Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators

Wiji Arulampalam and Mark B. Stewart

No 884

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators*

Wiji Arulampalam (University of Warwick & IZA)
and
Mark B. Stewart (University of Warwick)

December 2008

Abstract

This paper presents a convenient shortcut method for implementing the Heckman estimator of the dynamic random effects probit model and other dynamic nonlinear panel data models using standard software. It then compares the estimators proposed by Heckman, Orme and Wooldridge, based on three alternative approximations, first in an empirical model for the probability of unemployment and then in a set of simulation experiments. The results indicate that none of the three estimators dominates the other two in all cases. In most cases all three estimators display satisfactory performance, except when the number of time periods is very small.

Key words: Dynamic discrete choice models, initial conditions, dynamic probit, panel data, dynamic nonlinear panel data models.

JEL classification: C23, C25, C13, C51.

---

*Address: Economics Department, University of Warwick, Coventry, CV4 7AL, UK. Email: wiji.arulampalam@warwick.ac.uk, mark.stewart@warwick.ac.uk. We are grateful for financial support from the ESRC under research grants RES-000-22-0651 (Arulampalam) and RES-000-22-2611 (Stewart). Helpful comments were received from Stephen Jenkins and an anonymous referee.

1 Introduction

The initial conditions problem is well-recognised in the estimation of dynamic non-linear panel data models. Its cause is the presence of both the past value of the dependent variable and an unobserved heterogeneity term in the equation and the correlation between them. The strict exogeneity assumption for regressors, routinely used in static models in order to marginalise the likelihood function with respect to the unobserved heterogeneity, cannot be used in a dynamic setting due to the presence of the lagged dependent variable.

The standard estimator for the probit model in this context is that suggested by Heckman (1981a, 1981b), who was the first to explicitly address this problem.[1] His approach involves the specification of an approximation to the reduced form equation for the initial observation and maximum likelihood estimation using the full set of sample observations allowing cross-correlation between the main and initial period equations. However, use of the estimator has been limited by it requiring separate programming due to its absence from standard packages. This has led to the proposal of alternative estimators that have the advantage of requiring only standard software. The estimators suggested by Orme (1997, 2001) and Wooldridge (2005), based on alternative approximations, are commonly used in place of the Heckman estimator for this reason. The main merit claimed by both Orme and Wooldridge for

[1] Although Heckman discussed the issue in the context of the binary probit model, his suggested solution (as well as other suggested solutions discussed below) can also be applied to many other dynamic non-linear panel models, as we discuss later. This paper concentrates on fully parametric approaches to the estimation of these models. See Honoré (1993) and Honoré and Kyriazidou (2000) for semi-parametric estimators for this type of model.

their estimators relative to Heckman's is that theirs can be straightforwardly estimated using standard software.

These estimators have been widely used in many different applications. Some examples are as follows: labour force participation (Hyslop, 1999); unemployment spells (Arulampalam, et. al., 2000); unemployment and low-pay dynamics (Stewart, 2007); self-employment (Henley, 2004); well-being and income support receipts (Lee and Oguzoglu, 2007); social assistance and welfare participation (Andrén, 2007, and Cappellari and Jenkins, 2008); absence behaviour (Audas at al, 2004); self-reported health status (Contoyannis et al, 2004); health insurance (Propper, 2000); infant mortality (Arulampalam and Bhalotra, 2006); smoking behaviour (Dorsett, 1999, and Clark and Etilé, 2006); housing allowance and ownership (Chen and Entrom Ost, 2005); ownership of stocks and mutual funds (Alessie et al, 2004); firms' export behaviour (Bernard and Jensen, 2004); firms' dividend behaviour (Benito and Young, 2003, and Loudermilk, 2007); entry and exit of firms from foreign markets (Requena-Silvente, 2005); and debt relief (Chauvin and Kraay, 2007).

The majority of applications have been based on binary probit models, but some also use the estimators in the context of ordered probit models (e.g. Contoyannis et al, 2004, and Pudney, 2008) and Tobit models (e.g. Islam, 2007, and Loudermilk, 2007). Some applications have also used extensions to bivariate models (e.g. Alessie et al, 2004, and Clark and Etilé, 2006).

This paper sets up the three estimators in a common framework and presents a convenient shortcut method for implementing the Heckman estimator using standard software designed either for the estimation of static models with heteroskedastic random effects (available in Stata) or for constrained random coefficient models (available in Limdep). The dynamic random effects probit model is used as the example throughout the paper, since it is the most commonly used such model.

The increased ease and availability of the Heckman estimator that these shortcut methods provide removes some of the initial motivation for the simpler alternatives. However since the Heckman estimator is itself based on an approximation, this raises the question of the relative finite sample performance of these three approximation-based estimators. This paper therefore also provides an examination of the relative merits of the Heckman, Orme and Wooldridge estimators in the absence of the software issue. It examines differences between the three estimators first in the context of an empirical illustration using a model for the probability of unemployment and then presents a Monte Carlo investigation of their finite sample performance.[2] The Orme and Wooldridge estimators are found to perform as well as, and in some aspects better than, the Heckman estimator. However, none of the three estimators dominates the other two in all cases.

[2] The focus in this paper is on the comparison of the parameter estimates across the three estimators. "Partial effects" can also be estimated after use of these estimators in various ways. They can for example be evaluated for particular (real or hypothetical) individuals (in terms of unobservables as well as observables) or averaged across individuals. Wooldridge (2005) for example discusses easy calculation of "average partial effects" in the context of his approach. However, given the central focus of this paper, both the empirical illustration and the Monte Carlo investigation focus on comparison of the parameter estimates across the three estimators

2 Econometric Model and Estimators

Denote the conditional distribution for the observed dependent variable y_{it} by $D(y_{it}|y_{it-1}, x_{it}, \alpha_i)$, where i indexes independent cross section units and t indexes time periods, x_{it} is a vector of conditioning variables at time t which are assumed to be strictly exogenous, and α_i is the unobserved time-invariant heterogeneity.[3] Denote the parametric density associated with this conditional distribution by $f_t[y_{it}|y_{it-1}, x_{it}, \alpha_i; \delta_1]$ for $t = 1, \ldots, T$, where δ_1 is the associated vector of parameters. The density of $(y_{i1}, y_{i2}, \ldots, y_{iT}|y_{i0}, x_{it}, \alpha_i)$ is then given by

$$\prod_{t=1}^{T} f_t[y_{it}|y_{it-1}, x_{it}, \alpha_i; \delta_1] \tag{1}$$

A parametric specification for the distribution of the unobservables α in (1) would enable the researcher to integrate out the α from (1). However, in the absence of the start of the sample coinciding with the start of the stochastic process, y_0 will not be independent of α in (1). This requires some assumptions about the generation of the initial observation.[4] The three different estimators for estimating δ_1 from (1) proposed in the literature differ in terms of how the initial conditions problem in (1) is handled. These methods are detailed below.

The standard uncorrelated random effects model assumes additionally

[3] For notational convenience, a balanced panel data structure is assumed. The estimators and all the discussions of them below are easily generalisable to certain unbalanced cases.

[4] The assumption that the process has been in equilibrium for some time may also be used to solve the problem. However, this estimator is not easy to implement using standard software, and in most of the empirical applications this assumption is difficult to justify (Heckman, 1981b; Wooldridge, 2002; Hsiao, 2003).

that α_i is uncorrelated with x_{it}. Alternatively, following Mundlak (1978), correlation between α_i and the observed characteristics can be captured by including $x_i = (x_{i0}, \ldots, x_{iT})$, or alternatively averages of the x-variables over t, as additional regressors in the model. To simplify notation, this specification will not be used explicitly here; rather it should be understood that when the Mundlak correlated random effects (CRE) model is used, x_{it} in (1) implicitly subsumes a full set of period-specific versions of the (time-varying) x-variables (or their means).

2.1 The Initial Conditions Problem and Heckman's Estimator

The Heckman approach starts from the joint density of $(y_T, y_{T-1}, \ldots, y_0|x, \alpha)$ specified as

$$f(y_T, y_{T-1}, \ldots, y_0|x, \alpha) = f(y_T, y_{T-1}, \ldots, y_1|y_0, x, \alpha) f_0(y_0|x, \alpha) \qquad (2)$$

with the first term on the right hand side given by (1).[5] The unobservable α can be integrated out of the log likelihood by making a distributional assumption about the conditional density of the first observation $f_0(y_0|x, \alpha)$ and the density for α given x.

$$f(y_T, y_{T-1}, \ldots, y_0|x) = \int f(y_T, \ldots, y_1|y_0, x, \alpha) f_0(y_0|x, \alpha) g(\alpha|x) d\alpha \qquad (3)$$

Heckman suggested approximating the density $f_0(y_0|x, \alpha)$ using the same parametric form as the conditional density for the rest of the observations.

