robust, but more cumbersome, estimators such as those proposed by Arellano and Carrasco (1996) or Honoré and Kyriazidou (1996).

The plan of the paper is as follows: Section 2 introduces the notation and outlines the two-step procedure (derivation of the approximate artificial model is relegated to an Appendix). Evidence on the performance of the procedure is investigated using a Monte Carlo study and summary evidence is presented in Section 3. Section 4 concludes.

2 The Model and the Two-Step Procedure

2.1 The Model and Notation

The types of model under consideration are where the data consist of N independent observations, denoted $\mathbf{y}_i' = (y_{i1}, \ldots, y_{iT})$, $i = 1, \ldots, N$, on T dependent variables, where N is large and T is small. Conditional on strictly exogenous covariates $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$, the initial condition y_{i0}, and individual random effects u_i, $\mathbf{y}_i$ is modelled via a joint density, indexed by an unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. This density is the form

$$F_i(\mathbf{y}_i|u_i;\boldsymbol{\theta}) = \prod_{t=1}^{T} f\left(y_{it}; w_{it}\right), \tag{1}$$

where $w_{it} = m(\alpha y_{i,t-1}, \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma u_i)$ with $\boldsymbol{\theta}_1' = (\alpha, \boldsymbol{\beta}', \sigma)$; $\boldsymbol{\beta}'\mathbf{x}_{it}$ contains an intercept term, denoted β_1; and, $m(.,.)$ is some appropriate bivariate function. This specification includes censored normal linear models in which the latent (uncensored) variable is modelled as $y_{it}^* = \alpha y_{it-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma u_i + \theta_2 \varepsilon_{it}$ and $\sigma u_i + \theta_2 \varepsilon_{it}$ follows a strict error components structure; for example, a random effects probit/logit model ($\theta_2 \equiv 1$) or random effects tobit model. Some parametric models of multiple spell duration data can also be accommodated in this framework (see Lancaster (1990, p. 212)) as can the linear feedback count data model of Blundell, Griffith and Windmeijer (1999).

Such parametric models, in which strict exogeneity of $\mathbf{X}_i$ implies that its elements are uncorrelated with those of u_i, are distributionally restrictive but are nonetheless popular and are often used as the starting point in empirical work; see, for example, Vella and Verbeek (1998, 1999). Moreover, it is often assumed that the u_i are *iid* $N(0,1)$, $i = 1, ..., N$, which permits the use of Gaussian Quadrature to numerically integrate out this random effect. However, in general, account must be taken of the correlation between y_{i0} and u_i, otherwise "standard" random effects estimation is inconsistent. In the spirit of the discussion in Heckman (1981), such correlation is introduced by specifying a *conditional* model for y_{i0} as $h_i(y_{i0}|v_i; \boldsymbol{\lambda})$, indexed by an unknown parameter vector $\boldsymbol{\lambda}$, where (u, v) are standard bi-variate normal with correlation ρ. This conditional model is assumed to be of the form $k(y_{i0}; \boldsymbol{\gamma}'\mathbf{z}_i + \tau v_i)$, where $\boldsymbol{\lambda}$ includes $\boldsymbol{\gamma}$ and τ, yielding a *marginal* model for y_{i0} as

$$g_i(y_{i0}; \boldsymbol{\lambda}) = \int_{-\infty}^{\infty} k\left(y_{i0}; \boldsymbol{\gamma}'\mathbf{z}_i + \tau v\right)\phi(v)dv \tag{2}$$

in which $\mathbf{z}_i$ is a vector of exogenous covariates (or instruments), $\boldsymbol{\gamma}'\mathbf{z}_i$ is a regression function with intercept term, denoted γ_1 and $\phi(.)$ is the standard normal density function. For example, (i) if modelling y_{it} in a random effects probit model, y_{i0} would be modelled as a simple probit model with regression function $\boldsymbol{\gamma}'\mathbf{z}_i$ $(\tau = 1)$; (ii) Vella and Verbeek (1999) specify a random effects tobit model for y_{it} (weekly hours worked), whilst y_{i0} derives from a tobit model with regression function $\boldsymbol{\gamma}'\mathbf{z}_i$; (iii) for a random effects Poisson model, y_{i0} is (conditionally) distributed Poisson with mean $\mu_i = \exp(\boldsymbol{\gamma}'\mathbf{z}_i + \tau v)$ so that the marginal model for y_{i0} is an overdispersed Poisson variable with

$$g_i(y_{i0}; \boldsymbol{\lambda}) = \int_{-\infty}^{\infty} \frac{\mu_i^{y_{i0}}}{y_{i0}!}\exp(-\mu_i)\phi(v)dv, \qquad y_{i0} = 0, 1, 2, \ldots .$$

