TECHNICAL WORKING PAPER SERIES

MOMENT ESTIMATION WITH ATTRITION

John M. Abowd
Bruno Crépon
Francis Kramarz

Technical Working Paper 214

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1997

The authors thank Geert Ridder, Jean Marc Robin and participants at the CREST microeconometrics seminar for helpful comments on a preliminary draft. Abowd acknowledges the support of the National Science Foundation (SBR 96-18111 to the NBER). The data used in this study are confidential but the authors' access is not exclusive. An exact copy of the extract used in this paper is available from Francis Kramarz, INSEE-CREST, 15 bd. Gabriel Péri, 92245 Malakoff Cedex, France, (kramarz@ensae.fr). This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1997 by John M. Abowd, Bruno Crépon and Francis Kramarz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Moment Estimation with Attrition
John M. Abowd, Bruno Crépon and Francis Kramarz
NBER Technical Working Paper No. 214
August 1997
JEL Nos. C13, C33, D21
Labor Studies

ABSTRACT

We present a method that accommodates missing data in longitudinal datasets of the type usually encountered in economic and social applications. The technique uses various extensions of "missing at random" assumptions that we customize for dynamic models. Our method, applicable to longitudinal data on persons or firms, is implemented using the Generalized Method of Moments with reweighting that appropriately corrects for the attrition bias caused by the missing data. We apply the method to the estimation of dynamic labor demand models. The results demonstrate that the correction is extremely important.

John M. Abowd
Department of Economics
259 Ives Hall
Cornell University
Ithaca, NY 14853-3901
and NBER
john_abowd@cornell.edu

Bruno Crépon
Département de la recherche
CREST/INSEE
15, bd Gabriel Péri
92245 Malakoff Cedex
FRANCE

Francis Kramarz
Département de la recherche
CREST/INSEE
15, bd Gabriél Peri
92245 Malakoff Cedex
FRANCE
kramarz@ensae.fr

1. Introduction

Firms are born and die, plants open and close, individuals move from labor market to labor market. When economists and statisticians sample fragments of the lives of these entities, the impact of these movements, often manifested as unexplained attrition from the sample, on model estimation and inference can be devastating: most estimated coefficients are biased and inconsistent even under rather strong ancillary assumptions. The classical solution to this dilemma is to estimate the process of interest jointly with a model of the entry and exit process. We are going to focus primarily on attrition in this paper; hence, we can represent the process more specifically by letting s_{it} denote the survival process ($s_{it} = 1$ if entity i survives at date t), and s^*_{it} denote a latent (unobserved) variable governing the survival of the entity of interest. Then, exit or attrition is often represented as

$$s^*_{it} = \alpha' z_{it} + \varepsilon_{it}$$

where z_{it} is a vector of time varying observables, ε_{it} is an error term with zero mean, given z_{it}, and other properties that depend upon the model of interest, and α is a vector of unknown parameters, so that

$$\begin{cases} s_{it} &= 1, \text{ if } s^*_{it} \geq 0 \\ s_{it} &= 0 \text{ otherwise.} \end{cases}$$

If the economic or social process of interest is governed by the following equation

$$y_{it} = \gamma' x_{it} + u_{it}$$

where y_{it} and x_{it} are observable, u_{it} is an error term with properties that also depend upon the specific model, and γ is a vector of coefficients of interest, then consistent estimation of the system is theoretically straightforward.