[5] To simplify notation, parameters are not explicitly shown in the densities.

Consider the latent variable form of the random effects probit model for illustration. Let

$$y_{it} = \mathbf{1}[y_{it}^* > 0], \tag{4}$$

where

$$y_{it}^* = \gamma y_{it-1} + x_{it}'\beta + \theta_t\alpha_i + u_{it}, \qquad t = 1, \ldots, T \tag{5}$$

with $\theta_T = 1$ for identification, and the equation for the first period written using the error components structure as

$$y_{i0}^* = z_i'\lambda + \theta_0\alpha_i + u_{i0}, \tag{6}$$

where z_i is a vector of exogenous covariates which is expected to include x_{i0} and additional variables that can be viewed as "instruments" such as pre-sample variables. The u_{it} are independent of the α_i. The standard assumptions regarding the distributions of the u_{it} and α_i – that they are both normally distributed, the former with variance 1, the latter with variance σ_α^2 – are made. A test of $\theta_0 = 0$ provides a test of exogeneity of the initial condition in this model.

The above specifications are written in the spirit of the original Heckman (1981b) paper where his suggestion was to allow the error in the equation for the initial conditions $(\theta_0\alpha_i + u_{i0})$ to be freely correlated with the errors in the equations for the other periods $(\theta_t\alpha_i + u_{it})$. In addition, the above specification also relaxes the standard assumption of equi-correlated errors in periods $t = 1, \ldots, T$.[6] Most of the existing applications of this technique

[6] In the standard equi-correlated model, $\text{Covar}(\alpha_i + u_{it}, \alpha_i + u_{is}) = \sigma_\alpha^2$, for $t, s = 1, \ldots, T$, $t \neq s$. The correlation between two periods is therefore given by $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + 1)$.

have assumed fixed correlation between $(\theta_0\alpha_i+u_{i0})$ and the error terms in the equations for the other periods,[7] as well as within these latter, by specifying equation (5) as

$$y^*_{it} = \gamma y_{it-1} + x'_{it}\beta + \alpha_i + u_{it}, \qquad t = 1,\ldots,T \tag{7}$$

and equation (6) as

$$y^*_{i0} = z'_i\lambda + \theta\alpha_i + u_{i0}, \tag{8}$$

Equations (5) and (6) together specify a complete model for $(y_0, y_1, \ldots, y_T)$. The contribution to the likelihood function for individual i in this model is given by

$$L_i = \int \left\{ \Phi[(z'_i\lambda + \theta_0\alpha_i)(2y_{i0} - 1)] \prod_{t=1}^{T_i} \Phi[(x'_{it}\beta + \gamma y_{it-1} + \theta_t\alpha_i)(2y_{it} - 1)] \right\} g(\alpha_i)d\alpha_i \tag{9}$$

with $\theta_T = 1$, $g(\alpha)$ is the probability density function of the unobservable individual-specific heterogeneity and Φ is the standard normal cdf. In the standard case considered here, α is taken to be normally distributed and the integral in (9) can be evaluated using Gaussian-Hermite quadrature (Butler and Moffitt, 1982).

The approximation that is used in equation (6) for the initial period can also be derived as follows. Write the initial period latent equation as

$$y^*_{i0} = z'_i\lambda^* + v_{i0} \tag{10}$$

The initial condition problem is present because of correlation between v_{i0} and α_i. Assuming bivariate normality, $(v_{i0}, \alpha_i) \sim BVN(0, 0, \sigma^2_v, \sigma^2_\alpha, r)$, gives

[7] For an exception, see Andrén (2007).

$v_{i0}|\alpha_i \sim N[r(\sigma_v/\sigma_\alpha), \sigma_v^2(1-r^2)]$. Using this, (10) can be written as

$$y_{i0}^* = z_i'\lambda^* + r\frac{\sigma_v}{\sigma_\alpha}\alpha_i + (\sigma_v\sqrt{1-r^2})u_{i0} \tag{11}$$

where α_i and u_{i0} are orthogonal by construction and $u_{i0} \sim N(0,1)$. Rescaling this equation by $\sigma_v\sqrt{1-r^2}$ gives the equivalent of (6) with $\theta_0 = r/(\sigma_\alpha\sqrt{1-r^2})$ and the latent variable y_{i0}^* also rescaled by $\sigma_v\sqrt{1-r^2}$.

2.2 Shortcut Implementation of Heckman's Estimator

The simplified implementation procedure proposed here involves the creation of a set of $T+1$ dummy variables: $d_{it}^{(\tau)} = 1$ if $t = \tau$, $d_{it}^{(\tau)} = 0$ otherwise. Equations (5) and (6) in the model with "freely correlated" v_{it} can then be combined to give (with $\theta_T = 1$)

$$\Pr[y_{it} = 1|y_{it-1}, x_{it}, z_i, \alpha_i]$$

$$= \Phi[\{(\gamma y_{it-1} + x_{it}'\beta) * (1 - d_{it}^{(0)}) + (\theta_1 d_{it}^{(1)} + \ldots + \theta_T d_{it}^{(T)})\alpha_i\} + (z_i'\lambda + \theta_0\alpha_i) * d_{it}^{(0)}]$$

$$= \Phi[\gamma(1-d_{it}^{(0)})y_{it-1} + (1-d_{it}^{(0)})x_{it}'\beta + d_{it}^{(0)}z_i'\lambda + (d_{it}^{(T)} + \theta_{T-1}d_{it}^{(T-1)} + \ldots + \theta_0 d_{it}^{(0)})\alpha_i] \tag{12}$$

This is equivalent to a standard random effects specification, but with a heteroskedastic factor loading for the random effects. Software that allows this form of heteroskedasticity, such as the gllamm program in Stata, can be used to estimate this model. The more standard "equi-correlated" special case gives

$$\Pr[y_{it}=1|y_{it-1},x_{it},z_i,\alpha_i]=$$
$$\Phi[\gamma(1-d_{it}^{(0)})y_{it-1}+(1-d_{it}^{(0)})x_{it}'\beta+d_{it}^{(0)}z_i'\lambda+(1-d_{it}^{(0)}+\theta d_{it}^{(0)})\alpha_i] \tag{13}$$

Alternatively the model can be viewed as a constrained random coefficients model. The model with "equi-correlated" v_{it} can be rewritten as

$$\Pr[y_{it}=1|y_{it-1},x_{it},z_i,\alpha_i]=$$
$$\Phi[\alpha_i+\gamma(1-d_{it}^{(0)})y_{it-1}+(1-d_{it}^{(0)})x_{it}'\beta+d_{it}^{(0)}z_i'\lambda+(\theta-1)\alpha_i d_{it}^{(0)}] \tag{14}$$

This contains a random intercept term, α_i, and the coefficient on d is a second random coefficient, with a unit correlation with the random intercept, but a different variance. The specification can also be generalised to the "freely correlated" form as above with a different variance for each period. Software for estimating random coefficient models that allows this form of restriction, such as Limdep, can therefore also be used.

2.3 Orme's Two-step Estimator

Orme's two-step estimator is in the spirit of Heckman's two-step procedure for addressing the issue of endogenous sample selection. Since the cause of the initial conditions problem is the correlation between the regressor y_{it-1} and the unobservable α_i, Orme (1997, 2001) uses an approximation to substitute α_i with another unobservable component that is uncorrelated with the initial observation. Using the same assumption as in the derivation of the Heckman estimator, that $(v_{i0}, \alpha_i) \sim BVN(0, 0, \sigma_v^2, \sigma_\alpha^2, r)$, and writing now

$$\alpha_i | v_{i0} \sim N\left[r\frac{\sigma_\alpha}{\sigma_v}, \sigma_\alpha^2(1-r^2)\right]$$

means that we can write

$$\alpha_i = r\frac{\sigma_\alpha}{\sigma_v}v_{i0} + \sigma_\alpha\sqrt{(1-r^2)}w_i \tag{15}$$

where w_i is orthogonal to v_{i0} by construction and distributed as $N(0,1)$. Substituting for α_i in (5) gives

$$y_{it}^* = \gamma y_{it-1} + x_{it}'\beta + \theta_t\left[r\frac{\sigma_\alpha}{\sigma_v}v_{i0} + \sigma_\alpha\sqrt{(1-r^2)}w_i\right] + u_{it} \tag{16}$$

Equation (16) has two time-invariant unobserved components, v_{i0} and w_i. Since $E(w_i|y_{i0}) = 0$ by construction, the initial conditions problem can be addressed by allowing for the correlation of v_{i0} with y_{i0} in (16). As Orme notes, (10) and the assumption of bivariate normality for the joint distribution of (v_{i0}, α_i) implies that

$$e_i \equiv E(v_{i0}|y_{i0}) = (2y_{i0}-1)\sigma_v\phi(\lambda^{*\prime}z_i/\sigma_v)/\Phi((2y_{i0}-1)\lambda^{*\prime}z_i/\sigma_v) \tag{17}$$

where ϕ and Φ are the Normal density and distribution functions respectively. This is the generalised error from a first period probit equation, analogous to that used in Heckman's sample selection model estimator. Hence we can estimate (16) as a random effects probit model using standard software with v_{i0} replaced with an estimate of e_i after the estimation of (10) using a simple probit.

