DEPARTMENT OF ECONOMICS

Working Paper

A Comparison of Different Estimators for
Panel Data Sample Selection Models

Peter Jensen, Michael Rosholm and Mette Verner

Working Paper No. 2002-1

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

INSTITUT FOR ØKONOMI

AFDELING FOR NATIONALØKONOMI - AARHUS UNIVERSITET - BYGNING 350
8000 AARHUS C - ☎ 89 42 11 33 - TELEFAX 86 13 63 34

WORKING PAPER

A Comparison of Different Estimators for Panel Data Sample Selection Models

Peter Jensen, Michael Rosholm and Mette Verner

Working Paper No. 2002-1

DEPARTMENT OF ECONOMICS

SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

A Comparison of Different Estimators for Panel Data Sample Selection Models

Peter Jensen, Michael Rosholm, and Mette Verner*

CIM, CLS,
Department of Economics, Aarhus School of Business, and
Department of Economics, University of Aarhus

December 2001

Abstract

In this paper, we perform an extensive Monte Carlo study of the finite sample properties of different estimators for panel data sample selection models. The estimators investigated are various two-step estimators and maximum likelihood estimators with simultaneous equations for the sample selection process and the equation of interest. The main result of the Monte Carlo study is that the maximum likelihood estimators of random effects models in general perform better than the two-step estimators.

Keywords: Panel data, sample selection, individual-specific effects
JEL classification: C23, C33

*Corresponding author: Peter Jensen, Department of Economics, The Aarhus School of Business, Fuglesangs Allé 20, DK-8210 Aarhus V, Denmark, Phone: +45 89 48 64 82, Fax: +45 86 15 51 75, E-mail: pje@asb.dk

1 Introduction

When working with micro data sets, two problems invariably present themselves: the problem of sample selection and the problem of unobserved individual-specific effects. The problem of sample selection is that a selective sample - one obtained by employing a non-random sampling scheme; for instance, by sampling only individuals with an observed wage - leads to biased parameter estimates, unless an appropriate correction is made. Heckman (1979) pointed out the sample selection problem and introduced a method to make such a correction. Since then, numerous sample selection estimators have been introduced and applied in the literature (see Vella (1998) for a recent survey of sample selection models). The problem of unobserved individual-specific effects may be solved using panel data where each unit of analysis is observed more than once. A number of estimators are available for estimating the parameters of panel data models taking into account the presence of such unobserved effects (see Hsiao (1986) or Baltagi (1995) for overviews of panel data methods).

Sample selection models are frequently estimated in applied microeconometric work using cross-section data, but they are less frequently applied when panel data are available. It is often argued that if the sample selection process is constant over time, then some standard panel data estimators eliminate sample selection bias since they "difference out" both the unobserved individual-specific effect and the sample selection effect. However, in general there is no reason to believe that the sample selection process is time-invariant. Furthermore, estimators exploiting differencing will not be efficient. In general, unobservable individual-specific effects may occur in both the selection equation and the equation of interest, and they may exhibit a complex correlation structure. Hence, in applied microeconometric work one often needs estimation procedures for panel data sample selection models.

Recently, the econometrics literature has contained various suggestions on such estimation procedures. In general, two paths have been followed in the development of panel data sample selection model estimators: two-step estimators following the idea of Heckman (1979) and maximum likelihood estimators. The former includes the estimators suggested by Wooldridge (1995), Kyriazidou (1997), Vella and Verbeek (1999), Rochina-Barrachina (1999), and Lee (2001). The latter has been applied by Husted *et al.* (2001) and Verner (2001). The sample selection problem has also been considered

in the related setting of attrition in panel data by Hausman and Wise (1979), Ridder (1990), and Verbeek and Nijman (1996). The evidence on the performance of the various panel data sample selection estimators is rather sparse, although recently Dustmann and Rochina-Barrachina (2000) and Lee (2001) have reported on comparisons of some of these estimators.

The purpose of this paper is to undertake an extensive investigation of these different estimation methods for panel data sample selection models by a Monte Carlo study. Our viewpoint is that of a microeconometrician doing applied work. Hence, we will specify the data generating processes with a view towards "typical" processes encountered in empirical applications, namely by introducing various correlation structures of the unobserved components. In addition, we extend some of the previously proposed estimators to examine if there are simple ways to remedy violations of the basic assumptions. Specifically, we have chosen to investigate the properties of the following estimators:

- Selection process estimated by a conditional logit; the equation of interest estimated by a fixed effect model (the "naive" estimator which does not take sample selection into account).

- Selection process estimated by a conditional logit; the equation of interest estimated by use of the semiparametric estimator proposed by Kyriazidou (1997). Estimation of the equation of interest involves a correction for sample selection, based on the conditional logit.

- Selection process estimated by a panel data probit model; the equation of interest estimated by OLS with a correction calculated from the first step included, based on the two-step estimators proposed by Wooldridge (1995) and Vella and Verbeek (1999).

- Selection process estimated by a panel data probit model; the equation of interest estimated by a fixed effect method with a correction calculated from the first step included, an extension of the above estimator.

- Selection process estimated by a panel data probit model with "Mundlak correction" introduced; the equation of interest estimated by a fixed effect method with a correction calculated from the first step included, another extension of the above estimator.

- Parametric panel data random effects model, both equations estimated simultaneously by maximum likelihood.
- Parametric panel data random effects model, both equations estimated simultaneously by maximum likelihood, "Mundlak correction" introduced.

Each of these estimators may produce consistent estimates given that certain assumptions are satisfied. The consistency requirements are different for the various estimators and in many typical applications some of the assumptions are likely to be violated. From a practical point of view, this means that the choice of model specification and the choice of an appropriate estimator become intrinsically linked. If we have a correctly specified model, we would ideally like to choose a consistent and asymptotically efficient estimator, like e.g. the ML estimator for that model. However, in real-world applications we do not know the correct model and it is therefore important to have some knowledge about the performance of the estimators when the model is misspecified. In our Monte Carlo study, we provide such evidence by specifying various data generating processes and applying the estimators even if they are inconsistent. We are thus able to investigate the robustness against various forms of misspecification, such as for instance assuming that the error term in the selection equation follows a logistic or normal distribution or that the unobserved components have a certain correlation structure.

The results of our extensive Monte Carlo study show that, in general, the simultaneous maximum likelihood estimators of the random effects models have the best performance when both bias and variation are taken into account. However, these estimators are more computationally demanding than the two-step estimators. Furthermore, in a number of situations the two-step estimators are almost as well behaved as the maximum likelihood estimators, and the two-step estimators may therefore be preferred in cases where the computational burden represents a major obstacle.

The remainder of the paper is organized as follows. Section 2 presents the different estimation methods for panel data sample selection models that we investigate in this paper. Section 3 describes the Monte Carlo simulation design, and Section 4 contains the results of the Monte Carlo study and a discussion of the advantages and disadvantages of the estimators. Finally, Section 5 summarizes and concludes.

2 Estimation methods for panel data sample selection models

The model we consider can be formulated as follows:

$$\begin{aligned} y_{it}^* &= x_{it}'\beta + \alpha_i + \varepsilon_{it} \\ d_{it}^* &= w_{it}'\gamma + \eta_i + u_{it} \\ d_{it} &= 1 \text{ if } d_{it}^* > 0,\ 0 \text{ otherwise} \\ y_{it} &= y_{it}^* \cdot d_{it}, \end{aligned} \tag{1}$$

where i $(i = 1, ..., N)$ denotes the individual and t $(t = 1, ..., T)$ denotes the time period. We will restrict our attention to the case of $T = 2$. The equation of interest is the first one and the selection process is the second one in 1. Here, β and γ are unknown parameter vectors which we wish to estimate, and x_{it} and w_{it} are vectors of explanatory variables, possibly with common elements. The α_i and η_i are unobservable time-invariant individual-specific components which are possibly correlated with each other and with the explanatory variables, and ε_{it} and u_{it} are unobserved disturbances or idiosyncratic error terms (possibly correlated with each other). The variable y_{it}^* is only observable if the indicator variable $d_{it} = 1$, which gives rise to sample selectivity.

In the cross-sectional case the individual-specific components (or "incidental parameters") are absorbed into the error terms and it is rather simple to estimate the sample selection effect in a discrete choice model and insert it in the equation of interest. However, in the panel data case complications arise: First, estimation of γ requires more complicated estimation methods, since in a nonlinear model first-differencing the equation does not eliminate the incidental parameters. Second, the sample selection effect entering additively in the equation of interest is an unknown nonlinear function of the observed and unobserved time-varying regressors of the selection process and does not disappear when simple first-differencing is made. In the applied literature various more or less suitable methods have been used. We will investigate how well the following estimators perform:

Estimator 1 (CL/FE): Conditional logit / fixed effect

To consistently estimate the parameters of the selection process we use a conditional logit model. In that model the basic idea is to avoid the incidental

parameters problem by conditioning out the individual-specific components, η_i.

