#2007-007
The Behavior of the Maximum Likelihood Estimator of Dynamic
Panel Data Sample Selection Models

Wladimir Raymond, Pierre Mohnen, Franz Palm & Sybrand Schim van der Loeff

The Behavior of the Maximum Likelihood Estimator of Dynamic Panel Data Sample Selection Models[2]

Wladimir Raymond[3], Pierre Mohnen[4] , Franz Palm[5] and Sybrand Schim van der Loeff [6]

Abstract

This paper proposes a method to implement maximum likelihood estimation of the dynamic panel data type 2 and 3 tobit models. The likelihood function involves a two-dimensional indefinite integral evaluated using "two-step" Gauss-Hermite quadrature. A Monte Carlo study shows that the quadrature works well in finite sample for a number of evaluation points as small as two. Incorrectly ignoring the individual effects, or the dependence between the initial conditions and the individual effects results in an overestimation of the coefficients of the lagged dependent variables. An application to incremental and radical product innovations by Dutch business firms illustrates the method.

Maastricht Economic and social Research and training centre on Innovation and Technology, UNU-MERIT

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at the Centre to stimulate discussion on the issues raised.

[3]University of Maastricht, W.Raymond@KE.unimaas.nl
[4]University of Maastricht, MERIT and CIRANO,
MERIT, University of Maastricht, P.O. Box 616 6200 MD Maastricht, The Netherlands;
Tel: +31 43 388 3869; Fax: + 31 43 388 4905; P.Mohnen@MERIT.unimaas.nl
[5]University of Maastricht and CESifo fellow, F.Palm@KE.unimaas.nl
[6]University of Maastricht, S.Loeff@KE.unimaas.nl

1   Introduction

The ongoing development of the relevant econometric methodology for panel data has increased the choice for the applied researcher as to which model to choose in any given situation. Whereas this choice should be driven by the question which underlying assumptions are appropriate in the case at hand, optimality properties of the estimators are typically known only for large sample sizes. In order to have some guidance in the choice process it is important to know the implications, in terms of ease of computation, finite sample properties and the robustness to deviations of the assumptions, of employing a particular method. In this paper the behavior of the maximum likelihood (ML) estimator is investigated by means of Monte Carlo simulations in a general model that encompasses a wide set of models that have been considered in the literature. The generic form of the model consists of a regression equation, henceforth referred to as the equation of interest, and a selection equation. Both equations contain a lagged dependent variable and unobserved individual effects possibly correlated with the explanatory variables. The selection rule may be of the binary type or of the censoring type.[1] In terms of the Amemiya (1984) typology, the models considered in this paper are the dynamic panel data extensions of the type 2 and type 3 tobit models.[2] Variants of these types of models have been widely used, mostly in labor economics but also in other areas, but only in a static or "partial" dynamic framework (see Table 1).

This study contributes to the existing literature in a number of ways. Firstly, a practical method is provided to implement maximum likelihood estimation of the parameters of a panel sample selection model in which random individual effects and the lagged dependent variable are included in the equation of interest as well as in the selection equation. The selection rule may be either of the binary or of the censoring type.[3] To handle the initial conditions problem, Wooldridge's (2005) solution to this problem in single equation dynamic nonlinear panel data models is extended to models containing more than one equation. Gaussian quadrature is used to evaluate the resulting two-dimensional indefinite integral. Secondly, the paper reports on a Monte Carlo investigation into the behavior in finite samples of the proposed method. It is shown that using Wooldridge's solution to deal with the dependence between the initial conditions and the individual effects is successful in recovering the parameters of a process that generates the data according to a stationary process. For the likelihood functions generated in the Monte Carlo

---

[1]Following the literature the term censoring model will be used throughout the paper although it would, in line with Wooldridge (2005), be preferable to speak of corner solution models since the lagged dependent variable considered in this paper is the observed - rather than the latent - variable.

[2]The type 3 tobit differs from the type 2 tobit in that the selection into the sample is made through a censored (partially continuous) variable instead of a binary variable. The former model uses more information to estimate the equation of interest than the latter. Hence, we should expect the estimation of the equation of interest to be more accurate in the former model than in the latter.

[3]These are the types of selection rule that are mostly considered in the literature (see Table 1). Other types of selection rule are ordered, multinomial, and those based on multiple indices (Vella, 1998).

experiments, Gaussian quadrature, implemented as two successive Gauss-Hermite approximations, works well for a number of evaluation points as small as two.[4] Evidently, when using real data more than two evaluation points may be necessary to attain the required numerical accuracy. Nevertheless, the Monte Carlo results indicate that ML estimation of the parameters in a random-effects dynamic sample selection model is feasible and could conceivably be incorporated as a command in a general use software package such as Stata. Thirdly, the sensitivity of the ML estimator to misspecification is also investigated. It is shown that incorrectly ignoring the individual effects, or the dependence between the initial conditions and the individual effects results in an overestimation of the coefficients of the lagged dependent variable in both equations. Fourthly, the paper deals with the estimation of type 2 and type 3 tobit in a "full" dynamic framework, and provides for the first time an application of the dynamic type 3 tobit to the economics of innovation.

The remainder of the paper is organized as follows. We discuss the most cited studies on panel data sample selection models in Section 2. Section 3 describes the model and Section 4 describes its maximum likelihood estimation. The small sample behavior of the estimator is studied in a Monte Carlo study in Section 5, an application is provided in Section 6, and Section 7 concludes. The Gauss-Hermite quadrature is explained in Appendix A.

2 Literature

Most of the studies on panel data sample selection models are of the type 2 or type 3 tobit (see Table 1), which are estimated by ML or two-step least squares (Heckman, 1979), or by variants of the two-step Heckman type estimator.

For instance, Hausman and Wise (1979) describe a two-period model of attrition and estimate by ML the impact of attrition on earnings. Attrition is defined as the inability to observe the dependent variable of the equation of interest in the second period.[5] Ridder (1990) generalizes the model of Hausman and Wise to allow for more than two periods and attrition to occur also in the first period. Hence Ridder's model can be used as a standard attrition or sample selection model. Nijman and Verbeek (1992) apply this model to study the effect of nonresponse on Dutch households' consumption. Verbeek (1990) assumes random and fixed individual effects in the selection and regression equations respectively, and estimates by ML a transformed model with a selection equation written in levels and a within-transformed equation of interest. Verbeek's

---

[4]The use of Gauss-Hermite quadrature, instead of simulated likelihood, is motivated by the finding of Guilkey and Murphy (1993) that, for the same accuracy, the former method is 5 times as fast as the latter (see also Greene, 2004b).

[5]Attrition in the first period is usually referred to as sample selection.

model is criticized by Zabel (1992) who considers random individual effects correlated with the explanatory variables in both the selection equation and the equation of interest and estimates the resulting model by ML. All these studies are of limited use to the applied researcher who wants to implement the ML estimator as they present only the general expression of the likelihood function which involves a multiple integral that is yet to be calculated. Our study makes a step in this direction.

The two-step Heckman type estimator consists, in a first step, in estimating the selection equation and constructing an estimate of a selection correction term that is included as a regressor in the equation of interest which, in a second step, is estimated using ordinary least squares (OLS) regression. Wooldridge (1995) estimates an augmented equation of interest written in levels by pooled OLS where estimates of the selection correction terms are obtained, in a first step, using a probit model for each period. Kyriazidou (1997) and Honoré and Kyriazidou (2000) propose to estimate, by kernel-weighted least squares, a pairwise differenced equation of interest for individuals that are selected into the sample and have "the same" selection equation index in two different periods. Under a *conditional exchangeability* assumption, the sample selection is time-invariant for such individuals so that differencing the equation of interest over time wipes out, not only the individual effects, but also the sample selection effect. In order to construct the kernel weights, the parameters of the selection equation are estimated, in a first step, using conditional logit or smoothed conditional maximum score. Rochina-Barrachina (1999) applies OLS to a pairwise differenced equation of interest augmented with two selection correction terms for individuals that are selected into the sample in two different periods. Estimates of these selection correction terms are obtained, in a first step, using a bivariate probit for each combination of time periods. The studies cited so far consider sample selection models with a binary selection rule (type 2 tobit).

Heckman and MaCurdy (1980) estimate by ML a fixed-effects type 3 tobit and apply it to a wage equation with a labor supply (the number of hours worked) equation as the selection equation. Ai and Chen (1992) and Honoré and Kyriazidou (2000) estimate the same model using symmetrically trimmed least squares (STLS), and Wooldridge (1995) uses the same estimator as in his type 2 tobit case where estimates of the selection correction terms are now obtained, in a first step, using a tobit model for each period. A Monte Carlo study comparing the estimators of Wooldridge (1995), Honoré and Kyriazidou (2000) and a semiparametric first-difference estimator is provided by Lee (2001).

The above-mentioned studies on panel data sample selection models all assume strict exogeneity of the explanatory variables in the equation of interest. Models with censored endogenous explanatory variables are studied by Vella and Verbeek (1999) and Askildsen *et al.* (2003), while

Table 1: *Studies on sample selection models with panel data*

Study	Selection rule	Dynamic or static	Individual effects	Estimation method* of the equation of interest	Application	
					Selection equation	Equation of interest
Hausman and Wise (1979)	binary	static	random	ML	attrition equation	earnings equation
Heckman and MaCurdy (1980)	censored	static	fixed	ML	labor supply equation	wage equation
Ridder (1990)	binary	dynamic**	random	ML; OLS with SBC	n.a.	n.a.
Verbeek (1990)	binary	static	hybrid†	ML	n.a.	n.a.
Ai and Chen (1992)	censored	static	fixed	STLS	n.a.	n.a.
Nijman and Verbeek (1992)	binary	dynamic**	random	ML; FGLS with SBC	nonresponse equation	consumption equation
Zabel (1992)	binary	static	random	ML	n.a.	n.a.
Wooldridge (1995)	binary; censored	static	fixed	pooled OLS with SBC	n.a.	n.a.
Kyriazidou (1997)‡	binary	static	fixed	kernel weighted LS	n.a.	n.a.
Rochina-Barrachina (1999)‡	binary	static	fixed	OLS with SBC	n.a.	n.a.
Vella and Verbeek (1999)	binary; censored	dynamic**	random	OLS with SBC; CML	labor supply equation	wage equation
Dustmann and Rochina-Barrachina (2000)‡	binary	static	fixed	MD with SBC; kernel weighted IV; IV and GMM with SBC	labor supply participation decision	wage equation
Honoré and Kyriazidou (2000)‡	binary; censored	static	fixed	kernel weighted LS; STLS	n.a.	n.a.
Kyriazidou (2001)‡	binary	dynamic	fixed	kernel weighted GMM	n.a.	n.a.
Lee (2001)‡	censored	static	fixed	semiparametric FD	n.a.	n.a.
Askildsen et al. (2003)‡	binary	static	fixed	kernel weighted LS and IV	labor supply participation decision	labor supply equation
Semykina and Wooldridge (2005)	binary	static	fixed	fixed-effects and pooled 2SLS with SBC	labor supply participation decision	wage equation
Gayle and Viauroux (2007)	binary	dynamic	fixed	SIV; SMD	n.a.	n.a.

*ML, OLS, SBC, STLS, FGLS, LS, CML, MD, IV, GMM, FD, SIV, SMD and 2SLS respectively stand for maximum likelihood, ordinary least squares, selection bias correction, symmetrically trimmed least squares, feasible generalized least squares, least squares, conditional maximum likelihood, minimum distance, instrumental variable, generalized method of moments, first difference, sieve instrumental variable, sieve minimum distance and two-stage least squares. **Only the selection (or attrition) equation allows for dynamics. †The selection and the outcome equation are random and fixed respectively. ‡In these studies, the estimation is based on the pairwise differenced equation of interest.