Orme's method can easily be generalised to allow v_{i0} to be freely correlated with v_{it} in the spirit of Heckman, by including a set of time dummies interacted with the e_i as suggested by Orme. A potential problem is that the time-invariant error component in the second stage will be heteroscedastic. When v_{i0} is replaced by $e_i \equiv E(v_{i0}|y_{i0})$, a factor involving $[v_{i0} - e_i]$ gets incorporated into w_i, which is now heteroskedastic because it depends on the two conditional expectations involved in e_i. The extent of this heteroscedasticity declines as r does.

Although based on a local approximation for small r, Orme finds that the approximation works reasonably well even when this correlation is fairly different from zero. Since the Heckman and Orme estimators make the same distributional assumptions for (v_{i0}, α_i), the simplified implementation of the Heckman estimator in section 2.2 reduces the usefulness of the Orme estimator. However the Orme estimator offers dramatic savings in computing time relative to the Heckman estimator.

2.4 Wooldridge's Conditional ML estimator

The Heckman estimator approximates the joint probability of the full observed y sequence $(y_0, y_1, \ldots, y_T)$. Wooldridge (2005) on the other hand, has proposed an alternative Conditional Maximum Likelihood (CML) estimator that considers the distribution of $y_1, y_2, \ldots, y_T$ conditional on the initial period value y_0 (and exogenous variables).

The joint density for the observed sequence $(y_1, y_2, \ldots, y_T|y_0)$ is written as $f(y_T, y_{T-1}, \ldots, y_1|y_0, x, \alpha)$. In order to integrate out the unobservable α, Wooldridge specifies an approximation for the density of α conditional on the initial observation y_0. Thus a specification such as the following is assumed in the case of the random effects probit,

$$\alpha_i|y_{i0}, z_i \sim N(\varsigma_0 + \varsigma_1 y_{i0} + z_i'\varsigma, \sigma_a^2) \tag{18}$$

where

$$\alpha_i = \varsigma_0 + \varsigma_1 y_{i0} + z_i'\varsigma + a_i \tag{19}$$

in which z_i includes variables that are correlated with the unobservable α_i. The appropriate z may differ from that in the Heckman specification. The idea here is that the correlation between y_{i0} and α is handled by the use of (19) giving another unobservable individual-specific heterogeneity term a which is uncorrelated with the initial observation y_0. Wooldridge in fact specifies z_i to be x_i as in the Mundlak specification using information on periods 1 to T, but alternative specifications of it would also be possible.

Substituting (19) into (7) gives

$$\Pr(y_{it} = 1|a_i, y_{i0}) = \Phi[x_{it}'\beta + \gamma y_{it-1} + \varsigma_1 y_{i0} + z_i\varsigma + a_i] \qquad t = 1, \ldots, T \tag{20}$$

In this model, the contribution to the likelihood function for individual i is given by

$$L_i = \int \left\{ \prod_{t=1}^{T} \Phi\left[(x_{it}'\beta + \gamma y_{it-1} + \varsigma_1 y_{i0} + z_i'\varsigma + a_i)(2y_{it} - 1)\right] \right\} g^*(a_i) da_i \tag{21}$$

where $g^*(a)$ is the normal probability density function of the *new* unobservable individual-specific heterogeneity a_i given in (19). Since this is the standard random effects probit model likelihood contribution, one can proceed with the maximisation using standard software. Note that if x_i is used for z_i this means that the Wooldridge estimator for the uncorrelated random effects specification and for the Mundlak correlated random effects specification are the same, since x_i is already included in the model to be estimated. As for the other estimators, Wooldridge's method can also be easily generalised to allow the initial condition error to be freely correlated with the errors in the other periods in the spirit of Heckman, by including a set of time dummies interacted with the y_{i0}.

One useful way of contrasting the approaches used by Heckman and Wooldridge is in terms of the conditioning used and the implications that this has for the distributional approximations required. Both approaches share a common specification for $f(y_1, \ldots, y_T|y_0, \alpha)$. Heckman uses this to specify the joint density of $(y_0, y_1, \ldots, y_T)$ as in (3). This requires an assumption for the joint density of y_0 and α, which equals $f(y_0|\alpha)g(\alpha)$. Wooldridge in contrast uses it to specify the conditional density given y_0

$$f(y_1, \ldots, y_T|y_0) = \int f(y_1, \ldots, y_T|y_0, \alpha)h(\alpha|y_0)d\alpha \tag{22}$$

Thus while Heckman requires an approximation for the joint density of y_0 and α, Wooldridge only requires an assumption for the conditional density $h(\alpha|y_0)$. In practice in the context of the dynamic probit model, Heckman and Orme assume bivariate normality for (v_{i0}, α_i), while Wooldridge assumes normality of the conditional distribution of α_i given y_{i0}.

Another way of contrasting the Wooldridge estimator with those of Heckman and Orme is in terms of the implied specification of $E(\alpha_i|y_{i0})$. In the Heckman and Orme setups, $E(\alpha_i|y_{i0}) = \sigma_\alpha r e_i/\sigma_v$, with e_i given by (17). In the Wooldridge setup it is taken to be linear in y_{i0} and z_i.

2.5 Applications to Other Non-linear Models

Although the random effects probit model has been used for illustration, the basic principles of the three estimators are easily generalisable to other random effects dynamic non-linear models such as Tobit, Poisson etc. (see Wooldridge, 2005; Orme, 2001.) This generalisation also applies to the simplified implementation of the Heckman estimator provided in section 2.2.

Example 1: Dynamic random effects Tobit model

Equations (4) and (5) would become

$$y_{it} = \max[0, y_{it}^*] \tag{23}$$

where

$$y_{it}^* = \gamma q(y_{it-1}) + x_{it}'\beta + \theta_t \alpha_i + u_{it} \qquad t = 1, \ldots, T \tag{24}$$

with $\theta_T = 1$ and the latent initial condition equation again given by (6). The effect of the lagged observed response variable is specified in terms of the function $q(.)$. One can also allow separate effects according to whether the previous period's outcome was a corner solution or not. See Loudermilk (2007) and Islam (2007) for recent applications of this model. The other steps involved in the model estimation, using any of the three methods discussed earlier, go through.

Example 2: Dynamic random effects ordered Probit model

Equations (4), (5) and (6) would become

$$y_{it} = k \text{ if and only if } y^*_{it} \in [\Gamma_{k-1}, \Gamma_k), \qquad k = 1, \ldots, K \tag{25}$$

where

$$y^*_{it} = \gamma q(y_{it-1}) + x'_{it}\beta + \theta_t \alpha_i + u_{it} \qquad t = 1, \ldots, T \tag{26}$$

with $\theta_T = 1$ and the latent initial condition equation again given by (6). $\Gamma_1, \ldots, \Gamma_{K-1}$ are threshold parameters with $\Gamma_0 = -\infty$ and $\Gamma_K = +\infty$. The effect of the lagged observed response variable is specified in terms of the function $q(.)$, containing for example binary indicators for $K-1$ of the lagged potential outcomes. One can also allow separate effects according to the ordinal response in the previous period. Contoyannis et al. (2004) and Pudney (2008) use models of this form. The other steps involved in the model estimation, using any of the three methods discussed earlier, go through.

Example 3: Dynamic random effects Poisson model

Here the conditional mean of the y_{it} process is assumed to take the following form:

$$E(y_{it}|y_{it-1}, \ldots, y_{i0}, x_{it}, \alpha_i) = \alpha_i \exp[\gamma q(y_{it-1}) + x'_{it}\beta] \tag{27}$$

The effect of the lagged observed response variable is specified in terms of the function $q(.)$. One can allow separate effects according to the specific response in the previous period. The other steps involved in the model estimation, using any of the three methods discussed earlier, go through.