There are many reasons, however, that estimation of even simple models might be difficult. First, even when the simple model above is a reasonable description of the death or attrition process, the structure of the error processes, for example through person or firm-specific effects, can induce complicated correlation between the errors in the two equations that makes estimation cumbersome. Estimation using a maximum likelihood approach may require the computation of a T-fold integral (where T denotes the maximum number of periods for which an entity is observed.[1] Second, one generally does not know why a firm or an individual disappears from a sample. For instance, suppose that firms die either because they are not profitable enough or because they are very profitable and are, therefore, subject to a takeover. Then, the simple exit model above is inadequate. A model with both a ceiling and a floor would be a better description of the death process. John Sutton (1996) gives different examples of this possibility: "In looking at the U.K. steel casings industry, Charles Baden-Fuller (1989) found that many of the closing plants were not the least profitable ones. ... A similar theme emerges in the work of Martha Schary (1991) who appealed to a richer discussion of the determinants of exit, distinguishing between three routes by which firms 'disappeared' (bankruptcy, voluntary liquidation, or merger). Schary's paper examines the process of exit in the U.S. cotton textile industry, using a model in which firms make a series of decisions, considering each exit route in turn, in a pre-determined order." (pages 67 and 68). More generally, it seems difficult to summarize the attrition process without reference to a more structural and sector-specific model (see also Olley and Pakes (1996)).

We try to avoid these difficulties by taking a completely different route. Inspired by the statistical missing data literature (Rubin (1976), Robins, Rotnitzky, and Zhao (1994 and 1995), Imbens and Hellerstein (1996), and Hirano, Imbens, Ridder, and Rubin (1996)), we formulate simple assumptions on the stochastic process by which firms or individuals exit from samples. Then, based on these hypotheses, we show that a simple adaptation of the GMM framework allows consistent estimation of any equation using moment conditions. In particular, we propose a class of estimators based on a Weighted Generalized Method of Moments, in which weights are computed using the exit process, that leads to consistent estimators of the parameters of interest, γ, and shares the asymptotic properties of the Generalized Method of Moments.

After defining notation in Section 2, we analyze the implications of different assumptions for the attrition rule in Section 3. We provide the specific implementation details in Section 4. In Section 5, we apply our framework to the study of a labor demand equation using an unbalanced panel data set of firms. Finally, we conclude in section 6.

[1] See, among others, Meghir and Saunders (1988), Ridder (1990), Baltagi and Chang (1994), Dionne, Gagné, and Vanasse (1994), see also Heckman (1979) and Hausman and Wise (1979)).

2. Notation

Consider a population of N individuals. We focus on the following process $\underline{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$, with y_{it} a $K \times 1$ vector (the y and x's of the introduction), where $i = 1$ to N, denotes firms or individuals, t denotes time, and $t = 1$ is the date of birth, common to all individuals. This common date of birth is assumed to be known. Here, y_{it} is a vector which includes both dependent and explanatory variables of the process. Denote the conditions that relate these variables as follows:

$$Eg(\underline{y}_i, \theta) = 0 \tag{2.1}$$

or, denoting $\underline{y}_{it} = (y_{i1}, \ldots, y_{it})$,

$$Eg_t(\underline{y}_{it}, \theta) = 0 \text{ for } t = 1, \ldots, T \tag{2.2}$$

where $\theta \in \Theta$ denotes the parameters to be estimated and $g_t()$ denotes the appropriate subvector of $g()$. We assume that the set of parameters can be split into parameters of interest ($\alpha \in A$) and nuisance parameters ($\beta \in B$) with $\Theta = A \times B$. The nuisance parameters will only be mentioned when working with the moment conditions. Furthermore, we assume that there exists an additional $L \times 1$ vector z_{it} of time-varying covariates. These covariates do not enter the moment conditions 2.1.

Individuals are observed from date $t = 1$ until date d_i. Hence, $\underline{y}_{id_i} = (y_{i1}, y_{i2}, \ldots, y_{id_i})$ is observed but $\overline{y}_{id_i} = (y_{id_i+1}, y_{id_i+2}, \ldots, y_{iT})$ is not. The date d_i is the last date at which the variables are observed. Let $s_{it} \in \{0, 1\}$ be an indicator function equal to 1 whenever y_{it} is observed. Therefore, $s_{i1} = 1, \ldots s_{id_i} = 1$ and $s_{id_i+1} = 0, \ldots, s_{iT} = 0$. Finally, denote

$$q_{it} = P(s_{it} = 1 \mid \underline{y}_{it-1}, s_{it-1} = 1)$$

$$\pi_{it} = P(s_{it} = 1 \mid \underline{y}_{it-1})$$

$$\tilde{q}_{it} = P(s_{it} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1}), s_{it-1} = 1)$$

$$\tilde{\pi}_{it} = P(s_{it} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1}))$$

Notice that the above conditions in conjunction with equations (2.1) or (2.2) may have no empirical counterpart since data are observed at date t if and only if entity i survived until that date. Hence, estimation based on observed data may lead to biased estimates due to the attrition.