If a minimum sufficient statistic τ_i for η_i exists which is not dependent on γ, then the conditional density

$$f(d_i|\gamma, \tau_i) = \frac{f(d_i|\gamma, \eta_i)}{g(\tau_i|\gamma, \eta_i)}$$

for $g(\cdot) > 0$, does not depend on η_i.

Maximizing this conditional density will give a consistent estimator of the common parameter γ under mild regularity conditions. In the case of a binary choice problem, such a sufficient statistic can be $\sum_{t=1}^{2} d_{it}$. Since only the cases where$\sum_{t=1}^{2} d_{it} = 1$, that is, the cases where individuals change status between time-periods, contribute to the conditional likelihood function, only these individuals are used for estimation of the parameters. The conditional log-likelihood function becomes :

$$\ln\ L = \sum_{i \in B} \left[z_i \ln F\left((w_{i2} - w_{i1})' \gamma \right) + (1 - z_i) \ln(1 - F((w_{i2} - w_{i1})'\gamma)) \right],$$

where F, in the case of a logit model, is:

$$F(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}$$

and $z_i = 1$ if $(d_{i1}, d_{i2}) = (0, 1)$ and $z_i = 0$ if $(d_{i1}, d_{i2}) = (1, 0)$.

The parameters of the equation of interest are estimated by a "fixed effects" approach. This means that first-differences of the linear model are taken on an individual basis thereby eliminating the incidental parameters. Note, that only individuals where $d_{i1} = d_{i2} = 1$ can be used in this procedure. OLS is then performed on the first-differences. This "naive" estimator of β ignores sample selectivity and is therefore inconsistent. However, this estimator may be viewed as a "benchmark" against which we can compare the other estimators to assess the improvement that is obtained by controlling for sample selectivity.

Estimator 2 (KYRI): Two-step estimator proposed by Kyriazidou (1997)

This estimator follows the two-step approach proposed by Heckman (1979) in the case of parametric estimation of cross-section selection models.

In the panel data case, the sample selection effect, λ_{it}, can be defined as:

$$\begin{aligned} \lambda_{it} &\equiv E(\varepsilon_{it}|d_{i1}=1,d_{i2}=1,\zeta_i) \\ &= \Lambda_{it}(w'_{i1}\gamma+\eta_i, w'_{i2}\gamma+\eta_i,\zeta_i), \end{aligned} \tag{2}$$

where $\zeta_i \equiv (w_{i1}, w_{i2}, x_{i1}, x_{i2}, \alpha_i, \eta_i)$.

By subtracting this selection effect from the original error term of the equation of interest, we get a new error term, $v_{it} \equiv \varepsilon_{it} - \lambda_{it}$, which by construction satisfies $E(v_{it}|d_{i1}=1,d_{i2}=1,\zeta_i)=0$. We may now rewrite the equation of interest as:

$$y_{it} = x'_{it}\beta + \alpha_i + \lambda_{it} + v_{it} \tag{3}$$

Ideally, we would like to eliminate both the individual-specific component, α_i, and the selection effect, λ_{it}, by first-differencing the rewritten equation. Kyriazidou (1997) discusses in detail the assumptions needed for performing this "differencing out". There is no need for making strong distributional assumptions across individuals, since first-differences are taken on an individual basis. In particular, the functional form of $\Lambda_{it}(\cdot)$ may be allowed to vary across individuals. Under certain weak distributional assumptions (see Kyriazidou (1997) for further details), the sample selection effects λ_{it} will be the same in the two periods, when $w'_{i1}\gamma = w'_{i2}\gamma$. This suggests estimating β by OLS from a subsample consisting of those observations that have $w'_{i1}\gamma = w'_{i2}\gamma$ and $d_{i1} = d_{i2} = 1$ (presuming that γ is known). However, γ is unknown and it may be the case that $w'_{i1}\gamma \neq w'_{i2}\gamma$ for all individuals in our sample. The idea behind the estimator of β is thus to apply an estimation scheme that approximates the outlined procedure, based on pairs of observations for which $w'_{i1}\gamma$ and $w'_{i2}\gamma$ are "close".

The proposed two-step estimation procedure is as follows: In the first step, estimate the parameters of the selection equation, γ, consistently. In this paper, this is done by a conditional logit, using only the individuals who change status over time, i.e. $d_{i1}+d_{i2}=1$. Then in the second step, use these estimates to construct weights, ψ_{in}, to be inserted in a weighted least square regression. Let $D_i = 1\{d_{i1}=d_{i2}=1\} = d_{i1}d_{i2}$. This results in the following estimator

$$\widehat{\beta}_n = \left[\sum_{i=1}^{n} \widehat{\psi}_{in}\Delta x'_i \Delta x_i D_i\right]^{-1} \left[\sum_{i=1}^{n} \widehat{\psi}_{in}\Delta x'_i \Delta y_i D_i\right], \tag{4}$$

where ψ_{in} are constructed as "kernel" weights, declining to zero as $|w'_{i1}\widehat{\gamma}_n - w'_{i2}\widehat{\gamma}_n|$ increases:

$$\widehat{\psi}_{in} \equiv \frac{1}{h_n} K\left(\frac{\Delta w'_i \widehat{\gamma}_n}{h_n}\right)$$

K is a "kernel density" function, and h_n is a sequence of bandwidths which tends to zero as $n \to \infty$.

In the estimations performed in this paper, the kernel function is chosen to be a standard normal density function and the bandwidth is set to $h_n = h \cdot n^{-1/5}$.[1] The result of this estimation procedure is, in the case of a given number of observations, that observations with a lot of selection bias are given little weight, and asymptotically only individuals with no selection bias are used in the estimation procedure and the individual-specific component is eliminated from the equation of interest by taking first-differences.

It should be noted that the proposed estimator is asymptotically biased, and Kyriazidou provides a correction, resulting in an asymptotically unbiased, normally distributed estimator of β. This is done using $\widehat{\beta}_{n,\delta}$, which is an estimator of β estimated with a smaller bandwidth, namely $h_n = h \cdot n^{-\delta/5}$, and as suggested in her paper δ is set to 0.1.

The bias-corrected estimator of β becomes

$$\widehat{\widehat{\beta}}_{n,\delta} = \frac{\widehat{\beta}_n - n^{-(1-\delta)2/5}\widehat{\beta}_{n,\delta}}{1 - n^{-(1-\delta)2/5}}$$

which is the estimator used in the subsequent Monte Carlo experiments. This estimator is consistent for β, provided that a consistent estimator is used for γ.

Estimator 3 (TSE1): Two-step estimator based on Wooldridge (1995) and Vella and Verbeek (1999)

[1] For a given kernel density function and a given sample size the problem of choosing bandwidth reduces to choosing the constant h. Kyriazidou proposes a "plug-in" method for performing this choice, but her Monte Carlo study shows that this procedure results in less variation of the estimates, at the cost of a larger mean bias. Therefore, we have chosen not to complicate matters and to set h to 1, following the simple Monte Carlo estimations in Table II of her paper. This means that the bandwidth for our purpose is set to $h_n = n^{-1/5}$.

The basic idea of this estimator is also in the spirit of Heckman's (1979) method. Since our model is a special case of the very general model considered by Vella and Verbeek (1999), the two-step estimator they propose can also be applied to our case. This results in an estimator that is a slight generalization of the two-step estimator proposed by Wooldridge (1995), where we estimate the parameters of the selection equation by a panel method rather than by a cross-section method. First, estimate the selection equation by a panel data probit model where the individual-specific component is treated as a normally distributed random effect. The contribution to the likelihood function from a single individual for that model is

$$\mathcal{L} = \int_{-\infty}^{\infty} \prod_{t=1}^{2} \Phi\left[\left(2d_{it} - 1\right)\left(w_{it}'\gamma + \eta_i\right)\right] \cdot f\left(\eta_i\right) d\eta_i$$

Next, calculate the expected value of the error terms in this model:

$$\begin{aligned} e_{it} &= E\left[\eta_i + u_{it} | w_{i1}, w_{i2}, d_{i1}, d_{i2}\right] && (5) \\ &= \int \left[\eta_i + E\left[u_{it} | w_{i1}, w_{i2}, d_{i1}, d_{i2}, \eta_i\right]\right] f\left(\eta_i | w_{i1}, w_{i2}, d_{i1}, d_{i2}\right) d\eta_i && (6) \end{aligned}$$

where $E\left[u_{it} | w_{i1}, w_{i2}, d_{i1}, d_{i2}, \eta_i\right]$ is the cross-sectional generalized residual (Gourieroux *et al.*, 1987), that is,[2]

$$\begin{aligned} E\left[u_{it} | w_{i1}, w_{i2}, d_{i1}, d_{i2}, \eta_i\right] &= E\left[u_{it} | w_{it}, d_{it}, \eta_i\right] \\ &= \frac{\phi\left(w_{it}'\widehat{\gamma} + \eta_i\right)}{\Phi\left(w_{it}'\widehat{\gamma} + \eta_i\right)} d_{it} - \frac{\phi\left(w_{it}'\widehat{\gamma} + \eta_i\right)}{1 - \Phi\left(w_{it}'\widehat{\gamma} + \eta_i\right)} \left(1 - d_{it}\right) \end{aligned}$$

and $f\left(\eta_i | w_{i1}, w_{i2}, d_{i1}, d_{i2}\right)$ is calculated exploiting the following expression:

$$f\left(\eta_i | w_{i1}, w_{i2}, d_{i1}, d_{i2}\right) = \frac{f\left(d_{i1}, d_{i2}, \eta_i | w_{i1}, w_{i2}\right)}{f\left(d_{i1}, d_{i2} | w_{i1}, w_{i2}\right)}$$

The term in the denominator is identical to the likelihood contribution for an individual.