Dustmann and Rochina-Barrachina (2000) and Semykina and Wooldridge (2005) study models with continuous endogenous explanatory variables. Vella and Verbeek (1999) estimate by OLS with selection bias correction (SBC) a wage equation where labor supply enters the set of explanatory variables, and where estimates of the selection correction terms are obtained, in a first step, using a dynamic tobit model of labor supply. Askildsen *et al.* (2003) estimate the wage elasticity of labor supply for Norwegian nurses using an instrumental variable (IV) version of the Kyriazidou (1997) method. Dustmann and Rochina-Barrachina (2000) and Semykina and Wooldridge (2005) estimate the effect of experience, which is assumed to be endogenous, on wages. The former study extends the Wooldridge (1995), Kyriazidou (1997) and Rochina-Barrachina (1999) estimators in order to account for endogeneity in the explanatory variables, while the latter uses a pooled and a fixed-effects two-stage least squares estimator with SBC. Finally, sample selection models that allow for dynamics in both the selection equation and the equation of interest are studied by Kyriazidou (2001) who uses a two-step kernel weighted general method of moments (GMM) estimator, and Gayle and Viauroux (2007) who use a sieve instrumental variable (SIV) and a sieve minimum distance (SMD) estimator.

As Table 1 clearly shows, sample selection models with panel data are hardly studied in a "full" dynamic framework where both the selection equation and the equation of interest include a lagged dependent variable. Two exceptions are Kyriazidou (2001) and Gayle and Viauroux (2007) who use semiparametric estimators for a fixed-effects dynamic panel data type 2 tobit. This paper takes another route using ML to estimate a random-effects dynamic panel data type 2 tobit. Furthermore, while the literature considers only a static or "partial" dynamic panel data type 3 tobit, we consider the estimation by ML of the model in a "full" dynamic framework. An application of the model to study the dynamics of innovation in Dutch manufacturing is also provided for the first time.

3 The model

The dynamic panel data sample selection model studied in this paper consists of two latent dependent variables d_{it}^* and y_{it}^* written as

$$d_{it}^* = \rho d_{i,t-1} + \delta' \mathbf{w}_{it} + \eta_i + \epsilon_{1it}, \tag{1}$$

$$y_{it}^* = \gamma y_{i,t-1} + \beta' \mathbf{x}_{it} + \alpha_i + \epsilon_{2it}, \tag{2}$$

with observed counterparts d_{it} and y_{it}, and $i = 1, ...N$; $t = 1, ...T$. Equation (1) is the selection equation that determines whether individual i is included in the sample on which the estimation of the equation of interest (eq. (2)) is based at period t. It is a function of past selection outcome ($d_{i,t-1}$), strictly exogenous explanatory variables ($\mathbf{w}_{it}$), time-invariant unobserved individual effects (η_i) and other time-variant unobserved variables (ϵ_{1it}). The scalar ρ and the vector δ' capture respectively the effects of past selection outcome and the explanatory variables on the current selection process, and are to be estimated. The equation of interest depends on its past outcome ($y_{i,t-1}$), strictly exogenous explanatory variables ($\mathbf{x}_{it}$), time-invariant unobserved individual effects (α_i) and other time-variant unobserved variables (ϵ_{2it}), and is observed only when d_{it}^* is positive, i.e.

$$y_{it} = \mathbf{1}[d_{it}^* > 0]y_{it}^*, \tag{3}$$

where $\mathbf{1}[...]$ is the indicator function with value one if the expression between square brackets is true, and zero otherwise. The scalar γ and the vector β' capture respectively the effects of past outcome and explanatory variables on current outcome, and are to be estimated. Since a fully parametric approach is considered in this study, there is no exclusion restriction in the vector of strictly exogenous explanatory variables. In other words, $\mathbf{w}_{it}$ and $\mathbf{x}_{it}$ may be the same, totally different or may have common explanatory variables.

Two types of selection rule are considered in this study, namely binary and censored. When the selection rule is binary, only the sign of the current selection process is observed and the current selection outcome is defined as

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

while in the censored case, not only the sign but also the actual value of the current selection process is observed whenever it is positive and the current selection outcome is defined as

$$d_{it} = \mathbf{1}[d_{it}^* > 0]d_{it}^*. \tag{5}$$

Amemiya (1984) refers to the model described in equations (1), (2), (3) and (4) as type 2 tobit, and the one described in equations (1), (2), (3) and (5) as type 3 tobit. We now turn to the estimation technique.

4 Maximum likelihood estimation

Two difficulties arise when estimating dynamic panel data sample selection models, namely the presence of unobserved individual effects and the treatment of the initial observations.[6]

One way of handling the presence of unobserved individual effects is to create a dummy variable for each individual i and estimate the corresponding parameters η_i and α_i, together with the other parameters of the model, by maximum likelihood assuming a joint distribution for the error terms ϵ_{1it} and ϵ_{2it}. This approach is referred to as fixed-effects and has two shortcomings when the considered panel consists of a large N and a small T. The first one lies in the difficulty of computing the maximum likelihood estimator of the coefficients of possibly thousands of dummy variables. This computational problem can be overcome, for instance, by a two-step "zigzag" kind of likelihood maximization (Heckman and MaCurdy, 1980) or by "brute force" maximization (Greene, 2004a). These two computational methods cannot, however, solve the second problem of "incidental parameters" of the fixed-effects approach, namely the inconsistency of the maximum likelihood estimator of η_i and α_i when the number of periods T is small (Neyman and Scott, 1948). Unlike in the linear model, the inconsistency of the estimator of the individual effects carries over to the estimator of the slope parameters. Hence, the individual effects have to be conditioned out of the likelihood function so that the remaining parameters of the model can be consistently estimated by maximum likelihood. The resulting estimator is known as the conditional maximum likelihood estimator and is studied, for instance, by Chamberlain (1980) and Magnac (2004). The conditional likelihood approach is, however, restrictive for two reasons. First, there are very few nonlinear panel data models for which concentrating the likelihood with respect to the individual effects is possible. They are surveyed in the study by Lancaster (2000). Secondly, the approach works only under the assumption of strict exogeneity of the explanatory variables, which rules out the inclusion of lagged dependent variables as explanatory variables. The above-mentioned shortcomings of the fixed-effects approach may justify the use of a random-effects approach, where η_i and α_i are assumed to have a joint distribution.

A computational (and methodological) difficulty that arises in the random-effects approach is the so-called initial conditions problem. Two assumptions are often made on the initial conditions in the literature, namely they are exogenous or the process is in equilibrium. Neither assumption is satisfactory Hsiao (2003). Two approaches of handling the initial conditions problem are proposed by Heckman (1981b) and Wooldridge (2005) respectively. The first approach specifies a model for

---

[6]This study describes the difficulties that arise when estimating dynamic nonlinear panel data models and suggests solutions to them in a fully parametric framework. Details on how to handle the presence of unobserved individual effects and lagged dependent variables as regressors in a semiparametric framework are given in Arellano and Honoré (2001).

the initial conditions given the individual effects and the strictly exogenous explanatory variables. In empirical work, this model is often assumed to be similar to the model underlying the remaining process. For instance, if the underlying model is a dynamic panel data probit, the model at the initial period is assumed to be standard probit. A likelihood function which is marginal to both the individual effects and the initial conditions can be derived and maximized using standard numerical procedures. The second approach specifies a distribution for the individual effects conditional on the initial conditions and the strictly exogenous explanatory variables. In this case, the likelihood function is marginal to the individual effects but conditional on the initial conditions. Both approaches are legitimate and yield consistent estimates of the parameters of the model under the assumption of correct specification of the distribution of the errors. However, the Wooldridge approach is easier to implement, which has made it more popular in applied econometrics recently, and more flexible in the sense that it applies to a wide range of nonlinear dynamic panel data models and allows, unlike the Heckman approach, for individual effects to be correlated with the strictly exogenous explanatory variables (see Raymond, 2007). The likelihood function derived in the Wooldridge approach has the same structure for both the dynamic and the static versions of the nonlinear model. This study extends this approach to models that contain more than one equation. The approach is described as follows.

The individual effects are assumed, in each period, to be linear in the strictly exogenous explanatory variables and the initial conditions, i.e.

$$\eta_i = b_0^s + b_1^s d_{i0} + \mathbf{b}_2'^s \mathbf{w}_i + a_{1i}, \tag{6}$$

$$\alpha_i = b_0^r + b_1^r y_{i0} + \mathbf{b}_2'^r \mathbf{x}_i + a_{2i}, \tag{7}$$

where $\mathbf{w}_i' = (\mathbf{w}_{i1}', ..., \mathbf{w}_{iT}')$, $\mathbf{x}_i' = (\mathbf{x}_{i1}', ..., \mathbf{x}_{iT}')$, b_0^s, b_1^s, $\mathbf{b}_2'^s$, b_0^r, b_1^r and $\mathbf{b}_2'^r$ are to be estimated, and a_{1i} and a_{2i} are independent of $(d_{i0}, \mathbf{w}_i)$ and $(y_{i0}, \mathbf{x}_i)$ respectively.[7] The scalars b_1^s and b_1^r capture the dependence of the individual effects on the initial conditions. The vectors $(\epsilon_{1it}, \epsilon_{2it})'$ and $(a_{1i}, a_{2i})'$ are assumed to be independent of each other, and independently and identically distributed over time and across individuals following a normal distribution with mean zero and covariance matrices

$$\Omega_{\epsilon_1\epsilon_2} = \begin{pmatrix} \sigma_{\epsilon_1}^2 & \rho_{\epsilon_1\epsilon_2}\sigma_{\epsilon_1}\sigma_{\epsilon_2} \\ \rho_{\epsilon_1\epsilon_2}\sigma_{\epsilon_1}\sigma_{\epsilon_2} & \sigma_{\epsilon_2}^2 \end{pmatrix} \text{ and } \Omega_{a_1a_2} = \begin{pmatrix} \sigma_{a_1}^2 & \rho_{a_1a_2}\sigma_{a_1}\sigma_{a_2} \\ \rho_{a_1a_2}\sigma_{a_1}\sigma_{a_2} & \sigma_{a_2}^2 \end{pmatrix} \tag{8}$$

respectively. The parameters of the covariance matrices are also to be estimated. Hence, the

---

[7]The vectors of explanatory variables $\mathbf{w}_i$ and $\mathbf{x}_i$, in order to be included in equations (6) and (7), must be sufficiently time-variant, otherwise a collinearity problem will arise.

likelihood function of individual i, starting from $t = 1$ and conditional on the regressors and the initial conditions, is written as

$$L_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{t=1}^{T} L_{it}(d_{it}, y_{it}|d_{i0}, d_{i,t-1}, \mathbf{w}_i, y_{i0}, y_{i,t-1}, \mathbf{x}_i, a_{1i}, a_{2i}) g(a_{1i}, a_{2i}) da_{1i} da_{2i}, \qquad (9)$$

where $\prod_{t=1}^{T} L_{it}(d_{it}, y_{it}|d_{i0}, d_{i,t-1}, \mathbf{w}_i, y_{i0}, y_{i,t-1}, \mathbf{x}_i, a_{1i}, a_{2i})$ and $g(a_{1i}, a_{2i})$ denote respectively the likelihood function of individual i conditional on the individual effects, and the bivariate normal density function of (a_{1i}, a_{2i}). Define

$$A_{it} = \rho d_{i,t-1} + \delta' \mathbf{w}_{it} + b_0^s + b_1^s d_{i0} + \mathbf{b}_2'^s \mathbf{w}_i, \qquad (10)$$

$$B_{it} = \gamma y_{i,t-1} + \beta' \mathbf{x}_{it} + b_0^r + b_1^r y_{i0} + \mathbf{b}_2'^r \mathbf{x}_i, \qquad (11)$$

the individual likelihood function conditional on the individual effects is written as

$$\prod_{t=1}^{T} \Phi\left[-\left(\frac{A_{it} + a_{1i}}{\sigma_{\epsilon_1}}\right)\right]^{(1-d_{it})} \left[\frac{1}{\sigma_{\epsilon_2}} \phi\left(\frac{y_{it} - B_{it} - a_{2i}}{\sigma_{\epsilon_2}}\right)\right.$$
$$\left. \times \Phi\left(\frac{A_{it} + a_{1i} + \rho_{\epsilon_1\epsilon_2}\frac{\sigma_{\epsilon_1}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - a_{2i})}{\sigma_{\epsilon_1}\sqrt{1 - \rho_{\epsilon_1\epsilon_2}^2}}\right)^{d_{it}}\right], \qquad (12)$$

when the selection rule is binary and

$$\prod_{t=1}^{T} \Phi\left[-\left(\frac{A_{it} + a_{1i}}{\sigma_{\epsilon_1}}\right)\right]^{(1-s_{it})} \left[\frac{1}{\sigma_{\epsilon_2}} \phi\left(\frac{y_{it} - B_{it} - a_{2i}}{\sigma_{\epsilon_2}}\right)\right.$$
$$\left. \times \frac{1}{\sigma_{\epsilon_1}\sqrt{1 - \rho_{\epsilon_1\epsilon_2}^2}} \phi\left(\frac{d_{it} - A_{it} - a_{1i} - \rho_{\epsilon_1\epsilon_2}\frac{\sigma_{\epsilon_1}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - a_{2i})}{\sigma_{\epsilon_1}\sqrt{1 - \rho_{\epsilon_1\epsilon_2}^2}}\right)\right]^{s_{it}}, \qquad (13)$$

when the selection rule is censored.[8] In equations (12) and (13), ϕ and Φ denote respectively the univariate standard normal density and cumulative distribution functions, s_{it} is defined in the type 3 tobit as

$$s_{it} = \mathbf{1}[d_{it}^* > 0].^9 \qquad (14)$$

The double integral in equation (9) can be approximated, along the lines of Butler and Moffitt (1982), by "two-step" Gauss-Hermite quadrature (see Appendix A) so that the random-effects