3. Attrition Rules

3.1. Data Missing at Random

The Rubin "data missing at random" assumption can be expressed as follows. We say that data are missing at random if the date t observation indicator function, s_{it}, and the vector $\underline{y}_i$ are independent conditional on the history of the vector y, *i.e.*, conditional on $\underline{y}_{it-1} = (y_{i1}, y_{i2}, \ldots, y_{it-1})$. Then, if $l(a \mid b)$ denotes the distribution of a conditional on b (recall that for notational simplicity we do not write the parameters except in our moment conditions) missing at random means:

$$l(\underline{y}_i, s_{it} \mid \underline{y}_{it-1}) = l(\underline{y}_i \mid \underline{y}_{it-1}) l(s_{it} \mid \underline{y}_{it-1}) \tag{3.1}$$

The Rubin missing at random rule can be equivalently reformulated as:

$$P(s_{it} = 1 \mid \underline{y}_i) = P(s_{it} = 1 \mid \underline{y}_{it-1}) \tag{3.2}$$

which implies that

$$l(y_{it} \mid \underline{y}_{it-1}, s_{it} = 1) = l(y_{it} \mid \underline{y}_{it-1})$$

Therefore, given the chain rule for conditional probabilities, we have

$$l(\underline{y}_i) = l(y_{iT} \mid \underline{y}_{iT-1}) l(y_{iT-1} \mid \underline{y}_{iT-2}) \cdots l(y_{i1}),$$

and

$$l(\underline{y}_i) = l(y_{iT} \mid \underline{y}_{iT-1}, s_{iT} = 1) l(y_{iT-1} \mid \underline{y}_{iT-2}, s_{iT-1} = 1) \cdots l(y_{i1}, s_{i1} = 1),$$

Now, each conditional density is expressed in terms of observables so that the density is identified. However, its estimation requires a lot of computation and many observations. Hence, we apply this idea directly to our moment conditions rather than to the densities themselves.

Proposition 3.1. *Under Rubin's attrition rule (3.1), the following two equations hold*

$$E\left(g_t(\underline{y}_{it},\theta)\right) = E\left(\frac{g_t(\underline{y}_{it},\theta)s_{it}}{\pi_{it}}\right) \tag{3.3}$$

$$\pi_{it} = q_{it}\pi_{it-1} \tag{3.4}$$

Proof: Given the independence assumption of equation (3.1), we have

$$E\left(g_t(\underline{y}_{it},\theta)s_{it} \mid \underline{y}_{it-1}\right) = E\left(g_t(\underline{y}_{it},\theta) \mid \underline{y}_{it-1}\right) E\left(s_{it} \mid \underline{y}_{it-1}\right)$$

Notice that $E(g_t(\underline{y}_{it},\theta)s_{it} \mid y_{it-1})$ has a sample analog, since when $s_{it} = 0$ the function $g_t(\underline{y}_{it},\theta)s_{it}$ is identically equal to zero, and whenever $s_{it} = 1$, $\underline{y}_{it}$ is observed (by definition of s_{it}). Let $\pi_{it} = E(s_{it} \mid \underline{y}_{it-1}) = P(s_{it} = 1 \mid \underline{y}_{it-1})$. Furthermore, $q_{it} = E(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 1)$ is directly identified. On the other hand, π_{it} is not directly identified. It is not possible to estimate a model, say a probit or a logit, directly on s_{it} because some of the y_{it-1} are not observed. But, this becomes possible if done conditional on $s_{it-1} = 1$. This gives q_{it}. More precisely, we have the relation:

$$\begin{aligned} E\left(s_{it} \mid \underline{y}_{it-1}\right) &= E\left(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 1\right) P\left(s_{it-1} = 1 \mid \underline{y}_{it-1}\right) + \\ &\quad E\left(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 0\right) P\left(s_{it-1} = 0 \mid \underline{y}_{it-1}\right) \end{aligned}$$