Given this expression, the expression in 5 is calculated by (in the case of a normal mixing distribution) numerical integration. Once this is done, the following equation is estimated by OLS

$$y_{it} = x_{it}'\beta + e_{it}\theta + e_{i\cdot}\mu + \epsilon_{it} \qquad (7)$$

[2]In the present case, the u_{it}'s are assumed to be independent over time.

where $e_{i\cdot}$ is the individual-specific average value of the residual e_{it}, and ϵ_{it} now denotes the error term in the equation of interest (which may still contain an individual-specific component). If an individual-specific component is present in the error term and it is correlated with x_{it}, then the estimator of β will be inconsistent.

Estimator 4 (TSE2): Extension of two-step estimator above

For this estimator the first step is the same as above in estimator 3, that is, a panel data probit model. However, in the second step, the equation of interest is estimated in first-differences (no constant term included),

$$y_{i2} - y_{i1} = (x_{i2} - x_{i1})' \beta + (e_{i2} - e_{i1}) \theta + (\epsilon_{i2} - \epsilon_{i1})$$

Hence, the individual-specific average of the error term e_{it} is eliminated, as is any individual-specific component that might have been present in the equation of interest. Particularly, a component correlated with x_{it} would lead to inconsistency in the estimator of β used in estimator 3 above.

Estimator 5 (TSE3): Extension of two-step estimator above

Basically this estimator is a generalization of estimator 4, where the first step is augmented by introducing a "Mundlak correction" in the selection equation. The random effects formulation may be criticized on the grounds that it neglects the correlation that may exist between the random effect and the explanatory variables. If this correlation is ignored, the estimators of the parameters of interest (here γ and β) will be inconsistent for a finite number of observations per individual. Mundlak (1978) proposes a way to take this correlation into account. He approximates $E(\eta_i|w)$ by a linear function and includes the individual means of the explanatory variables in the linear predictor.[3] A simple F-test will then determine whether correlation is present. We make the assumption that η_i is correlated with the individual-specific average, $\overline{w_i} = (w_{i1} + w_{i2})/2$. Hence, we assume

$$\eta_i = \overline{w_i}\omega + \upsilon_i,$$

where $\upsilon_i \sim N(0, \sigma_\upsilon^2)$. In the second step, the parameters of the equation of interest is estimated by first-differencing, as in estimator 4 above. If the

[3] While this approach works well in the linear regression model, there is no guarantee that it does so in a nonlinear estimation problem.

distributional assumptions for the selection equation are satisfied, then this estimator will be consistent for γ and β.

Estimator 6 (MLE1): Maximum Likelihood Estimator

In this estimation procedure the selection process and the equation of interest are estimated simultaneously. For this purpose it is necessary to specify the joint distribution of the individual-specific components treated as random effects in the selection equation and the equation of interest. Specifically, we make the following assumptions for the model. The idiosyncratic error terms are assumed to follow a bivariate normal distribution

$$(\varepsilon_{it}, u_{it}) \sim N(0, \Sigma), \text{where } \Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1 \end{bmatrix}$$

Furthermore, we make the following assumptions concerning the random effects and their interactions with the idiosyncratic errors:

$$\begin{aligned} E[\alpha_i] &= E[\eta_i] = 0 \\ \varepsilon_{it}, u_{it} &\perp \alpha_i, \eta_i \end{aligned}$$

Thus, the individual-specific components in the selection equation and the equation of interest may be correlated, but they are assumed to be uncorrelated with the idiosyncratic error terms. The likelihood of a single observation, conditional on the random effects, is then

$$\begin{aligned} \mathcal{L}_{it}(\gamma, \beta, \Sigma | \alpha_i, \eta_i) &= f(\varepsilon_{it}, u_{it} | \alpha_i, \eta_i) \\ &= \left[\int_{-w_{it}'\gamma - \eta_i}^{\infty} \phi_{\varepsilon u}(y_{it} - x_{it}'\beta - \alpha_i, u) du \right]^{d_{it}} \\ &\quad \times \left[\int_{-\infty}^{-w_{it}'\gamma - \eta_i} \int_{-\infty}^{\infty} \phi_{\varepsilon u}(\varepsilon, u) d\varepsilon du \right]^{1-d_{it}} \\ &= \left[\int_{-w_{it}'\gamma - \eta_i}^{\infty} \phi_{u|\varepsilon}(u | y_{it} - x_{it}'\beta - \alpha_i) \cdot \phi_\varepsilon(y_{it} - x_{it}'\beta - \alpha_i) du \right]^{d_{it}} \\ &\quad \times \left[\int_{-\infty}^{-w_{it}'\gamma - \eta_i} \phi_u(u) du \right]^{1-d_{it}} \\ &= \left[\left(1 - \Phi_{u|\varepsilon}(-w_{it}'\gamma - \eta_i | y_{it} - x_{it}'\beta - \alpha_i) \right) \cdot \phi_\varepsilon(y_{it} - x_{it}'\beta - \alpha_i) \right]^{d_{it}} \\ &\quad \times \left[\Phi_u(-w_{it}'\gamma - \eta_i) \right]^{1-d_{it}}, \end{aligned}$$

where the conditional distribution $u|\varepsilon \sim N\left(\frac{\rho\varepsilon}{\sigma_\varepsilon}, (1-\rho^2)\right)$. When a distribution is specified for the random effects it is straightforward to integrate them out of the likelihood function. If (α_i, η_i) is distributed according to the distribution function G$(\cdot)$ we have:

$$\mathcal{L}_i(\gamma, \beta, \Sigma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i} f(\varepsilon_{it}, u_{it}|\alpha_i, \eta_i)\right] dG(\alpha_i, \eta_i)$$

In the estimations of this paper, $G(\cdot)$ is specified as a bivariate discrete distribution with 2×2 points of support.[4] If the distributional assumptions are satisfied, then this estimator will be consistent for γ and β, but it does not allow for correlation between the observed and unobserved variables.

Estimator 7 (MLE2): Extension of Maximum Likelihood Estimator

This estimator is the same as the one described above, except that in the selection equation, we assume that

$$\eta_i = \overline{w_i}\omega + \upsilon_i,$$

that is, we introduce once again the Mundlak correction. In addition, we make the assumption that the observed and unobserved variables of the main equation may be correlated

$$\alpha_i = \overline{x_i}\varphi + \varsigma_i,$$

and we finally assume that $(\upsilon_i, \varsigma_i)$ follows a bivariate discrete distribution with 2×2 points of support. If the distributional assumptions are satisfied, then this estimator will be consistent for γ and β, and in particular, it allows for correlation of the observed and unobserved variables.

3 Monte Carlo simulation design

The data generating processes (DGP) are designed to mimic situations common in applied microeconometric work. There are two explanatory variables

[4] Clearly, the precision of this estimator depends on the number of support points in the discrete distribution of the individual-specific effects. Our interest in this paper is on estimators that are simple to apply and compute. Hence, we have restricted our attention to the case with 2×2 points of support. However, we briefly investigate the properties of the estimator when more support points are allowed; see Appendix 2.

in the selection equation: w_1 and w_2. They are independent and normally distributed, with mean 0 and variance 1. There is one variable, x, in the main equation, and $x = w_1$. We thus have one variable which is excluded from the main equation, as it is standard in order to obtain nonparametric identification of the parameters of the model.

The idiosyncratic error term in the selection equation, u_{it}, follows either a logistic distribution or a normal distribution, in both cases with mean 0 and variance 1. We perform one set of Monte Carlo experiments with u_{it} generated by the logistic distribution and another set with u_{it} generated by the normal distribution. The results from the former are reported in the following section, whereas the results from the latter are reported in Appendix 1. The idiosyncratic error term in the equation of interest is defined as $\varepsilon_{it} = 0.6u_{it}+0.8\xi_1$, with ξ_1 an independent standard normal variable. With this specification the correlation between u_{it} and ε_{it} is 0.6. In addition, we define three random variables: ξ_2, ξ_3, and ξ_4. These are individual-specific variables. They are all independent and standard normally distributed. This defines the common set-up of all the different DGPs that we use in this paper.

