Discussion Papers No. 330, September 2002
Statistics Norway, Research Department

Arvid Raknerud

Identification, Estimation and Testing in Panel Data Models with Attrition: The Role of the Missing at Random Assumption

Abstract:
This paper discusses identification, estimation and testing in panel data models with attrition. We focus on a situation which often occurs in the analysis of firms: Attrition (exit) is endogenous and depends on the outcomes of an observed stochastic process and the interest-parameters characterizing this process. Thus attrition is non-ignorable even if selection is based only on observed variables - that is, even if the missing items are missing at random (MAR). The likelihood function obtained by ignoring the attrition mechanism is a pseudo likelihood function. Assuming that the MAR condition holds, this paper establishes conditions for identification and consistent estimation based on the pseudo likelihood function. It is also shown that the MAR hypothesis has testable implications in many situations that are encountered in practice. Simulations suggest that in the case of the autoregressive model with random effects, the efficiency of the pseudo likelihood estimator (based on normality) is not much affected even by strong departures from normality. In a variety of simulation models, the pseudo likelihood estimator clearly outperforms the moment estimators - even when the latter are consistent.

Keywords: Missing at random, non-ignorable attrition, unbalanced panel data, identification, pseudo likelihood, martingale.

JEL classification: C13, C23, C33.

Acknowledgement: An earlier version of this paper was presented at the 2001 Econometric Society European Meeting (ESEM 2001) in Lausanne. Comments from Torbjørn Hægeland, Tor Jakob Klette, John Dagsvik, Tom Kornstad, and seminar participants are highly appreciated.

Address: Arvid Raknerud, Statistics Norway, Research Department. E-mail: arvid.raknerud @ssb.no

Discussion Papers comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

In panel data, the same unit is observed repeatedly over time. As is well documented, this enables us to estimate models with complex behavioral relationships; e.g. about consumer and firm behavior. On the other hand, missing data problems are severe in most panel surveys. A particular problem, which is the theme of this paper, is that a unit which initially is in the sample may drop out before the survey period is over. This phenomenon is called *attrition.* Examples include firms which close down due to bankruptcy, households who stop responding to consumer surveys, and patients who die during the test of an AIDS drug. A traditional "fix" is to retain a balanced sub-sample of the original sample. In most cases this leads to severely biased inference – unless attrition is independent of the endogenous variables; an unreasonable assumption in most econometric applications.

A typical situation is described in Hirano et al. (1998): In the Dutch Transportation Panel, households are asked to fill out a travel diary over one week each year in subsequent years. The burden of responding to the questionnaire depends on the total number of travels. Therefore, non-response is highly correlated with an endogenous variable. In their approach, Hirano et al. suggest replacing non-responding units with so-called refreshment samples to make inferences about the non-response mechanism. This is done in order to impute missing data. Unfortunately, their method has limited applicability in complex models or when attrition is not due to non-response, such as is the case with e.g. firm exit.

The study of attrition has, of course, a long history in econometrics. A classical model is due to Hausman and Wise (1979), who allow the probability of attrition to depend on unobserved contemporary variables – but not on lagged endogenous variables. Another well-known approach where self-selection is based on unobserved endogenous variables is described in Heckman (1979).

This paper focuses on methods and models of attrition based on the missing at random (MAR) hypothesis: Given the full history of observed variables on an observation unit, the probability of exit in the current period is independent of unobserved (contemporary and future) variables (see Little and Rubin, 1987). Moffitt, Fitzgerald and Gottschalk (1999) study the role of the MAR assumption in econometrics, and use the term *selection on* *observables* to characterize this situation. Another recent contribution to the econometric literature about attrition and the MAR hypothesis is Abowd, Kramarz and Crepon (2001).

The MAR assumption may be particularly relevant in the analysis of firm behavior. For example, if a firm's decision of whether to exit or not is made at the end of the year, and is based on expected profits given this years observed profit and historically observed profits, then an exited firm will be MAR. However, if the firm makes predictions about future profitability based on information about random variables which are unknown to the econometrician, the MAR assumption may fail.

If (i) the MAR assumption holds and (ii) the exit mechanism is independent of the interest parameters, attrition is said to be *ignorable* with respect to likelihood based inference (see Little and Rubin, 1987). In many situations condition (i) may be reasonable, while (ii) is too restrictive. An example is the so-called passive learning model of Jovanovic (1982), which fits naturally into the MAR framework, but where firm exit depends on interest-parameters. I shall return to this example at the end of the next section.

If MAR holds, but attrition is non-ignorable, we obtain a *partial likelihood* if we ignore the attrition mechanism when setting up the likelihood function. The term partial likelihood is often associated with a conditional likelihood or a profile likelihood (see Cox, 1975). However, that is not the case in the present situation. Therefore, I shall call this likelihood a *pseudo likelihood* (see Gourieroux and Montfort, 1984). The term pseudo likelihood is also motivated by the desire to investigate the properties of the implied estimators outside model conditions, e.g. when distributional assumptions fail.

While there may be a loss of efficiency associated with pseudo likelihood based inference, we shall see that there is no asymptotic bias if the MAR condition is fulfilled. Moreover, the general inference results in White (1982) and Gourieroux and Montfort (1984) are valid, thus providing tools for assessing the accuracy of estimates (e.g. constructing confidence intervals) and testing hypotheses.

The MAR assumption is substantially more general than what is needed for consistency of traditional generalized method of moments (GMM) estimators. However, interesting attempts to rescue GMM under the MAR assumption have been put forward. Abowd et al. (2001) propose to weigh orthogonality conditions implied by the econometric model by the inverse sampling probability. Their weighted GMM (WGMM) estimator is consistent under the MAR assumption. But there is a drawback to their method: Sampling probabilities are nuisance parameters which have to be modelled and estimated from the data. Moreover, simulations presented in this paper suggest that their method may not perform well in practice (see Section 5).

The aimed contributions of the present paper are threefold: (i) to provide rigorous regularity conditions for identification of interest parameters when the MAR assumption holds but attrition is non-ignorable; (ii) to show that the MAR assumption has testable implications in parametric models – a fact which has been widely overlooked (see e.g. Horowitz and Manski, 1998); and (iii), in the particular case of the linear-normal autoregressive model, to present Monte Carlo results about the performance of the pseudo likelihood (PL) and moment estimators under different attrition rules and error distributions. The simulations show that the MAR assumption is critical for the good performance of the PL estimator. On the other hand, normality is not: The PL estimator (derived from normality assumptions) strikingly outperforms the moment estimators also in simulation models with non-normal errors.

2 Basic assumptions

We assume that a variable x_t^i is observed on each unit $i = 1,..,N$ from some initial observation date ι^i, and then each period until a stopping date τ^i. Both ι^i and τ^i are random variables. The stopping date τ^i may be the last year of the survey period (T), or the date of exit. It is assumed that there are no wholes in the data between ι^i and τ^i, and that ι^i is an ancillary random variable. That is, ι^i is independent of θ in distribution (see Barndorff-Nielsen and Cox (1994)).

The econometrician is not interested in the process that determines death or birth per se, but in the law that governs the state process x_t^i. This law is assumed known up to some parameter vector θ. However, since the x_t^i-process is subject to attrition, it is well known that inference about θ may be severely biased if we ignore self-selection.

Formally, for unit $i = 1,..,N$, we observe the sequence $(\iota^i, x_{\iota^i}^i,..., x_{\tau^i}^i, \tau^i)$, where $\iota^i \geq 1$ is the (exogenous) birth or entry date and $\tau^i \leq T$ is the (endogenous) exit date. To simplify notation, we will drop the i-superscript from now on unless needed to avoid ambiguity.

Next, define the exit-indicator variable

$$z_t = \begin{cases} 1 & t \geq \tau \\ 0 & \text{else} \end{cases}$$

and the sequence

$$Z_t = \{\iota, z_0, z_1, ..., z_t\}.$$

Thus, Z_t contains all information about the life-span of the observation unit up until time t.

The underlying probability model concerns the joint distribution of $(\iota, x_\iota, ..., x_\tau, \tau)$. Let $\mathcal{F}$ denote the corresponding σ-field, i.e. $\mathcal{F} = \sigma(\iota, x_\iota, ..., x_\tau, \tau)$. (See Billingsley (1986) for a construction of this σ-field). We now state some regularity conditions regarding this probability space:

ASSUMPTION A.1: (a) The collections of random variables $\iota, x_\iota, ..., x_\tau, \tau$ are independent across observation units $i = 1, ..., N$, with a joint probability measure P on a measurable space $(\Omega, \mathcal{F})$. *(b)* The distribution of $x_s, .., x_t$ conditional on $\iota = s$ for $1 \leq s \leq t \leq T$ has a density function $f(x_s, , .., x_t|\iota = s; \theta)$ with respect to a given measure $\nu(dx_s, .., dx_t)$ for every θ in Θ, a compact subset of a p-dimensional Euclidean space.