---

[8]The individual likelihood function conditional on the individual effects is the product over time of individual cross-sectional likelihood functions. These latter functions are derived for the type 2 and 3 tobit models in Amemiya (1984).

[9]In the type 2 tobit model, equations (4) and (14) are equivalent so that $s_{it} = d_{it}$, and σ_{ϵ_1} cannot be identified and is set to 1 in equation (12).

individual likelihood functions of the type 2 and type 3 tobit models become respectively

$$
L_i \simeq \frac{\sqrt{1-\rho_{a_1 a_2}^2}}{\pi} \sum_{p=1}^{P} w_p \left\{ \prod_{t=1}^{T} \left[\frac{1}{\sigma_{\epsilon_2}} \phi \left(\frac{y_{it} - B_{it} - a_p \sigma_{a_2} \sqrt{2(1-\rho_{a_1 a_2}^2)}}{\sigma_{\epsilon_2}} \right) \right]^{d_{it}} \right.
$$
$$
\times \sum_{m=1}^{M} w_m \left\{ e^{2\rho_{a_1 a_2} a_p a_m} \prod_{t=1}^{T} \Phi \left[- \left(A_{it} + a_m \sigma_{a_1} \sqrt{2(1-\rho_{a_1 a_2}^2)} \right) \right]^{(1-d_{it})} \right.
$$
$$
\left. \left. \times \ \Phi \left(\frac{N_{it} + \frac{\rho_{\epsilon_1 \epsilon_2}}{\sigma_{\epsilon_2}} (y_{it} - B_{it} - a_p \sigma_{a_2} \sqrt{2(1-\rho_{a_1 a_2}^2)})}{\sqrt{1-\rho_{\epsilon_1 \epsilon_2}^2}} \right)^{d_{it}} \right\} \right\} \tag{15}
$$

and

$$
L_i \simeq \frac{\sqrt{1-\rho_{a_1 a_2}^2}}{\pi} \sum_{p=1}^{P} w_p \left\{ \prod_{t=1}^{T} \left[\frac{1}{\sigma_{\epsilon_2}} \phi \left(\frac{y_{it} - B_{it} - a_p \sigma_{a_2} \sqrt{2(1-\rho_{a_1 a_2}^2)}}{\sigma_{\epsilon_2}} \right) \right]^{s_{it}} \right.
$$
$$
\times \sum_{m=1}^{M} w_m \left\{ e^{2\rho_{a_1 a_2} a_p a_m} \prod_{t=1}^{T} \Phi \left[- \left(\frac{A_{it} + a_m \sigma_{a_1} \sqrt{2(1-\rho_{a_1 a_2}^2)}}{\sigma_{\epsilon_1}} \right) \right]^{(1-s_{it})} \right.
$$
$$
\left. \left. \times \left[\frac{1}{\sigma_{\epsilon_1} \sqrt{1-\rho_{\epsilon_1 \epsilon_2}^2}} \phi \left(\frac{d_{it} - N_{it} - \rho_{\epsilon_1 \epsilon_2} \frac{\sigma_{\epsilon_1}}{\sigma_{\epsilon_2}} (y_{it} - B_{it} - a_p \sigma_{a_2} \sqrt{2(1-\rho_{a_1 a_2}^2)})}{\sigma_{\epsilon_1} \sqrt{1-\rho_{\epsilon_1 \epsilon_2}^2}} \right) \right]^{s_{it}} \right\} \right\}, \tag{16}
$$

where w_m, w_p, a_m and a_p are respectively the weights and abscissas of the first- and second-step Gauss-Hermite quadrature with M and P being the first- and second-step total number of integration points, and N_{it} is written as

$$
N_{it} = A_{it} + a_m \sigma_{a_1} \sqrt{2(1-\rho_{a_1 a_2}^2)}. \tag{17}
$$

The product over i of the approximate likelihood functions (15) and (16) can be maximized using standard numerical procedures to obtain estimates of the parameters of the dynamic type 2 and type 3 tobit models.

Our approach handles the estimation, by maximum likelihood, of a wide range of linear and nonlinear panel data models. Indeed, equations (15) and (16) encompass the likelihood functions of the models which are obtained by restricting the values of certain parameters of the dynamic sample selection models (see Table 2). For instance, if $\rho_{a_1 a_2} = \rho_{\epsilon_1 \epsilon_2} = 0$ and the remaining parameters are unrestricted, the likelihood functions are the product of the likelihood functions of a dynamic random-effects probit and a dynamic random-effects linear regression (eq. (15)), and that of the likelihood functions of a dynamic random-effects type 1 tobit and a dynamic random-effects linear regression (eq. (16)). In other words, when there is no selection bias, estimating a dynamic type 2 (type 3) tobit amounts to estimating separately a dynamic probit (type 1 tobit)

Restricted model	Parameter restrictions		
	Individual effects*	Initial conditions**	Sample selection
The general model is dynamic type 2 (type 3) tobit†			
Dynamic††type 2 (type3) tobit	$\sigma_{a_1} \neq 0$; $\sigma_{a_2} \neq 0$; present	$b_1^s = b_1^r = 0$; exogenous	$\rho_{\epsilon_1 \epsilon_2} \neq 0$; $\rho_{a_1 a_2} \neq 0$; selection bias correction
Dynamic type 2 (type3) tobit	$\sigma_{a_1} = \sigma_{a_2} = 0$; absent	$b_1^s = b_1^r = 0$; exogenous	$\rho_{\epsilon_1 \epsilon_2} \neq 0$; $\rho_{a_1 a_2} \neq 0$; selection bias correction
Dynamic probit (type 1 tobit) + dynamic linear regression	$\sigma_{a_1} \neq 0$; $\sigma_{a_2} \neq 0$; present	$b_1^s \neq 0$; $b_1^r \neq 0$; endogenous	$\rho_{\epsilon_1 \epsilon_2} = \rho_{a_1 a_2} = 0$; no selection bias correction
Dynamic probit (type 1 tobit) + dynamic linear regression	$\sigma_{a_1} \neq 0$; $\sigma_{a_2} \neq 0$; present	$b_1^s = b_1^r = 0$; exogenous	$\rho_{\epsilon_1 \epsilon_2} = \rho_{a_1 a_2} = 0$; no selection bias correction
Dynamic probit (type 1 tobit) + dynamic linear regression	$\sigma_{a_1} = \sigma_{a_2} = 0$; absent	$b_1^s = b_1^r = 0$; exogenous	$\rho_{\epsilon_1 \epsilon_2} = \rho_{a_1 a_2} = 0$; no selection bias correction
Static††type 2 (type 3) tobit	$\sigma_{a_1} \neq 0$; $\sigma_{a_2} \neq 0$; present	n.a.	$\rho_{\epsilon_1 \epsilon_2} \neq 0$; $\rho_{a_1 a_2} \neq 0$; selection bias correction
Static type 2 (type 3) tobit	$\sigma_{a_1} = \sigma_{a_2} = 0$; absent	n.a.	$\rho_{\epsilon_1 \epsilon_2} \neq 0$; $\rho_{a_1 a_2} \neq 0$; selection bias correction
Static probit (type 1 tobit) + static linear regression	$\sigma_{a_1} \neq 0$; $\sigma_{a_2} \neq 0$; present	n.a.	$\rho_{\epsilon_1 \epsilon_2} = \rho_{a_1 a_2} = 0$; no selection bias correction
Static probit (type 1 tobit) + static linear regression	$\sigma_{a_1} = \sigma_{a_2} = 0$; absent	n.a.	$\rho_{\epsilon_1 \epsilon_2} = \rho_{a_1 a_2} = 0$; no selection bias correction

*The individual effects, when present, may be correlated $(\mathbf{b}_2'^s \neq 0; \mathbf{b}_2'^r \neq 0)$ or uncorrelated $(\mathbf{b}_2'^s = \mathbf{b}_2'^r = 0)$ with the explanatory variables. **When the model is static $(\rho = \gamma = 0)$ the initial conditions problem is not an issue, so $b_1^s = b_1^r = 0$. †The general model contains individual effects and the initial conditions are endogenous and correlated with the individual effects. ††"Hybrid" models with a static selection and a dynamic equation of interest (and vice versa) can also be estimated using our approach.

and a dynamic linear regression. The parameter restrictions listed in Table 2 can be tested using a standard likelihood ratio or Wald test.

We now present Monte Carlo simulations to study the small sample behavior of the maximum likelihood estimator of the type 2 and type 3 tobit models.

5   Monte Carlo study

The data are generated according to the following "true" model

$$d_{it}^* = \rho d_{i,t-1} + \delta w_{it} + \eta_i + \epsilon_{1it}, \tag{18}$$

$$y_{it}^* = \gamma y_{i,t-1} + \beta x_{it} + \alpha_i + \epsilon_{2it}, \tag{19}$$

where w_{it} and x_{it} are generated according to the standard normal distribution and are independent of each other, $(\epsilon_{1it}, \epsilon_{2it})$ is generated according to the bivariate normal distribution with mean zero and covariance matrix defined as in equation (8) (first matrix). The individual effects (η_i, α_i)

are generated independently of $(\epsilon_{1it}, \epsilon_{2it})$ according to the bivariate normal distribution with mean (d_{i0}, y_{i0}) and covariance matrix defined as in equation (8) (second matrix). The dependent variables of the equation of interest and the selection rules are defined as in equations (3-5), i.e., both the type 2 and type 3 tobit are considered. The true parameter values are $\delta = \beta = 1$, $\sigma_{a_1} = \sigma_{a_2} = \sigma_{\epsilon_2} = 0.5$, $\sigma_{\epsilon_1} = 1$, $\rho_{a_1 a_2} = 0.5$, $\rho_{\epsilon_1 \epsilon_2} = 0.8$, and two sets of values are chosen for ρ and γ, namely $\rho = \gamma = 0$ when the model is static, and $\rho = \gamma = 0.5$ when the model is dynamic. All dynamic and static data generating processes considered include individual effects.

Given these values of the parameters, sample selection is not constant over time, i.e. the selection outcome d_{it} (or s_{it}) has a rather high within standard deviation (around 0.40) for the (dynamic and static) type 2 and type 3 tobit. Furthermore, the censoring rate over the whole period, for both models, is approximately 40% and 50% in the dynamic and the static settings respectively. Finally, the correlation between the individual effects and the initial conditions for the type 2 and type 3 tobit is 0.70 and 0.80 in the selection equation, and 0.80 for both models in the equation of interest.