3 Empirical Illustration

The empirical illustration uses data from the first six waves of the British Household Panel Survey (BHPS), covering the period 1991-1996, to examine the unemployment dynamics of British men.[8] The data used are a subsample of those used in Stewart (2007). The sample is restricted to those who were in the labour force (employed or unemployed) at each of the six waves. The ILO/OECD definition of unemployment is used, under which a man is unemployed if he does not have a job, but had looked for work in the past four weeks and is available for work.

Results for different estimators for a model for the probability of unemployment of the form of equation (7) above are given in Table 1. The standard model that assumes equi-correlated errors over periods 1 to T is estimated to keep the illustration simple. Column [1] gives the pooled probit estimates. Additional education, more labour market experience and being married reduce the probability of unemployment. Being in poor health or living in a travel to work area with a high unemployment-vacancy ratio raise the probability. Being unemployed at $t-1$ strongly increases the probability of being unemployed at t.

Column [2] gives the equivalent standard random effects probit estimates, treating lagged unemployment as exogenous. The coefficients on all the x-variables are increased, while that on y_{t-1} is reduced relative to the pooled

[8] The BHPS contains a nationally representative sample of households whose members are re-interviewed each year. The sample used here contains only Original Sample Members, is restricted to those aged 18-64 and excludes full-time students.

probit estimates. However the random effects probit and pooled probit models involve different normalizations. To compare coefficients those from the random effects estimator need to be multiplied by an estimate of $\sqrt{1-\rho}$, where ρ is the constant cross-period error correlation given by $\rho = \sigma_{\alpha}^{2}/(\sigma_{\alpha}^{2}+1)$ (see Arulampalam, 1999). The scaled coefficient estimate on unemployment at $t-1$ in column [2] is 1.35. Compared with the pooled probit estimator, the estimate of γ is reduced by a quarter in the random effects model, but remains strongly significant.

The corresponding results for the Heckman estimator are given in column [3], with the initial period equation including two exogenous pre-labour market instruments and the full set of period-specific versions of the time-varying x-variables. (Only the married, poor health and local unemployment-vacancy ratio variables are treated as time-varying. There are very few changes in the years of education variable in the sample.) The estimate of θ is 0.88, significantly greater than zero, rejecting the exogeneity of the initial conditions. (In fact θ is insignificantly different from 1.) Compared to the random effects estimator treating the initial conditions as exogenous, the Heckman estimator shows a fall in the estimate of γ of about a third and a near doubling in the estimate of ρ. In terms of scaled coefficient estimates, $\gamma(1-\rho)^{1/2}$, the standard random effects probit with the initial conditions treated as exogenous gives 1.35, while the Heckman estimator gives 0.79.

The Orme two-step estimates for the same model are given in column [4]. The estimator uses two exogenous pre-labour market instruments in conjunction with x_{it} for all time periods in z_i in the initial period equation as in the Heckman estimator. Relative to the Heckman estimator, the Orme estimator gives a slightly higher estimate of γ: 1.11 compared with 1.05 and a slightly lower estimate of ρ: 0.35 compared with 0.43.

The corresponding Wooldridge CML estimates are given in column [5]. The equation estimated contains x_{it} for all time periods. This gives an estimate of γ of 1.06, between the other two estimates and close to the Heckman estimate, and an estimate of ρ of 0.36, also between the other two estimates and close to the Orme estimate. In terms of scaled coefficient estimates, $\gamma(1-\rho)^{1/2}$, the Wooldridge estimator gives 0.85, about half way between 0.79 for the Heckman estimator and 0.89 for the Orme estimator. However all three of these estimates are fairly close together. The Wooldridge estimates of the elements of β corresponding to education, experience and the local unemployment/vacancy ratio are fairly similar to those from the other estimators. However this is not the case for the coefficients on married and health limits. The latter is cut by about half, the former by about two-thirds. Their standard errors are also appreciably higher than for the other estimators and both are insignificantly different from zero with this estimator. The likely reason for this is seen in the next paragraph.

Estimates for the corresponding correlated random effects model, using the Mundlak specification, are given in Table 2. This results in the full set of period-specific versions of the time-varying x-variables being added to the main equation (in addition to already being in the initial period specification). Recall that the Wooldridge estimator is the same in both cases. The estimates of γ using the Heckman and Orme estimators both fall slightly when this specification is used. The estimates of the coefficients on education and experience are little changed, but those on the (time-varying) married and health limits variables fall considerably and now match closely those from using the Wooldridge estimator.

As indicated above, other specifications of both the z-vector and the relationship between α and the x-variables have been proposed and can be used as alternatives. However the contenders considered here have little effect on the estimates in Tables 1 and 2. To illustrate, using only x_{i1} rather than the whole of x_i in the initial period equation (in addition to the two exogenous pre-labour market instruments) reduces the Heckman estimate of γ in Table 1 from 1.048 to 1.047 and increases the estimate of ρ from 0.430 to 0.433. Replacing the full x_i by the time means changes the estimate of γ to 1.049 and that of ρ to 0.431. Similar very small differences are found for the elements of β, for the other estimators and for the correlated random effects estimates in Table 2.

4 Simulation Illustration

In this section we present the results from a set of Monte Carlo simulation experiments, to provide a comparison of these estimators in a set of situations where the true values of the parameters are known.

For the baseline experiment we consider the data generation process used by Heckman (1981b) and Orme (2001), but then consider a fuller set of variants of, and deviations from, this baseline experiment (as well as investigating all three estimators). The setup for the baseline experiment is as follows. The latent variable is generated as

$$y_{it}^{*} = \gamma y_{i,t-1} + \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it} \qquad t = 1, \ldots, T \tag{28}$$

with $y_{it} = \mathbf{1}[y_{it}^{*} > 0]$, where u_{it} is generated as *iid* $N(0,1)$ and α_i as *iid* $N(0, \sigma_\alpha^2)$. (The inter-period error correlation is therefore given by $\rho = \sigma_\alpha^2/(1+\sigma_\alpha^2)$.) The start of the process is assumed for the baseline experiment to be at t=-25, i.e. there are 25 unobserved time periods before the observed "initial condition" period at t=0. Only observations from periods $t = 0, \ldots, T$ are used in the estimation.

The exogenous regressor is taken to be generated by a Nerlove process of the form $x_{it} = 0.1t + 0.5x_{i,t-1} + U(-0.5, 0.5)$ with $x_{i,-25} \sim U(-3, 2)$.[9] The N individual x_{it} sequences are held fixed across replications. In the first set of experiments $y_{i,-25}^{*}$ is generated as a standard normal random variate. For the baseline experiment samples with $N = 200$ and $T = 3$ are used to match those in Heckman (1981b) and Orme (2001), but an extensive range of alternative values for these were also examined in further experiments. In the baseline experiment the parameter values were set at $\gamma = 0.5$, $\beta_1 = -1$, $\beta_0 = 4$, $\sigma_\alpha = 1$. Different experiments are then conducted for different values of T, N, σ_α and γ. Each of the experiments is based on 1000 Monte

[9] Originally used by Nerlove (1971), this process is used to approximate trended variables commonly found in, for example, labour market micro data (see Heckman, 1981b).

Carlo replications.[10] The Heckman and Orme estimators used x_{i0} alone in the specification for the initial condition.

Table 3 gives the average bias (in percentage terms relative to the true value) and the root mean square error for the estimates of γ and β using the Heckman, Wooldridge and Orme estimators in each of the first set of experiments. In the baseline experiment, the Heckman estimator has the largest relative bias in the estimator of γ, in excess of 12% of the true value, while the Wooldridge estimator has the smallest relative bias at around 4%. However in contrast the Wooldridge estimator of γ has a slightly larger root mean square error than the other two estimators and the Heckman estimator a slightly smaller one. In addition the Heckman estimator of β has the smallest bias at around 2% and the Wooldridge estimator the largest at around 6%. The standard errors of the estimated percentage relative biases are about 1.7% for γ and 0.8% for β for all three estimators. Thus the estimated biases for the baseline experiment are all significantly different from zero.