Writing $v = \rho u + \sqrt{1-\rho^2}\varepsilon$, it is then possible to establish the conditional distribution for y_{i0} given $u_i = u$ as

$$g_i(y_{i0}|u; \boldsymbol{\zeta}) = \int_{-\infty}^{\infty} k\left(y_{i0}; \boldsymbol{\gamma}'\mathbf{z}_i + \tau\rho u + \tau\sqrt{1-\rho^2}\varepsilon\right)\phi(\varepsilon)d\varepsilon \tag{3}$$

indexed by parameter vector $\boldsymbol{\zeta}' = (\boldsymbol{\lambda}, \rho)$. The marginal distribution for y_{i0} can be expressed as $g_i(y_{i0}; \boldsymbol{\lambda}) = \int_{-\infty}^{\infty} g_i(y_{i0}|u; \boldsymbol{\zeta})\phi(u)du$, which is independent of ρ. A typical likelihood contribution, derived from the conditional probability of observing the sequence of outcomes $\mathbf{y}_i$ given y_{i0} is therefore

$$\begin{aligned} P_i^c(\mathbf{y}_i|y_{i0}; \boldsymbol{\psi}) &= \frac{\int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u; \boldsymbol{\theta})g_i(y_{i0}|u; \boldsymbol{\zeta})\phi(u)du}{\int_{-\infty}^{\infty} g_i(y_{i0}|u; \boldsymbol{\zeta})\phi(u)du} \\ &= \frac{\int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u; \boldsymbol{\theta})g_i(y_{i0}|u; \boldsymbol{\zeta})\phi(u)du}{g_i(y_{i0}; \boldsymbol{\lambda})}. \end{aligned} \tag{4}$$

which is indexed by $\boldsymbol{\psi}' = (\boldsymbol{\theta}', \boldsymbol{\lambda}', \rho)$. Exogenous y_{i0} is implied by $\rho = 0$, in which case $g_i(y_{i0}|u;\boldsymbol{\zeta})|_{\rho=0} = g_i(y_{i0};\boldsymbol{\lambda})$ and (4) becomes $\int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du$.

Assuming (1) and (3) are correct, (4) provides the basis for the construction of a sample log-likelihood although, in general, standard packages do not accommodate such a model. However, if the correlation between y_{i0} and u_i is 'weak' (i.e., ρ is small), then an accessible two-step procedure can be derived by exploiting a locally equivalent alternative model specification which agrees with (4) to $O(\rho)$. Although, locally equivalent alternative methodology is often used to derive score-type tests (Godfrey (1988, pp.70-75)), in this context it not only affords the applied worker a simple check of $H_0 : \rho = 0$, it also provides bias corrected MLE's for small $\rho \neq 0$.

The intuition for the formal local approximation, given in the next section, can be explained as follows. Although $E[u] = 0$, $E[u|y_{i0}] \neq 0$ when $\rho \neq 0$ and this will cause inconsistency in MLE's derived from the model

$$\int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du,$$

which ignores the initial condition problem. However, the correlation between the unobservable and y_{i0} can be removed by writing $u = E[u|y_{i0}] + u^*$, so that $E[u^*|y_{i0}] = 0$ by construction, and considering the "model" $F_i^a(\mathbf{y}_i|u^*;\boldsymbol{\psi}) = \prod_{t=1}^{T} f(y_{it}; w_{it}^a)$, where $w_{it}^a = m(\alpha y_{i,t-1}, \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma E[u|y_{i0}] + \sigma u^*)$. The conditional expectation, which acts as a correction term, is given by

$$E[u|y_{i0}] = \frac{\int_{-\infty}^{\infty} u\, g_i(y_{i0}|u;\boldsymbol{\zeta})\phi(u)du}{\int_{-\infty}^{\infty} g_i(y_{i0}|u;\boldsymbol{\zeta})\phi(u)du}$$

which simplifies to $\tau\rho\dfrac{\partial \ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1}$, where $g_i(y_{i0};\boldsymbol{\lambda})$ is defined at (2). To see this, note that integration by parts yields

$$\begin{aligned}
\int_{-\infty}^{\infty} u\, g_i(y_{i0}|u;\boldsymbol{\zeta})\phi(u)du &= \int_{-\infty}^{\infty} \frac{dg_i(y_{i0}|u;\boldsymbol{\zeta})}{du}\phi(u)du \\
&= \tau\rho \int_{-\infty}^{\infty} \frac{\partial g_i(y_{i0}|u;\boldsymbol{\zeta})}{\partial\gamma_1}\phi(u)du \\
&= \tau\rho \frac{\partial}{\partial\gamma_1}\int_{-\infty}^{\infty} g_i(y_{i0}|u;\boldsymbol{\zeta})\phi(u)du \\
&= \tau\rho \frac{\partial g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1}.
\end{aligned}$$

Let $e_{i0} = \dfrac{\partial \ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1}$, which we shall call the generalised error from from the marginal model $g_i(y_{i0};\boldsymbol{\lambda})$, since it is proportional to $E[u|y_{i0}]$.