To summarize, the data generating process (DGP) at the initial period is standard type 2 (type 3) tobit while the remaining period DGP is dynamic type 2 (type 3) tobit with endogenous initial conditions correlated with the individual effects, and accounting for selection bias correction (SBC). The static DGP is obtained by setting the true values of ρ, γ to 0.

The estimation results of ρ, γ, δ, β, σ_{a_1}, σ_{a_2}, $\rho_{a_1 a_2}$ and $\rho_{\epsilon_1 \epsilon_2}$ based on 200 replications are shown in Tables 3-8 for the following sample sizes: $N = 500$ and $T = 4$, $N = 500$ and $T = 7$, and $N = 800$ and $T = 4$. We report the mean and standard deviation over the replications of these estimates.[10] In each replication, the estimation is based on $T-1$ periods, the first one being the initial period. We study the small sample behavior of the maximum likelihood estimator (MLE) of the above-mentioned parameters when the model is both correctly and incorrectly specified. More specifically, when the DGP is dynamic, misspecification includes a dynamic model with no individual effects but SBC; with individual effects, exogenous initial conditions and SBC; with individual effects, endogenous initial conditions but no SBC; and a static model with individual effects and SBC. When the DGP is static, misspecification includes a static model with no individual effects but SBC; with individual effects but no SBC; and a dynamic model with no individual effects but SBC, and with individual effects and endogenous initial conditions correlated with the individual effects and SBC.

We now discuss the Monte Carlo results.

---

[10] We do not report the estimate results of σ_{ϵ_1} (for the type 3 tobit), σ_{ϵ_2}, and the additional parameters of equations (6) and (7).

5.1 The MLE behavior when the model is correctly specified

The mean and standard deviation of the MLE of ρ, γ, δ, β, σ_{a_1}, σ_{a_2}, $\rho_{a_1 a_2}$ and $\rho_{\epsilon_1 \epsilon_2}$, when the model is correctly specified, are reported in Table 3.

The MLE of δ and β in both the dynamic and static type 2 and type 3 tobit is biased *towards* zero. The bias does not exceed 2% and 0.7% for δ and β in the type 2 tobit, and 3.5% and 0.8% in the type 3 tobit. There is no clear pattern of bias reduction as either T or N increases. In all cases but the dynamic type 2 tobit with $N = 800$ and $T = 4$, the estimate of β is on average closer to the true value than that of δ. In both the type 2 and type 3 tobit, the estimates of δ and β are closer to the true values in the static case than in the dynamic case, except in the dynamic type 2 tobit with $N = 800$ and $T = 4$ where the pattern shows up the other way round in the estimate of δ. From this, we can conclude that it is more difficult to estimate accurately the coefficients of the strictly exogenous explanatory variables in both the selection equation and the equation of interest of dynamic panel data sample selection models than those of their static counterparts.

The MLE of ρ and γ is also biased *towards* zero in both the dynamic type 2 and type 3 tobit. The bias does not exceed 2% and 4.7% for ρ and γ in the type 2 tobit, and 3.5% and 4.2% in the type 3 tobit. Like δ and β, there is no clear pattern of bias reduction as either T or N increases, and the MLE bias of ρ and γ is overall larger than that of δ and β. As a result, we can state that it is more difficult to estimate accurately the coefficients of the lagged dependent variables than those of the strictly exogenous variables in both the selection equation and the equation of interest of dynamic panel data sample selection models.

In both models, σ_{a_1} and σ_{a_2} are much less accurately estimated than ρ, γ, δ and β, even though their bias remains *towards* zero. For both models, the bias is smaller in the static case than in the dynamic case. In all cases but, for σ_{a_1} when $N = 500$ and $T = 7$, the bias is smaller in the type 2 tobit than in the type 3 tobit. The two models exhibit different patterns in the bias reduction as T increases. Indeed, the type 2 tobit shows an increase in the MLE bias of σ_{a_1} and σ_{a_2} which is more pronounced for σ_{a_1} than it is for σ_{a_2}. The increase in the MLE bias of σ_{a_1}, as T increases, confirms the results by Rabe-Hesketh *et al.* (2005) that show that the MLE of the standard deviation of the individual effects in random-effects probit models is biased *away* from zero for large cluster size T, with a bias that ranges from 19% for T as small as 10 to 178% for T as large as 500. The type 3 tobit on the other hand shows a decrease in the bias of the MLE of σ_{a_1} and σ_{a_2} which is more pronounced for σ_{a_1} than it is for σ_{a_2}. As N increases, the bias reduction pattern is not clear cut, especially in the type 3 tobit.

Finally, in both models, the MLE of $\rho_{a_1 a_2}$ is biased *away* from zero with a bias of about 25%,

Table 3: *ML estimates based on 200 replications and 2-point Gauss-Hermite quadrature: Data generating process*

(True value)*	Estimate	Estimated model							
		Static type 2 tobit		Dynamic type 2 tobit		Static type 3 tobit		Dynamic type 3 tobit	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
		N=500; T=4							
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	0.480	(0.086)	-	-	0.474	(0.023)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.541	(0.018)	-	-	0.540	(0.016)
$(\delta = 1.0)$	$\widehat{\delta}$	0.991	(0.050)	1.016	(0.055)	0.997	(0.027)	1.024	(0.025)
$(\beta = 1.0)$	$\widehat{\beta}$	1.001	(0.021)	1.007	(0.019)	1.002	(0.018)	1.008	(0.016)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.521	(0.077)	0.557	(0.096)	0.434	(0.070)	0.412	(0.075)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.448	(0.021)	0.409	(0.058)	0.421	(0.018)	0.371	(0.108)
$(\rho_{a_1a_2} = 0.5)$	$\widehat{\rho}_{a_1a_2}$	0.278	(0.014)	0.283	(0.041)	0.256	(0.017)	0.236	(0.070)
$(\rho_{\epsilon_1\epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1\epsilon_2}$	0.682	(0.047)	0.603	(0.072)	0.709	(0.024)	0.740	(0.022)
		N=500; T=7							
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	0.489	(0.058)	-	-	0.465	(0.015)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.545	(0.012)	-	-	0.539	(0.011)
$(\delta = 1.0)$	$\widehat{\delta}$	0.999	(0.041)	1.020	(0.043)	0.998	(0.021)	1.035	(0.018)
$(\beta = 1.0)$	$\widehat{\beta}$	1.000	(0.014)	1.005	(0.013)	1.000	(0.011)	1.005	(0.009)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.581	(0.053)	0.593	(0.065)	0.526	(0.052)	0.532	(0.054)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.446	(0.014)	0.397	(0.058)	0.429	(0.013)	0.391	(0.014)
$(\rho_{a_1a_2} = 0.5)$	$\widehat{\rho}_{a_1a_2}$	0.308	(0.014)	0.308	(0.045)	0.276	(0.016)	0.256	(0.021)
$(\rho_{\epsilon_1\epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1\epsilon_2}$	0.685	(0.038)	0.618	(0.054)	0.714	(0.017)	0.750	(0.013)
		N=800; T=4							
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	0.480	(0.062)	-	-	0.477	(0.016)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.541	(0.015)	-	-	0.542	(0.013)
$(\delta = 1.0)$	$\widehat{\delta}$	0.986	(0.038)	1.005	(0.042)	0.993	(0.021)	1.020	(0.019)
$(\beta = 1.0)$	$\widehat{\beta}$	1.000	(0.017)	1.006	(0.014)	0.999	(0.014)	1.008	(0.012)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.516	(0.064)	0.542	(0.068)	0.437	(0.054)	0.407	(0.059)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.447	(0.017)	0.409	(0.057)	0.419	(0.014)	0.387	(0.014)
$(\rho_{a_1a_2} = 0.5)$	$\widehat{\rho}_{a_1a_2}$	0.280	(0.010)	0.287	(0.041)	0.256	(0.011)	0.248	(0.013)
$(\rho_{\epsilon_1\epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1\epsilon_2}$	0.680	(0.035)	0.604	(0.053)	0.707	(0.019)	0.740	(0.017)

* $\rho = \gamma = 0$ in the static data generating process.

while that of $\rho_{\epsilon_1\epsilon_2}$ is biased *away* from zero in the type 2 tobit and *towards* zero in the type 3 tobit. This bias decreases only slightly as T increases, while there is no clear pattern of reduction as N increases.

To summarize, we can state that, when the model is correctly specified, the MLE bias of the parameters of panel data sample selection models is very small for a sample size as large as the one considered in the Monte Carlo study. The "two-step" Gauss-Hermite quadrature used to approximate the likelihood function works very well, even for a number of integration points, in each step, as small as 2. However, the standard deviations of the individual effects and the sample selection terms are less accurately estimated than the remaining parameters of Table 3. Accuracy in the estimates can be gained by either increasing the number of integration points or using adaptive Gauss-Hermite quadrature in the spirit of Rabe-Hesketh *et al.* (2005).[11]

We now discuss the behavior of the MLE when the model is incorrectly specified.

5.2 The MLE behavior when the model is incorrectly specified

The mean and standard deviation of the MLE of ρ, γ, δ, β, σ_{a_1}, σ_{a_2}, $\rho_{a_1a_2}$ and $\rho_{\epsilon_1\epsilon_2}$, when the model is incorrectly specified, are reported in Tables 4-8. More specifically, we study the finite sample bias of the MLE when we fail to account for individual effects, when we assume exogenous initial conditions and when we fail to correct for selection bias. We also study the case where a static model is estimated under a dynamic DGP and vice versa. All DGPs considered include individual effects.

5.2.1 No individual effects

Table 4 reports the mean and standard deviation of the MLE of ρ, γ, δ, β and $\rho_{\epsilon_1\epsilon_2}$ when the individual effects, while being present in the DGP, are ignored in the estimation of the model, i.e., σ_{a_1}, σ_{a_2} and $\rho_{a_1a_2}$ are all assumed to be zero.

The MLE of β remains biased *towards* zero in all cases. In other words, the MLE bias of the coefficient of the strictly exogenous explanatory variable in the equation of interest is unaffected if we fail to account for individual effects. This result holds regardless of the type of the selection rule, and regardless of whether the model is dynamic or static. The MLE of δ also remains biased *towards* zero in the type 3 tobit and in the static type 2 tobit with a larger bias than that of β. In the dynamic type 2 tobit, however, the MLE of δ is biased *away* from zero with a bias that is

---

[11] In the study by Rabe-Hesketh *et al.* (2005), adaptive Gauss-Hermite quadrature is shown to work better than normal Gauss-Hermite quadrature in static random-effects probit models when T or the equicorrelation is very large. The use of adaptive Gauss-Hermite quadrature can be generalized to sample selection models but is beyond the scope of this study.