Another worthwhile comparison is with the standard random effects probit estimator, i.e. treating the initial condition as exogenous. This gives an average estimate of γ of 1.37 compared with the true value of 0.5 and hence a percentage relative bias of 174%. Thus all three of the estimators examined here do a good job of dramatically reducing this bias. Looking at the asymptotic t-statistics for the null hypothesis that γ equals its true value and using

[10] The Monte Carlo simulations were conducted using Stata V10 (StataCorp. 2008). The Heckman estimator used the gllamm command (http://www.gllamm.org).

a nominal 5% significance level, the three estimators (in the order used in the table) give rejection rates of 5.6%, 5.3% and 5.4%. So all fairly close to the nominal level. This contrasts with a 98% rejection rate for the standard random effects probit estimator treating the initial condition as exogenous.[11]

The relative biases and the root mean square errors all decline when T is increased to 5 in experiment 2 and mostly decline further when T is increased to 8 in experiment 3. Figure 1 plots the percentage relative biases in the estimates of γ for each value of T from 2 to 12. For T of 4 and above this shows bias of about 3% or less, markedly lower than for $T = 2$ or 3. These are however significantly different from zero at the 5% level for at least one of the estimators for $T = 4$, 5 and 10. As for T=3, the rejection rates for the tests of γ equal to its true value are fairly close to the nominal 5% for all $T \geq 4$. However for T=2, these rejection rates are far too high: 18%, 14% and 18% for the three estimators.

The next two experiments reported in Table 3 are for higher values of N than in the baseline experiment. The three estimators show slight improvements in the bias in $\widehat{\gamma}$ when N is increased from 200 to 500 and considerable improvement when it is increased to 1,000. The picture is less clear for $\widehat{\beta}$. Figure 2 plots the percentage relative biases in the estimates of γ for values of N between 200 and 1,000. The biases are small from about N=800 upwards. Experiments 4 and 5 and Figure 2 are all based on T=3. The biases are smaller for T=5. This is shown further in Figure 3, which gives the cor-

[11] Details of results referred to in the text but not tabulated are available from the authors on request.

responding plot to Figure 2 but for T=5. In this case the biases are small for slightly lower N too. Experiment 5 indicates that the biases are small with N=1,000 even for T=3. Figure 4 plots the percentage relative biases in the estimates of γ for N=1,000 and T between 2 and 12. The percentage relative biases are reasonably small for all T except T=2.

The first five experiments were conducted in parallel with $\overline{x}_i$ included in the initial condition specification, with $\{x_{it}\}$ included, and with neither included. The biases in $\widehat{\gamma}$ and $\widehat{\beta}$ are similar across these three specifications for all these experiments. In fact for each of the three estimators, the inclusion of $\{x_{it}\}$ worsens the bias more often than it improves it.

Experiments 6 and 7 examine the impact of different values of σ_α, lower than that in the baseline experiment in experiment 6 and higher in experiment 7. A change in σ_α has more than one effect on the model and the estimators. In experiment 6, σ_α is reduced from 1.0 to 0.5. This of course reduces the variance of the combined error term and hence increases the "explanatory power" of x_t and y_{t-1} in the latent equation. Second, it reduces the cross-period error correlation, given by $\rho = \sigma_\alpha^2/(\sigma_\alpha^2+1)$, from 0.5 to 0.2. Third, it reduces r, the correlation between v_{i0} and α_i, from 0.59 to 0.35. Fourth, it reduces $\overline{y}_0$, the mean of the outcome variable in the initial condition period, from 0.31 to 0.25 (and also the means in subsequent periods). Despite the reduction in r, this worsens the bias in all three estimators of γ and in two of the estimators of β.

In experiment 7, σ_α is increased from 1.0 to 1.5. This reduces the "explanatory power" of x_t and y_{t-1} in the latent equation and increases ρ from 0.5 to 0.69, r from 0.59 to 0.70, and $\overline{y}_0$ from 0.31 to 0.36. Despite the increase in r, this reduces the bias in two of the estimators of γ. However the bias in the estimation of β is increased for all three estimators.

Experiments 8 and 9 examine the impact of different values of γ, lower than that in the baseline experiment in experiment 8 and higher in experiment 9. Again changing γ has knock-on effects. In experiment 8, γ is reduced from 0.5 to 0.25. This reduces the "explanatory power" of the latent equation. However it also slightly reduces r, from 0.59 to 0.56, and $\overline{y}_0$ from 0.31 to 0.28. This worsens the bias in all three estimators of γ and for two of the estimators of β. In experiment 9, γ is increased from 0.5 to 0.75. This increases the "explanatory power" of the latent equation. However it also slightly increases r, from 0.59 to 0.62, and $\overline{y}_0$ from 0.31 to 0.35. This reduces the bias in all three estimators of γ and for two of the estimators of β (although only very slightly).

Experiments 10–13 repeat experiments 6–9 with T=5 and experiments 14–17 do so with T=8. For the three experiments with any biases in excess of 10% (namely experiments 1, 6 and 8) the Wooldridge estimator has a smaller bias in $\widehat{\gamma}$ than the other two and the Orme estimator slightly smaller than the Heckman estimator. The Wooldridge estimator does not have a bias of 10% in any of the experiments. However for each of these experiments, the Wooldridge estimator of γ has the largest bias of the three estimators when T is increased to 5 and there is nothing to choose between them when

T is further increased to 8.

In summary, when T and/or N are reasonably large (e.g. $T \geq 6$, $N \geq 800$), the relative bias in $\widehat{\gamma}$ is generally fairly small for all three estimators. For smaller samples, the biases are larger, but none of the estimators dominates the other two in all experiments.

More light can potentially be thrown on the differences in performance of the estimators by the next group of experiments, which take the initial condition period to coincide with the start of the process and can then vary r directly without changing other parameters. (The data generation process used in experiments 1-17 allowed the process to settle down prior to the estimation period by discarding the first 25 periods prior to the estimation sample.) The results for this group of experiments are given in Table 4. (They all also use T=5.) In experiment 18, r is set to zero and hence the initial condition is exogenous. The initial observation y_{i0}^* was drawn from $N(-0.45, 1.0)$ and β_0 was set equal to -1.0. These intercepts were chosen to give period-by-period sample means for the observed y similar to the base experiment. The rest of the parameters are the same as in the base experiment.

Relative to experiment 2, all three estimators of γ show an improvement in the relative bias, which is almost zero for all three estimators. However, despite this, the root mean square error worsens for all three estimators. In contrast there is a slight increase in the relative bias in the estimators of β, to about 4% for all three estimators, but a slight reduction in the root mean square error.

In the remaining experiments reported in Table 4 the initial condition is treated as endogenous by allowing correlation between the error term in the initial observation equation and that in the equation for the subsequent time periods. This is achieved by generating the initial observation using $y_{i0}^* = -0.45 + r\alpha_i + \sqrt{1-r^2}u_{i0}$ where $u_{i0} \sim N(0,1)$. (The exogenous initial condition case is given by $r = 0$.)[12] For small values of r, there are no significant differences between the three estimators. However, as r is increased, while all three estimators of γ deteriorate, the Heckman estimator worsens slightly more than the other two in terms of bias. The root mean square errors for the three estimators though are virtually identical. For $\widehat{\beta}$ the relative bias and root mean square error change relatively little as r is increased and the differences between the three estimators in both of these are narrow.

In summary, as the initial conditions problem becomes more serious (as measured by the correlation between the equation errors in the initial period and later periods), the Heckman estimator deteriorates somewhat more than the other two in terms of the relative biases. However, the root mean square errors for all three estimators are very similar.

Judged across the full set of experiments conducted, none of the three estimators dominates the other two in all cases, or even in a majority of cases.

[12] When x was included in the generation process for the initial observation, the results were very similar to those given here.

5 Conclusions

This paper presents a convenient shortcut method for implementing the Heckman estimator of the dynamic random effects probit model using standard software. This removes the need for separate programming and puts this estimator on a similar footing to the simpler estimators suggested by Orme and Wooldridge based on alternative approximations. The choice between these estimators can therefore be based on performance rather than availability or ease of use. An empirical illustration has been presented in section 3 and a set of simulation experiments in section 4. The former suggests that it is advantageous to allow for correlated random effects using the approach of Mundlak (1978), but that once this is done, the three estimators provide similar results. The simulation experiments suggest that none of the three estimators dominates the other two in all cases. In most cases all three estimators display satisfactory performance, except when T is very small.

References

Alessie, R. Hochguertel, S. and van Soest, A. (2004). 'Ownership of Stocks and Mutual Funds: A Panel Data Analysis', *Review of Economics and Statistics*, Vol. 86, pp. 783-796.

Andrén, T. (2007). 'The persistence of welfare participation', Göteborg University, Working Papers in Economics, No. 266.