2.2 Local Approximations

Now consider an *approximate* (or artificial) conditional model defined by likelihood contributions of the form

$$P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = \int_{-\infty}^{\infty} F_i^a(\mathbf{y}_i|u;\boldsymbol{\psi})\phi(u)du, \tag{5}$$

where $F_i^a(\mathbf{y}_i|u) = \prod_{t=1}^{T} f\left(y_{it}; w_{it}^a\right)$, denotes an *augmented* model in which $w_{it}^a = m_1(\alpha y_{it-1}, \boldsymbol{\beta}'\mathbf{x}_{it} + \delta e_{i0} + \sigma u)$ with $\delta = \sigma\tau\rho$; i.e., $F^a(\mathbf{y}_i|u;\boldsymbol{\psi})$ is of the same form as $F_i(\mathbf{y}_i|u;\boldsymbol{\theta})$, defined at (1), except that the regressor set, $\mathbf{X}_i$, has been augmented with the artificial regressor $e_{i0} = \dfrac{\partial \ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial \gamma_1}$, which is indexed by the parameter $\boldsymbol{\lambda}$. It is easily verified that $P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})$ is a proper density (in that the implied probabilities are positive, for all possible $\mathbf{y}_i'$, and integrate/sum to unity). Furthermore, it is shown in the Appendix that $P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})$ is a locally equivalent alternative to $P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})$, since it has the same Taylor series expansion to $O(\rho)$; $P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) + o(\rho)$. It is this result and the following theoretical arguments which lead to the simple two-step test procedure and bias-corrected estimator.

Let $L^c(\boldsymbol{\psi})$ be the true log-likelihood derived from (4), maximised by $\hat{\boldsymbol{\psi}}$, with $L^a(\boldsymbol{\psi})$ the log-likelihood derived from (5), maximised by $\tilde{\boldsymbol{\psi}}$. If ρ is local to zero, $\rho = O(\frac{1}{\sqrt{N}})$, the application of standard asymptotic local theory reveals that $\tilde{\boldsymbol{\psi}} - \hat{\boldsymbol{\psi}} = O_p(N^{-1})$. This means that all standard inferences based on the approximate log-likelihood, $L^a(\boldsymbol{\psi})$, are asymptotically equivalent to those associated with the true log-likelihood, $L^c(\boldsymbol{\psi})$. (The idea of using a sequence of 'local to zero' parameter values is not unusual; see Staiger and Stock (1997) who employ it to obtain the asymptotic distribution of the instrumental variable estimator when the partial correlations between the instruments and the included endogenous variables are weak.) Moreover, a little algebra reveals that $\hat{\boldsymbol{\lambda}}$ will be asymptotically independent of $\left(\hat{\boldsymbol{\theta}}', \hat{\rho}\right)$ under $\rho = O(\frac{1}{\sqrt{N}})$ so that $(\tilde{\boldsymbol{\theta}}', \tilde{\rho})$ and $\tilde{\boldsymbol{\lambda}}$ must also will be asymptotically independent. Therefore, if $\tilde{\boldsymbol{\lambda}}$ denotes *any* consistent estimator for $\boldsymbol{\lambda}$, maximisation of the 'concentrated' (approximate) log-likelihood $\tilde{L}^a(\boldsymbol{\theta},\rho) = L^a(\boldsymbol{\theta}, \tilde{\boldsymbol{\lambda}},\rho)$, yields estimators which are asymptotically equivalent to $\left(\hat{\boldsymbol{\theta}}', \hat{\rho}\right)$; see, for example Cox and Reid (1987). Under the assumption that $\rho = O(\frac{1}{\sqrt{N}})$, a suitable choice for $\tilde{\boldsymbol{\lambda}}$ is simply the MLE for the model defined by $g_i(y_{i0};\boldsymbol{\lambda})$ and this leads to the following two-step procedure:[1]

[1]The procedure is also reminiscent of the two-step estimator proposed by Vella (1993) and is in the spirit of Blundell and Smith (1991).

1. Estimate the marginal model for y_{i0}, using $g_i(y_{i0};\boldsymbol{\lambda})$, and obtain the generalised residual $\tilde{e}_{i0}$.
2. Augment the regressor set, $\mathbf{X}_i$, with $\tilde{e}_{i0}$. Estimate the artificial random effects model $\int_{-\infty}^{\infty}\tilde{F}_i^a(\mathbf{y}_i|u;\boldsymbol{\psi})\phi(u)du$ by maximum likelihood where, for the purposes of the two-step procedure,

$$\tilde{F}_i^a(\mathbf{y}_i|u;\boldsymbol{\psi}) = \prod_{t=1}^{T} f\left(y_{it};\tilde{w}_{it}^a\right),$$

with $\tilde{w}_{it}^a = m_1(\alpha y_{it-1}, \boldsymbol{\beta}'\mathbf{x}_{it}+\delta\tilde{e}_{i0}+\sigma u_i)$. That is, maximise a *pseudo log-likelihood* which incorporates the dynamic term y_{it-1} but which treats y_{i0} as exogenous. (The inclusion of $\tilde{e}_{i0}$ provides a test of, and an approximate control for, the initial conditions problem.)

Locally equivalent alternative methodology ensures that the standard *t-test* of the significance of adding $\tilde{e}_{i0}$ provides a simple asymptotically valid test of $\rho = 0$, whilst the obtained (two-step) estimator, $\tilde{\boldsymbol{\theta}}$, provides an approximate bias-corrected estimator for $\boldsymbol{\theta}$.[2] Consequently, $\tilde{\boldsymbol{\theta}}$ shall be referred to as the *pseudo maximum likelihood estimator* (PMLE). Also, note that the (approximate) asymptotic independence between $\tilde{\boldsymbol{\lambda}}$ and the remaining parameter estimates affords the possibility of exploring an adequate specification for the (reduced form) model for y_{i0}, prior to implementation of the second step in the PML procedure.