Table 4: *ML estimates based on 200 replications and 2-point Gauss-Hermite quadrature: No individual effects*

(True value)*	Estimate	Estimated model							
		Static type 2 tobit		Dynamic type 2 tobit		Static type 3 tobit		Dynamic type 3 tobit	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
		N=500; T=4							
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	1.106	(0.061)	-	-	0.989	(0.019)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	1.046	(0.020)	-	-	1.047	(0.019)
$(\delta = 1.0)$	$\widehat{\delta}$	0.901	(0.042)	0.841	(0.045)	1.002	(0.026)	1.049	(0.036)
$(\beta = 1.0)$	$\widehat{\beta}$	1.002	(0.023)	1.002	(0.031)	1.001	(0.020)	1.002	(0.026)
$(\rho_{\epsilon_1\epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1\epsilon_2}$	0.664	(0.031)	0.473	(0.065)	0.666	(0.021)	0.535	(0.028)
		N=500; T=7							
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	1.036	(0.040)	-	-	0.907	(0.015)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.963	(0.012)	-	-	0.963	(0.014)
$(\delta = 1.0)$	$\widehat{\delta}$	0.900	(0.034)	0.828	(0.035)	1.002	(0.021)	1.057	(0.025)
$(\beta = 1.0)$	$\widehat{\beta}$	1.000	(0.017)	1.000	(0.018)	1.001	(0.014)	1.002	(0.018)
$(\rho_{\epsilon_1\epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1\epsilon_2}$	0.665	(0.025)	0.428	(0.052)	0.663	(0.017)	0.534	(0.023)
		N=800; T=4							
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	1.103	(0.043)	-	-	0.989	(0.014)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	1.046	(0.015)	-	-	1.048	(0.016)
$(\delta = 1.0)$	$\widehat{\delta}$	0.896	(0.033)	0.832	(0.036)	0.997	(0.021)	1.048	(0.029)
$(\beta = 1.0)$	$\widehat{\beta}$	1.001	(0.018)	1.000	(0.021)	0.999	(0.014)	1.002	(0.021)
$(\rho_{\epsilon_1\epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1\epsilon_2}$	0.664	(0.024)	0.473	(0.049)	0.664	(0.018)	0.531	(0.022)

* $\rho = \gamma = 0$ in the static data generating process. The individual effects, while being present in the data generating process, are ignored in the estimation of the model.

always larger than 15%. Hence, unlike β, when we fail to account for individual effects, the MLE behavior of the coefficient of the strictly exogenous explanatory variable in the selection equation differs according to the type of the selection rule and according to whether the model is dynamic or static.

In the dynamic version of the model, the MLE of ρ and γ is biased *away* from zero and *upwards* in all cases, with a bias that gets as large as 61% in the type 2 tobit, and 55% in the type 3 tobit. This bias is reduced fairly substantially (by about 9%) as T increases but remains very high, and is unchanged as N increases. In other words, when estimating the dynamic version of the model and failing to account for individual effects, the coefficients of the lagged dependent explanatory variables, in both the selection equation and the equation of interest, are overestimated. In the dynamic panel data discrete choice model, this phenomenon is known as *spurious state dependence* (Heckman, 1981a), i.e. too much credit is attributed to past event as a determinant of current event if intertemporal correlation in the unobservables is not accounted for.

Finally, the MLE of $\rho_{\epsilon_1\epsilon_2}$ is biased *away* from zero with a bias of about 25% in the static case, and about 35% in the dynamic case. There is no clear pattern of bias reduction as either T or N increases.

5.2.2 Exogenous initial conditions

Table 5 reports the mean and standard deviation of the MLE of ρ, γ, δ, β, σ_{a_1}, σ_{a_2}, $\rho_{a_1a_2}$ and $\rho_{\epsilon_1\epsilon_2}$ when the initial conditions, while being endogenous and correlated with the individual effects in the DGP, are assumed to be exogenous, i.e., b_1^s and b_1^r are assumed to be zero.

In both the type 2 and type 3 tobit, the MLE bias of δ and β increases by about 6% with respect to the case of the DGP but remains fairly small, less than 9% in the type 2 tobit, and less than 11% in the type 3 tobit. As for the MLE of ρ and γ, their bias remains substantially large compared to the case where the individual effects are ignored. Accounting for individual effects but assuming exogenous initial conditions also results in a spurious state dependence situation.[12] Hence, we can state that, in order for the coefficients of the lagged dependent explanatory variables to be accurately estimated, the correlation between the initial conditions and the individual effects must be taken into account: controlling for individual effects only is not sufficient. As T increases, the MLE bias of ρ and γ, in both the type 2 and type 3 tobit, decreases by about 10% but remains substantially large, while, as N increases, the bias remains the same. Finally, the MLE bias of $\rho_{a_1a_2}$ and $\rho_{\epsilon_1\epsilon_2}$ is higher than in the DGP case, and that of σ_{a_1} and σ_{a_2} is comparable to that of

---

[12]Although the term spurious state dependence was used in the context of discrete choice models, we use it in our study to explain the overestimation of the coefficients of the lagged dependent explanatory variables in both the selection equation and the equation of interest.

the same parameters in the DGP case.

Table 5: ML estimates based on 200 replications and 2-point Gauss-Hermite quadrature: Exogenous initial conditions

(True value)*	Estimate	Mean	Std. Dev.	Mean	Std. Dev.
			Estimated model		
		Dynamic type 2 tobit		Dynamic type 3 tobit	
		N=500; T=4			
$(\rho = 0.5)$	$\widehat{\rho}$	1.115	(0.069)	0.924	(0.020)
$(\gamma = 0.5)$	$\widehat{\gamma}$	0.967	(0.026)	0.973	(0.024)
$(\delta = 1.0)$	$\widehat{\delta}$	0.923	(0.051)	1.086	(0.035)
$(\beta = 1.0)$	$\widehat{\beta}$	1.062	(0.030)	1.071	(0.025)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.571	(0.080)	0.583	(0.078)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.458	(0.175)	0.482	(0.177)
$(\rho_{a_1 a_2} = 0.5)$	$\widehat{\rho}_{a_1 a_2}$	0.229	(0.088)	0.214	(0.080)
$(\rho_{\epsilon_1 \epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1 \epsilon_2}$	0.411	(0.095)	0.560	(0.034)
		N=500; T=7			
$(\rho = 0.5)$	$\widehat{\rho}$	1.030	(0.045)	0.829	(0.017)
$(\gamma = 0.5)$	$\widehat{\gamma}$	0.863	(0.016)	0.869	(0.019)
$(\delta = 1.0)$	$\widehat{\delta}$	0.915	(0.040)	1.102	(0.024)
$(\beta = 1.0)$	$\widehat{\beta}$	1.048	(0.017)	1.051	(0.016)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.632	(0.058)	0.718	(0.075)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.532	(0.076)	0.538	(0.159)
$(\rho_{a_1 a_2} = 0.5)$	$\widehat{\rho}_{a_1 a_2}$	0.266	(0.041)	0.240	(0.073)
$(\rho_{\epsilon_1 \epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1 \epsilon_2}$	0.403	(0.080)	0.586	(0.027)
		N=800; T=4			
$(\rho = 0.5)$	$\widehat{\rho}$	1.113	(0.480)	0.924	(0.017)
$(\gamma = 0.5)$	$\widehat{\gamma}$	0.968	(0.019)	0.974	(0.020)
$(\delta = 1.0)$	$\widehat{\delta}$	0.914	(0.039)	1.085	(0.029)
$(\beta = 1.0)$	$\widehat{\beta}$	1.060	(0.020)	1.071	(0.019)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.563	(0.057)	0.587	(0.058)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.487	(0.027)	0.503	(0.102)
$(\rho_{a_1 a_2} = 0.5)$	$\widehat{\rho}_{a_1 a_2}$	0.247	(0.012)	0.224	(0.046)
$(\rho_{\epsilon_1 \epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1 \epsilon_2}$	0.406	(0.069)	0.557	(0.025)

*The initial conditions are endogenous and correlated with the individual effects in the data generating process, but assumed to be exogenous in the estimation of the model.

5.2.3 No selection bias correction

The mean and standard deviation of the MLE of ρ, γ, δ, β, σ_{a_1} and σ_{a_2} are reported in Table 6 when we do not correct for the selection bias, i.e., we assume $\rho_{a_1 a_2}$ and $\rho_{\epsilon_1 \epsilon_2}$ to be zero.

In both the type 2 and type 3 tobit and regardless of whether the model is dynamic or static, the MLE bias of the coefficients of the lagged dependent and strictly exogenous explanatory variables is unaffected. This result always holds for δ and β when the strictly exogenous explanatory variables in the selection equation and in the equation of interest are not at all or only slightly correlated. When they are highly correlated, Monte Carlo results not reported here show that the MLE of the coefficients in the equation of interest are biased *downwards* while that of the coefficients in

Table 6: ML estimates based on 200 replications and 2-point Gauss-Hermite quadrature: No selection bias correction

(True value)*	Estimate	Static type 2 tobit		Dynamic type 2 tobit		Static type 3 tobit		Dynamic type 3 tobit	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
				Estimated model					
				N=500; T=4					
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	0.475	(0.091)	-	-	0.415	(0.031)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.547	(0.018)	-	-	0.545	(0.020)
$(\delta = 1.0)$	$\widehat{\delta}$	1.007	(0.051)	1.016	(0.057)	1.003	(0.033)	1.042	(0.033)
$(\beta = 1.0)$	$\widehat{\beta}$	1.001	(0.022)	1.008	(0.020)	1.001	(0.022)	1.007	(0.020)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.709	(0.070)	0.714	(0.079)	0.711	(0.050)	0.737	(0.050)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.463	(0.021)	0.372	(0.121)	0.460	(0.020)	0.398	(0.021)
				N=500; T=7					
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	0.486	(0.057)	-	-	0.408	(0.018)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.550	(0.012)	-	-	0.548	(0.015)
$(\delta = 1.0)$	$\widehat{\delta}$	1.006	(0.042)	1.014	(0.048)	1.001	(0.024)	1.063	(0.025)
$(\beta = 1.0)$	$\widehat{\beta}$	1.000	(0.015)	1.006	(0.012)	1.000	(0.016)	1.006	(0.015)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.710	(0.045)	0.713	(0.054)	0.708	(0.034)	0.770	(0.043)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.459	(0.014)	0.370	(0.095)	0.461	(0.014)	0.388	(0.014)
				N=800; T=4					
$(\rho = 0.5)$	$\widehat{\rho}$	-	-	0.478	(0.072)	-	-	0.416	(0.024)
$(\gamma = 0.5)$	$\widehat{\gamma}$	-	-	0.546	(0.015)	-	-	0.548	(0.017)
$(\delta = 1.0)$	$\widehat{\delta}$	1.003	(0.041)	1.002	(0.045)	0.999	(0.028)	1.043	(0.025)
$(\beta = 1.0)$	$\widehat{\beta}$	1.000	(0.019)	1.006	(0.015)	1.000	(0.018)	1.007	(0.016)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.704	(0.056)	0.710	(0.065)	0.711	(0.038)	0.739	(0.039)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	0.461	(0.016)	0.380	(0.093)	0.459	(0.015)	0.396	(0.016)

* $\rho_{a_1 a_2} = 0.5$ and $\rho_{\epsilon_1 \epsilon_2} = 0.8$ in the data generating process but are assumed to be 0 in the estimation of the model.

19

the selection equation remains unaffected. The most notable change occurs in the MLE of σ_{a_1} which, in all cases, becomes biased *away* from zero, while that of σ_{a_2} remains, in all cases but in the dynamic type 2 tobit, biased *towards* zero.

5.2.4 Static model estimated under a dynamic DGP

The MLE of δ, β, σ_{a_1}, σ_{a_2}, $\rho_{a_1 a_2}$ and $\rho_{\epsilon_1 \epsilon_2}$ are reported in Table 7 when a static model is estimated under a dynamic DGP, i.e., ρ and γ are assumed to be zero.

In both the type 2 and type 3 tobit, the MLE of δ is biased *away* from zero, with a bias that is no less than 30%, while the MLE of β remains biased *towards* zero but with a larger bias than in the DGP case. Furthermore, the standard deviations of the individual effects are overestimated and their MLE bias gets as large as 200% for σ_{a_1}, and 100% for σ_{a_2}. In other words, when past event is a determinant of current event and is left out of the explanatory variables, too much credit is attributed to the individual effects as a determinant of current event. This is the reverse phenomenon of the spurious state dependence. Finally, $\rho_{a_1 a_2}$ and $\rho_{\epsilon_1 \epsilon_2}$ are biased *away* from zero with a bias of approximately 25% and 45% respectively.