Arulampalam, W. (1999). 'A Note on estimated effects in random effect probit models', *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 597-602.

Arulampalam, W. and Bhalotra, S. (2006). 'Sibling death clustering in India: genuine scarring versus unobserved heterogeneity', *Journal of the Royal Statistical Society Series A*, Vol. 169, pp. 829-848.

Arulampalam, W., Booth, A.L. and Taylor, M.P. (2000). 'Unemployment persistence', *Oxford Economic Papers*, Vol. 52, pp. 24-50.

Audas, R., Barmby, T. and Treble, J. (2004). 'Luck, Effort, and Reward in an Organizational Hierarchy', *Journal of Labor Economics*, Vol. 22, pp. 379-395.

Benito, A. and Young, G. (2003). 'Hard times or great expectations? Dividend omissions and dividend cuts by UK firms', *Oxford Bulletin of Economics and Statistics*, Vol. 65, pp. 531-555.

Bernard, A.B. and Jensen J.B. (2004). 'Why some firms export', *Review of Economics and Statistics*, Vol. 86, pp. 561-569.

Butler, J.S. and Moffitt, R. (1982). 'A computationally efficient quadrature procedure for the one-factor multinomial probit model', *Econometrica*, Vol. 50, pp. 761-764.

Cappellari, L. and Jenkins, S.P. (2008). 'The dynamics of social assistance receipt: Measurement and modelling issues, with an application to Britain', ISER Essex, Working Paper No. 2008-34.

Chauvin, N. and Kraay, A. (2007). 'Who gets debt relief?', *Journal of the European Economic Association*, Vol. 5, pp. 333-342.

Chen, J. and Entrom Ost, C. (2005). 'Housing allowance and the recipient's homeownership: Evidence from a panel data study in Sweden', *Housing Studies*, Vol. 20, pp. 605-625.

Clark A.E. and Etilé, F. (2006). 'Don't give up on me baby: Spousal correlation in smoking behaviour', *Journal of Health Economics*, Vol. 25, pp. 958-978

Contoyannis, P., Jones, A.M. and Rice, N. (2004). 'The dynamics of health in the British Household Panel Survey', *Journal of Applied Econometrics*, Vol. 19, pp. 473-503.

Dorsett, R. (1999). 'An econometric analysis of smoking prevalence among lone mothers', *Journal of Health Economics*, Vol. 18, pp. 429-441.

Heckman, J.J. (1981a). 'Heterogeneity and state dependence,' in S. Rosen (ed.), *Studies in Labor Markets*, Chicago Press: Chicago, IL.

Heckman, J.J. (1981b). 'The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process', in C.F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press: Cambridge, MA, pp. 114-178.

Henley, A. (2004). 'Self-employment status: The role of state dependence and initial circumstances', *Small Business Economics*, Vol. 22, pp. 67-82.

Honoré, B.E. (1993). 'Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables', *Journal of Econometrics*, Vol. 59, pp. 35-61.

Honoré, B.E. and Kyriazidou, E. (2000). 'Panel data discrete choice models with lagged dependent variables', *Econometrica*. Vol. 68, pp. 839-874.

Hsiao, C. (2003). *Analysis of Panel Data*. Second Edition. Cambridge University Press: Cambridge, UK.

Hyslop, D.E. (1999). 'State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women', *Econometrica*, Vol. 67, pp. 1255-1294.

Islam, N. (2007). 'A dynamic Tobit model of female labour supply'. Göteborg University, Discussion Paper No. 259.

Lee, W. and Oguzoglu, U. (2007). 'Income support and stigma effects for young Australians', *Australian Economic Review*, Vol. 40, pp. 369-384.

Loudermilk, MS. (2007). 'Estimation of fractional dependent variables in dynamic panel data models with an application to firm dividend policy', *Journal of Business and Economic Statistics*, Vol. 25, pp. 462-472.

Mundlak, Y. (1978). 'On the pooling of time series and cross section data', *Econometrica*, Vol. 46, pp. 69-85.

Nerlove, M. (1971). 'Further evidence on the estimation of dynamic economic relations from a time series of cross sections', *Econometrica*, Vol. 39, pp. 359-382.

Orme, C.D. (1997). 'The initial conditions problem and two-step estimation in discrete panel data models', mimeo, University of Manchester.

Orme, C.D. (2001). 'Two-step inference in dynamic non-linear panel data models', mimeo, University of Manchester.

Propper, C. (2000). 'The demand for private health care in the UK', *Journal of Health Economics*, Vol. 19, pp. 855-876.

Pudney, S. (2008). 'The dynamics of perception: Modelling subjective well-being in a short panel', *Journal of the Royal Statistical Society, Series A*, Vol. 171, pp. 21-40.

Requena-Silvente, F. (2005). 'The decision to enter and exit foreign markets: Evidence from U.K. SMEs', *Small Business Economics*, Vol. 25, pp. 237-53.

StataCorp. (2008). Stata Statistical Software: Release 10, College Station, TX: StataCorp LP.

Stewart, M.B. (2007). 'Inter-related dynamics of unemployment and low-wage employment', *Journal of Applied Econometrics*, Vol. 22, pp. 511-531.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Wooldridge, J. (2005). 'Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity', *Journal of Applied Econometrics*, Vol. 20, pp. 39-54.

Table 1
Unemployment probability model: Alternative estimators

	[1] Probit	[2] RE probit	[3] Heckman	[4] Orme	[5] Wooldridge
Unemp(t-1)	1.837	1.536	1.048	1.107	1.062
	[0.074]	[0.122]	[0.130]	[0.115]	[0.115]
Education	-0.043	-0.050	-0.058	-0.054	-0.055
	[0.011]	[0.014]	[0.017]	[0.016]	[0.017]
Experience	-0.048	-0.068	-0.072	-0.064	-0.066
	[0.030]	[0.037]	[0.045]	[0.043]	[0.045]
Married	-0.186	-0.236	-0.309	-0.280	-0.092
	[0.066]	[0.082]	[0.100]	[0.093]	[0.227]
Health limits	0.429	0.503	0.585	0.569	0.289
	[0.093]	[0.114]	[0.133]	[0.126]	[0.185]
Local u/v	0.654	0.849	0.941	0.919	0.880
	[0.229]	[0.268]	[0.306]	[0.292]	[0.396]
$\widehat{e}$				0.459	
				[0.076]	
Unemp(0)					1.016
					[0.161]
ρ		0.225	0.430	0.354	0.357
		[0.065]	[0.063]	[0.044]	[0.043]
θ			0.882		
			[0.189]		
LogL	-1052.00	-1044.81	-1341.14	-1024.24	-1014.01

Estimators:

1. Pooled Probit
2. Standard Random Effects Probit (initial condition taken to be exogenous)
3. Heckman estimator, with x in all periods and 2 exogenous instruments in initial period equation
4. Orme estimator, with x in all periods and 2 exogenous instruments in initial period equation
5. Wooldridge estimator, with x in all periods included in z

Notes:

1. Full sample size = 10,092. (N = 1,682, T = 5.)
2. LogL in [3] is for joint model for all periods 0 to T. Those in other columns are for periods 1 to T only.

Table 2
Unemployment probability model: Alternative estimators with Mundlak correction for correlated individual effects

	[1] Probit	[2] RE probit	[3] Heckman	[4] Orme	[5] Wooldridge
Unemp(t-1)	1.811	1.500	1.009	1.074	1.062
	[0.075]	[0.124]	[0.130]	[0.115]	[0.115]
Education	-0.044	-0.052	-0.060	-0.056	-0.055
	[0.012]	[0.015]	[0.018]	[0.017]	[0.017]
Experience	-0.050	-0.072	-0.077	-0.070	-0.066
	[0.031]	[0.040]	[0.048]	[0.045]	[0.045]
Married	-0.041	-0.063	-0.095	-0.090	-0.092
	[0.194]	[0.212]	[0.231]	[0.226]	[0.227]
Health limits	0.211	0.254	0.299	0.287	0.289
	[0.158]	[0.174]	[0.189]	[0.185]	[0.185]
Local u/v	0.633	0.900	0.896	0.873	0.880
	[0.338]	[0.378]	[0.406]	[0.396]	[0.396]
$\widehat{e}$				0.469	
				[0.076]	
Unemp(0)					1.016
					[0.161]
ρ		0.232	0.439	0.357	0.357
		[0.066]	[0.063]	[0.044]	[0.043]
θ			0.885		
			[0.189]		
LogL	-1044.03	-1044.81	-1332.14	-1015.40	-1014.01

Notes:

1. Estimators as in Table 1 with x in all periods added to main equation.
2. Full sample size = 10,092. (N = 1,682, T = 5.)
3. LogL in [3] is for joint model for all periods 0 to T. Those in other columns are for periods 1 to T only.