The interest is in the parameters θ characterizing the probability distribution of the time series x_t, i.e. $f(\cdot; \theta)$, not in the stopping times ι or τ. On the other hand, the joint probability measure P may also depend on nuisance parameters, say φ, characterizing the entry and exit process.

We cannot disregard the stopping times ι and τ when making inferences about θ, as ι and τ determine the "window" through which we observe x_t. But we shall assume that ι is exogenous, so that we can always condition on entry. On the other hand, we cannot condition on τ because we would then need an explicit model of attrition. Note that the marginal distribution of τ depends on θ, i.e. τ is not an ancillary statistic.

The fundamental MAR assumption is stated next:

ASSUMPTION A.2: (MAR) The distribution of x_{t+1} conditional on Z_t and $x_\iota, ..., x_t$ has a density with respect to $\nu(dx_{t+1})$ which satisfies

$$f(x_{t+1}|Z_t, x_\iota, .., x_t) = f(x_{t+1}|\iota, x_\iota, .., x_t; \theta) \quad \text{for } t = \iota, ..., T-1. \tag{1}$$

That is, x_{t+1} is independent of the life-span information Z_t given $(\iota, x_\iota, .., x_t)$. Note that (1) must hold also for $t \geq \tau$: There is no conditioning on x_{t+1} actually being observed in the definition of MAR in (1).

It is easily verified that Assumption A.2 is equivalent to the following, more usual, formulation of the MAR condition:

$$x_{\tau+1}, .., x_T \perp \tau \mid \iota, x_\iota, .., x_\tau,$$

which says that the unobserved variables are independent of attrition (the exit date) given the observed variables. It should be noted here that the transition equation $f(x_{t+1}|\iota, x_\iota, .., x_t; \theta)$ may depend explicitly both on the "cohort" ι and on calendar time t. E.g. the model may contain both time- and cohort-specific dummies.

Panel data models are typically formulated in terms of latent variables which are specific to each observation unit, say v_t^i. It is then useful to formulate a slightly different version of MAR (where we again drop the i-superscript):

$$f(x_{t+1}, v_{t+1}|Z_t, x_\iota, .., x_t) = f(x_{t+1}, v_{t+1}|\iota, x_\iota, .., x_t; \theta). \tag{2}$$

That is; x_{t+1} and v_{t+1} are *jointly* independent of Z_t, given $(\iota, x_\iota, .., x_t)$. In particular, this implies that, having made predictions about the latent variable v_{t+1} based on observation of $(\iota, x_\iota, .., x_t)$, the *additional* information that the x_t-process is subject to attrition is irrelevant for predicting of v_{t+1}. Clearly, (2) implies (1).

ASSUMPTION A.3: *(a)* For every $1 \leq s \leq t \leq T$, $f(x_s, .., x_t|\iota = s; \theta)$ is continuous in θ and positive. *(b)* (Identification) If $\theta \neq \theta_0$ there exist integers $s \leq k$ and a set of sequences $(x_s, ..., x_k)$ with positive υ-measure, so that for every $(x_s, ..., x_k)$ in this set $f(x_s, .., x_k|\iota = s; \theta) \neq f(x_s, .., x_k|\iota = s; \theta_0)$, $P(\tau = k|\iota = s, x_s, .., x_k) > 0$ and $P(\iota = s) > 0$. (c) $E\{\ln f(x_\iota, .., x_\tau|\iota; \theta_0)\}$ exists. *(d)*$|\ln f(x_\iota, .., x_\tau|\iota; \theta_0)| \leq m(\iota, x_\iota, ..., x_\tau)$ for all θ in Θ for a function $m(\cdot)$ integrable with respect to P.

It follows from A.3.(a) that $F(A) \equiv \int_A f \, d\upsilon = 0$ implies $\upsilon(A) = 0$, regardless of θ and ι. Hence the support of F is determined by υ. Furthermore, A.3.(b) is the ordinary identification condition supplemented by an observability condition: Identification must not depend upon outcomes that cannot be observed. In the next section we shall investigate identification of θ based on a pseudo likelihood function which ignores the attrition mechanism. I show in Section 3 that θ is identified under Assumption A.1-3.

Motivating example: An interesting illustration of the setup described above is the so-called passive learning model of Jovanovic (1982). In this model firm i is equipped at birth with some productivity parameter v^i. The productivity parameter is unobserved by the firm, but the firm knows the stochastic model which has generated v^i:

$$v^i = \beta + \eta^i, \quad \eta^i \sim \mathcal{N}(0, \delta^2),$$

where β and δ^2 are parameters known to the firm. As a by-product of operation, the firm observes a variable x_t^i, and each year it updates the conditional distribution of v^i based on the observation equation:

$$x_t^i = v^i + \varepsilon_t^i, \quad \varepsilon_t^i \sim \mathcal{N}(0, \sigma^2),$$

for known σ^2. Here $\theta = (\beta, \delta^2, \sigma^2)$, $v_t^i = v^i$, and $\iota^i = 1$.

The firm chooses output so as to maximize expected discounted profits given its current update of the distribution of v^i. The firm decides to close down if the value of remaining operative is lower than the "scrap value", or alternative value, of the firm. The structure of the problem is such that the firm exits at the end of year t if the posterior mean $E\{v^i|x_1^i, ..., x_t^i\}$ falls short of a time-varying threshold (depending on deterministic prices). The Jovanovic model is thus a model with non-ignorable selection. On the other hand, if x_t^i is observed by the econometrician, the MAR condition is satisfied.

3 Identification

This section is concerned with identification. I present detailed conditions which are sufficient for identification and consistent estimation based on the pseudo likelihood function obtained by ignoring the attrition mechanism. The main result is stated in Proposition 2. This, and other, results rely on a martingale property of the pseudo likelihood ratio under the MAR assumption established in Proposition 1.

In the general case, the likelihood of the complete set of observations $\{x_t, z_t, t = \iota, ..., \tau\}$ can be written

$$L(\theta, \varphi) = P(z_\iota|\iota, x_\iota; (\theta, \varphi))f(x_\iota|\iota; \theta) \prod_{t=\iota+1}^{\tau} P(z_t|Z_{t-1}, x_\iota, .., x_t; (\theta, \varphi))f(x_t|Z_{t-1}, x_\iota, .., x_{t-1}; (\theta, \varphi))$$

where θ are the interest parameters and φ are the nuisance parameters. The fundamental question is whether we can make inferences about θ based on N independent realizations (for $i = 1, .., N$) of the function:

$$\widetilde{L}(\theta) = f(x_\iota|\iota;\theta) \prod_{t=\iota+1}^{\tau} f(x_t|\iota, x_\iota, .., x_{t-1};\theta), \tag{4}$$

which ignores attrition and only depends on θ. If (i) the MAR assumption (1) holds and (ii) $P(z_t|Z_{t-1}, x_\iota, ..., x_t; (\theta, \varphi))$ is independent of θ in distribution, $L(\theta, \varphi)$ and $\widetilde{L}(\theta)$ are identical (except for an uninteresting proportionality constant which only depends on φ). If any of these two conditions fail, $\widetilde{L}(\theta)$ differs from $L(\theta, \varphi)$. Note that, regardless of the attrition mechanism, $\widetilde{L}(\theta)$ satisfies the conditions of a likelihood because it is a probability density function. I will therefore refer to (4) as the pseudo likelihood function (PL). We shall now investigate the properties of the PL estimator when condition (i) holds but not (ii).

Let $E_t\{\cdot\}$ denote the expectation conditional on the σ-field $\mathcal{F}_t$, where

$$\mathcal{F}_t = \begin{cases} \sigma(Z_t, x_\iota, ..., x_{t\wedge\tau}) & t = \iota, .., T \\ \sigma(Z_t) & t = 0, .., \iota - 1, \end{cases}$$

and let $E\{\cdot\}$ denote the unconditional expectation (both under the true distribution P). Define:

$$\widetilde{L}_t(\theta) = f(x_\iota|\iota;\theta) \prod_{t=\iota+1}^{t\wedge\tau} f(x_t|\iota, x_\iota, .., x_{t-1};\theta).$$

Furthermore, let

$$Q_t = \frac{\widetilde{L}_t(\theta)}{\widetilde{L}_t(\theta_0)} \text{ for } t \geq \iota \text{ and } Q_t = 1 \text{ for } t < \iota.$$

We shall first see that Q_t is a martingale under the true model.