For censored normal linear models, the above analysis assumes that the initial condition is uncorrelated with innovations; i.e., $E\left[y_{i0}\varepsilon_{it}\right] = 0$, for all i and t. A similar two-step estimator still applies if a less restrictive covariance structure between y_{i0} and ε_{it} is adopted. In this case the augmented regression function for the latent model would be $\alpha y_{it-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \delta_t\tilde{e}_{i0}$.

2.3 The Quality of the Local Approximation

In the Appendix it is shown that

$$P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) - P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = \frac{\rho^2\tau^2}{2} r_{i0}\int_{-\infty}^{\infty}(u^2-1)F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du + o(\rho^2)$$

as ρ passes to zero, implying that

$$|P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) - P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})| \leq \frac{\rho^2\tau^2}{2}\left|r_{i0}\right|\left|\int_{-\infty}^{\infty}(u^2-1)F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du\right| + \left|o(\rho^2)\right|,$$

[2]Although, strictly speaking, an initial conditions problem should induce a positive correlation this constraint is not binding for the PMLE approach. Thus a two-sided $t-test$ of $\xi = 0$ is asymptotically valid.

where $r_{i0} = \dfrac{\partial^2 \ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial \gamma_1^2}$. Now for the dynamic probit random effects model, with $g_i(y_{i0};\boldsymbol{\lambda})$ being a simple probit model in which $\tau = 1$, $0 < F_i(\mathbf{y}_i|u;\boldsymbol{\theta}) < 1$ and $|r_{i0}| < 1$. Then, because $\int_{-\infty}^{\infty}(u^2-1)\phi(u)du = 0$, it follows that

$$\begin{aligned}
\left|P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) - P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})\right| &< \frac{\rho^2}{2}\left|\int_{-\infty}^{\infty}(u^2-1)F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du\right| + \left|o(\rho^2)\right| \\
&< \frac{\rho^2}{2}\int_{-\infty}^{\infty}\left|u^2-1\right|\phi(u)du + \left|o(\rho^2)\right| \\
&= \rho^2\int_{-1}^{1}(1-u^2)\phi(u)du + \left|o(\rho^2)\right| \\
&= 2\rho^2\phi(1) + \left|o(\rho^2)\right| \\
&\cong 0.484\rho^2 + \left|o(\rho^2)\right|
\end{aligned}$$

3 Monte Carlo Experiments

This section investigates the finite sample reliability of the two-step procedure in a *random effects logit* setting using Monte Carlo experimentation. Two groups of experiments were undertaken, differing merely in how the initial observations were generated. Common features of the data generation process (DGP) in both experiments are:

$$y_{it}^* = \alpha y_{i,t-1} + \beta_1 + \beta_2 z_{it} + \sigma u_i + \varepsilon_{it}, \tag{6}$$

where ε_{it} were *iid* standard logit, with the random effects term, u_i being *iid* $N(0,1)$ and $y_{it} = 1$ if $y_{it}^* > 0$ and $y_{it} = 0$, otherwise. In *Experiment 1*, the process is started at $t = -24$ with $y_{i,-25}^*$ being a random draw from a standard logit distribution, uncorrelated with u_i. In *Experiment 2*, y_{i0} derives from $y_{i0}^* = \gamma_0 + z_{i0}\gamma_1 + v_i$, where $v_i = \xi u_i + \sqrt{1-\xi^2}\varepsilon_{i0}$ and ε_{i0} *iid* $N(0,1)$. Following Heckman (1981a), the exogenous regressor was generated as $z_{it} = 0.1t + 0.5z_{i,t-1} + U(-0.5, 0.5)$ with $z_{i,-25} \sim U(-3,2)$. A different sequence, $\{z_{it}\}$, was employed for each individual, but held fixed over repeated sampling. For the purposes of estimation, in both cases, only observations from $t = 0, \ldots, T$ are retained. Various parameter configurations and differing N and T, were assessed, but only the results for $\alpha = 0.5$, $\beta_1 = 1$, $\beta_2 = -1$, $\sigma = 1$, $N = 200$ and $T = 3$ are reported in detail here. (A full set of results is available on request.)

In all experiments, 1000 replications of sample data were used in order to summarise the properties of estimators for $(\alpha, \beta_1, \beta_2)$. These estimators were:

(a) the standard binary random effects MLE, based on (1) and treating y_{i0} as exogenous; and, (b) the (two-step) PML estimator, which augments the regressor set in (6) with $\tilde{e}_{i0}$ and then employs standard random effects estimation treating y_{i0} as exogenous (denoted PMLE). The generalised residual, $\tilde{e}_{i0}$, used in the PML procedure was obtained from fitting an initial probit specification for y_{i0} with the regressors (instruments) simply being $(1, z_{i0})$. Applied workers may fit a logit model for y_{i0}, although there is little to choose between these specifications. In practice, the inclusion of additional instruments such as $z_{i1}, \ldots, z_{iT}$ might also be appropriate; see Blundell and Bond (1995). However, due to the first order process generating z_{it}, one might expect the extra instruments to be of little help in the present situation. Indeed Monte Carlo results, not reported here, support this conjecture.