Table 7: *ML estimates based on 200 replications and 2-point Gauss-Hermite quadrature: Static model estimated under a dynamic data generating process*

(True value)*	Estimate	Mean	Std. Dev.	Mean	Std. Dev.
		Estimated model			
		\multicolumn{2}{c}{Static type 2 tobit}	\multicolumn{2}{c}{Static type 3 tobit}	
		N=500; T=4			
$(\delta = 1.0)$	$\widehat{\delta}$	0.682	(0.047)	1.338	(0.072)
$(\beta = 1.0)$	$\widehat{\beta}$	0.909	(0.057)	0.919	(0.060)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.762	(0.102)	2.075	(0.193)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	1.408	(0.075)	1.410	(0.079)
$(\rho_{a_1 a_2} = 0.5)$	$\widehat{\rho}_{a_1 a_2}$	0.248	(0.016)	0.237	(0.018)
$(\rho_{\epsilon_1 \epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1 \epsilon_2}$	0.345	(0.115)	0.365	(0.077)
		N=500; T=7			
$(\delta = 1.0)$	$\widehat{\delta}$	0.688	(0.041)	1.381	(0.057)
$(\beta = 1.0)$	$\widehat{\beta}$	0.921	(0.038)	0.931	(0.041)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.865	(0.077)	2.509	(0.196)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	1.541	(0.068)	1.583	(0.078)
$(\rho_{a_1 a_2} = 0.5)$	$\widehat{\rho}_{a_1 a_2}$	0.279	(0.019)	0.266	(0.018)
$(\rho_{\epsilon_1 \epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1 \epsilon_2}$	0.389	(0.133)	0.423	(0.073)
		N=800; T=4			
$(\delta = 1.0)$	$\widehat{\delta}$	0.679	(0.034)	1.347	(0.054)
$(\beta = 1.0)$	$\widehat{\beta}$	0.908	(0.044)	0.915	(0.047)
$(\sigma_{a_1} = 0.5)$	$\widehat{\sigma}_{a_1}$	0.752	(0.086)	2.079	(0.142)
$(\sigma_{a_2} = 0.5)$	$\widehat{\sigma}_{a_2}$	1.402	(0.057)	1.403	(0.056)
$(\rho_{a_1 a_2} = 0.5)$	$\widehat{\rho}_{a_1 a_2}$	0.249	(0.013)	0.236	(0.015)
$(\rho_{\epsilon_1 \epsilon_2} = 0.8)$	$\widehat{\rho}_{\epsilon_1 \epsilon_2}$	0.338	(0.105)	0.365	(0.063)

*$\rho = \gamma = 0.5$ in the data generating process but are assumed to be 0 in the estimation of the model.

5.2.5 Dynamic model estimated under a static DGP

We estimate two types of dynamic models under a static DGP. The first one ignores the individual effects, and the second one accounts for individual effects that are correlated with endogenous initial conditions. In other words, the first type assumes σ_{a_1}, σ_{a_2} and $\rho_{a_1 a_2}$ to be all zero while the second one makes no restriction assumptions on the parameters of the model. The mean and standard deviation of the MLE of ρ, γ, δ, β, σ_{a_1}, σ_{a_2}, $\rho_{a_1 a_2}$ and $\rho_{\epsilon_1 \epsilon_2}$ are reported in Table 8.

In both cases, the MLE of δ and β are biased *towards* zero in the type 2 and type 3 tobit. However, the bias for δ is larger when we do not account for the individual effects than when we account for them. The most notable difference between the two cases lies in the estimates of the coefficients of the lagged dependent explanatory variables. They are positive, around 0.10 for ρ and 0.15 for γ, and significantly different from zero when we do not account for the individual effects. In other words, ignoring the individual effects results in attributing credit to past event as a determinant of current event while the former does not condition at all the latter. This is an extreme case of the spurious state dependence mentioned earlier. When we account for individual effects that are correlated with endogenous initial conditions, these estimates become very close to zero, their true value, and the standard deviations of the individual effects are biased *towards* zero. Hence, when estimating a dynamic panel data model, regardless of whether the DGP is dynamic or static, the issues of properly accounting for the individual effects and treating the initial conditions must be taken care of. As for $\rho_{a_1 a_2}$, it is biased *away* from zero in both the type 2 and type 3 tobit, while $\rho_{\epsilon_1 \epsilon_2}$ is biased *away* from zero in the type 2 tobit, and in the type 3 tobit with no individual effects. When individual effects that are correlated with endogenous initial conditions are accounted for, $\rho_{\epsilon_1 \epsilon_2}$ is biased *towards* zero in the type 3 tobit.

6   Application

To illustrate our method we proceed to estimate a dynamic type 3 tobit model that explains the amount of product innovation. Innovation surveys distinguish two types of product innovations, those new to the firm (incremental) and those new to the market (radical). The importance of each type of product innovation is measured by the share in total sales of innovative sales. But from the way the questionnaire of the survey is formulated, we only observe the latter if we observe the former. Equation (1) determines whether enterprise i is an incremental product innovator at period t ($d_{it}^* > 0$). In this case, the actual share of sales of incremental product innovations (eq. (5)) is positive, and the share of sales of radical product innovations (eq. (3)) and the full set of regressors included in the vector $\mathbf{x}_{it}$ are observed. When enterprise i is not an incremental product

Table 8: ML estimates based on 200 replications and 2-point Gauss-Hermite quadrature: Dynamic model estimated under a static data generating process

(True value) Estimate	Dynamic type 2 tobit with no individual effects Mean	Std. Dev.	Dynamic type 2 tobit with individual effects and endogenous initial conditions Mean	Std. Dev.	Dynamic type 3 tobit with no individual effects Mean	Std. Dev.	Dynamic type 3 tobit with individual effects and endogenous initial conditions Mean	Std. Dev.
N=500; T=4								
($\rho=0.0$) $\hat{\rho}$	0.106	(0.050)	-0.001	(0.074)	0.099	(0.026)	0.019	(0.028)
($\gamma=0.0$) $\hat{\gamma}$	0.158	(0.025)	0.043	(0.026)	0.147	(0.021)	0.050	(0.021)
($\delta=1.0$) $\hat{\delta}$	0.910	(0.042)	1.004	(0.050)	0.985	(0.027)	0.994	(0.028)
($\beta=1.0$) $\hat{\beta}$	1.001	(0.023)	1.005	(0.021)	1.001	(0.019)	1.007	(0.018)
($\sigma_{a_1}=0.5$) $\hat{\sigma}_{a_1}$	-	-	0.537	(0.080)	-	-	0.430	(0.069)
($\sigma_{a_2}=0.5$) $\hat{\sigma}_{a_2}$	-	-	0.405	(0.102)	-	-	0.393	(0.017)
($\rho_{a_1 a_2}=0.5$) $\hat{\rho}_{a_1 a_2}$	-	-	0.255	(0.065)	-	-	0.243	(0.019)
($\rho_{\epsilon_1 \epsilon_2}=0.8$) $\hat{\rho}_{\epsilon_1 \epsilon_2}$	0.659	(0.034)	0.660	(0.051)	0.664	(0.021)	0.709	(0.025)
N=500; T=7								
($\rho=0.0$) $\hat{\rho}$	0.103	(0.031)	-0.002	(0.044)	0.097	(0.018)	0.010	(0.019)
($\gamma=0.0$) $\hat{\gamma}$	0.155	(0.018)	0.045	(0.017)	0.146	(0.015)	0.051	(0.013)
($\delta=1.0$) $\hat{\delta}$	0.908	(0.034)	1.010	(0.041)	0.986	(0.020)	1.000	(0.021)
($\beta=1.0$) $\hat{\beta}$	1.000	(0.016)	1.004	(0.015)	1.001	(0.013)	1.004	(0.011)
($\sigma_{a_1}=0.5$) $\hat{\sigma}_{a_1}$	-	-	0.593	(0.055)	-	-	0.528	(0.053)
($\sigma_{a_2}=0.5$) $\hat{\sigma}_{a_2}$	-	-	0.414	(0.015)	-	-	0.400	(0.014)
($\rho_{a_1 a_2}=0.5$) $\hat{\rho}_{a_1 a_2}$	-	-	0.293	(0.016)	-	-	0.263	(0.019)
($\rho_{\epsilon_1 \epsilon_2}=0.8$) $\hat{\rho}_{\epsilon_1 \epsilon_2}$	0.661	(0.025)	0.664	(0.034)	0.664	(0.016)	0.714	(0.016)
N=800; T=4								
($\rho=0.0$) $\hat{\rho}$	0.103	(0.040)	0.000	(0.061)	0.098	(0.019)	0.019	(0.023)
($\gamma=0.0$) $\hat{\gamma}$	0.154	(0.012)	0.044	(0.021)	0.143	(0.016)	0.047	(0.017)
($\delta=1.0$) $\hat{\delta}$	0.905	(0.034)	0.998	(0.043)	0.980	(0.021)	0.989	(0.021)
($\beta=1.0$) $\hat{\beta}$	1.002	(0.018)	1.006	(0.017)	0.999	(0.014)	1.005	(0.013)
($\sigma_{a_1}=0.5$) $\hat{\sigma}_{a_1}$	-	-	0.528	(0.072)	-	-	0.430	(0.054)
($\sigma_{a_2}=0.5$) $\hat{\sigma}_{a_2}$	-	-	0.412	(0.041)	-	-	0.394	(0.014)
($\rho_{a_1 a_2}=0.5$) $\hat{\rho}_{a_1 a_2}$	-	-	0.261	(0.034)	-	-	0.244	(0.013)
($\rho_{\epsilon_1 \epsilon_2}=0.8$) $\hat{\rho}_{\epsilon_1 \epsilon_2}$	0.661	(0.025)	0.654	(0.042)	0.663	(0.017)	0.707	(0.018)

Estimated model

innovator, the shares of sales of incremental (d_{it}) and radical (y_{it}) product innovations are equal to zero, and only the set of regressors included in the vector $\mathbf{w}_{it}$ is observed.

The dynamic type 3 tobit is implemented using the same data as in Raymond *et al.* (2006). They are collected by the *Centraal Bureau voor de Statistiek* (CBS) and stem from three waves of the Dutch Community Innovation Survey, CIS 2 (1994-1996), CIS 2.5 (1996-1998) and CIS 3 (1998-2000), merged with data from the Production Survey (PS). The population of interest consists of Dutch manufacturing enterprises with at least ten employees and positive sales at the end of the period covered by the innovation survey. We consider enterprises that existed in 1994, survived (at least) until 2000 and took part in the three innovation surveys, resulting in a balanced panel of 861 enterprises.

Table 9: Descriptive statistics

Variable	Description	Overall Mean	Overall Std. Dev.	Between Std. Dev.	Within Std. Dev.
Dependent variables					
Intensity of incremental innovation[*]	innovative sales (of products new to the firm)/total sales (for product innovators)	0.290	0.258	0.218	0.147
Intensity of radical innovation[*]	innovative sales (of products new to the market)/total sales (for product innovators)	0.065	0.136	0.101	0.092
Explanatory variables					
Demand pull	1 if product-oriented innovation objectives are very important (for product innovators)	0.640	0.480	0.370	0.337
Technology push	1 if innovation sources are from universities or other institutes (for product innovators)	0.213	0.410	0.298	0.271
Innovation cooperation	1 if there is any type of cooperation (for product innovators)	0.349	0.477	0.379	0.293
Non-R&D performers	1 if not performing R&D (for product innovators)	0.196	0.397	0.380	0.214
Continuous R&D performers	1 if performing continuous R&D (for R&D performers)	0.754	0.431	0.399	0.247
Subsidies	1 if being subsidized at least once (for product innovators)	0.577	0.494	0.424	0.284
R&D intensity[**]	R&D expenditures/total sales (for R&D performers)	0.046	0.078	0.064	0.037
Size[†]	number of employees	209.962	539.248	534.722	71.292
Relative size[††]	total sales/sales of industry	0.006	0.020	0.020	0.004
# of observations		2583			

[*]A logit transformation; [**]ln(R&D/total sales); [†] ln(number of employees); [††] ln(total sales/sales of industry) are used in the estimation.

Table 9 shows descriptive statistics and the description of the dependent and explanatory variables of the model. For instance, the average share in total sales of products new to the firm is rather small (29%), and that of products new to the market is even smaller (6.5%). These

two variables are logit-transformed in order to make them lie within the set of real numbers.[13] The specification of the model is as follows. The current share in total sales of products new to the firm is explained by the lagged share in total sales of products new to the firm, lagged size and relative size.[14] Besides lagged size, the current share in total sales of products new to the market is explained by the lagged share in total sales of products new to the market and dummy variables capturing demand pull, proximity to science, innovation cooperation, non-R&D performers, subsidies, and a continuous R&D intensity variable.