Proposition 1 *(The Martingale property of Q_t). Given Assumption A.1-3*

$$E_t\{Q_{t+1}\} = Q_t \text{ a.s..} \tag{5}$$

Proof. If $t \geq \tau$ or $t < \iota - 1$ (5) obviously holds. If $t = \iota - 1$, then $Q_t = 1$ and

$$E_t\{Q_{t+1}\} = \int \frac{f(x_\iota|\iota;\theta)}{f(x_\iota|\iota;\theta_0)} \times f(x_\iota|\iota;\theta_0)\nu(dx_\iota) = Q_t.$$

Finally, if $\iota \leq t < \tau$:

$$\begin{aligned} E_t\{Q_{t+1}\} &= E_t\left\{Q_t \frac{f(x_{t+1}|\iota, x_\iota, .., x_t;\theta)}{f(x_{t+1}|\iota, x_\iota, .., x_t;\theta_0)}\right\} \\ &= Q_t \int \frac{f(x_{t+1}|\iota, x_\iota, .., x_t;\theta)}{f(x_{t+1}|\iota, x_\iota, .., x_t;\theta_0)} \times f(x_{t+1}|Z_t, x_\iota, .., x_t)\nu(dx_{t+1}) \\ &= Q_t \int \frac{f(x_{t+1}|\iota, x_\iota, .., x_t;\theta)}{f(x_{t+1}|\iota, x_\iota, .., x_t;\theta_0)} \times f(x_{t+1}|\iota, x_\iota, .., x_t;\theta_0)\nu(dx_{t+1}) \\ &= Q_t, \end{aligned}$$

where Assumption A.2 was used in the third equation. ■

The next proposition shows that θ_0 is identified from the pseudo-likelihood function given that it is identified in the family $f(\cdot;\theta)$.

Proposition 2 *(Identification) Given A.1-3, if $\theta \neq \theta_0$, $E\{\ln \widetilde{L}(\theta_0)\} > E\{\ln \widetilde{L}(\theta)\}$.*

Proof. It follows from the martingale property of Q_t that $E_0\{Q_T\} = E_0\{Q_1\} = 1$ and thus, averaging over ι,

$$E\{Q_T\} = 1. \tag{6}$$

Furthermore,

$$\begin{aligned}
& E\{Q_T\} \\
= & \sum_{s=1}^{T}\sum_{k=s}^{T}\int_{[\tau=k\cap\iota=s]}\frac{f(x_s,..,x_k|\iota=s;\theta)}{f(x_s,..,x_k|\iota=s;\theta_0)}dP \\
= & \sum_{s=1}^{T}\sum_{k=s}^{T}\int\frac{f(x_s,..,x_k|\iota=s;\theta)}{f(x_s,..,x_k|\iota=s;\theta_0)}P(\tau=k|\iota=s,x_s,..,x_k)f(x_s,..,x_k|\iota=s;\theta_0)\times \\
& P(\iota=s)\nu(dx_s,..,dx_k).
\end{aligned}$$

Under A.3.(a)-(b), if $\theta \neq \theta_0$, the integrand $Q_T = \frac{f(x_\iota,..,x_\tau|\iota;\theta)}{f(x_\iota,..,x_\tau|\iota;\theta_0)}$ will differ from 1 on a set with positive probability. Hence, taking the logarithm on both sides of (6), and using Jensen's inequality yields:

$$E\{\ln \widetilde{L}_T(\theta) - \ln \widetilde{L}_T(\theta_0)\} < 0.$$

Using A.3.(c), and the identity $\widetilde{L}(\theta) = \widetilde{L}_T(\theta)$, we obtain

$$E\{\ln \widetilde{L}(\theta_0)\} > E\{\ln \widetilde{L}(\theta)\}$$

if $\theta \neq \theta_0$. ■

We have just established identification of θ as the unique maximizer of the pseudo likelihood $\widetilde{L}(\theta)$, regardless of whether the exit mechanism is ignorable or not.

4 Estimation and testing

4.1 The pseudo likelihood estimator

Let $\widetilde{L}^i(\theta)$ be the realization of $\widetilde{L}(\theta)$ on observation unit i, i.e. based on $(\iota^i, x^i_{\iota^i},..., x^i_{\tau^i}, \tau^i)$. Define

$$l_N(\theta) = N^{-1}\sum_{i=1}^{N}\ln \widetilde{L}^i(\theta).$$

Under standard regularity conditions, such as A.4-A.6 in White (1982), it can now be shown that (a) $l_N(\theta)$ converges almost surely, uniformly on the parameter space Θ, to $l_\infty(\theta) = E\{\ln \widetilde{L}^i(\theta)\}$, (b) the pseudo likelihood estimator $\widehat{\theta}_N$ defined as

$$\widehat{\theta}_N = \arg\max_{\theta\in\Theta} l_N(\theta)$$

will be a consistent estimator of θ_0, (c) $\sqrt{N}(\widehat{\theta}_N - \theta_0)$ converges in distribution to $\mathcal{N}(0, J^{-1}IJ^{-1})$ where

$$\begin{aligned} J &= E\left\{-\frac{\partial^2 \ln \widetilde{L}^i(\theta_0)}{\partial\theta\partial\theta'}\right\} \\ I &= E\left\{\frac{\partial \ln \widetilde{L}^i(\theta_0)}{\partial\theta}\frac{\partial \ln \widetilde{L}^i(\theta_0)'}{\partial\theta}\right\}. \end{aligned}$$

The proofs of these results can be taken directly from the proofs of Theorem 2.2 and 3.2 in White (1982). My assumptions A.1-3 together with the identification result in Proposition 2, ensure that the regularity conditions A.1-3 in White is fulfilled.

In exact likelihood inference, it is well known that $I = J$. This is the so-called information equality. Typically $I \neq J$ in pseudo likelihood based inference (even when the estimator is consistent), but Cox (1975) shows that the information equality also holds for partial likelihood, and therefore in our case. For completeness, this result is established in Proposition 3.

Proposition 3 *(The information equality). Given Assumption A.1-3 in Section 2, and Assumption A.4-6 in White (1982):*

$$E\left\{-\frac{\partial^2 \ln \widetilde{L}^i(\theta_0)}{\partial\theta\partial\theta'}\right\} = E\left\{\frac{\partial \ln \widetilde{L}^i(\theta_0)}{\partial\theta}\frac{\partial \ln \widetilde{L}^i(\theta_0)'}{\partial\theta}\right\}$$

Proof. Differentiating equation (6) with respect to θ, yields

$$\frac{\partial^2 E\{Q_T^i\}}{\partial\theta\partial\theta'} = 0$$

By assumption, we can interchange the order of integration and differentiation. This yields

$$\sum_{s=1}^{T}\sum_{k=s}^{T}\int_{[\tau=k\cap\iota=s]} \frac{\partial^2 f(x_s,..,x_k|\iota=s;\theta)}{\partial\theta\partial\theta'} \times \frac{1}{f(x_s,..,x_k|\iota=s;\theta_0)} dP = 0 \qquad (7)$$

For $\theta = \theta_0$, (7) is equivalent to

$$\int \left(\frac{\partial^2 \ln \widetilde{L}^i(\theta_0)}{\partial\theta\partial\theta'} + \frac{\partial \ln \widetilde{L}^i(\theta_0)}{\partial\theta}\frac{\partial \ln \widetilde{L}^i(\theta_0)'}{\partial\theta}\right) dP = 0,$$

and the conclusion follows. ■

4.2 Moment estimators

The results established in this section show that the MAR assumption is sufficient for the validity of pseudo likelihood based methods. It is interesting to compare with GMM-methods, which have a dominant position in the econometric literature. We shall do so in relation to a concrete example.

Our starting point will be the autoregressive AR(1) model with random effects:

$$\begin{aligned} x_t^i &= \phi x_{t-1}^i + (1-\phi)v^i + \varepsilon_t^i \text{ for } \quad t = 2, ..., T \\ x_1^i &= v^i + \varepsilon_1^i, \end{aligned} \tag{8}$$

where v^i is a random effect, with $E\{v^i\} = 0$, $E\{v^i \varepsilon_t^i\} = 0$, and $Var(v^i) = \sigma_v^2$. Furthermore, ε_t^i is white noise: $E\{\varepsilon_t^i\} = 0$, $E\{\varepsilon_t^i \varepsilon_s^i\} = 0$ for $s \neq t$, $Var(\varepsilon_1^i) = \sigma_1^2$, and $Var(\varepsilon_t^i) = \sigma_\varepsilon^2$ for $t = 2, ..., T$. Finally, it is assumed that $E\{x_1^i \varepsilon_t^i\} = 0$ for $t > 1$. This model is weakly stationary if $|\phi| < 1$ and $\sigma_1^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}$. On the other hand, x_t^i is a pure random walk when $\phi = 1$. The interest parameter is ϕ – the autoregressive coefficient.