All estimated standard errors were obtained from the negative inverse hessian matrix and integrals appearing in the likelihoods were evaluated, numerically, using 8 point *Gauss-Hermite Quadrature*.[3]

3.1 The Results

Tables I and II summarise the evidence on the efficacy of the proposed two-step test procedure and associated estimator under *Experiments 1* and *2*, respectively, using various summary statistics. (Definitions appearing at the foot of Table I.)

Insert Table I about here

The rejection rate reported for δ is that of the t-statistic on the generalised residual in the augmented model using a nominal 5% significance level. Under *Experiment 1*, exogenous y_{i0} means that this initial observation is drawn from a standard logit model independently of u_i, and in this case the test is slightly undersized (although in other experiments with $T = 5$ and 10, not reported here, rejection rates of 6.1% and 4.3%, respectively, show closer agreement with the nominal significance level). Under exogenous y_{i0}, there is little to choose between the standard MLE and the bias-correcting PMLE and classical asymptotic theory appears to provide a reasonable guide to finite sample behaviour insofar that the individual t-ratios have acceptable rejection rates (as reported in the final column of the table). The lower half of the Table reports the corresponding results under endogenous y_{i0}. The t-test

[3]When undertaking a Monte Carlo study a relatively high degree of automation is required. In a number of cases the estimation routine failed to find an interior maximum, in which case that particular data set was discarded and a fresh set generated.

on the generalised residual indicates substantial misspecification in the standard random effects model (exhibiting rejection rate of 69.2%, which rises to 95.3% and 99.1% when $T = 5$ and 10, respectively) and inferences based on the standard MLE now become unreliable. Not only is the standard MLE for α is biased upwards, but the true sampling variance is underestimated leading to a substantial over-rejection (41.8%) of the correct null hypothesis that $\alpha = 0.5$. These results confirm Heckman's (1981b) finding that previous event history can be dramatically overstated if the initial conditions problem is ignored. In contrast, however, the PMLE is quite effective at bias correcting and provides estimated standard errors which are quite close to the sampling standard deviations. The PMLE for α is clearly superior to that of the standard MLE in terms of *root mean square error*, whilst the t-ratios on all three coefficients have acceptable rejection rates $(5.0\% - 6.7\%)$ under the correct individual null hypotheses that $\beta_1 = 1, \beta_2 = -1, \alpha = 0.5$, respectively.

Since the justification for the PML approach is a small correlation approximation, it seems appropriate to investigate what happens as ξ passes from zero towards unity. Controlling the correlation is difficult to achieve under *Experiment 1*, but is easily done under the data generation process of *Experiment 2*. Table II reports the summary statistics under this sampling scheme for various values of ξ, the correlation parameter.

Insert Table II about here

The t-statistic on the generalised residual in the artificial model has good size properties and exhibits increasing power as ξ increases, as would be hoped. The standard MLE's and the PMLE's have larger sampling variances under this scheme than under *Experiment 1*, although the MLE's exhibit the same directional bias which increases as ξ moves towards unity. The standard MLE's are slightly more efficient than the PMLE's under exogenous y_{i0}, but as ξ increases the sampling variance of the MLE for α is underestimated and the corresponding t-ratio has an increasing tendency to reject $\alpha = 0.5$. This rejection rate is 57% at $\xi = 0.8$ whilst the root mean square error of the PMLE for α is 56% that of the MLE. On the other hand, the PMLE's prove effective as bias correcting the MLE, their estimated standard errors appear to remain reasonable estimates of their standard deviations and the rejection rates of the t-ratios are close to 5% for all ξ.

4 Concluding Remarks

The procedure developed in this paper yields a simple inferential tool which may be used to detect, and correct for, the so-called initial condition problem in fully parametric, non-linear, dynamic random effects models. The distributional assumptions are theoretically restrictive but reflect the approach taken in recent empirical work. Although the implied two-step estimator is predicated on the notion of a key correlation parameter being local to zero, Monte Carlo results suggest that it can still perform quite well when the correlation between the initial condition and the random effect is quite strong. In particular, for the binary random effects model considered, it effectively corrects the standard MLE for the impact of previous event history. Applied workers can use the artificial model and two-step estimator as an exploratory device to inform a more sophisticated estimation strategy or to determine whether such effort is warranted.

Financial support was received from the ESRC, grant H51955006. My thanks to Wiji Arulampalam, Len Gill and Mark Stewart for helpful comments, and the hospitality of the University of Warwick Summer Workshop in Economics. The usual disclaimer applies.

Appendix: Local Equivalence of $P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})$ and $P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})$

Under the assumption that $\rho = 0$ (independence of u and v), the marginal model for y_{i0} will be identical to the conditional model; i.e., $g_i(y_{i0}|u;\boldsymbol{\zeta})|_{\rho=0} = g_i(y_{i0};\boldsymbol{\lambda})$. Then, from (3) we can write

$$\begin{aligned}
\frac{\partial^j g_i(y_{i0}|u;\boldsymbol{\zeta})}{\partial\rho^j}|_{\rho=0} &= \int_{-\infty}^{\infty}\frac{\partial^j k(.)}{\partial\rho^j}|_{\rho=0}\phi(\varepsilon)d\varepsilon \\
\frac{\partial k(.)}{\partial\rho}|_{\rho=0} &= \tau u\left(\frac{\partial k(.)}{\partial\gamma_1}|_{\rho=0}\right) \\
\frac{\partial^2 k(.)}{\partial\rho^2}|_{\rho=0} &= \tau^2u^2\left(\frac{\partial^2 k(.)}{\partial\gamma_1^2}|_{\rho=0}\right) - \tau\varepsilon\left(\frac{\partial k(.)}{\partial\gamma_1}|_{\rho=0}\right).
\end{aligned}$$