In the following, we specify three different DGPs, which differ according to the correlation structure between observed and unobserved variables and the correlation structure across equations.

Model A: Here we assume that the two individual-specific components α and η are correlated. More specifically, let

$$\eta_i = \xi_2 + \xi_4 \tag{8}$$
$$\alpha_i = \xi_3 + \xi_4 \tag{9}$$

Model B: In this case we assume that observed and unobserved variables are correlated, but that there is no correlation between individual-specific components across equations

$$\eta_i = \xi_2 + \frac{w_{2,i1} + w_{2,i2}}{2} \tag{10}$$
$$\alpha_i = \xi_3 + \frac{x_{i1} + x_{i2}}{2} \tag{11}$$

Note, however, that this implies that α and w_1 are positively correlated, since $x = w_1$.

Model C: This is the most general model. Here we allow for both types of correlation structures, between the two individual-specific components and

between individual-specific components and observed variables.

$$\eta_i = \xi_2 + \frac{w_{2,i1} + w_{2,i2}}{2} + \xi_4 \tag{12}$$

$$\alpha_i = \xi_3 + \frac{x_{i1} + x_{i2}}{2} + \xi_4 \tag{13}$$

The parameter values are all set to 1 ($\beta = \gamma_1 = \gamma_2$). Given these values, we generate samples of two different sizes $N = 1000,\ 4000$ individuals[5] observed in two subsequent periods, i.e. $T = 2$.

We use the GAUSS programme to generate the random numbers (procedure RNDN for normal random variables, procedure RNDU for uniform random variables). All random numbers are drawn independently over individuals and time. We use the same draws for all three (six) DGPs to minimize sources of variation across the experiments.

Each Monte Carlo experiment is performed 500 times.

4 Results

In this section the results of the Monte Carlo experiments are presented. Tables 1, 2, and 3 present the mean bias, the median bias, the standard error, the root mean squared error, and the median absolute deviation of the various estimation techniques for both the γ and the β parameters. We have chosen to report the results for the parameters of both the selection equation and the equation of interest, since they may all be of interest to the researcher. As indicated by the terminology, the main interest is often on the parameters of the equation of interest, but in several applications the selection equation may also represent an important part of the model.

Again it is important to stress that our viewpoint is that of a microeconometrician doing applied work. Our main interest is thus in finding an estimation procedure that works well in practical applications, where typically we do not know the correct model (the DGP). In our Monte Carlo experiments, we apply all seven estimation procedures to the specified DGPs regardless of the theoretical properties of the estimators. This allows us to provide evidence on the performance of the estimators in the presence of

[5] We have also performed all experiments with $N = 250$ individuals, but the results of these experiments are not reported in the paper due to space considerations. They are available from the authors on request.

misspecification of the model or violations of consistency requirements. The results reported and discussed in this section are all based on the DGPs with u_{it} generated by a logistic distribution. This means that only Kyriazidou's estimator is consistent for γ and β. As mentioned in the previous section, we report results from another set of experiments with u_{it} generated by a normal distribution in Appendix 1. There the TSE3 and the MLE2 estimators are consistent for γ and β, whereas the remaining estimators are not.

Table 1 shows the results from the seven different estimation procedures when the data generating process is specified by Model A with the error term in the selection equation generated by a logistic distribution. This is the model where the two individual-specific components are correlated, i.e. there is sample selection on the individual-specific components.

Table 1 about here

For the parameters of the selection equation, the consistent estimator based on the conditional logit model (CL/FE and KYRI) is moderately biased for these finite samples, whereas the standard error shows that this estimator has a considerable variance, resulting in a considerable rmse. Clearly, this is a result of the estimation procedure where only individuals with a change in status contribute to the conditional likelihood function. The estimators from the other two-step models have a lower bias than the conditional logit estimator and are also more efficient, with a standard error of about half the size. Among the other two-step estimators the small loss of efficiency when the "Mundlak correction" is included leads to the conclusion that the TSE1/TSE2 estimator performs best. The ML estimators suffer from a large negative bias and this leads, despite a low standard error, to a very large rmse. These problems are most pronounced in the model without Mundlak correction. So, if the parameters of interest are the γ's, then the estimator based on the random effects probit model in the two-step procedure does best.

Turning to the β parameter, we see a considerable bias of the estimator from the simple fixed effects model (CL/FE), which is also reflected in the rmse, whereas the standard error is rather low. Since this estimator does not take the sample selection into account, this result is as expected. Kyriazidou's estimator is virtually unbiased, but this is at the expense of a considerable loss in efficiency, which is reflected by the large standard error and the correspondingly large rmse. In the other two-step models, the unobserved

heterogeneity generates a large bias in the simple version, whereas the fixed effects version performs very well both in terms of bias and variation. The inclusion of the Mundlak correction does not make much difference. The ML estimators do slightly worse than the TSE2 and TSE3 estimators in terms of bias, but slightly better in terms of variance. The inclusion of the Mundlak correction term leads to a slightly lower bias. In total, the TSE2 and TSE3 estimators have the best performance when both bias and variation are taken into account, closely followed by the ML estimators.

If both the γ and the β parameters are of interest, the ML estimators should be avoided due to the large bias in the estimation of the γ parameters. Instead, a reasonable choice of estimator would be the TSE2 estimator, since it performs well for both sets of parameters.

Table 2 shows the results when the data generating process is specified by Model B with the error term in the selection equation generated by a logistic distribution. This is the model where the two individual-specific effects are uncorrelated, but with correlation between observables and unobservables.

Table 2 about here

Again, first considering the parameters of the selection equation, we see that the consistent conditional logit estimator (CL/FE and KYRI) has the lowest bias of the estimators considered, but it still has a considerable variation, especially when the sample size is small. The other two-step estimators have very large biases, in particular for γ_2 which is the coefficient of the observed variable that is correlated with the unobserved individual-specific effect. The introduction of the Mundlak correction (in TSE3) reduces this bias, but it is still considerable and the correction actually increases the bias for γ_1. The ML estimator is also biased, but here the Mundlak correction improves the estimator (MLE2) by removing most of the bias which in combination with a low standard error means that this estimator has the lowest rmse. So, if the parameters of interest are the γ's, then the preferred estimators are the conditional logit (which does best in terms of bias) and the MLE2 (which does best in terms of variation and rmse).

For the equation of interest, Kyriazidou's estimator is again virtually unbiased, but at least for the sample sizes considered here it is quite inefficient, leaving it with a very large rmse. Compared with model A, the absence of correlation between the individual-specific effects has made the performance of the simple two-step estimator (TSE1) much better, since it has a very

low bias and a small standard error. Particularly for the small sample size, $N = 1000$, it can compete with the other estimators in terms of rmse. The TSE2 estimator is unattractive due to a fairly large bias. The ML estimator is seriously biased (which is not a big surprise, since it is also inconsistent), but the Mundlak correction removes this bias entirely, yielding an estimator that is virtually unbiased (MLE2). This does not come as a surprise since this estimator is specifically constructed to correct for the correlation between observables and unobservables. Actually, also the efficiency of this estimator is very good, which leads to the conclusion that the preferred estimator for β is MLE2, followed by TSE1 and KYRI.

When observables and unobservables are correlated, it turns out that the MLE2 estimator is uniformly superior in terms of rmse for all parameters of the model, and in particular for the β parameter since it is unbiased.

Table 3 shows the results when the data generating process is specified by Model C with the error term in the selection equation generated by a logistic distribution. This is the model where both types of correlation are present.

Table 3 about here

For the parameters of the selection equation, we see again that the consistent conditional logit estimator (CL/FE and KYRI) is moderately biased, at a level comparable with the bias observed for Models A and B. However, its standard error is still quite large. For the two-step estimator (TSE1/TSE2), the correlation between observables and unobservables still creates a large bias for the γ_2 parameter, although the estimator shows a good performance for the γ_1 parameter. The inclusion of the Mundlak correction (TSE3) eliminates the bias almost entirely, which in combination with a very small standard error makes this estimator the estimator preferred for the γ parameters. Apparently, the presence of both types of correlation is important for this estimator, as can be seen by comparing with its performance for Model B where the Mundlak correction was not able to remove the bias. For the ML estimators, we find the same pattern as we did for Model A, i.e. they cannot handle the correlation between unobservables across the two equations which causes a large bias, regardless of the Mundlak correction[6].

[6]However, the performance of the MLE2 estimator for the γ parameters may be considerably improved by increasing the number of support points for the heterogeneity distribution in the selection equation. Results regarding this aspect are reported in Appendix 2.