Table 3: Simulation results

Experiment	Estimator	γ Relative Bias (%)	γ RMSE	β Relative Bias (%)	β RMSE
1) Base	Heckman	-12.63	0.264	1.85	0.257
	Wooldridge	-3.96	0.280	-5.95	0.262
	Orme	-8.48	0.274	-2.42	0.257
2) T=5	Heckman	-1.98	0.164	-1.57	0.170
	Wooldridge	-3.09	0.165	-3.55	0.172
	Orme	-0.65	0.164	-1.77	0.170
3) T=8	Heckman	0.34	0.154	-0.66	0.132
	Wooldridge	-0.46	0.155	-2.08	0.133
	Orme	0.66	0.155	-0.86	0.132
4) N=500	Heckman	7.60	0.177	-0.46	0.169
	Wooldridge	2.73	0.177	-6.92	0.181
	Orme	8.23	0.179	-1.30	0.169
5) N=1000	Heckman	0.04	0.117	4.99	0.128
	Wooldridge	0.07	0.122	-1.02	0.118
	Orme	0.20	0.119	3.99	0.125
6) σ_α=0.5	Heckman	-15.06	0.247	4.85	0.252
	Wooldridge	-6.18	0.248	-0.36	0.240
	Orme	-13.02	0.245	3.58	0.247
7) σ_α=1.5	Heckman	-3.28	0.276	-2.93	0.283
	Wooldridge	-5.00	0.321	-7.67	0.303
	Orme	-6.36	0.321	-6.81	0.299
8) γ=0.25	Heckman	-20.73	0.269	0.65	0.254
	Wooldridge	-6.31	0.283	-6.55	0.259
	Orme	-14.25	0.277	-3.21	0.253
9) γ=0.75	Heckman	-8.21	0.270	2.56	0.265
	Wooldridge	-1.11	0.288	-5.78	0.268
	Orme	-4.58	0.282	-2.11	0.262

Table 3 (continued): Simulation results

Experiment	Estimator	γ Relative Bias (%)	γ RMSE	β Relative Bias (%)	β RMSE
10) σ_α=0.5,	Heckman	-1.53	0.170	-2.07	0.169
$T = 5$	Wooldridge	-3.25	0.171	-3.27	0.170
	Orme	-1.17	0.170	-2.24	0.170
11) σ_α=1.5,	Heckman	-1.77	0.180	-1.44	0.178
$T = 5$	Wooldridge	-2.45	0.182	-3.29	0.182
	Orme	-0.07	0.183	-1.42	0.180
12) γ=0.25,	Heckman	-5.01	0.172	-1.71	0.167
$T = 5$	Wooldridge	-7.30	0.173	-3.64	0.168
	Orme	-3.18	0.173	-1.88	0.167
13) γ=0.75,	Heckman	-0.66	0.167	-1.25	0.171
$T = 5$	Wooldridge	-1.54	0.167	-3.26	0.173
	Orme	0.36	0.167	-1.49	0.171
14) σ_α=0.5,	Heckman	0.72	0.167	-0.54	0.138
$T = 8$	Wooldridge	-0.07	0.168	-1.67	0.138
	Orme	1.10	0.168	-0.72	0.138
15) σ_α=1.5,	Heckman	3.34	0.152	-3.71	0.136
$T = 8$	Wooldridge	-2.21	0.157	-2.57	0.137
	Orme	-1.38	0.158	-1.48	0.137
16) γ=0.25,	Heckman	0.88	0.157	-0.61	0.130
$T = 8$	Wooldridge	-0.55	0.156	-2.04	0.131
	Orme	1.27	0.157	-0.85	0.130
17) γ=0.75,	Heckman	-0.44	0.155	-0.50	0.135
$T = 8$	Wooldridge	-1.15	0.155	-1.90	0.136
	Orme	-0.19	0.155	-0.66	0.136

Notes:

1. 1000 Monte Carlo replications used in each experiment.

2. In the baseline experiment the parameter values were set at γ=0.5, β_1=-1, β_0=4, σ_α=1. The process starts at t=-25, the observed "initial condition" period is t=0, T=3 and N=200.

Table 4
Simulation results for alternative data generation process

Experiment	Estimator	γ Relative Bias (%)	γ RMSE	β Relative Bias (%)	β RMSE
18)$r = 0.0$	Heckman	0.04	0.185	3.79	0.146
	Wooldridge	-0.10	0.184	3.85	0.146
	Orme	-0.15	0.183	3.85	0.146
19) $r = 0.2$	Heckman	-0.73	0.183	3.84	0.146
	Wooldridge	-0.71	0.184	3.88	0.147
	Orme	-0.79	0.184	3.86	0.146
20) $r = 0.4$	Heckman	-1.57	0.184	4.00	0.149
	Wooldridge	-1.28	0.185	3.87	0.148
	Orme	-1.38	0.184	3.88	0.148
21) $r = 0.6$	Heckman	-2.96	0.185	4.14	0.151
	Wooldridge	-2.14	0.186	3.78	0.149
	Orme	-2.27	0.186	3.81	0.149
22) $r = 0.8$	Heckman	-4.37	0.193	4.26	0.153
	Wooldridge	-1.91	0.193	3.43	0.149
	Orme	-2.04	0.192	3.52	0.149

Notes:

1. 1000 Monte Carlo replications used in each experiment.

2. In experiments 18-22 there were no run-in periods, i.e. the process was started at t=0. The initial observation y^*_{i0} was generated in the exogenous initial condition experiment (18) as $N(-0.45, 1)$ and in the endogenous initial conditions experiments (19-22) as $y^*_{i0} = -0.45 + r\alpha_i + \sqrt{1-r^2}u_{i0}$ where $u_{i0} \sim N(0,1)$. β_0 was set equal to -1.0. These values were chosen to gives averages for y similar to the base experiment. T=5, N=200 for all experiments in this table.