To estimate ϕ it is common to apply the generalized method of moments (GMM) using instrumental variables. The traditional set of instruments is obtained by differencing (8) to eliminate v^i (assuming, for simplicity of notation, that $\iota^i = 1$ for all i):

$$x_t^i - x_{t-1}^i - \phi(x_{t-1}^i - x_{t-2}^i) = \varepsilon_t^i - \varepsilon_{t-1}^i \text{ for } \ t = 3, ..., T.$$

We then obtain the following orthogonality conditions:

$$E\{(\varepsilon_t^i - \varepsilon_{t-1}^i)x_{t-k}^i\} = 0 \text{ for } k = 2, ..., t-1 \text{ and } t = 3, ..., T. \tag{9}$$

That is, the x_{t-k}^i are instruments for the differenced equations (see Arellano and Bover, 1995; Ahn and Schmidt, 1995). Another set of instruments has been studied by Blundell and Bond (1998) and Hahn (1999):

$$E\{u_t^i \Delta x_{t-1}^i\} = 0 \text{ for } t = 3, ..., T, \tag{10}$$

where $u_t^i = (1-\phi)v^i + \varepsilon_t^i$. Thus, the Δx_{t-1}^i are instruments for equations in levels.

Since we do not observe x_t^i for $t > \tau^i$, extending (9) and (10) to the case with attrition would require:

$$\begin{aligned} E\left\{(\varepsilon_t^i - \varepsilon_{t-1}^i)x_{t-k}^i | t \leq \tau^i\right\} &= 0 \\ E\left\{u_t^i \Delta x_{t-1}^i | t \leq \tau^i\right\} &= 0, \end{aligned} \tag{11}$$

or equivalently:

$$\begin{aligned} E\left\{(\varepsilon_t^i - \varepsilon_{t-1}^i)x_{t-k}^i I(t \leq \tau^i)\right\} &= 0 \\ E\left\{u_t^i \Delta x_{t-k}^i I(t \leq \tau^i)\right\} &= 0, \end{aligned} \tag{12}$$

where $I(t \leq \tau^i)$ is the indicator function which is 1 if $t \leq \tau^i$ and 0 otherwise. Unfortunately, equations (12) are not implied by the MAR assumption because the event $t \leq \tau^i$ could depend on all lagged realizations $x_{t-1}^i, \ldots, x_1^i$ and hence on $\varepsilon_{t-1}^i, \ldots, \varepsilon_1^i$ and v^i. For example, the conditional expectation of ε_{t-1}^i given that $t \leq \tau^i$ will in general differ from its unconditional expectation $E\{\varepsilon_{t-1}^i\} = 0$.

It is possible to rescue the GMM estimator for this model by applying the weighting procedure proposed in Abowd et al. (2001): The orthogonality conditions in (9)-(10) have the form $E\left(g_t(x_1^i, \ldots, x_t^i)\right) = 0$. Abowd et al. show that the weighted orthogonality conditions:

$$E\left(\frac{g_t(x_1^i, \ldots, x_t^i)I(t \leq \tau^i)}{\pi_t^i}\right) = 0, \text{ with } \pi_t^i = P(t \leq \tau^i | x_1^i, \ldots, x_{t-1}^i),$$

hold under the MAR assumption. We shall return to questions regarding implementation and performance of this weighted GMM estimator in Section 5.

4.3 Testing the MAR assumption

The MAR assumption has been criticized because it does not imply any testable restrictions. For example, Horowitz and Manski (1998) writes:

"*Survey non-response is problematic for identification of population parameters. Whether nonresponse takes the form of particular missing items or entire missing interviews, the only way to identify population parameters is to make assumptions about ...[what]... determine the ... distribution of missing data. A basic problem ... is that such assumptions are not testable.*"

Nevertheless, as is frequently overlooked, the MAR assumption does not only imply restrictions on the distribution of the missing data, but also on the *observed* data. Moreover, the formulation of the MAR assumption in (2) suggests a very simple way to test its validity.

Let us consider the common situation where $f(x_{t+1}^i | v_{t+1}^i, \iota^i, x_{\iota^i}^i, .., x_t^i; \theta)$ is formulated in terms of an explicit transition equation. Equation (2) then implies that no information about the life-span of observation unit i contained in Z_t^i should help us to predict x_{t+1}^i given v_{t+1}^i, ι^i, $x_{\iota^i}^i, ..., x_t^i$. In the particular case of the linear-normal autoregressive model with random effects (8), we have $v_t^i = v^i$ and $\theta = (\phi, \sigma_1^2, \sigma_v^2, \sigma_\varepsilon^2)$. Then (2) implies that the parameter vector θ entering the transition equation is "invariant" with respect to the survival time ("age") A_t^i of observation unit i at time t, where $A_t^i = (t \wedge \tau^i) - \iota^i + 1$. Note that A_t^i is a random variable and a function of Z_t^i.

For any component θ_j of θ, we can write:

$$\theta_j = \theta_j^0 + \sum_{s=2}^{t} \theta_{js} I(A_t^i = s),$$

where $I(A_t^i = s)$ is the indicator function which is 1 if the survival time of observation unit i at time t is s (and 0 otherwise), while θ_{js}, for $s \geq 2$, are auxiliary parameters. If the MAR assumption holds, the pseudo true value of $\theta_{js} = 0$ for all s. That is, $\theta_j = \theta_j^0$ independently of the "age" A_t^i of the observation unit at the time of the transition. Furthermore, we can test the hypothesis that $\theta_{js} = 0$ by estimating these auxiliary parameters and then apply the results about pseudo likelihood based inference established above.

This procedure requires that θ_{js} is identified, which is a non-trivial requirement. For example, assume that the model (8) contains time-dummies:

$$\begin{aligned} x_t^i &= \phi x_{t-1}^i + \mu_t + (1-\phi)v^i + \varepsilon_t^i \quad \text{for} \quad t = 2, ..., T \\ x_1^i &= v^i + \varepsilon_1^i. \end{aligned}$$

If we assume, as in Abowd et al. (2001), that attrition is related to the *level* of the endogenous variables, it is natural to test whether the survival time A_t^i affects the estimates of the intercepts μ_t. Hence, let $\theta_j = \mu_t$ (for some j and t). In this case, *if all units have entered the sample at the same date* $\iota^i = 1$, identification of θ_{js} will fail because all observation units will have the same survival time: For all i, $I(A_t^i = s) = 1$ when $s = t$ and 0 when $s \neq t$. To discriminate between time-effects and self-selection effects, we therefore need to have sufficient "cohort" variation in the sample. In general, therefore, the possibility of testing the MAR assumption is facilitated by an approperiate sampling design.

5 Monte Carlo results

In this section we analyze pseudo likelihood and moment estimators of the autoregressive parameter ϕ in (8) using Monte Carlo experiments. We shall consider different attrition mechanisms and distributions of error terms. Since moment estimators have been mostly developed for balanced panel data sets, I shall first present these estimators in some detail. Then, an explicit form for the pseudo likelihood function which is convenient for estimation purposes will be derived. Throughout we assume that $\iota^i = 1$.

The GMM and WGMM estimators: We first define the diagonal matrix:

$$D^i = \begin{bmatrix} I(3 \leq \tau^i) & 0 & \cdots \\ 0 & \ddots & \cdots \\ \vdots & \cdots & I(T \leq \tau^i) \end{bmatrix}$$

and the $(T-2) \times \frac{1}{2}(T-2)(T-1)$ matrix Y_i:

$$Y_i = \begin{bmatrix} y_1^i & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & y_1^i & y_2^i & 0 & \cdots & \cdots & 0 \\ . & . & . & \cdots & & & \\ 0 & 0 & 0 & \cdots & y_1^i & \cdots & y_{T-2}^i \end{bmatrix}.$$

The GMM estimator of ϕ solves the linear equation

$$\left(\sum_{i=1}^{N} \widetilde{y}_{-1}^i{}' D_i^+ Y_i^+ \right) W \left(\sum_{i=1}^{N} Y_i^+{}' D_i^+ \left(\widetilde{y}^i - \widetilde{y}_{-1}^i \phi \right) \right) = 0 \tag{13}$$

where

$$Y_i^+ = \begin{bmatrix} Y_i & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_2^i & 0 & \cdots & 0 \\ . & . & . & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \Delta y_{T-1}^i \end{bmatrix}$$

$$D_i^+ = \begin{bmatrix} D_i & 0 \\ 0 & D_i \end{bmatrix}$$

$$\begin{aligned} \widetilde{y}^i &= [\Delta y_3^i, \ldots\ldots, \Delta y_T^i, y_3^i, \ldots\ldots, y_T^i]' \\ \widetilde{y}_{-1}^i &= [\Delta y_2^i, \ldots\ldots, \Delta y_{T-1}^i, y_2^i, \ldots\ldots, y_{T-1}^i]' \end{aligned}$$

and W is a weight matrix.