Given that

$$E\left(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 0\right) = 0$$

$$P\left(s_{it-1} = 1 \mid \underline{y}_{it-1}\right) \doteq P\left(s_{it-1} = 1 \mid \underline{y}_{it-2}\right) = \pi_{t-1},$$

and equation (3.1), we have:

$$\pi_{it} = q_{it}\pi_{it-1}$$

Thus, given $\pi_{i1} = q_{i1}$

$$\pi_{it} = q_{it} \cdots q_{i1}$$

and it follows that π_{it} is identifiable. Q.E.D.

Equations (3.3) and (3.4) have several implications. Consider the case of GMM estimation. There is a set of orthogonality conditions which is satisfied on the whole population given by equation (2.1). Decompose the function g into

$g = (g_1, g_2, \ldots, g_T)$ with g_t a function of $(y_{i1}, \ldots, y_{it})$. Then, the equation (3.3) can be applied to each g_t. This provides the following orthogonality conditions:

$$E\left(\frac{g_t(\underline{y}_{it}, \theta) s_{it}}{\pi_{it}}\right) = 0 \tag{3.5}$$

which have an empirical counterpart.

Consider next the Chamberlain method (see Chamberlain (1984)). To apply this methodology, it is necessary to compute the moment matrix $\Sigma_{t,s} = E\left(y_{it} y'_{is}\right)$. Once more, use the previous result to $g_t = y_{it} y'_{is}$ with $s \leq t$. This yields:

$$E\left(\frac{y_{it} y'_{is} s_{it}}{\pi_{it}} - y_{it} y'_{is} \middle| \underline{y}_{it-1}\right) = 0$$

and, in particular,

$$E\left(\frac{y_{it} y'_{is} s_{it}}{\pi_{it}}\right) = \Sigma_{t,s} \tag{3.6}$$

Hence, it becomes possible to compute the moment matrix of the variables, and to make inferences using Chamberlain's framework.

In our first version of Rubin's data missing at random assumption, the conditioning variable was $\underline{y}_{it-1}$: the same variables that enter the moment conditions. We could slightly alter this rule by including the z variables as follows

$$l((\underline{y}_i, \underline{z}_i), s_{it} \mid (\underline{y}_{it-1}, \underline{z}_{it-1})) = l((\underline{y}_i, \underline{z}_i) \mid (\underline{y}_{it-1}, \underline{z}_{it-1})) l(s_{it} \mid (\underline{y}_{it-1}, \underline{z}_{it-1})) \tag{3.7}$$

The statistical assumption is formally identical, data are missing at random, but the conditioning set of variables is larger. This situation and the previous one intersect but none is included in the other. From this variant of the missing at random assumption, we derive the following result.

Proposition 3.2. *Under Rubin's attrition rule (3.7), the following two equations hold*

$$E\left(g_t(\underline{y}_{it}, \theta)\right) = E\left(\frac{g_t(\underline{y}_{it}, \theta) s_{it}}{\tilde{\pi}_{it}}\right) \tag{3.8}$$

$$\tilde{\pi}_{it} = \tilde{q}_{it} \tilde{\pi}_{it-1} \tag{3.9}$$

Proof: As for proposition 3.1. Q.E.D.