Therefore a second order Taylor series expansion of, $g_i(y_{i0}|u;\boldsymbol{\zeta})$ about $\rho = 0$ yields

$$\begin{aligned}
g_i(y_{i0}|u;\boldsymbol{\zeta}) &= g_i\left(y_{i0};\boldsymbol{\lambda}\right) + \rho\tau u\int_{-\infty}^{\infty}\left(\frac{\partial k(.)}{\partial\gamma_1}|_{\rho=0}\right)\phi(\varepsilon)d\varepsilon + \frac{\rho^2\tau^2u^2}{2}\int_{-\infty}^{\infty}\left(\frac{\partial^2 k(.)}{\partial\gamma_1^2}|_{\rho=0}\right)\phi(\varepsilon)d\varepsilon \\
&\quad -\frac{\rho^2\tau}{2}\int_{-\infty}^{\infty}\left(\frac{\partial k(.)}{\partial\gamma_1}|_{\rho=0}\right)\varepsilon\phi(\varepsilon)d\varepsilon + o\left(\rho^2\right).
\end{aligned}$$

But

$$\begin{aligned}
\int_{-\infty}^{\infty}\left(\frac{\partial k(.)}{\partial\gamma_1}|_{\rho=0}\right)\phi(\varepsilon)d\varepsilon &= \frac{\partial}{\partial\gamma_1}\int_{-\infty}^{\infty}k(y_{i0};\boldsymbol{\gamma}'\mathbf{z}_i+\tau\varepsilon)\phi(\varepsilon)d\varepsilon \\
&= \frac{\partial g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1}
\end{aligned}$$

$$\int_{-\infty}^{\infty}\left(\frac{\partial^2 k(.)}{\partial\gamma_1^2}|_{\rho=0}\right)\phi(\varepsilon)d\varepsilon = \frac{\partial^2 g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1^2}$$

and, intergrating by parts,

$$\begin{aligned}
\int_{-\infty}^{\infty}\left(\frac{\partial k(.)}{\partial\gamma_1}|_{\rho=0}\right)\varepsilon\phi(\varepsilon)d\varepsilon &= \int_{-\infty}^{\infty}\frac{\partial k(y_{i0};\boldsymbol{\gamma}'\mathbf{z}_i+\tau\varepsilon)}{\partial\gamma_1}\varepsilon\phi(\varepsilon)d\varepsilon \\
&= \frac{1}{\tau}\int_{-\infty}^{\infty}\frac{dk(y_{i0};\boldsymbol{\gamma}'\mathbf{z}_i+\tau\varepsilon)}{d\varepsilon}\varepsilon\phi(\varepsilon)d\varepsilon \\
&= \frac{1}{\tau}\left\{-\left[\frac{dk(y_{i0};\boldsymbol{\gamma}'\mathbf{z}_i+\tau\varepsilon)}{d\varepsilon}\phi(\varepsilon)\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\frac{d^2k(y_{i0};\boldsymbol{\gamma}'\mathbf{z}_i+\tau\varepsilon)}{d\varepsilon^2}\phi(\varepsilon)d\varepsilon\right\} \\
&= \tau\frac{\partial^2 g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1^2}.
\end{aligned}$$

Therefore

$$\begin{aligned}
g_i(y_{i0}|u;\boldsymbol{\zeta}) &= g_i(y_{i0};\boldsymbol{\lambda})+\rho\tau u\frac{\partial g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1}+\frac{\rho^2\tau^2u^2}{2}\frac{\partial^2 g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1^2} \\
&\quad -\frac{\rho^2\tau^2}{2}\frac{\partial^2 g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1^2}+o\left(\rho^2\right) \\
&= g_i(y_{i0};\boldsymbol{\lambda})\left\{1+\rho\tau u e_{i0}+\frac{\rho^2\tau^2(u^2-1)}{2}d_{i0}\right\}+o(\rho^2)
\end{aligned}$$

where

$$\begin{aligned}
d_{i0} &= \frac{1}{g_i(y_{i0};\boldsymbol{\lambda})}\frac{\partial^2 g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1^2} \\
&= \frac{\partial^2\ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1^2}+\left(\frac{\partial\ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial\gamma_1}\right)^2
\end{aligned}$$

Using this, a second order Taylor expansion of $P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi})$ gives

$$P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = \int_{-\infty}^{\infty}F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\left\{1+\rho\tau u e_{i0}+\frac{\rho^2\tau^2(u^2-1)}{2}d_{i0}\right\}\phi(u)du+o(\rho^2).$$