For $t = 3, .., T$, each element in row number $t-2$ and $T+t-4$ in Y_i^+, $\widetilde{y}^i$ and $\widetilde{y}_{-1}^i$ are multiplied by the indicator $I(t \leq \tau^i)$, and hence replaced by 0 if the corresponding variable is unobserved. The estimator (13) is implemented in the popular software package DPD (see Arellano and Bond, 1998).

Under the assumption that all missing data are MCAR, (12) holds and

$$\left(\frac{1}{N}\sum_{i=1}^{N} Y_i^{+\prime} D_i^{+}\left(\widetilde{y}^i - \widetilde{y}^i_{-1}\phi\right)\right) \xrightarrow{P} 0.$$

Thus the GMM estimator will be consistent, regardless of the choice of W.

On the other hand, the weighted GMM (WGMM) method solves the equation:

$$\left(\sum_{i=1}^{N} \widetilde{y}^{i\prime}_{-1}\Pi_i^{-1} D_i^{+} Y_i^{+}\right) W \left(\sum_{i=1}^{N} Y_i^{+\prime} D_i^{+}\Pi_i^{-1}\left(\widetilde{y}^i - \widetilde{y}^i_{-1}\phi\right)\right) = 0,$$

with

$$\Pi_i = \begin{bmatrix} \pi_3^i & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \pi_T^i & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \pi_3^i & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \pi_T^i \end{bmatrix}$$

where

$$\pi_t^i = P(\tau^i \geq t | x_1^i, ..., x_{t-1}^i).$$

The results in Abowd et al. (2001) show that this estimator will be consistent under the MAR hypothesis and that it can be applied in practice by using the recursive formula:

$$\begin{aligned} \pi_{t+1}^i &= (1 - q_t^i)\pi_t^i, \text{ with } \pi_1^i = 1 \\ q_t^i &= P(\tau^i = t | \tau^i \geq t, x_1^i, ..., x_t^i), \end{aligned}$$

where the conditional exit probability q_t^i (i.e. the conditional probability that the last observation of unit i will be at t) can be estimated from the sample of observation units who were observed until (at least) t. In practice, q_t^i might be a logistic or a probit function, as suggested in Abowd et al. (2001) – implicitly assuming that attrition occurs when (a function of) the endogenous variables meet some threshold.

It is easy to verify that the GMM and WGMM estimators are identical when the π_t^i do not depend upon i (but only on t) – and thus are independent of the endogenous variables. Thus, in the MCAR case the two methods give identical results. Another interesting situation occurs when π_{it} becomes zero with positive probability. In this case

WGMM breaks down. An interesting example of this is when the event $[\tau^i = t]$ depends deterministically on the history $x_1^i, ..., x_t^i$. Thus, π_{it} is one when $t \leq \tau^i$ and zero else.

For both moment estimators, the choice of W is critical for the efficiency of the estimators. I follow Arellano and Bond (1998) and choose $W = (\frac{1}{N}\sum_{i=1}^{N} Y_i^{+\prime} D_i^+ H^+ D_i^+ Y_i^+)^{-1}$ for the GMM estimator – and its obvious modification in the WGMM case – where H^+ is the $2T-4$ matrix:

$$H^+ = \begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix},$$

(I is the identity matrix of order $T-2$) and

$$H = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

In my experience, if H is replaced by I the performance of GMM and WGMM deteriorate sharply as the sample becomes more unbalanced. Two-step estimators (see Blundell and Bond (1998) and Abowd et al. (2001)) was partially tested in simulations. Despite being computationally costly, I found no improvement in performance compared to the much simpler one-step estimators.

The PL estimator: Let

$$u_t^i = (1-\phi)v^i + \varepsilon_t^i \quad \text{for } t = 2, ..., \tau^i. \tag{14}$$

The first equation in (8) can then be written

$$x_t^i = \phi x_{t-1}^i + u_t^i \quad \text{for } t = 2, .., \tau^i. \tag{15}$$

It is easily verified that

$$f(x_1^i, .., x_{\tau^i}^i) \propto f(\widetilde{u}_2^i, .., \widetilde{u}_{\tau^i}^i | x_1^i) f(x_1^i),$$

where

$$\widetilde{u}_t^i = \begin{cases} \sqrt{\frac{t-1}{t}}(\varepsilon_{t+1}^i - \frac{1}{t-1}\sum_{v=2}^{t} \varepsilon_v^i) & t = 2, .., \tau^i - 1 \ (\tau^i > 2) \\ (1-\phi)v^i + \frac{1}{\tau^i - 1}\sum_{v=2}^{\tau^i} \varepsilon_v^i & t = \tau^i \ (\tau^i > 1). \end{cases}$$

Note that the $\widetilde{u}_t^i$ are linear functions of the unknown parameters:

$$\widetilde{u}_t^i = \begin{cases} \sqrt{\frac{t-1}{t}}\left((x_{t+1}^i - \frac{1}{t-1}\sum_{v=2}^t x_v^i) - \phi(x_t^i - \frac{1}{t-1}\sum_{v=2}^t x_{v-1}^i)\right) & t = 2, .., \tau^i - 1 \\ \frac{1}{\tau^i}\sum_{v=1}^{\tau^i}(x_v^i - \phi x_{v-1}^i) & t = \tau^i. \end{cases}$$

Hence, the PL estimator is easy to obtain from the distribution of $\widetilde{u}_t^i$:

$$\widetilde{u}_t^i \sim \mathcal{IN}(0, \sigma_\varepsilon^2) \text{ for } t = 2, ..., \tau^i - 1.$$

Furthermore, $(\widetilde{u}_2^i, .., \widetilde{u}_{\tau^i-1}^i)$ are independent of $(\widetilde{u}_{\tau^i}^i, x_1^i)$. The likelihood can therefore be factorized as:

$$f(x_1^i, .., x_{\tau^i}^i) \propto f(x_1^i) f(\widetilde{u}_{\tau^i}^i | x_1^i) \prod_{t=2}^{\tau^i - 1} f(\widetilde{u}_t^i).$$

From the relations

$$\begin{aligned} E\{u_{\tau^i}^i | x_1^i\} &= (1-\phi) E\{v^i | x_1^i\} \\ Var\{u_{\tau^i}^i | x_1^i\} &= (1-\phi)^2 Var\{v^i | x_1^i\} + \frac{\sigma_\varepsilon^2}{\tau^i - 1} \end{aligned}$$

and

$$\begin{aligned} E\{v^i | x_1^i\} &= \frac{x_1^i}{\sigma_1^2}\left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_v^2}\right]^{-1} \\ Var\{v^i | x_1^i\} &= \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_v^2}\right]^{-1} \end{aligned}$$

we obtain:

$$\begin{aligned} \widetilde{u}_{\tau^i}^i | x_1^i &\sim \mathcal{N}\left(\beta x_0^i, \omega^2 + \frac{\sigma_\varepsilon^2}{\tau^i - 1}\right) \\ x_1^i &\sim \mathcal{N}\left(0, \omega_1^2\right), \end{aligned}$$

where

$$\begin{aligned} \beta &= \frac{1-\phi}{\sigma_1^2}\left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_v^2}\right]^{-1} \\ \omega^2 &= (1-\phi)^2\left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_v^2}\right]^{-1} \\ \omega_1^2 &= \sigma_1^2 + \sigma_v^2. \end{aligned}$$

Exit rule	$\tau^i = t$ iff t is the first time the following event occurs
MCAR	$\xi_t^i < c$
MAR	$x_t^i + \gamma\xi_t^i < c$
HW	$x_{t+1}^i + \gamma\xi_t^i < c$
HYBRID	$\frac{1}{\sqrt{2}}(x_t^i + x_{t+1}^i) + \gamma\xi_t^i < c$

Table 1: **Attrition rules used in simulations**

Attrition rules: We focus on three main types of attrition mechanisms which all have been extensively discussed in the literature:

(i) $\tau^i \perp x_1^i, .., x_T^i$ (MCAR)

(ii) $x_{\tau^i+1}^i, .., x_T^i \perp \tau^i \mid x_1^i, .., x_{\tau^i}^i$ (MAR)

(iii) $x_1^i, .., x_{\tau^i}^i \perp \tau^i \mid x_{\tau^i+1}^i, .., x_T^i$ (HW)

Recall that the exit time τ^i is the last period the unit is observed (and not the first time it is missing).