One may wonder whether a weakening of the previous assumptions leads to an equivalent result. More specifically, let us assume that

$$l(\underline{y}_i, s_{it} \mid (\underline{y}_{it-1}, \underline{z}_{it-1})) = l(\underline{y}_i \mid (\underline{y}_{it-1}, \underline{z}_{it-1}))l(s_{it} \mid (\underline{y}_{it-1}, \underline{z}_{it-1})) \tag{3.10}$$

Hence $\underline{y}_i$ is orthogonal to s_{it} given $(\underline{y}_{it-1}, \underline{z}_{it-1})$. We call this rule "conditional" Rubin. Then, we can state the following result.

Proposition 3.3. *Under Rubin's attrition rule (3.10), the following two equations hold*

$$E\left(g_t(\underline{y}_{it}, \theta)\right) = E\left(\frac{g_t(\underline{y}_{it}, \theta)s_{it}}{\tilde{\pi}_{it}}\right) \tag{3.11}$$

$$\tilde{\pi}_{it} \neq \tilde{q}_{it}\tilde{\pi}_{it-1} \tag{3.12}$$

Proof: For the first equation, the demonstration is identical to the proof of proposition 3.1. However,

$$\tilde{\pi}_{it} = P(s_{it} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1}))$$

$$\tilde{\pi}_{it} = P(s_{it} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1}), s_{it-1} = 1)P(s_{it-1} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1}))$$

$$\tilde{\pi}_{it} = \tilde{q}_{it}P(s_{it-1} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1}))$$

but, we do not have

$$P(s_{it-1} = 1 \mid (\underline{y}_{it-1}, \underline{z}_{it-1})) = P(s_{it-1} = 1 \mid (\underline{y}_{it-2}, \underline{z}_{it-2}))$$

and, therefore, we do not have $\tilde{\pi}_{it} = \tilde{q}_{it}\tilde{\pi}_{it-1}$. Indeed to have the equality above, condition (3.7) is sufficient but not necessary. Q.E.D.

Suppose now that there is a set of variables, w, which are always observed (or do not time-vary) even after firm's i death. Then,

$$l(\underline{y}_i, s_{it} \mid (\underline{y}_{it-1}, \underline{w}_i) = l(\underline{y}_i \mid (\underline{y}_{it-1}, \underline{w}_i))l(s_{it} \mid (\underline{y}_{it-1}, \underline{w}_i)) \tag{3.13}$$

Then, denoting

$$\bar{q}_{it} = P(s_{it} = 1 \mid (\underline{y}_{it-1}, \underline{w}_i), s_{it-1} = 1)$$

$$\bar{\pi}_{it} = P(s_{it} = 1 \mid (\underline{y}_{it-1}, \underline{w}_i))$$

we have

Proposition 3.4. *Under Rubin's attrition rule (3.13), the following two equations hold*

$$E\left(g_t(\underline{y}_{it},\theta)\right) = E\left(\frac{g_t(\underline{y}_{it},\theta)s_{it}}{\overline{\pi}_{it}}\right)$$

$$\overline{\pi}_{it} = \overline{q}_{it}\overline{\pi}_{it-1}$$

Proof: Contrary to the result in propostion 3.3, we have

$$P(s_{it-1} = 1 \mid (\underline{y}_{it-1}, \underline{w}_i)) = P(s_{it-1} = 1 \mid (\underline{y}_{it-2}, \underline{w}_i))$$

Q.E.D.

The case of missing at random given w. also mentioned in Robins *et al.* (1995), is encountered when some variables do not time-vary or when there are multiple data sources. For example, variables known at the start of the sample like industry and location are potential w variables. A second example occurs when one has two different datasets, the first giving information on balance-sheets, which is always present for all firms in all years, and the second giving information on research and development expenditures, the variables of interest for the model, for which some firms have missing data. In this example, the balance sheet variables are candidate w variables.

3.2. Robins, Rotniztky, and Zhao' Attrition Rule

Robins, Rotnitzky. and Zhao (1994) and (1995) propose the following attrition rule:

$$P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_i) = P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_{it-1}) \qquad (3.14)$$

with $P(s_{i1} = 1 \mid \underline{y}_i) = P(s_{i1} = 1) = 1$ (date 1 is the first date of the sample period).