Note that integration by parts reveals that

$$\begin{aligned}\int_{-\infty}^{\infty} uF_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du &= \int_{-\infty}^{\infty} \frac{dF_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{du}\phi(u)du \\ &= \sigma\int_{-\infty}^{\infty} \frac{\partial F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1}\phi(u)du,\end{aligned}$$

where β_1 is the intercept term in $\boldsymbol{\beta}'\mathbf{x}_{it}$, and

$$\int_{-\infty}^{\infty}(u^2-1)F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du = \sigma^2\int_{-\infty}^{\infty} \frac{\partial^2 F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1^2}\phi(u)du.$$

A Taylor expansion of $P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = \int_{-\infty}^{\infty} F_i^a(\mathbf{y}_i|u)\phi(u)du$ about $\rho = 0$, yields

$$\begin{aligned}P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) &= \int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du + \rho\sigma\tau e_{i0}\int_{-\infty}^{\infty} \frac{\partial F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1}\phi(u)du \\ &\quad + \frac{\rho^2\sigma^2\tau^2}{2}e_{i0}^2\int_{-\infty}^{\infty} \frac{\partial^2 F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1^2}\phi(u)du + o(\rho^2),\end{aligned}$$

from which we conclude that

$$P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) - P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = \frac{\rho^2\tau^2}{2}r_{i0}\int_{-\infty}^{\infty}(u^2-1)F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du + o(\rho^2),$$

where $r_{i0} = d_{i0} - e_{i0}^2 = \dfrac{\partial^2 \ln g_i(y_{i0};\boldsymbol{\lambda})}{\partial \gamma_1^2}$.

Extending the analysis, consider a Taylor series expansion of $P_i^b(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = \int_{-\infty}^{\infty} F_i^b(\mathbf{y}_i|u)\phi(u)du$ about $\rho = 0$, where $F_i^b(\mathbf{y}_i|u;\boldsymbol{\psi}) = \prod_{t=1}^{T} f\left(y_{it}; w_{it}^b\right)$, denotes an *augmented* model in which $w_{it}^b = m_1(\alpha y_{it-1}, \boldsymbol{\beta}'\mathbf{x}_{it} + \sigma\tau\rho e_{i0} + \sigma(1 + \frac{\tau^2\rho^2}{2}r_{i0})u)$:

$$\begin{aligned}P_i^b(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) &= \int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du + \rho\sigma\tau e_{i0}\int_{-\infty}^{\infty} \frac{\partial F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1}\phi(u)du \\ &\quad + \frac{\rho^2\sigma^2\tau^2}{2}e_{i0}^2\int_{-\infty}^{\infty} \frac{\partial^2 F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1^2}\phi(u)du \\ &\quad + \frac{\rho^2\tau^2}{2}\sigma r_{i0}\int_{-\infty}^{\infty} u\frac{\partial F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1}\phi(u)du + o(\rho^2), \\ &= \int_{-\infty}^{\infty} F_i(\mathbf{y}_i|u;\boldsymbol{\theta})\phi(u)du + \rho\sigma\tau e_{i0}\int_{-\infty}^{\infty} \frac{\partial F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1}\phi(u)du \\ &\quad + \frac{\rho^2\sigma^2\tau^2}{2}d_{i0}\int_{-\infty}^{\infty} \frac{\partial^2 F_i(\mathbf{y}_i|u;\boldsymbol{\theta})}{\partial \beta_1^2}\phi(u)du + o(\rho^2).\end{aligned}$$

Thus, $P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) - P_i^b(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = o(\rho^2)$ whilst $P_i^c(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) - P_i^a(\mathbf{y}_i|y_{i0};\boldsymbol{\psi}) = o(\rho)$.

References

[1] Arellano, M. and R. Carrasco. 1996. Binary Choice Panel Data Models with Predetermined Variables. CEMFI Working Paper, No. 9618.

[2] Blundell, R.W. and R.J. Smith. 1991. Conditions Initiales et Estimation Efficace dans les Modèles Dynamiques sur Données de Panel. Annales d'Economie et de Statistique 20/21, 109-123.

[3] Blundell, R.W. and S. Bond. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. Journal of Econometrics 87, 115-143.

[4] Blundell, R.W., R. Griffith and F. Windmeijer. 1999. Individual Effects and Dynamics in Count Data Models. The Institute for Fiscal Studies, Working Paper Series No. W99/3.

[5] Chesher, A.D. 1991. The Effect of Measurement Error. Biometrika 78, 451-462.

[6] Chesher, A.D. 1997. Non-Normal Variation and Regression to the Mean. Statistical Methods in Medical Research 6, 147-166.

[7] Chesher, A.D. 1998. Measurement Error and Bias Reduction. University of Bristol, Department of Economics Discussion Paper No. 98/449.

[8] Chesher, A.D. and J. Santos-Silva. 1995. Taste Variation in Discrete Choice Models. University of Bristol, Department of Economics Discussion Paper No. 95/396.

[9] Cox, D.R. and N. Reid. 1987. Parameter Orthogonality and Approximate Conditional Inference. Journal of the Royal Statistical Society B 49, 1-39.

[10] Godfrey, L.G. 1988. Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches. Cambridge University Press.

[11] Gourieroux, C. and M. Visser. 1997. A Count Data Model with Unobserved Heterogeneity. Journal of Econometrics 79, 247-268.

[12] Heckman, J.J. 1981a. The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data

Stochastic Process, in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C.F. Manski and D.McFadden, pp.179-95. Cambridge: MIT Press.