Turning to the equation of interest, we see the same picture as for Model B. Kyriazidou's estimator is unbiased for the large sample size, but it has a very large standard error. The two-step estimators TSE1, TSE2, and TSE3, are reasonably well-behaved with the simple version TSE1 actually performing best in terms of rmse. The correlation between observables and unobservables still causes the ML estimator to be very biased, but this is handled by the Mundlak correction. So for the β parameter, the MLE2 estimator is preferred.

4.1 Summary and comparison

Above we have discussed the performance of the various estimators for the three different data generating processes. This has been done for both the γ parameters and the β parameters, and our focus has been on bias and efficiency in finite samples. Clearly, to be efficient an estimator needs to have a small bias and a relatively small standard error. However, these two aspects may often conflict like we saw in a number of cases above. In the same manner, a given estimator may not have good properties for the parameters of both equations at the same time. An additional aspect that deserves attention is the computing time of the various estimators. In the following, we will briefly summarize our results for each estimator and make a final comparison where all these aspects will be considered.

CL/FE: This estimator is consistent for the γ parameters, but is moderately biased in finite samples and suffers from a large standard error. Thus, it is quite inefficient for estimating the γ parameters. It is obviously not a good choice for estimating the β parameter, since it is not consistent.

KYRI: For the γ parameters, this estimator is identical to CL/FE. For the β parameter, it is virtually unbiased, but it has a very large standard error, which makes it inefficient for the sample sizes considered here.

TSE1: It performs well for the γ parameters when there is no correlation between observables and unobservables, but it becomes seriously biased when there is correlation. In that situation, however, it typically becomes a good estimator of the β parameter.

TSE2: For the γ parameters, this estimator is identical to TSE1. For the β parameter, it is moderately biased and has a relatively small standard error, but it is only the preferred choice of estimator (together with TSE3) when there is no correlation between observables and unobservables.

TSE3: It performs well for the γ parameters when the unobservables

are correlated across equations, but not otherwise. For the β parameter, its performance is similar to TSE2, although it is slightly more rmse efficient.[7]

MLE1: Unless there is no correlation between observables and unobservables, this estimator is a complete waster. In the case with no correlation between observables and unobservables, it is still unable to provide good estimates of the γ parameters, but it does a nice job estimating the β parameter.

MLE2: In the case where the unobservables are not correlated across equations, this estimator is efficient for the γ parameters despite a moderate bias. Otherwise, it is quite biased for the γ parameters (although we are able to remove most of the bias as shown in Appendix 2). For the β parameter, it is the preferred choice when observables and unobservables are correlated, and otherwise it is still a strong contender.

It is easy to rank the seven different estimators according to the computing time needed to obtain the estimates. The naive estimator (CL/FE) requires least computing time, followed by Kyriazidou's estimator, and followed again by the other two-step estimators (TSE1, TSE2, TSE3) which use approximately the same amount of computing time. The simultaneous maximum likelihood estimators (MLE1, MLE2) are computationally much more demanding than the other estimators, so judged by a computing time criterion the ML estimators are the least attractive.

Obviously, the three criteria applied here - bias, efficiency (rmse), and computing time - do not capture all aspects of the performance of the estimators, but in our opinion they represent the most important aspects. Thus, they form an appropriate basis on which to choose between the various estimators.

However, in real-world applications the true data generating processes are unknown, and we need to bear that in mind when choosing an estimator. Thus, there may be sample selection and/or correlation between observed and unobserved variables, in which case the choice of estimator narrows down to a close race between the extended version of the two-step estimator, TSE3, and the simultaneous maximum likelihood estimator with Mundlak correction, MLE2. Here, the ultimate choice must depend on the type of the main parameters of interest - the parameters of the selection process or the para-

[7]The results reported in Appendix 1 show that the performance of the TSE3 estimator is improved when the DGP has a normal error term instead of a logistic error term. This reflects that the estimator is consistent in this case. The choice of preferred estimator may thus be altered in favour of TSE3, but if it is important to safeguard against misspecification of the error term distribution, MLE2 is more robust.

meters of the equation of interest - and the sample size. When the sample size is very large, the MLE2 estimator requires a lot of computing time, especially when there are many parameters.[8]

5 Conclusion

In this paper, we have examined the consequences of the choice between different estimation methods for panel data sample selection models suggested in the literature. We have done this by use of Monte Carlo experiments with data generated by three (six) different data generating processes chosen in a way that mimics "real" data encountered in empirical applications, and in a way that makes it possible to illustrate what happens when basic model assumptions are violated.

When we look at the results of the Monte Carlo study it is important to bear in mind which parameters are the parameters of interest. In many microeconometric applications only the parameters of the equation of interest, β, are of interest, and in these cases the problems of estimating the parameters of the selection process may be ignorable. However, in general we should be able to obtain satisfactory estimates of all parameters.

The results show that, in general, the "Mundlak corrected" simultaneous maximum likelihood estimator, MLE2, has the best performance when both bias and variation are taken into account. However, this estimator is more computationally demanding than the two-step estimators. Furthermore, in a number of situations the two-step estimators are almost as well behaved as the maximum likelihood estimator, and the two-step estimators may therefore be preferred in cases where the computational burden represents a major obstacle. The estimator proposed by Kyriazidou performs very well in terms of bias, but the variation is quite considerable compared with the other estimators investigated in this paper. This large variation is reduced when the number of individuals increases (this is also shown by our results comparing $N = 1000$ with $N = 4000$), so for very large data sets this estimator may become a good choice, but in moderately sized data sets it is not a preferable choice.

A number of extensions of this study could be topics for future research:

[8] For example, with 50,000 observations, 50 parameters, and running on a decently sized Pentium III with sufficient amounts of RAM, the estimation will typically take a day (using GAUSS software).

Firstly, we have limited ourselves to the case of $T = 2$, but it is clearly relevant to analyse longer panels as well. Secondly, the specified DGPs could be generalized, mainly by changing the distribution of the error term in the selection process. In this paper, we have applied two different distributions, the logistic and the normal. Thirdly, our results show that semiparametric estimators may perform very well, and this could of course be investigated in more detail.

6 Appendix 1: Simulation results with normal error term

In this Appendix, we present the results of the Monte Carlo experiments with the error term in the selection equation, u_{it}, generated by the normal distribution. The results reported in the main text of the paper are from a set of experiments where the error term is generated by the logistic distribution. The simulation design with a logistic error term was chosen because some of our estimators of the parameters of the selection equation are based on a conditional logit model (CL/FE and KYRI). On the other hand, we are also examining estimators where we apply a probit model to estimate the parameters of the selection equation. Hence, not to treat these estimators unfairly, we have chosen also to use a simulation design with a normal error term. In addition, this also allows us to investigate the sensitivity of the estimators to the distribution of the error term.

Table A1 shows the results when the data generating process is specified by Model A with the error term in the selection equation generated by a normal distribution.

Table A1 about here

For the γ parameters, the bias of the conditional logit estimator is reduced considerably compared with the case in Table 1, at least for the large sample size. Also the standard deviation is reduced, although not as much as the bias. The bias of the other two-step estimators is also reduced remarkably, such that these estimators are now virtually unbiased. This may be taken as an indication that the bias seen in Table 1 originates from the misspecification of the distribution of error term in the selection equation. When the DGP has a normally distributed error term, the estimators based on the probit model

are more satisfactory. The ML estimators do slightly worse than in Table 1, both in terms of bias and efficiency, but the larger rmse is almost exclusively due to the increased bias. The overall conclusion has not changed compared with the results in Table 1; the two-step estimators are still preferred.

For the β parameter, there are no major changes compared with the results in Table 1. The fixed effects estimator still has a considerable bias and now a slightly larger standard error. Kyriazidou's estimator has a negligible bias, but its standard error has actually increased slightly, making it even more unattractive in terms of rmse. The extended two-step estimators are improved and so are the ML estimators. The qualitative conclusions regarding the estimators are unchanged.

Table A2 shows the results when the data generating process is specified by Model B with the error term in the selection equation generated by a normal distribution.

Table A2 about here

Compared with the results in Table 2, the performance of the TSE-based two-step estimators of the γ parameters is basically unchanged; they still have a very large bias. The conditional logit estimator has improved, mainly in terms of a reduced bias, and in terms of relative efficiency it is now the best estimator of the γ parameters. The MLE2 estimator has a larger bias, which causes it to have a larger rmse than the conditional logit estimator, thereby interchanging their ranking compared with Table 2.

For the β parameter, Kyriazidou's estimator and the MLE2 estimator have a slightly larger bias, whereas the extended two-step estimators (TSE2 and TSE3) have a smaller bias compared with Table 2. In fact, the TSE3 estimator is now virtually unbiased (reflecting the consistency of this estimator with normally distributed error terms), but this does not alter the ranking of the estimators; the preferred estimator is still MLE2, followed by TSE1.