Lemma 3.5. *Under Robins et al.'s attrition rule (3.14), one has for any k, $1 \leq k \leq t$,*

$$P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i) = P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})$$

Proof: The proof is by induction. Consider any k, $1 \leq k \leq t$. And consider the following equality (for a given k),

$$P(s_{i\tau} = 1 \mid s_{ik} = 1, \underline{y}_i) = P(s_{i\tau} = 1 \mid s_{ik} = 1, \underline{y}_{i\tau-1})$$

This equality is obviously true for $\tau = k$. Suppose now that this equality is true for $\tau = t-1$. We show now that it is also true for $\tau = t$. For any conditioning vector z, we have

$$\begin{aligned} P(s_{it} &= 1 \mid s_{ik} = 1, z) = P(s_{it} = 1 \mid s_{ik} = 1, s_{it-1} = 1, z) \times P(s_{it-1} = 1 \mid s_{ik} = 1, z) \\ +P(s_{it} &= 1 \mid s_{ik} = 1, s_{it-1} = 0, z) \times P(s_{it-1} = 0 \mid s_{ik} = 1, z) \end{aligned}$$

But $P(s_{it} = 1 \mid s_{ik} = 1, s_{it-1} = 0, z) = 0$, since there is no reentry (a dead firm never reappears). Hence,

$$P(s_{it} = 1 \mid s_{ik} = 1, z) = P(s_{it} = 1 \mid s_{ik} = 1, s_{it-1} = 1, z) \times P(s_{it-1} = 1 \mid s_{ik} = 1, z)$$

Furthermore, notice that $P(s_{it} = 1 \mid s_{ik} = 1, s_{it-1} = 1, z) = P(s_{it} = 1 \mid s_{it-1} = 1, z)$ for the same reason as above. Now, apply these facts to $z = \underline{y}_i$ and $z = \underline{y}_{it-1}$. This yields

$$P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i) = P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_i) \times P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_i)$$

and

$$P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) = P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_{it-1}) \times P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})$$

Now, using Robins *et al.* attrition rule, we derive that

$$\frac{P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i)}{P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})} = \frac{P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_i)}{P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})}$$

Use the induction hypothesis $(\tau = t-1)$ *i.e.*

$$P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_i) = P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-2})$$

But, we know that

$$\begin{aligned} P(s_{it-1} &= 1 \mid s_{ik} = 1, \underline{y}_{it-1}) = \\ \int P(s_{it-1} &= 1 \mid s_{ik} = 1, \underline{y}_i) f(y_{iT}, y_{iT-1}, \ldots, y_{it} \mid s_{ik} = 1, \underline{y}_{it-1}) dy_{iT} \ldots dy_{it} \end{aligned}$$

where $f(.)$ denotes the distribution function of our variables. Apply the previous equality to get:

$$\begin{aligned} P(s_{it-1} &= 1 \mid s_{ik} = 1, \underline{y}_{it-1}) = \\ \int P(s_{it-1} &= 1 \mid s_{ik} = 1, \underline{y}_{it-2}) f(y_{iT}, y_{iT-1}, \ldots, y_{it} \mid s_{ik} = 1, \underline{y}_{it-1}) dy_{iT} \ldots dy_{it} \end{aligned}$$

The first part under the integral is a constant and therefore can be taken out of the integral to give

$$\begin{aligned} P(s_{it-1} &= 1 \mid s_{ik} = 1, \underline{y}_{it-1}) = \\ P(s_{it-1} &= 1 \mid s_{ik} = 1, \underline{y}_{it-2}) \int f(y_{iT}, y_{iT-1}, \ldots, y_{it} \mid s_{ik} = 1, \underline{y}_{it-1}) dy_{iT} \ldots dy_{it} \end{aligned}$$

Hence,

$$P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) = P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-2})$$

by definition of a distribution function. Apply this last equality to yield

$$\frac{P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i)}{P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})} = \frac{P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_i)}{P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})} \frac{P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_i)}{P(s_{it-1} = 1 \mid s_{ik} = 1, \underline{y}_{it-2})}$$

But, by the induction hypothesis the right-handside is equal to 1. So, we have shown that the equality is also true for $\tau = t$. Q.E.D.