[13] Heckman, J.J. 1981b. Heterogeneity and State Dependence', in *Studies in Labor Markets*, edited by S. Rosen, pp. 91-139. University of Chicago Press.

[14] Honoré, B.E. and E. Kyriazidou. 1996. Panel Data Discrete Choice Models with Lagged Dependent Variables. Unpublished manuscript, Department of Economics, Princeton University

[15] Hsiao, C. 1986. Analysis of Panel Data. Cambridge University Press.

[16] Lancaster, T. 1990. The Econometric Analysis of Transition Data. Cambridge University Press.

[17] Staiger, D. and J. Stock. 1997. Instrumental Variables in Regression with Weak Instruments. Econometrica 65, 557-586.

[18] Vella, F. 1993. A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors. International Economic Review 34, 441-457.

[19] Vella, F. and M. Verbeek. 1998. Whose Wages do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men. Journal of Applied Econometrics 13, 163-183.

[20] Vella, F. and M. Verbeek. 1999. Two-step Estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias. Journal of Econometrics 90, 239-263.

TABLE I: Random Effects Logit Model
$N = 200$, $T = 3$,
$\beta_1 = 1$, $\beta_2 = -1$, $\alpha = 0.5$, $\sigma = 1$,
Experiment 1
1000 Replications

summary statistics of estimators

		ae	**pb**	**ase**	**sde**	**rmse**	**rejt**
		exog y_0					
MLE	β_1	1.0372	3.72	0.2775	0.2790	0.2815	4.9
	β_2	−1.0045	−0.45	0.3079	0.3087	0.3087	4.2
	α	0.4733	−5.34	0.2756	0.2736	0.2788	5.1
PMLE	β_1	1.0396	3.96	0.2995	0.2967	0.2993	4.1
	β_2	−1.0030	−0.30	0.3086	0.3081	0.3080	4.1
	α	0.4701	−5.98	0.3052	0.3021	0.3035	4.3
	δ	−	−	−	−	−	3.5
		endog y_0					
MLE	β_1	0.5472	−45.28	0.3268	0.3654	0.5819	40.0
	β_2	−0.8886	11.14	0.2948	0.2892	0.3100	7.3
	α	0.9696	93.92	0.3077	0.3414	0.5806	41.8
PMLE	β_1	1.0235	2.35	0.3951	0.3983	0.3990	6.7
	β_2	−1.0162	−1.62	0.3209	0.3174	0.3178	5.0
	α	0.4900	−2.99	0.3675	0.3614	0.3615	5.4
	δ	−	−	−	−	−	69.2

Notes:

ae = *average value of estimate*
pb = *percentage bias* = $\frac{\textbf{ae}\text{ - true value}}{|\text{true value}|} \times 100$
ase = *average value of estimated standard error*
sde = *standard deviation of estimate*
rmse = *root mean square error of estimate*
rejt = *percentage of samples for which:-* $\left|\frac{\text{estimate - true value}}{\text{standard error}}\right| > 1.96$.

TABLE II: Random Effects Logit Model
$N = 200$, $T = 3$,
$\beta_1 = 1$, $\beta_2 = -1$, $\alpha = 0.5$, $\sigma = 1$,
Experiment 2
1000 Replications

summary statistics of MLE

	ae	**pb**	**ase**	**sde**	**rmse**	**rejt**
			$\xi = 0.0$			
β_1	1.0387	3.87	0.3900	0.3883	0.3903	5.4
β_2	−1.0105	−1.05	0.3291	0.3361	0.3363	4.5
α	0.4673	−6.54	0.3392	0.3680	0.3689	5.3
			$\xi = 0.4$			
β_1	0.7452	−25.48	0.3934	0.4128	0.4851	18.0
β_2	−0.9222	7.78	0.3168	0.3241	0.3334	6.7
α	0.7460	49.20	0.3512	0.3762	0.3836	16.9
			$\xi = 0.8$			
β_1	0.3793	−62.07	0.3309	0.3814	0.7285	56.5
β_2	−0.8092	19.08	0.2925	0.2921	0.3488	11.6
α	1.0999	119.98	0.3131	0.3667	0.7031	57.3

summary statistics of PMLE

	ae	**pb**	**ase**	**sde**	**rmse**	**rejt**
			$\xi = 0.0$			
β_1	1.0220	2.20	0.4148	0.4213	0.4219	4.9
β_2	−1.0188	−1.88	0.3322	0.3334	0.3339	4.0
α	0.4882	−2.36	0.3675	0.3735	0.3738	5.3
δ	–	–	–	–	–	4.6
			$\xi = 0.4$			
β_1	1.0189	1.89	0.4244	0.4202	0.4206	5.0
β_2	−1.0169	−1.69	0.3318	0.3359	0.3363	4.5
α	0.4847	−3.06	0.3792	0.3746	0.3748	4.7
δ	–	–	–	–	–	36.5
			$\xi = 0.8$			
β_1	1.0316	3.16	0.4384	0.4287	0.4299	5.9
β_2	−1.0214	−2.14	0.3309	0.3321	0.3328	4.4
α	0.4538	−9.24	0.3984	0.3910	0.3937	4.8
δ	–	–	–	–	–	95.1

For definitions, see Table I