Table 10 presents the estimation results of four versions of the type 3 tobit, namely a static model with unobserved individual effects, a dynamic model with no unobserved individual effects, a dynamic model with individual effects and endogenous initial conditions but ignoring the selection bias, and a dynamic model with individual effects and endogenous initial conditions correcting for selection bias.[15] The hypotheses of the first three models are tested against those of the fourth one, and rejected using a likelihood ratio test at 1% level of significance. Hence, the preferred model is dynamic with individual effects and endogenous initial conditions correcting for selection bias. The results suggest that, in all four versions of the model, lagged size affects positively and significantly the current share of sales of incremental innovations. The current share of sales of radical innovations is positively and significantly affected by R&D intensity, performing R&D continuously, demand pull, innovation cooperation, subsidies while, *ceteris paribus*, non-R&D performers are less successful than R&D performers in terms of generating innovative sales from radical innovations. Sample selection operates through the individual effects and the idiosyncratic errors in the static case, and only through the idiosyncratic errors in the dynamic case. In the static case, too much of a role is attributed to the individual effects in that their standard deviations, as well as the correlation between them, are overestimated. Finally, the coefficients associated with the lagged dependent variables are positively and statistically significantly estimated in the three dynamic cases. However, they are overestimated when the individual effects are not accounted for or when the individual effects are accounted for but exogenous initial conditions are assumed.

A dynamic type 2 tobit was also estimated using information only from a binary variable, indicating whether a firm is an incremental product innovator or not, for the selection equation

---

[13]The share of sales of products new to the firm takes on the values 1 for innovators that are newly established. They are replaced by 0.9999 in the logit transformation. Furthermore, the share of sales of products new to the market takes on the value 0 for some incremental product innovators. They are replaced by 0.0001 in the logit transformation.

[14]In the innovation survey, firms are asked whether they have product and/or process innovations, or incomplete and abandoned innovation activities during the whole period under study, while size and sales are measured only at the end of the period. Hence, all the cross-sectional studies on the determinants of innovation have this unsatisfactory feature of explaining the probability of being an innovator (during the whole period) by size or relative size (at the end of the period). Taking lagged size or relative size as an explanatory variable in the panel data context overcomes this problem.

[15]A dynamic model with unobserved individual effects and exogenous initial conditions was also estimated. The results, not reported in the paper, are similar to those of the dynamic model with no individual effects.

Table 10: ML estimates of the type 3 tobit: 5-point Gauss-Hermite quadrature

| Variable | Static model with | | Dynamic model with | | | | | |
| | Unobserved individual effects | | No unobserved individual effects | | Unobserved individual effects
No selection bias correction | | Unobserved individual effects and endogenous initial conditions
Selection bias correction | |
	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
Current share of sales of incremental product innovations (d_{it} in logit)								
Past share of incremental innov. sales ($d_{i,t-1}$)	-	-	0.226**	(0.020)	0.109**	(0.033)	0.113**	(0.033)
Lagged size (in log)	0.374*	(0.169)	0.304*	(0.153)	0.351*	(0.160)	0.354*	(0.159)
Lagged relative size (in log)	0.204†	(0.124)	0.094	(0.113)	0.070	(0.118)	0.066	(0.117)
Intercept	-2.126	(1.479)	-1.793	(1.341)	-2.136	(1.404)	-2.172	(1.395)
Current share of sales of radical product innovations (y_{it} in logit)								
Past share of radical innov. sales ($y_{i,t-1}$)	-	-	0.217**	(0.029)	0.114†	(0.063)	0.120*	(0.060)
Lagged size (in log)	0.139	(0.114)	0.068	(0.104)	0.031	(0.105)	0.057	(0.105)
R&D intensity (in log)	0.249**	(0.091)	0.199*	(0.086)	0.217*	(0.087)	0.193*	(0.087)
Non-R&D performers	-1.415*	(0.580)	-1.111*	(0.558)	-1.049†	(0.564)	-1.010†	(0.563)
Continuous R&D performers	0.950**	(0.301)	0.786**	(0.294)	0.816**	(0.300)	0.797**	(0.298)
Demand pull	0.664**	(0.222)	0.702**	(0.219)	0.753**	(0.219)	0.688**	(0.219)
Technology push	-0.093	(0.254)	-0.162	(0.246)	-0.077	(0.248)	-0.104	(0.247)
Cooperation in innovation	0.573*	(0.230)	0.539*	(0.223)	0.554*	(0.224)	0.530*	(0.223)
Subsidies	0.697**	(0.242)	0.562*	(0.236)	0.532*	(0.238)	0.555*	(0.237)
Intercept	-6.976**	(0.731)	-5.353**	(0.695)	-4.978**	(0.721)	-5.165**	(0.726)
Extra parameters								
Initial share of incremental innov. sales (d_{i0})	-	-	-	-	0.127**	(0.031)	0.123**	(0.030)
Initial share of radical innov. sales (y_{i0})	-	-	-	-	0.122*	(0.050)	0.125*	(0.050)
σ_{a_1}	1.779**	(0.155)	-	-	1.139**	(0.241)	1.102**	(0.243)
σ_{a_2}	1.452**	(0.180)	-	-	0.593	(0.692)	0.597	(0.681)
σ_{ϵ_1}	3.110**	(0.084)	3.463**	(0.073)	3.272**	(0.098)	3.283**	(0.098)
σ_{ϵ_2}	2.994**	(0.096)	3.247**	(0.070)	3.181**	(0.143)	3.179**	(0.142)
$\rho_{a_1 a_2}$	0.237*	(0.115)	-	-	-	-	0.137	(0.337)
$\rho_{\epsilon_1 \epsilon_2}$	0.073†	(0.040)	0.107**	(0.029)	-	-	0.101**	(0.037)
Number of observations	1722							
Log-likelihood	-5978.255		-5915.979		-5910.001		-5904.049	

Significance levels : †: 10%   *: 5%   **: 1%

and using the same equation of interest. As expected, the results (not reported in the paper) show that the parameters of the equation of interest are (slightly) less accurately estimated but similar to those of the type 3 tobit.

7 Conclusion

We have proposed a simple method to implement the maximum likelihood estimator of dynamic panel data sample selection models. The method consists in writing the likelihood function conditional on the individual effects which are then "integrated out" with respect to their joint normal distribution. We have handled the two-dimensional indefinite integral involved in the likelihood function using "two-step" Gauss-Hermite quadrature, and solved the initial conditions problem using the approach of Wooldridge (2005). The resulting likelihood function encompasses a wide range of likelihood functions of panel data models whose estimation can be achieved by simply making restriction assumptions on the parameters of the dynamic panel data sample selection model. These assumptions can be tested using a standard likelihood ratio or Wald test.

In order to assess the quality of the Gauss-Hermite quadrature, and hence that of the maximum likelihood estimator, we have conducted a Monte Carlo study and obtained the following results. First, under the data generating process, the maximum likelihood estimator works very well, even for a number of integration points, in each step of the Gauss-Hermite quadrature, as small as 2. The estimator bias of the coefficients of the lagged dependent and strictly exogenous explanatory variables, as well as that of the standard deviations of the individual effects is very small for a sample size as large as the one considered in the Monte Carlo study. The sample selection terms, however, are less accurately estimated when the number of integration points is very small. Secondly, too much of a role is attributed to the lagged dependent variables as a determinant of the current dependent variables when either the individual effects or the correlation between the initial outcome and the individual effects are not taken into account. Thirdly, the selection bias really matters when the explanatory variables in the selection equation and in the equation of interest are highly correlated, otherwise it can be safely ignored. Fourthly, when the lagged dependent variables condition the current dependent variables and are left out of the explanatory variables, too much of a role is attributed to the individual effects as a determinant of the current dependent variables. Finally, when estimating a dynamic panel data model, regardless of whether the true model is dynamic or static, the individual effects and the correlation between the initial outcome and the individual effects must be taken into account, otherwise a situation of spurious state dependence will arise even under its extreme form. In this latter case, credit is attributed

to the lagged dependent variables as a determinant of the current dependent variables while the former do not condition at all the latter.

We have applied the method to estimate a dynamic type 3 tobit model of incremental and radical product innovations. The estimation results confirm those of the Monte Carlo study. More specifically, the effect of past incremental and radical product innovations on current incremental and radical product innovations is overestimated when either the individual effects or the correlation between the initial conditions and the individual effects are not accounted for. We also find that too much importance is attributed to the individual effects as a determinant of current incremental and radical product innovations when past incremental and radical product innovations do not enter the set of explanatory variables. Finally, the estimation results of the model with no selection bias correction are similar to those of the model with a selection bias correction, even though the latter model is shown, using a likelihood ratio test, to be the preferred one.

Our approach can be, in a straightforward manner, extended to estimate by maximum likelihood panel data sample selection models with more than two equations, that include a lagged dependent explanatory variable and individual effects in each equation. For instance, our approach can handle the estimation of dynamic panel data type 4 and type 5 tobit models with individual effects.

Appendix A "Two-step" Gauss-Hermite quadrature

Consider the dynamic panel data type 2 tobit model defined by equations (1-4) and (6-7), and consider expressions (10) and (11). Assume that the vectors $(\epsilon_{1it}, \epsilon_{2it})'$ and $(a_{1i}, a_{2i})'$ are independent of each other, and independently and identically distributed over time and across individuals following a normal distribution with mean zero and covariance matrix given in equation (8). The likelihood function of individual i, starting from $t = 1$ and conditional on the regressors and the initial conditions, is obtained by "integrating out" the individual effects (equation (9)). The likelihood function of individual i conditional on the individual effects is written in equation (12) with σ_{ϵ_1} being normalized to 1 for identification reasons. The expression of the bivariate normal distribution of a_{1i} and a_{2i} is written as

$$g(a_{1i}, a_{2i}) = (2\pi\sigma_{a_1}\sigma_{a_2})^{-1} \left(1 - \rho_{a_1 a_2}^2\right)^{\frac{-1}{2}} e^{\frac{-1}{2(1-\rho_{a_1 a_2}^2)}\left(\frac{a_{1i}^2}{\sigma_{a_1}^2} - 2\rho_{a_1 a_2}\frac{a_{1i}}{\sigma_{a_1}}\frac{a_{2i}}{\sigma_{a_2}} + \frac{a_{2i}^2}{\sigma_{a_2}^2}\right)}. \tag{20}$$

Hence, equation (9) can be written as

$$L_i = \int_{-\infty}^{\infty} e^{\frac{-a_{2i}^2}{2\sigma_{a_2}^2(1-\rho_{a_1 a_2}^2)}} \prod_{t=1}^{T} \left[\frac{1}{\sigma_{\epsilon_2}}\phi\left(\frac{y_{it} - B_{it} - a_{2i}}{\sigma_{\epsilon_2}}\right)\right]^{d_{it}} H(a_{2i})da_{2i}, \tag{21}$$

where $H(a_{2i})$ is written as

$$H(a_{2i}) = (2\pi\sigma_{a_1}\sigma_{a_2})^{-1} \left(1 - \rho_{a_1 a_2}^2\right)^{\frac{-1}{2}} \int_{-\infty}^{\infty} e^{\frac{-a_{1i}^2}{2\sigma_{a_1}^2(1-\rho_{a_1 a_2}^2)}} e^{\frac{1}{(1-\rho_{a_1 a_2}^2)}\left(\rho_{a_1 a_2}\frac{a_{1i}}{\sigma_{a_1}}\frac{a_{2i}}{\sigma_{a_2}}\right)}$$

$$\times \prod_{t=1}^{T} \Phi\left[-(A_{it} + a_{1i})\right]^{(1-d_{it})} \left[\Phi\left(\frac{A_{it} + a_{1i} + \frac{\rho_{\epsilon_1 \epsilon_2}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - a_{2i})}{\sqrt{1 - \rho_{\epsilon_1 \epsilon_2}^2}}\right)\right]^{d_{it}} da_{1i}. \tag{22}$$

Equation (21) can be approximated using "two-step" Gauss-Hermite quadrature which states that

$$\int_{-\infty}^{\infty} e^{-z^2} f(z)dz \simeq \sum_{m=1}^{M} w_m f(a_m), \tag{23}$$

where w_m and a_m are respectively the weights and abscissas of the Gauss-Hermite quadrature, the tables of which are formulated in mathematical textbooks (e.g. Abramovitz and Stegun, 1964), and M is the total number of integration points. The larger M, the more accurate the Gauss-Hermite quadrature. The "two-step" Gauss-Hermite quadrature consists, in the first step, in approximating equation (22) using equation (23). In the second step, a second approximation is applied to equation (21) where $H(a_{2i})$ is replaced by its first-step Gauss-Hermite approximation. The approach is described as follows.