Table A3 shows the results when the data generating process is specified by Model C with the error term in the selection equation generated by a normal distribution.

Table A3 about here

Generally, not much has changed compared with Table 3 and the ranking of the estimators is still the same. For the γ parameters, the TSE3 estimator is clearly preferable, especially since its bias and standard error have

decreased a bit. For the β parameter, the MLE2 estimator still shows the best performance in terms of both bias and rmse.

Summarizing the results, we find that the sensitivity of the estimators to the distribution of the error term is moderate. In general, there are no major changes in the performance of the estimators when we use a normal error term instead of a logistic error term. Some smaller changes occur and in most cases they are as expected; the performance of the estimators are improved when the correct distributional assumptions are applied in the estimation procedure. Most importantly, the choice of error term distribution does not have any major impact on the ranking of the estimators.

7 Appendix 2: Sensitivity to the discrete heterogeneity distribution

For the simultaneous maximum likelihood estimators, we have chosen to model the joint distribution of the random effects (the individual-specific components) as a bivariate discrete distribution with 2×2 points of support. Clearly, the precision of the ML estimators depends on the number of support points. Since our interest in the paper is on estimators that are relatively simple to apply and compute, we have restricted our attention in the main text to the case with 2×2 points of support. In this Appendix, we present some additional results obtained by increasing the number of support points in the selection equation. Specifically, we have estimated the parameters of the model where the DGP is specified by Model C and the MLE2 estimator is used with 2×2, 3×2, and 4×2 points of support, respectively. The experiments are only performed for sample size $N = 1000$. The results are shown in Table A4.

Table A4 about here

The results for 2×2 points of support are identical to the results in Table 3 in the main text of the paper. The estimator is unbiased for the β parameter, but has a moderate bias for the γ parameters. By increasing the number of support points in the selection equation, we are able to reduce this bias quite substantially; approximately by a factor 4 when going from 2 to 3 points and approximately by a factor 2 when going from 3 to 4 points. However, this decrease in bias occurs at the expense of an increased standard

deviation, but the rmse of the estimator is reduced by one third, bringing it to a level where it may compete with the TSE3 estimator (see Table 3). It should also be noted that the rmse is basically the same for 3 and 4 points, i.e. the reduction in bias is almost exactly counteracted by an increase in the standard deviation. Finally, note that the increased number of support points only has a negligible influence on the estimator of the β parameter.

8 Acknowledgements

We would like to thank Martin Browning, Claus Birn Jensen, Michael Svarer, and Allan Würtz for helpful comments. We also appreciate the research assistance from Claus Houmann Frederiksen. Financial support from the Danish Research Agency (the FREJA and SUE grants) and the Danish Social Science Research Council (the grant to the Graduate School for Integration, Production and Welfare) is gratefully acknowledged.

9 References

Baltagi, B., 1995, Econometric Analysis of Panel Data (John Wiley and Sons, New York).

Dustmann, C., and M. E. Rochina-Barrachina, 2000, Selection Correction in Panel Data Models: An Application to Labour Supply and Wages, IZA Discussion Paper 162, IZA-Bonn.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon, 1987, Generalised Residuals, Journal of Econometrics 34, 5-32.

Hausman, J.A., and D.A. Wise, 1979, Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, Econometrica 47, 455-473.

Heckman, J., 1979, Sample selection bias as a specification error, Econometrica 47, 153-161.

Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

Husted, L., H. S. Nielsen, M. Rosholm, and N. Smith, 2001, Employment and Wage Assimilation of Male First-Generation Immigrants in Denmark, International Journal of Manpower 22, 39-68.

Kyriazidou, E., 1997, Estimation of a panel data sample selection model, Econometrica 65, 1335-1364.

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

Mundlak, Y., 1978, On the Pooling of Time Series and Cross-Sectional Data, Econometrica 46, 69-86.

Ridder, G., 1990, Attrition in Multi-Wave Panel Data, in: J. Hartog, G. Ridder, and J. Theeuwes, eds., Panel Data and Labor Market Studies (Elsevier Science Publishers B.V., Amsterdam).

Rochina-Barrachina, M. E., 1999, A new Estimator for Panel Data Sample Selection Models, Annales d'Économie et de Statistique 55/56, 153-181.

Vella, F., 1998, Estimating Models with Sample Selection Bias: A Survey, Journal of Human Resources 33, 127-169.

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

Verbeek, M., and T. Nijman, 1996, Incomplete Panels and Selection Bias, in: L. Mátyás and P. Sevestre, eds., The Econometrics of Panel Data (Klüwer Academic Publishers, New York).

Verner, M., 2001, The Consequences of Unemployment in Wage Determination, in: Causes and Consequences of Interruptions in the Labour Market, PhD-thesis 2001-4, Department of Economics, University of Aarhus.

Wooldridge, J. M., 1995, Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions, Journal of Econometrics 68, 115-132.

Table 1. Results from Model A, logistic error term.

	2000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0240	0.0240	0.0138	0.0138	0.0141	-0.1155	-0.1139
	medianbias	0.0099	0.0099	0.0115	0.0115	0.0118	-0.1180	-0.1155
	std err	0.1307	0.1307	0.0664	0.0664	0.0664	0.0635	0.0637
	rmse	0.1327	0.1327	0.0678	0.0678	0.0678	0.1318	0.1305
	mad	0.0820	0.0820	0.0440	0.0440	0.0435	0.1185	0.1167
γ_2	meanbias	0.0277	0.0277	0.0138	0.0138	0.0171	-0.1105	-0.0973
	medianbias	0.0165	0.0165	0.0132	0.0132	0.0167	-0.1136	-0.0975
	std err	0.1296	0.1296	0.0647	0.0647	0.0720	0.0628	0.0696
	rmse	0.1324	0.1324	0.0661	0.0661	0.0739	0.1271	0.1196
	mad	0.0780	0.0780	0.0434	0.0434	0.0487	0.1136	0.1027
β	meanbias	-0.1099	-0.0085	-0.4851	-0.0046	-0.0050	-0.0264	-0.0201
	medianbias	-0.1125	-0.0258	-0.4861	-0.0019	-0.0019	-0.0282	-0.0193
	std err	0.0615	0.2719	0.0310	0.0753	0.0749	0.0681	0.0736
	rmse	0.1259	0.2717	0.4862	0.0753	0.0750	0.0729	0.0762
	mad	0.1126	0.1831	0.4861	0.0501	0.0506	0.0488	0.0597
	8000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0148	0.0148	0.0109	0.0109	0.0108	-0.1203	-0.1191
	medianbias	0.0145	0.0145	0.0110	0.0110	0.0107	-0.1197	-0.1187
	std err	0.0642	0.0642	0.0322	0.0322	0.0323	0.0288	0.0292
	rmse	0.0658	0.0658	0.0339	0.0339	0.0340	0.1237	0.1226
	mad	0.0448	0.0448	0.0227	0.0227	0.0227	0.1197	0.1191
γ_2	meanbias	0.0163	0.0163	0.0114	0.0114	0.0122	-0.1141	-0.1015
	medianbias	0.0113	0.0113	0.0104	0.0104	0.0123	-0.1158	-0.1008
	std err	0.0622	0.0622	0.0325	0.0325	0.0374	0.0313	0.0360
	rmse	0.0642	0.0642	0.0344	0.0344	0.0393	0.1183	0.1077
	mad	0.0394	0.0394	0.0225	0.0225	0.0246	0.1158	0.1017
β	meanbias	-0.1079	0.0042	-0.4855	-0.0034	-0.0035	-0.0287	-0.0219
	medianbias	-0.1093	0.0046	-0.4856	-0.0062	-0.0064	-0.0301	-0.0210
	std err	0.0292	0.1342	0.0155	0.0380	0.0379	0.0342	0.0348
	rmse	0.1118	0.1341	0.4857	0.0381	0.0381	0.0446	0.0411
	mad	0.1093	0.0876	0.4856	0.0270	0.0270	0.0326	0.0330

Table 2. Results from Model B, logistic error term.