In (i) exit is independent of the endogenous variables. Missing items are therefore MCAR. Type (ii) is the MAR-case: Exit is independent of the unobserved endogenous variables given the observed ones. Type (iii) is the Hausman-Wise (HW) model: Exit is independent of the observed endogenous variables given the unobserved ones. In addition, we will consider an attrition rule (HYBRID) which is neither MAR or HW: Exit depends on both observed and unobserved endogenous variables.

Table 1 specifies the exit rules which are employed in the simulation study. All exit rules say that attrition occurs if a certain lower threshold, c, is met. In Table 1, γ is a scale parameter and $\xi_t^i \sim \mathcal{N}(0,1)$ is white noise – independently distributed of x_t^i and v^i. The threshold c is a number chosen to keep $E(\tau^i)$ constant in all simulations with the same T . We shall consider the cases $T = 6$, with $E(\tau^i) = 5$, and $T = 11$, with $E(\tau^i) = 7.5$. Throughout, the number of units is fixed at $N = 500$. We shall also assume weak stationarity: $\sigma_\varepsilon^2 = \sigma_v^2 = 1$ and $\sigma_1^2 = (1 - \phi^2)^{-1}$.

The attrition rule MCAR says that exit only depends on noise ξ_t^i. The rule MAR says that exit occurs at t depending on the outcomes of ξ_t^i and the endogenous variable x_t^i. The rule HW says that exit occurs at t depending ξ_t^i and the future variable x_{t+1}^i. The rule HYBRID is a combination between MAR and HW: Exit occurs at t depending on both x_t^i, x_{t+1}^i, and noise ξ_t^i.

Exit rule:			MCAR				MAR					
Estimator:			GMM		PL		GMM		WGMM		PL	
T	γ	ϕ	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
6	0	.5	-	-	-	-	-.086	.096	-	-	.001	.036
		.9	-	-	-	-	-.090	.104	-	-	-.007	.029
		.99	-	-	-	-	-.031	.061	-	-	-.017	.037
	σ_1	.5	-.002	.042	-.001	.035	-.054	.070	-.016	.057	-.001	.037
		.9	-.007	.049	-.007	.028	-.048	.068	-.022	.061	-.008	.030
		.99	-.014	.049	-.017	.036	-.021	.055	-.015	.052	-.018	.039
11	0	.5	-	-	-	-	-.041	.052	-	-	-.000	.023
		.9	-	-	-	-	-.057	.067	-	-	-.002	.018
		.99	-	-	-	-	-.031	.045	-	-	-.009	.019
	σ_1	.5	-.007	.036	-.001	.028	-.033	.044	-.037	.056	.000	.022
		.9	-.017	.035	-.003	.017	-.042	.053	-.050	.068	-.002	.017
		.99	-.020	.038	-.010	.020	-.027	.042	-.028	.046	-.009	.020

Table 2: **Estimates of ϕ. Simulation results for attrition rules MCAR and MAR.** $N = 500$; normal error terms; $\sigma_v^2 = \sigma_\varepsilon^2 = 1$, $\sigma_1^2 = \frac{1}{1-\phi^2}$.

The scale parameter γ determines the relative importance of the noise ξ_t^i relative to the endogenous variables in the exit rules specified in Table 1. As γ increases, all scenarios will approach MCAR. In the simulations, two cases are considered: (i) $\gamma = 0$; exit depends deterministically on the endogenous variables and (ii) $\gamma = \sigma_1$; the noise has, roughly, the same impact on the exit decision as the endogenous variables. These scenarios, when varying γ and ϕ, span a wide variety of relevant attrition rules.

The simulation results for normal error terms and with $\phi = 0.5, \phi = 0.9$, and $\phi = 0.99$ are depicted in Tables 2-3. In Table 4 simulation results with non-normal random variables are presented. In the latter case, as in Blundell and Bond (1998), we use the highly non-normal $\chi^2(1)$-distribution: v^i, ε_1^i, and ε_t^i $(t \geq 2)$ are distributed as $\sigma \left(\chi^2(1) - 1\right)/\sqrt{2}$ for $\sigma = \sigma_v, \sigma_1$, and σ_ε, respectively.

Results: The results in Table 2 show that when data are missing completely at random (MCAR), both the PL and the GMM estimator perform quite well – although there is some negative bias as ϕ approaches one. Overall, the root mean square error (RMSE) of the GMM estimator is 50% higher than for the PL estimator (which in this case coincides with the maximum likelihood estimator). The results shift dramatically when we turn to the MAR exit rule. While the performance of the PL estimator remains virtually

Exit rule:			HW						HYBRID					
Estimator:			GMM		WGMM		PL		GMM		WGMM		PL	
T	γ	ϕ	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
6	0	.5	-.067	.078	-	-	-.067	.074	-.091	.100	-	-	-.077	.084
		.9	-.079	.092	-	-	-.069	.084	-.082	.096	-	-	-.050	.065
		.99	-.036	.063	-	-	-.044	.069	-.033	.058	-	-	-.032	.056
	σ_1	.5	-.048	.062	-.043	.058	-.044	.055	-.067	.078	-.060	.073	-.049	.059
		.9	-.052	.071	-.037	.063	-.030	.045	-.057	.074	-.041	.064	-.027	.042
		.99	-.019	.054	-.014	.053	-.023	.045	-.021	.053	-.017	.052	-.023	.045
11	0	.5	-.060	.066	-	-	-.065	.069	-.080	.084	-	-	-.077	.080
		.9	-.075	.081	-	-	-.068	.073	-.078	.084	-	-	-.050	.056
		.99	-.039	.050	-	-	-.026	.039	-.041	.052	-	-	-.020	.031
	σ_1	.5	-.041	.050	-.050	.058	-.041	.046	-.058	.064	-.080	.086	-.049	.054
		.9	-.048	.057	-.056	.067	-.028	.036	-.053	.062	-.065	.075	-.023	.032
		.99	-.026	.040	-.030	.046	-.012	.022	-.027	.041	-.032	.048	-.011	.021

Table 3: **Simulation results for attrition rules HW and HYBRID**. $N = 500$; normal error terms; $\sigma_v^2 = \sigma_\varepsilon^2 = 1$, $\sigma_1^2 = \frac{1}{1-\phi^2}$.

Estimator:		GMM		WGMM		PL	
Exit rule:	ϕ	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
MCAR	.5	-.004	.030	-	-	-.000	.023
	.9	-.015	.035	-	-	-.002	.018
	.99	-.020	.037	-	-	-.009	.021
MAR	.5	-.025	.036	-.006	.031	-.052	.022
	.9	-.045	.053	-.023	.040	.002	.015
	.99	-.030	.042	-.020	.036	-.008	.018
HW	.5	-.026	.037	-.017	.032	-.012	.024
	.55	-.046	.053	-.030	.042	-.007	.018
	.9	-.033	.045	-.027	.042	-.011	.022
HYBRID	.5	-.036	.044	-.023	.037	-.012	.025
	.9	-.052	.058	-.038	.048	-.007	.017
	.99	-.036	.049	-.029	.043	-.011	.022

Table 4: **Simulation results with $\chi^2(1)$-distributed random variables**. $N = 500$; $\sigma_v^2 = \sigma_\varepsilon^2 = 1$, $\sigma_1^2 = \frac{1}{1-\phi^2}$; $\gamma = \sigma_1$.

unchanged, the GMM estimator exhibits substantial negative bias; varying between $-.02$ and $-.09$. On average, the RMSE of the GMM estimator is 125% higher than for the PL estimator in the MAR simulations. The presence of noise in the MAR exit rule ($\gamma \neq 0$) improves the performance of GMM relative to the deterministic case, i.e. $\gamma = 0$, as do a high value of ϕ compared to a small ϕ.

The performance of the WGMM method in the MAR case is somewhat disappointing: The weighting method succeeds in reducing the bias compared to GMM when $T = 6$, but their RMSE is roughly the same. When $T = 11$, the weighting method is actually counterproductive. One explanation for this could be that the weighting method is sensitive to errors in the estimates of q_{it}, which are magnified when more terms are multiplied together to obtain π_{it}. Although the correct exit probability model was estimated: q_{it} =Probit$(c/\gamma - x_t^i/\gamma)$, the estimated parameters do, of course, differ from the true ones due to estimation error. As noted above, when $\gamma = 0$ the WGMM estimator is not well-defined.

Turning to the non-MAR scenarios HW and HYBRID (Table 3), the PL and GMM estimators perform much more evenly. GMM performs similarly for these two exit rules, while its RSME is about 15 percent higher than for the PL estimator in the HW case, and 20 percent higher in the HYBRID case. Because the simulation scenario HYBRID is closer to MAR than to HW, this relative difference is not surprising. As expected, there is no substantial difference between GMM and WGMM in these cases.