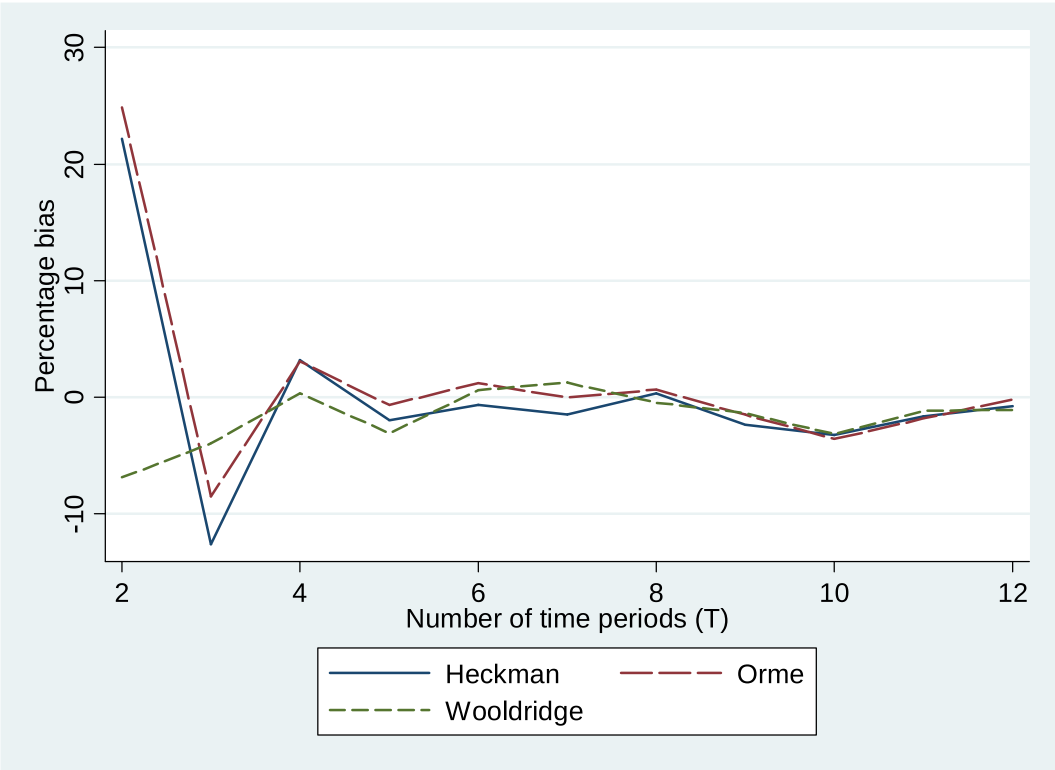

Figure 1: Percentage Bias in Estimates of γ by *T* for *N*=200

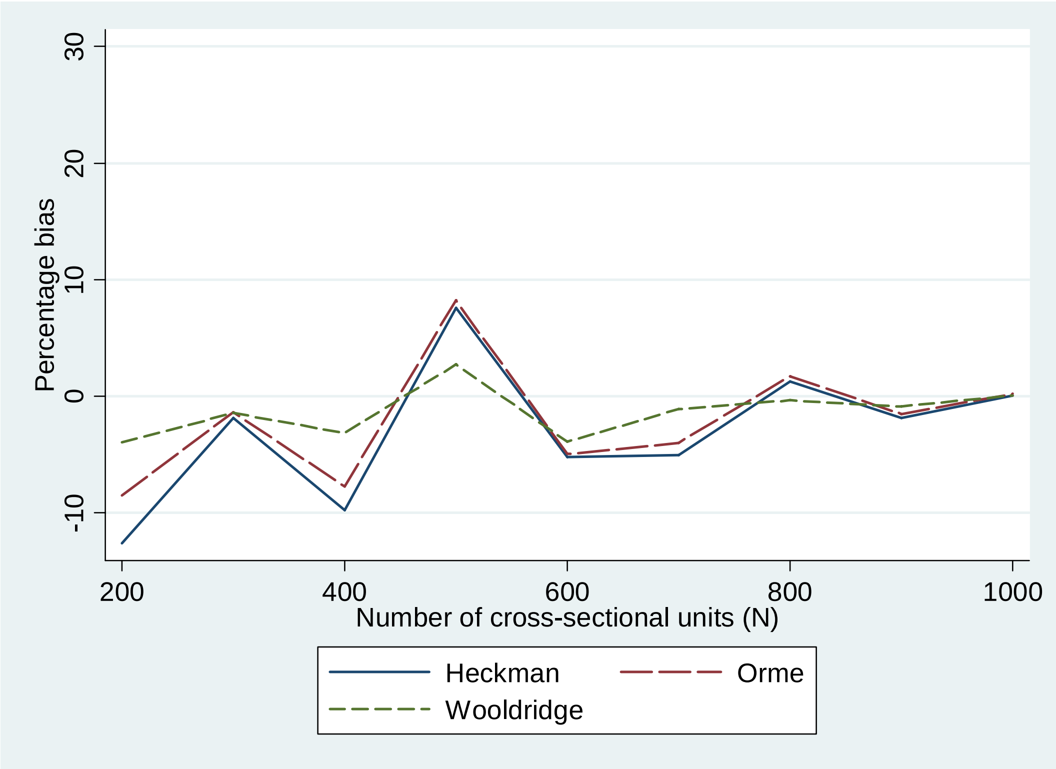

Figure 2: Percentage Bias in Estimates of γ by *N* for *T*=3

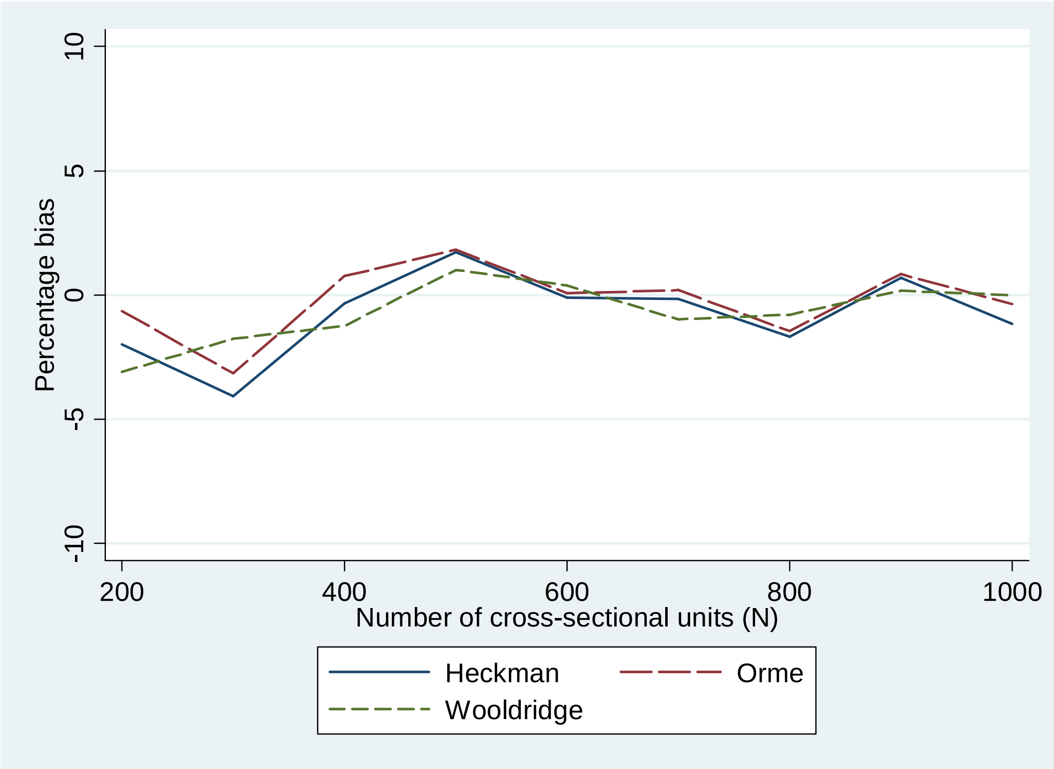

Figure 3: Percentage Bias in Estimates of γ by *N* for *T*=5

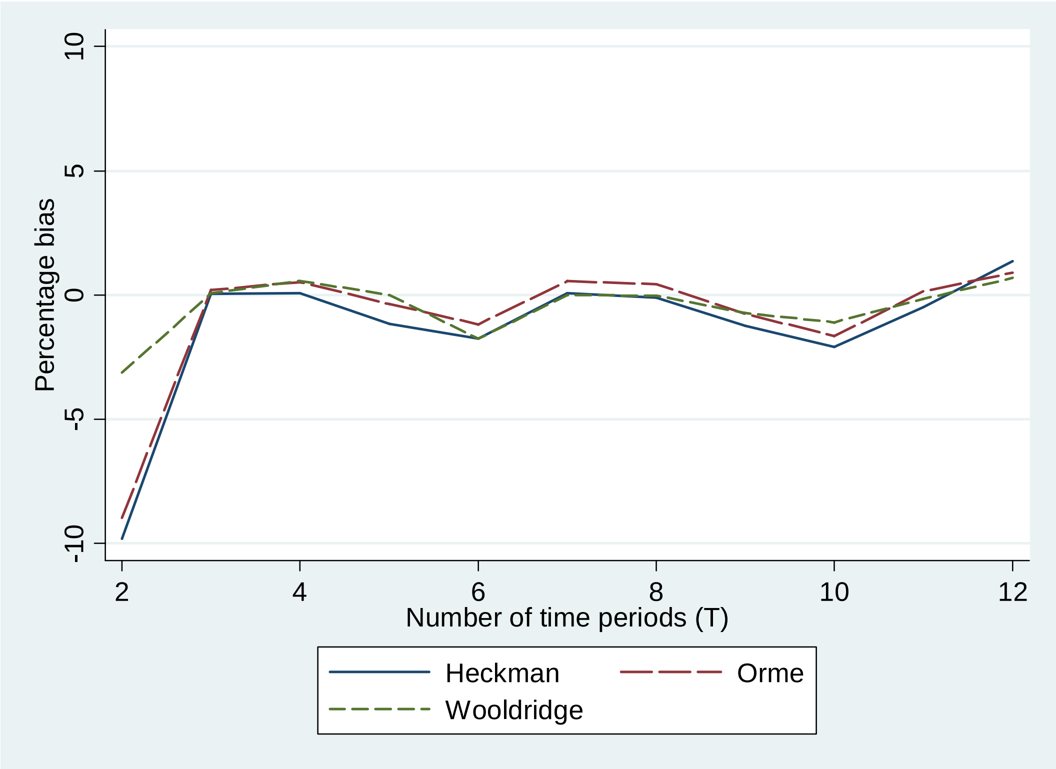

Figure 4: Percentage Bias in Estimates of γ by *T* for *N*=1000