	2000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0273	0.0273	0.0958	0.0958	0.1479	-0.0507	-0.0434
	medianbias	0.0038	0.0038	0.0911	0.0911	0.1457	-0.0490	-0.0463
	std err	0.1411	0.1411	0.0687	0.0687	0.0750	0.0715	0.0710
	rmse	0.1436	0.1436	0.1178	0.1178	0.1658	0.0876	0.0832
	mad	0.0899	0.0899	0.0913	0.0913	0.1457	0.0716	0.0688
γ_2	meanbias	0.0346	0.0346	0.4746	0.4746	0.1460	0.2644	-0.0434
	medianbias	0.0175	0.0175	0.4711	0.4711	0.1428	0.2570	-0.0473
	std err	0.1387	0.1387	0.0866	0.0866	0.0879	0.0786	0.0819
	rmse	0.1429	0.1429	0.4824	0.4824	0.1703	0.2758	0.0927
	mad	0.0795	0.0795	0.4711	0.4711	0.1428	0.2644	0.0749
β	meanbias	-0.1028	-0.0132	0.0023	-0.0315	-0.0056	0.7587	0.0018
	medianbias	-0.1031	-0.0146	0.0041	-0.0299	-0.0055	0.7597	-0.0031
	std err	0.0605	0.2649	0.0469	0.0673	0.0711	0.0701	0.0572
	rmse	0.1192	0.2650	0.0469	0.0743	0.0712	0.7619	0.0572
	mad	0.1031	0.1700	0.0317	0.0491	0.0458	0.7587	0.0462
	8000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0129	0.0129	0.0869	0.0869	0.1389	-0.0636	-0.0507
	medianbias	0.0098	0.0098	0.0867	0.0867	0.1392	-0.0644	-0.0515
	std err	0.0672	0.0672	0.0336	0.0336	0.0363	0.0336	0.0338
	rmse	0.0684	0.0684	0.0932	0.0932	0.1435	0.0719	0.0609
	mad	0.0454	0.0454	0.0867	0.0867	0.1392	0.0645	0.0532
γ_2	meanbias	0.0162	0.0162	0.4676	0.4676	0.1382	0.2570	-0.0485
	medianbias	0.0129	0.0129	0.4677	0.4677	0.1373	0.2555	-0.0483
	std err	0.0668	0.0668	0.0406	0.0406	0.0414	0.0364	0.0373
	rmse	0.0687	0.0687	0.4693	0.4693	0.1443	0.2595	0.0612
	mad	0.0428	0.0428	0.4677	0.4677	0.1373	0.2570	0.0519
β	meanbias	-0.1002	-0.0014	0.0036	-0.0293	-0.0043	0.7633	0.0017
	medianbias	-0.0994	-0.0011	0.0043	-0.0278	-0.0029	0.7624	0.0018
	std err	0.0298	0.1386	0.0242	0.0334	0.0353	0.0369	0.0281
	rmse	0.1045	0.1385	0.0244	0.0444	0.0355	0.7642	0.0282
	mad	0.0994	0.0924	0.0164	0.0296	0.0231	0.7633	0.0227

Table 3. Results from Model C, logistic error term.

	2000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0294	0.0294	-0.0183	-0.0183	0.0129	-0.1635	-0.1365
	medianbias	0.0158	0.0158	-0.0191	-0.0191	0.0142	-0.1634	-0.1352
	std err	0.1494	0.1494	0.0650	0.0650	0.0668	0.0664	0.0623
	rmse	0.1521	0.1521	0.0674	0.0674	0.0679	0.1765	0.1501
	mad	0.0880	0.0880	0.0444	0.0444	0.0411	0.1640	0.1371
γ_2	meanbias	0.0354	0.0354	0.2939	0.2939	0.0135	0.1574	-0.1281
	medianbias	0.0191	0.0191	0.2888	0.2888	0.0112	0.1556	-0.1298
	std err	0.1445	0.1445	0.0768	0.0768	0.0791	0.0723	0.0725
	rmse	0.1486	0.1486	0.3038	0.3038	0.0802	0.1732	0.1471
	mad	0.0865	0.0865	0.2888	0.2888	0.0533	0.1579	0.1313
β	meanbias	-0.0910	-0.0235	0.0135	-0.0250	-0.0054	0.7214	-0.0015
	medianbias	-0.0898	-0.0183	0.0131	-0.0235	-0.0051	0.7200	0.0015
	std err	0.0578	0.2851	0.0492	0.0662	0.0706	0.0778	0.0645
	rmse	0.1078	0.2858	0.0510	0.0707	0.0708	0.7255	0.0645
	mad	0.0903	0.1895	0.0320	0.0459	0.0485	0.7214	0.0510
	8000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0162	0.0162	-0.0228	-0.0228	0.0094	-0.1719	-0.1446
	medianbias	0.0132	0.0132	-0.0229	-0.0229	0.0080	-0.1722	-0.1451
	std err	0.0728	0.0728	0.0334	0.0334	0.0343	0.0335	0.0313
	rmse	0.0745	0.0745	0.0404	0.0404	0.0355	0.1751	0.1480
	mad	0.0497	0.0497	0.0293	0.0293	0.0234	0.1722	0.1446
γ_2	meanbias	0.0178	0.0178	0.2930	0.2930	0.0098	0.1524	-0.1371
	medianbias	0.0154	0.0154	0.2905	0.2905	0.0100	0.1495	-0.1385
	std err	0.0680	0.0680	0.0383	0.0383	0.0389	0.0362	0.0364
	rmse	0.0702	0.0702	0.2955	0.2955	0.0401	0.1566	0.1418
	mad	0.0462	0.0462	0.2905	0.2905	0.0271	0.1495	0.1371
β	meanbias	-0.0905	-0.0031	0.0132	-0.0256	-0.0067	0.7230	-0.0008
	medianbias	-0.0896	0.0022	0.0114	-0.0266	-0.0074	0.7241	-0.0010
	std err	0.0283	0.1318	0.0257	0.0327	0.0349	0.0410	0.0306
	rmse	0.0948	0.1317	0.0289	0.0415	0.0355	0.7242	0.0306
	mad	0.0896	0.0860	0.0186	0.0304	0.0256	0.7241	0.0244

Table A1. Results from Model A, normal error term.

	2000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0255	0.0255	0.0056	0.0056	0.0063	-0.1236	-0.1223
	medianbias	0.0177	0.0177	0.0037	0.0037	0.0040	-0.1267	-0.1251
	std err	0.1199	0.1199	0.0607	0.0607	0.0608	0.0566	0.0565
	rmse	0.1224	0.1224	0.0609	0.0609	0.0611	0.1359	0.1347
	mad	0.0798	0.0798	0.0402	0.0402	0.0407	0.1267	0.1232
γ_2	meanbias	0.0276	0.0276	0.0052	0.0052	0.0059	-0.1195	-0.1078
	medianbias	0.0186	0.0186	0.0019	0.0019	0.0045	-0.1225	-0.1103
	std err	0.1301	0.1301	0.0630	0.0630	0.0748	0.0622	0.0715
	rmse	0.1329	0.1329	0.0631	0.0631	0.0749	0.1347	0.1294
	mad	0.0865	0.0865	0.0436	0.0436	0.0511	0.1225	0.1130
β	meanbias	-0.1071	-0.0073	-0.4850	0.0004	0.0005	-0.0289	-0.0203
	medianbias	-0.1062	-0.0200	-0.4854	0.0054	0.0048	-0.0303	-0.0197
	std err	0.0603	0.2750	0.0305	0.0772	0.0774	0.0652	0.0684
	rmse	0.1228	0.2748	0.4860	0.0772	0.0774	0.0712	0.0712
	mad	0.1062	0.1759	0.4854	0.0509	0.0519	0.0460	0.0568
	8000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0036	0.0036	-0.0009	-0.0009	-0.0011	-0.1327	-0.1310
	medianbias	0.0026	0.0026	-0.0004	0.0004	0.0001	-0.1330	-0.1317
	std err	0.0608	0.0608	0.0311	0.0311	0.0310	0.0287	0.0291
	rmse	0.0609	0.0609	0.0311	0.0311	0.0310	0.1358	0.1342
	mad	0.0435	0.0435	0.0216	0.0216	0.0216	0.1330	0.1310
γ_2	meanbias	0.0028	0.0028	-0.0024	-0.0024	-0.0031	-0.1293	-0.1172
	medianbias	0.0012	0.0012	-0.0006	-0.0006	-0.0042	-0.1294	-0.1162
	std err	0.0628	0.0628	0.0314	0.0314	0.0361	0.0311	0.0355
	rmse	0.0628	0.0628	0.0315	0.0315	0.0362	0.1330	0.1224
	mad	0.0426	0.0426	0.0212	0.0212	0.0247	0.1294	0.1172
β	meanbias	-0.1069	-0.0021	-0.4845	0.0009	0.0009	-0.0252	-0.0197
	medianbias	-0.1061	-0.0020	-0.4852	0.0008	0.0010	-0.0261	-0.0197
	std err	0.0303	0.1368	0.0162	0.0397	0.0397	0.0352	0.0357
	rmse	0.1111	0.1366	0.4848	0.0397	0.0397	0.0433	0.0407
	mad	0.1061	0.0883	0.4852	0.0269	0.0275	0.0300	0.0332

Table A2. Results from Model B, normal error term.