The HW and HYBRID attrition models lead to a negative bias in the range of $-.01$ and $-.09$ for both the PL and the GMM estimator. This bias can be explained by a consideration of standard regression arguments: The forward-looking exit rules imply that a unit is observed at $t + 1$ if *either* x_t^i *or* u_{t+1}^i is large, thus inducing a negative correlation between the "regressor" x_t^i and the error term ε_{t+1}^i. Consequently, as the exit rules become more noisy (large γ), the performance of both estimators improve. The bias (and RMSE) decrease as ϕ approaches 1. This result can be explained by the same type of argument: When ϕ increases, $Var(x_t^i)$ also increases, but not the covariance between x_t^i and ε_{t+1}^i. Thus the relative importance of the bias-inducing covariance decreases.

Some of the most striking results are found in Table 4, which compares the estimators when the error terms in the autoregressive model are $\chi^2(1)$-distributed (re-scaled and re-centered to have mean zero and variance one). We see that the performance of the PL estimator changes very little compared to normality. In particular, there is virtually no bias in the MAR and MCAR attrition models, while the performance under the attrition rules HW and HYBRID are actually somewhat better than before. The results for the GMM and WGMM methods are also similar to those in Table 2-3. Averaging over all cases reported in Table 4, the RMSE of the GMM method is more than 2 times as high as for the PL estimator.

In econometrics it is often argued that GMM is preferable to likelihood based methods, because one does not have to specify the distributions of the random variables (However, see Sims (2000) for a different view). The results in Table 4 yield little merit to this argument: The PL estimator outperforms GMM (and WGMM) under normality as well as under the highly non-normal $\chi^2(1)$-distribution. The kind of departures from model assumptions which are most critical for inference are related to the nature of the attrition mechanism. A forward looking attrition rule that depends on the outcomes of future variables (as in the HW and HYBRID attrition models) is problematic for inference. Our simulations indicate that these problems are particularly important when there is little noise in the attrition rule relative to the variance in the endogenous variables. Departure from normality is a much lesser concern.

6 Conclusions

This paper has discussed identification, estimation and testing in panel data models with attrition. In the situation where attrition is endogenous and depends on the outcomes of an observed stochastic process *and* the interest-parameter characterizing this process, attrition is non-ignorable even if selection is based only on observed variables – that is, even if the missing items are *missing at random* (MAR). The likelihood function obtained by ignoring the attrition mechanism is a pseudo likelihood function. Assuming that MAR holds, this paper has established conditions for identification of interest parameters based on the pseudo likelihood function.

In contrast to a widely held opinion, the MAR hypothesis has testable implications in many situations which are encountered in practice: It implies that at any point in time information about the life-span of the observation unit up until that point in time is irrelevant for predicting future variables, given the complete history of the observed variables. Hence, augmenting transition equations with auxiliary parameters which measure effects of such information, should not lead to a significant increase in the pseudo likelihood.

While MAR is a sufficient condition for validity of pseudo likelihood based inference, it is not sufficient for the validity of GMM-methods. Traditional GMM-estimators require that attrition is independent of the endogenous variables. This is a very unreasonable assumption in many applications. On the other hand, the MAR hypothesis allows attrition to depend arbitrarily on observed endogenous variables, and may therefor accommodate much more realistic attrition mechanisms. In many panel data studies, this provides a strong rationale for using likelihood based methods instead of moment estimators.

References

[2001] Abowd, J.M., F. Kramarz, and B.Crepon (2001). "Moment Estimation with Attrition: An Application to Economic Models." *Journal of the American Statistical Association,* 96(456).

[1995] Ahn, S.C. and Schmidt, P. (1995). "Efficient Estimation of Models for Panel Data". *Econometrics*, 68, 5-27.

[1998] Arellano, M. and S. Bond (1998). "Dynamic Panel data estimation using DPD98 for Gauss: A guide for users." Manuscript. Institute for Fiscal Studies, London.

[1995] Arellano, M. and Bover, O. (1995). "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68, 29-51.

[1994] Barndorff-Nielsen, O.E., and Cox, D. R. (1994). *Inference and Asymptotics.* Chapman and Hall, London.

[1986] Billingsley, P. (1986). *Probability and Measure.* John Wiley, New York.

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

[1975] Cox, D.R. (1975). "Partial Likelihood." *Biometrika*, 62, 269-76.

[1984] Gourieroux, C. and A. Montfort (1984). "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52,681-700.

[1999] Hahn, J. (1999). "How informative is the initail condition in the dynamic panel model with fixed effects?". *Journal of Econometrics* 93, 309-26.

[1979] Hausman, J.A. and D.A. Wise (1979). "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment." *Econometrica*, 47,455-73.

[1979] Heckman, J.J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica*, 47, 153-61.

[1998] Hirano, K, G.W. Imbens, G. Ridder and D. Rubin (1998). "Combining Panel Data Sets and Refreshment Samples". *Econometrica,* forthcoming.

[1998] Horowitz, J.L. and C. Manski (1998). "Censoring Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation using Wights and Imputations." *Journal of Econometrics*, 84, 37-58..

[1982] Jovanovic, B. (1982). "Selection and Evolution of Industries." *Econometrica*, 50, 649-70.

[1987] Little, R.J.A. and D. Rubin. *Statistical Analysis with Missing Dat*a. John Wiley, New York.

[1999] Moffitt, R., J. Fitzgerald, P. Gottschalk (1999). "Sample Attrition in Panel Data: The Role of Selection on Observables". *Annales-d'Economie-et-de-Statistique*, 0(55-56). 129-52.

[2000] Sims, C. (2000). "Using a likelihood perspective to sharpen econometric discourse Three examples." *Journal of Econometrics*, 95, 443-62.

[1982] White, H (1982). "Maximum Likelihood Estimation of Misspecified Models". Econometrica, 50,1-25.

Recent publications in the series Discussion Papers

237 K-G. Lindquist (1998): The Response by the Norwegian Aluminium Industry to Changing Market Structure

238 J.K. Dagsvik, A.S. Flaatten and H. Brunborg: A Behavioral Two-Sex Model

239 K.A. Brekke, R.B. Howarth and K. Nyborg (1998): Are there Social Limits to Growth?

240 R. B. Howarth and K. A. Brekke (1998): Status Preferences and Economic Growth

241 H. Medin, K. Nyborg and I. Bateman (1998): The Assumption of Equal Marginal Utility of Income: How Much Does it Matter?

242 B. Bye (1998): Labour Market Rigidities and Environmental Tax Reforms: Welfare Effects of Different Regimes

243 B.E. Naug (1999): Modelling the Demand for Imports and Domestic Output

244 J. Sexton and A. R. Swensen (1999): ECM-algorithms that converge at the rate of EM

245 E. Berg, S. Kverndokk and K.E. Rosendahl (1999): Optimal Oil Exploration under Climate Treaties

246 J.K. Dagsvik and B.H. Vatne (1999): Is the Distribution of Income Compatible with a Stable Distribution?

247 R. Johansen and J.K. Dagsvik (1999): The Dynamics of a Behavioral Two-Sex Demographic Model

248 M. Søberg (1999): Asymmetric information and international tradable quota treaties. An experimental evaluation

249 S. Grepperud, H. Wiig and F.A. Aune (1999): Maize Trade Liberalization vs. Fertilizer Subsidies in Tanzania: A CGE Model Analysis with Endogenous Soil Fertility

250 K.A. Brekke and Nils Chr. Stenseth (1999): A Bio-Economic Approach to the study of Pastoralism, Famine and Cycles. Changes in ecological dynamics resulting from changes in socio-political factors

251 T. Fæhn and E. Holmøy (1999): Welfare Effects of Trade Liberalisation in Distorted Economies. A Dynamic General Equilibrium Assessment for Norway

252 R. Aaberge (1999): Sampling Errors and Cross-Country Comparisons of Income Inequality

253 I. Svendsen (1999): Female labour participation rates in Norway – trends and cycles

254 A. Langørgen and R. Aaberge: A Structural Approach for Measuring Fiscal Disparities

255 B. Halvorsen and B.M. Larsen (1999): Changes in the Pattern of Household Electricity Demand over Time

256 P. Boug (1999): The Demand for Labour and the Lucas Critique. Evidence from Norwegian Manufacturing

257 M. Rege (1999): Social Norms and Private Provision of Public Goods: Endogenous Peer Groups

258 L. Lindholt (1999): Beyond Kyoto: CO_2 permit prices and the markets for fossil fuels

259 R. Bjørnstad and R. Nymoen (1999): Wage and Profitability: Norwegian Manufacturing 1967-1998

260 T.O. Thoresen and K.O. Aarbu (1999): Income Responses to Tax Changes – Evidence from the Norwegian Tax Reform

261 B. Bye and K. Nyborg (1999): The Welfare Effects of Carbon Policies: Grandfathered Quotas versus Differentiated Taxes