Consider the variable change $z_{1i} = \frac{a_{1i}}{\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)}}$ so that $da_{1i} = dz_{1i}\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)}$. Equation (22) is then written as

$$
\begin{aligned}
H(a_{2i}) = & \frac{\sqrt{2}}{2\pi\sigma_{a_2}} \int_{-\infty}^{\infty} e^{-z_{1i}^2} e^{\frac{1}{1-\rho_{a_1a_2}^2}\left(\rho_{a_1a_2} z_{1i} \sqrt{2(1-\rho_{a_1a_2}^2)}\frac{a_{2i}}{\sigma_{a_2}}\right)} \\
& \times \prod_{t=1}^{T} \Phi\left[-\left(A_{it} + z_{1i}\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)}\right)\right]^{(1-d_{it})} \\
& \times \left[\Phi\left(\frac{A_{it} + z_{1i}\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)} + \frac{\rho_{\epsilon_1\epsilon_2}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - a_{2i})}{\sqrt{1-\rho_{\epsilon_1\epsilon_2}^2}}\right)\right]^{d_{it}} dz_{1i}
\end{aligned}
\tag{24}
$$

and can be approximated using equation (23) by

$$
\begin{aligned}
H(a_{2i}) \simeq & \frac{\sqrt{2}}{2\pi\sigma_{a_2}} \sum_{m=1}^{M} w_m \left\{ e^{\frac{1}{1-\rho_{a_1a_2}^2}\left(\rho_{a_1a_2} a_m \sqrt{2(1-\rho_{a_1a_2}^2)}\frac{a_{2i}}{\sigma_{a_2}}\right)} \right. \\
& \times \prod_{t=1}^{T} \Phi\left[-\left(A_{it} + a_m\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)}\right)\right]^{(1-d_{it})} \\
& \left. \times \left[\Phi\left(\frac{A_{it} + a_m\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)} + \frac{\rho_{\epsilon_1\epsilon_2}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - a_{2i})}{\sqrt{1-\rho_{\epsilon_1\epsilon_2}^2}}\right)\right]^{d_{it}} \right\},
\end{aligned}
\tag{25}
$$

where w_m and a_m are the weights and abscissas of the first-step Gauss-Hermite quadrature with M being the total number of integration points.

Consider the second variable change $z_{2i} = \frac{a_{2i}}{\sigma_{a_2}\sqrt{2(1-\rho_{a_1a_2}^2)}}$ so that $da_{2i} = dz_{2i}\sigma_{a_2}\sqrt{2(1-\rho_{a_1a_2}^2)}$. Replacing equation (25) into equation (21) yields

$$
\begin{aligned}
L_i \simeq & \frac{\sqrt{1-\rho_{a_1a_2}^2}}{\pi} \int_{-\infty}^{\infty} e^{-z_{2i}^2} \prod_{t=1}^{T}\left[\frac{1}{\sigma_{\epsilon_2}}\phi\left(\frac{y_{it} - B_{it} - z_{2i}\sigma_{a_2}\sqrt{2(1-\rho_{a_1a_2}^2)}}{\sigma_{\epsilon_2}}\right)\right]^{d_{it}} \\
& \times \sum_{m=1}^{M} w_m \left\{ e^{2\rho_{a_1a_2} z_{2i} a_m} \prod_{t=1}^{T} \Phi\left[-\left(A_{it} + a_m\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)}\right)\right]^{(1-d_{it})} \right. \\
& \left. \times \left[\Phi\left(\frac{A_{it} + a_m\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)} + \frac{\rho_{\epsilon_1\epsilon_2}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - z_{2i}\sigma_{a_2}\sqrt{2(1-\rho_{a_1a_2}^2)})}{\sqrt{1-\rho_{\epsilon_1\epsilon_2}^2}}\right)\right]^{d_{it}} \right\} dz_{2i}
\end{aligned}
\tag{26}
$$

which can be approximated using again equation (23) by

$$
\begin{aligned}
L_i \simeq & \frac{\sqrt{1-\rho_{a_1a_2}^2}}{\pi} \sum_{p=1}^{P} w_p \left\{ \prod_{t=1}^{T}\left[\frac{1}{\sigma_{\epsilon_2}}\phi\left(\frac{y_{it} - B_{it} - a_p\sigma_{a_2}\sqrt{2(1-\rho_{a_1a_2}^2)}}{\sigma_{\epsilon_2}}\right)\right]^{d_{it}} \right. \\
& \times \sum_{m=1}^{M} w_m \left\{ e^{2\rho_{a_1a_2} a_p a_m} \prod_{t=1}^{T} \Phi\left[-\left(A_{it} + a_m\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)}\right)\right]^{(1-d_{it})} \right. \\
& \left.\left. \times \Phi\left(\frac{A_{it} + a_m\sigma_{a_1}\sqrt{2(1-\rho_{a_1a_2}^2)} + \frac{\rho_{\epsilon_1\epsilon_2}}{\sigma_{\epsilon_2}}(y_{it} - B_{it} - a_p\sigma_{a_2}\sqrt{2(1-\rho_{a_1a_2}^2)})}{\sqrt{1-\rho_{\epsilon_1\epsilon_2}^2}}\right)^{d_{it}} \right\} \right\},
\end{aligned}
\tag{27}
$$

where w_p and a_p are the weights and abscissas of the second-order Gauss-Hermite quadrature with P being the total number of integration points.

Using equations (9), (13), (14) and (20), the random-effects individual likelihood function of the type 3 tobit can be derived similarly to yield equation (16).

References

Abramovitz M, Stegun I. 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards Applied Mathematics, US Government Printing Office: Washington, D.C.

Ai C, Chen C. 1992. Estimation of a fixed effects bivariate censored regression model. *Economics Letters* **40**: 403–406.

Amemiya T. 1984. Tobit models: A survey. *Journal of Econometrics* **24**: 3–62.

Arellano M, Honoré B. 2001. Panel data models: Some recent developments. In *Handbook of Econometrics*, Heckman JJ, Leamer E (eds). North Holland: Amsterdam.

Askildsen JE, Baltagi BH, Holmås TH. 2003. Wage policy in the health care sector: A panel data analysis of nurses' labour supply. *Health Economics* **12**: 705–719.

Butler JS, Moffitt R. 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* **50**: 761–764.

Chamberlain G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* **47**: 225–238.

Dustmann Christian, Rochina-Barrachina María Engracia. 2000. Selection correction in panel data models: An application to labour supply and wages. IZA Discussion Paper No. 162.

Gayle G-L, Viauroux C. 2007. Root-n consistent semiparametric estimators of a dynamic panel-sample-selection model. *Journal of Econometrics* . Forthcoming.

Greene W. 2004a. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal* **7**: 98–119.

Greene W. 2004b. Convenient estimators for the panel probit model: Further results. *Empirical Economics* **29**: 21–47.

Guilkey DK, Murphy JL. 1993. Estimation and testing in the random effects probit model. *Journal of Econometrics* **59**: 301–317.

Hausman JA, Wise DA. 1979. Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica* **47**: 455–474.

Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* **47**: 153–162.

Heckman JJ. 1981a. Heterogeneity and state dependence. In *Studies in Labor Markets*, Rosen S (ed). University of Chicago Press.

Heckman JJ. 1981b. 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*, Manski CF, McFadden D (eds). MIT Press: Cambridge, MA.

Heckman JJ, MaCurdy TE. 1980. A life cycle model of female labour supply. *Review of Economic Studies* **47**: 47–74.

Honoré BE, Kyriazidou E. 2000. Estimation of tobit-type models with individual specific effects. *Econometric Reviews* **19**: 341–366.

Hsiao C. 2003. *Analysis of Panel Data.* 2nd edn. Cambridge University Press.

Kyriazidou E. 1997. Estimation of a panel data sample selection model. *Econometrica* **65**: 1335–1364.

Kyriazidou E. 2001. Estimation of dynamic panel data sample selection models. *Review of Economic Studies* **68**: 543–572.

Lancaster T. 2000. The incidental parameter problem since 1948. *Journal of Econometrics* **95**: 391–413.

Lee M-J. 2001. First-difference estimator for panel censored-selection models. *Economics Letters* **70**: 43–49.

Magnac T. 2004. Panel binary variables and sufficiency: Generalizing conditional logit. *Econometrica* **72**: 1859–1876.

Neyman J, Scott EL. 1948. Consistent estimates based on partially consistent observations. *Econometrica* **16**: 1–32.

Nijman T, Verbeek M. 1992. Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics* **7**: 243–257.

Rabe-Hesketh S, Skrondal A, Pickles A. 2005. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* **128**: 301–323.

Raymond W. 2007. *The dynamics of innovation and firm performance: An econometric panel data analysis*. Ph.D. thesis, Maastricht University. forthcoming.

Raymond W, Mohnen P, Palm F, Schim van der Loeff S. 2006. Persistence of innovation in Dutch manufacturing: Is it spurious? CESifo Working Paper No. 1681.

Ridder G. 1990. Attrition in multi-wave panel data. In *Panel Data and Labor Market Studies*, Hartog J, Ridder G, Theeuwes J (eds). North-Holland, Amsterdam.

Rochina-Barrachina ME. 1999. A new estimator for panel data sample selection models. *Annales d'Economie et de Statistiques* **55-56**: 153–182.

Semykina A, Wooldridge JM. 2005. Estimating panel data models in the presence of endogeneity and selection: Theory and application. unpublished working paper, Michigan State University.

Vella F. 1998. Estimating models with sample selection bias: A survey. *The Journal of Human Resources* **33**: 127–169.

Vella F, Verbeek M. 1999. Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics* **90**: 239–263.

Verbeek M. 1990. On the estimation of a fixed effects model with selectivity bias. *Economic Letters* **34**: 267–270.

Wooldridge JM. 1995. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* **68**: 115–132.

Wooldridge JM. 2005. Simple solutions to the initial conditions problem in dynamic nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**: 39–54.

Zabel JE. 1992. Estimating fixed and random effects models with selectivity. *Economics Letters* **40**: 269–272.

The UNU-MERIT WORKING Paper Series

2007-01   *Developing science, technology and innovation indicators: what we can learn from the past* by Christopher Freeman & Luc Soete

2007-02   *The impact of innovation activities on productivity and firm growth: evidence from Brazil* by Micheline Goedhuys

2007-03   *Estimations of US debt dynamics: Growth cum debt and the savings glut in Kouri's model* by Thomas Ziesemer

2007-04   *States and Firms on the Periphery: The Challenges of a Globalising World* by Gabriel R.G. Benito & Rajneesh Narula

2007-05   *How Do Consumers Make Choices? A Summary of Evidence from Marketing and Psychology* by *Zakaria Babutsidze*

2007-06   *Inter-firm Technology Transfer: Partnership-embedded Licensing or Standard Licensing Agreements?* by John Hagedoorn, Stefanie Lorenz-Orlean & Hans Kranenburg

2007-07   *The Behavior of the Maximum Likelihood Estimator of Dynamic Panel Data Sample Selection Models* by Wladimir Raymond, Pierre Mohnen, Franz Palm & Sybrand Schim van der Loeff