	2000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0187	0.0187	0.0797	0.0797	0.1307	-0.0697	-0.0603
	medianbias	0.0044	0.0044	0.0762	0.0762	0.1272	-0.0749	-0.0634
	std err	0.1360	0.1360	0.0650	0.0650	0.0703	0.0649	0.0668
	rmse	0.1372	0.1372	0.1028	0.1028	0.1484	0.0952	0.0899
	mad	0.0902	0.0902	0.0792	0.0792	0.1272	0.0778	0.0748
γ_2	meanbias	0.0253	0.0253	0.4612	0.4612	0.1348	0.2494	-0.0542
	medianbias	0.0081	0.0081	0.4552	0.4552	0.1334	0.2476	-0.0574
	std err	0.1400	0.1400	0.0811	0.0811	0.0855	0.0744	0.0766
	rmse	0.1421	0.1421	0.4683	0.4683	0.1596	0.2602	0.0938
	mad	0.0888	0.0888	0.4552	0.4552	0.1334	0.2476	0.0763
β	meanbias	-0.0981	0.0072	0.0035	-0.0254	-0.0006	0.7623	0.0054
	medianbias	-0.0981	0.0059	0.0028	-0.0259	-0.0012	0.7684	0.0063
	std err	0.0576	0.2534	0.0489	0.0649	0.0703	0.0730	0.0543
	rmse	0.1138	0.2533	0.0490	0.0696	0.0703	0.7658	0.0545
	mad	0.0981	0.1644	0.0313	0.0452	0.0472	0.7684	0.0434
	8000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0075	0.0075	0.0764	0.0764	0.1265	-0.0752	-0.0623
	medianbias	-0.0003	-0.0003	0.0755	0.0755	0.1254	-0.0746	-0.0609
	std err	0.0627	0.0627	0.0320	0.0320	0.0342	0.0320	0.0325
	rmse	0.0631	0.0631	0.0828	0.0828	0.1310	0.0817	0.0703
	mad	0.0394	0.0394	0.0755	0.0755	0.1254	0.0746	0.0629
γ_2	meanbias	0.0088	0.0088	0.4531	0.4531	0.1260	0.2418	-0.0595
	medianbias	0.0067	0.0067	0.4519	0.4519	0.1257	0.2420	-0.0590
	std err	0.0659	0.0659	0.0371	0.0371	0.0396	0.0339	0.0364
	rmse	0.0664	0.0664	0.4546	0.4546	0.1321	0.2442	0.0698
	mad	0.0443	0.0443	0.4519	0.4519	0.1257	0.2420	0.0629
β	meanbias	-0.1005	0.0085	0.0052	-0.0259	-0.0009	0.7655	0.0035
	medianbias	-0.0998	0.0147	0.0054	-0.0261	-0.0007	0.7668	0.0025
	std err	0.0291	0.1447	0.0240	0.0337	0.0360	0.0390	0.0281
	rmse	0.1046	0.1448	0.0246	0.0425	0.0360	0.7665	0.0282
	mad	0.0998	0.1001	0.0165	0.0298	0.0253	0.7668	0.0228

Table A3. Results from Model C, normal error term.

	2000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0297	0.0297	-0.0288	-0.0288	0.0018	-0.1782	-0.1465
	medianbias	0.0107	0.0107	-0.0306	-0.0306	-0.0004	-0.1806	-0.1490
	std err	0.1439	0.1439	0.0630	0.0630	0.0669	0.0679	0.0632
	rmse	0.1468	0.1468	0.0692	0.0692	0.0668	0.1907	0.1595
	mad	0.0913	0.0913	0.0473	0.0473	0.0456	0.1806	0.1476
γ_2	meanbias	0.0370	0.0370	0.2878	0.2878	0.0059	0.1481	-0.1358
	medianbias	0.0257	0.0257	0.2871	0.2871	0.0025	0.1427	-0.1427
	std err	0.1450	0.1450	0.0730	0.0730	0.0769	0.0709	0.0730
	rmse	0.1495	0.1495	0.2969	0.2969	0.0771	0.1641	0.1541
	mad	0.0933	0.0933	0.2871	0.2871	0.0507	0.1427	0.1389
β	meanbias	-0.0876	0.0088	0.0140	-0.0208	-0.0015	0.7222	0.0007
	medianbias	-0.0878	0.0000	0.0118	-0.0173	0.0008	0.7227	0.0027
	std err	0.0573	0.2726	0.0488	0.0660	0.0705	0.0806	0.0603
	rmse	0.1047	0.2725	0.0507	0.0691	0.0705	0.7266	0.0603
	mad	0.0878	0.1837	0.0338	0.0445	0.0480	0.7227	0.0479
	8000 obs.	CL/FE	KYRI	TSE1	TSE2	TSE3	MLE1	MLE2
γ_1	meanbias	0.0109	0.0109	-0.0327	-0.0327	-0.0018	-0.1833	-0.1543
	medianbias	0.0066	0.0066	-0.0331	-0.0331	-0.0028	-0.1830	-0.1555
	std err	0.0672	0.0672	0.0320	0.0320	0.0338	0.0324	0.0304
	rmse	0.0680	0.0680	0.0457	0.0457	0.0338	0.1861	0.1573
	mad	0.0441	0.0441	0.0351	0.0351	0.0245	0.1830	0.1543
γ_2	meanbias	0.0113	0.0113	0.2804	0.2804	-0.0024	0.1376	-0.1477
	medianbias	0.0113	0.0113	0.2789	0.2789	-0.0031	0.1363	-0.1479
	std err	0.0676	0.0676	0.0355	0.0355	0.0359	0.0335	0.0352
	rmse	0.0685	0.0685	0.2826	0.2826	0.0360	0.1416	0.1518
	mad	0.0451	0.0451	0.2789	0.2789	0.0249	0.1363	0.1477
β	meanbias	-0.0881	0.0065	0.0149	-0.0204	-0.0014	0.7244	0.0015
	medianbias	-0.0897	0.0096	0.0157	-0.0200	-0.0018	0.7234	0.0016
	std err	0.0295	0.1380	0.0261	0.0345	0.0368	0.0406	0.0314
	rmse	0.0929	0.1380	0.0301	0.0401	0.0368	0.7255	0.0314
	mad	0.0897	0.0940	0.0200	0.0278	0.0251	0.7234	0.0254

Table A4. Results from Model C estimated with MLE2, for various numbers of support points in the selection equation.

	2000 obs.	$2x2$ support points	$3x2$ support points	$4x2$ support points
γ_1	meanbias	-0.1365	-0.0351	-0.0176
	medianbias	-0.1352	-0.0391	-0.0249
	std err	0.0623	0.0848	0.0904
	rmse	0.1501	0.0917	0.0920
	mad	0.1371	0.0744	0.0505
γ_2	meanbias	-0.1281	-0.0329	-0.0193
	medianbias	-0.1298	-0.0334	-0.0241
	std err	0.0725	0.0956	0.0994
	rmse	0.1471	0.1010	0.1013
	mad	0.1313	0.0816	0.0808
β	meanbias	-0.0015	0.0043	0.0050
	medianbias	0.0015	0.0051	0.0061
	std err	0.0645	0.0651	0.0645
	rmse	0.0645	0.0652	0.0647
	mad	0.0510	0.0518	0.0505

Working Paper

2000-19 Torben M. Andersen: Nominal Rigidities and the Optimal Rate of Inflation.

2001-1 Jakob Roland Munch and Michael Svarer: Mortality and Socio-economic Differences in a Competing Risks Model.

2001-2 Effrosyni Diamantoudi: Equilibrium Binding Agreements under Diverse Behavioral Assumptions.

2001-3 Bo Sandemann Rasmussen: Partial vs. Global Coordination of Capital Income Tax Policies.

2000-4 Bent Jesper Christensen and Morten Ø. Nielsen: Semiparametric Analysis of Stationary Fractional Cointegration and the Implied-Realized Volatility Relation in High-Frequency.

2001-5: Bo Sandemann Rasmussen: Efficiency Wages and the Long-Run Incidence of Progressive Taxation.

2001-6: Boriss Siliverstovs: Multicointegration in US consumption data.

2001-7: Jakob Roland Munch and Michael Svarer: Rent Control and Tenancy Duration.

2001-8: Morten Ø. Nielsen: Efficient Likelihood Inference in Non-stationary Univariate Models.

2001-9: Effrosyni Diamantoudi: Stable Cartels Revisited.

2001-16: Bjarne Brendstrup, Svend Hylleberg, Morten Nielsen, Lars Skipper and Lars Stentoft: Seasonality in Economic Models.

2001-17: Martin Paldam: The Economic Freedom of Asian Tigers - an essay on controversy.

2001-18: Celso Brunetti and Peter Lildholt: Range-based covariance estimation with a view to foreign exchange rates.

2002-1: Peter Jensen, Michael Rosholm and Mette Verner: A Comparison of Different Estimators for Panel Data Sample Selection Models.