262 T. Kornstad and T.O. Thoresen (1999): Means-testing the Child Benefit

263 M. Rønsen and M. Sundström (1999): Public Policies and the Employment Dynamics among new Mothers – A Comparison of Finland, Norway and Sweden

264 J.K. Dagsvik (2000): Multinomial Choice and Selectivity

265 Y. Li (2000): Modeling the Choice of Working when the Set of Job Opportunities is Latent

266 E. Holmøy and T. Hægeland (2000): Aggregate Productivity and Heterogeneous Firms

267 S. Kverndokk, L. Lindholt and K.E. Rosendahl (2000): Stabilisation of CO_2 concentrations: Mitigation scenarios using the Petro model

268 E. Biørn, K-G. Lindquist and T. Skjerpen (2000): Micro Data On Capital Inputs: Attempts to Reconcile Stock and Flow Information

269 I. Aslaksen and C. Koren (2000): Child Care in the Welfare State. A critique of the Rosen model

270 R. Bjørnstad (2000): The Effect of Skill Mismatch on Wages in a small open Economy with Centralized Wage Setting: The Norwegian Case

271 R. Aaberge (2000): Ranking Intersecting Lorenz Curves

272 J.E. Roemer, R. Aaberge , U. Colombino, J, Fritzell, S.P. Jenkins, I. Marx, M. Page, E. Pommer, J. Ruiz-Castillo, M. Jesus SanSegundo, T. Tranaes, G.G.Wagner and I. Zubiri (2000): To what Extent do Fiscal Regimes Equalize Opportunities for Income Acquisition Among citizens?

273 I. Thomsen and L.-C. Zhang (2000): The Effect of Using Administrative Registers in Economic Short Term Statistics: The Norwegian Labour Force Survey as a Case Study

274 I. Thomsen, L.-C. Zhang and J. Sexton (2000): Markov Chain Generated Profile Likelihood Inference under Generalized Proportional to Size Non-ignorable Non-response

275 A. Bruvoll and H. Medin (2000): Factoring the environmental Kuznets curve. Evidence from Norway

276 I. Aslaksen, T. Wennemo and R. Aaberge (2000): "Birds of a feather flock together". The Impact of Choice of Spouse on Family Labor Income Inequality

277 I. Aslaksen and K.A. Brekke (2000): Valuation of Social Capital and Environmental Externalities

278 H. Dale-Olsen and D. Rønningen (2000): The Importance of Definitions of Data and Observation Frequencies for Job and Worker Flows - Norwegian Experiences 1996-1997

279 K. Nyborg and M. Rege (2000): The Evolution of Considerate Smoking Behavior

280 M. Søberg (2000): Imperfect competition, sequential auctions, and emissions trading: An experimental evaluation

281 L. Lindholt (2000): On Natural Resource Rent and the Wealth of a Nation. A Study Based on National Accounts in Norway 1930-95

282 M. Rege (2000): Networking Strategy: Cooperate Today in Order to Meet a Cooperator Tomorrow

283 P. Boug, Å. Cappelen and A.R. Swensen (2000): Expectations in Export Price Formation: Tests using Cointegrated VAR Models

284 E. Fjærli and R. Aaberge (2000): Tax Reforms, Dividend Policy and Trends in Income Inequality: Empirical Evidence based on Norwegian Data

285 L.-C. Zhang (2000): On dispersion preserving estimation of the mean of a binary variable from small areas

286 F.R. Aune, T. Bye and T.A. Johnsen (2000): Gas power generation in Norway: Good or bad for the climate? Revised version

287 A. Benedictow (2000): An Econometric Analysis of Exports of Metals: Product Differentiation and Limited Output Capacity

288 A. Langørgen (2000): Revealed Standards for Distributing Public Home-Care on Clients

289 T. Skjerpen and A.R. Swensen (2000): Testing for long-run homogeneity in the Linear Almost Ideal Demand System. An application on Norwegian quarterly data for non-durables

290 K.A. Brekke, S. Kverndokk and K. Nyborg (2000): An Economic Model of Moral Motivation

291 A. Raknerud and R. Golombek: Exit Dynamics with Rational Expectations

292 E. Biørn, K-G. Lindquist and T. Skjerpen (2000): Heterogeneity in Returns to Scale: A Random Coefficient Analysis with Unbalanced Panel Data

293 K-G. Lindquist and T. Skjerpen (2000): Explaining the change in skill structure of labour demand in Norwegian manufacturing

294 K. R. Wangen and E. Biørn (2001): Individual Heterogeneity and Price Responses in Tobacco Consumption: A Two-Commodity Analysis of Unbalanced Panel Data

295 A. Raknerud (2001): A State Space Approach for Estimating VAR Models for Panel Data with Latent Dynamic Components

296 J.T. Lind (2001): Tout est au mieux dans ce meilleur des ménages possibles. The Pangloss critique of equivalence scales

297 J.F. Bjørnstad and D.E. Sommervoll (2001): Modeling Binary Panel Data with Nonresponse

298 Taran Fæhn and Erling Holmøy (2001): Trade Liberalisation and Effects on Pollutive Emissions and Waste. A General Equilibrium Assessment for Norway

299 J.K. Dagsvik (2001): Compensated Variation in Random Utility Models

300 K. Nyborg and M. Rege (2001): Does Public Policy Crowd Out Private Contributions to Public Goods?

301 T. Hægeland (2001): Experience and Schooling: Substitutes or Complements

302 T. Hægeland (2001): Changing Returns to Education Across Cohorts. Selection, School System or Skills Obsolescence?

303 R. Bjørnstad: (2001): Learned Helplessness, Discouraged Workers, and Multiple Unemployment Equilibria in a Search Model

304 K. G. Salvanes and S. E. Førre (2001): Job Creation, Heterogeneous Workers and Technical Change: Matched Worker/Plant Data Evidence from Norway

305 E. R. Larsen (2001): Revealing Demand for Nature Experience Using Purchase Data of Equipment and Lodging

306 B. Bye and T. Åvitsland (2001): The welfare effects of housing taxation in a distorted economy: A general equilibrium analysis

307 R. Aaberge, U. Colombino and J.E. Roemer (2001): Equality of Opportunity versus Equality of Outcome in Analysing Optimal Income Taxation: Empirical Evidence based on Italian Data

308 T. Kornstad (2001): Are Predicted Lifetime Consumption Profiles Robust with respect to Model Specifications?

309 H. Hungnes (2001): Estimating and Restricting Growth Rates and Cointegration Means. With Applications to Consumption and Money Demand

310 M. Rege and K. Telle (2001): An Experimental Investigation of Social Norms

311 L.C. Zhang (2001): A method of weighting adjustment for survey data subject to nonignorable nonresponse

312 K. R. Wangen and E. Biørn (2001): Prevalence and substitution effects in tobacco consumption. A discrete choice analysis of panel data

313 G.H. Bjertnær (2001): Optimal Combinations of Income Tax and Subsidies for Education

314 K. E. Rosendahl (2002): Cost-effective environmental policy: Implications of induced technological change

315 T. Kornstad and T.O. Thoresen (2002): A Discrete Choice Model for Labor Supply and Child Care

316 A. Bruvoll and K. Nyborg (2002): On the value of households' recycling efforts

317 E. Biørn and T. Skjerpen (2002): Aggregation and Aggregation Biases in Production Functions: A Panel Data Analysis of Translog Models

318 Ø. Døhl (2002): Energy Flexibility and Technological Progress with Multioutput Production. Application on Norwegian Pulp and Paper Industries

319 R. Aaberge (2002): Characterization and Measurement of Duration Dependence in Hazard Rate Models

320 T. J. Klette and A. Raknerud (2002): How and why do Firms differ?

321 J. Aasness and E. Røed Larsen (2002): Distributional and Environmental Effects of Taxes on Transportation

322 E. Røed Larsen (2002): The Political Economy of Global Warming: From Data to Decisions

323 E. Røed Larsen (2002): Searching for Basic Consumption Patterns: Is the Engel Elasticity of Housing Unity?

324 E. Røed Larsen (2002): Estimating Latent Total Consumption in a Household.

325 E. Røed Larsen (2002): Consumption Inequality in Norway in the 80s and 90s.

326 H.C. Bjørnland and H. Hungnes (2002): Fundamental determinants of the long run real exchange rate:The case of Norway.

327 M. Søberg (2002): A laboratory stress-test of bid, double and offer auctions.

328 M. Søberg (2002): Voting rules and endogenous trading institutions: An experimental study.

329 M. Søberg (2002): The Duhem-Quine thesis and experimental economics: A reinterpretation.

330 A. Raknerud (2002): Identification, Estimation and Testing in Panel Data Models with Attrition: The Role of the Missing at Random Assumption