Remark: Applying our lemma to $k = 1$ shows that Robins *et al.* attrition rule is a special case of data missing at random (as noted in Robins *et al.* (1994)).

Robins *et al.*'s attrition rule is helpful when the date of birth is unobserved. Indeed, denoting $\pi_{it}^{k+1} = q_{it}q_{it-1}\cdots q_{ik+1}$, we have the following result.

Proposition 3.6. *Under Robins et al.'s attrition rule (3.14), one has for any k,*

$$E\left(\frac{g_t(\underline{y}_{it})s_{it}}{\pi_{it}^{k+1}} - g_t(\underline{y}_{it}) \,|s_{ik} = 1\right) = 0. \tag{3.15}$$

Proof: First we show that

$$P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i) = P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) = q_{it}q_{it-1}\cdots q_{ik+1}.$$

Indeed, we have

$$\begin{aligned} H &= P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) \\ &= P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}, s_{it-1} \times \cdots \times s_{ik+1} = 1) \\ &\times P(s_{it-1} \times \cdots \times s_{ik+1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) \\ &+ P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}, s_{it-1} \times \cdots \times s_{ik+1} = 0) \\ &\times P(s_{it-1} \times \cdots \times s_{ik+1} = 0 \mid s_{ik} = 1, \underline{y}_{it-1}) \end{aligned}$$

$P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}, s_{it-1} \times \cdots \times s_{ik+1} = 0) = 0$ because entities never reappear. Thus,

$$\begin{aligned} H &= P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) \\ &= P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}, s_{it-1} \times \cdots \times s_{ik+1} = 1) \\ &\times P(s_{it-1} \times \cdots \times s_{ik+1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) \\ &= P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_{it-1}) P(s_{it-1} \times \cdots \times s_{ik+1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) \\ &= P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_{it-1}) P(s_{it-1} = 1 \mid s_{it-2} = 1, \underline{y}_{it-1}) \\ &\times \cdots P(s_{ik+1} = 1 \mid s_{ik} = 1, \underline{y}_{it-1}) \end{aligned}$$

Lemma 3.5 implies that:

$$P(s_{it-j} = 1 \mid s_{it-j-1} = 1, \underline{y}_{it-1}) = P(s_{it-j} = 1 \mid s_{it-j-1} = 1, \underline{y}_{it-j-1})$$

which implies that $H = q_{it} q_{it-1} \cdots q_{ik+1}$. Note also that nothing is changed in $\underline{y}_i$ since $H = P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i)$. Our lemma 3.5 has shown that under Robins *et al.*'s attrition rule:

$$P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_i) = P(s_{it} = 1 \mid s_{ik} = 1, \underline{y}_{it-1})$$

Hence,

$$\begin{aligned} E(s_{it} g_t(\underline{y}_{it}) \mid \underline{y}_{it-1}, s_{ik} = 1) &= E(s_{it} \mid \underline{y}_{it-1}, s_{ik} = 1) E(g_t(\underline{y}_{it}) \mid \underline{y}_{it-1}, s_{ik} = 1) \\ &= q_{it} q_{it-1} \cdots q_{ik+1} E(g_t(\underline{y}_{it}) \mid \underline{y}_{it-1}, s_{ik} = 1) \end{aligned}$$

which yields the desired result. Q.E.D.

Remark: Under the stronger assumptions that (i) a finite number of lags of y_{it}, say m, has to be introduced in q_{it} *i.e.*

$$P(s_{it} = 1 \mid s_{it-1} = 1, \underline{y}_{it-1}) = P(s_{it} = 1 \mid s_{it-1} = 1, y_{it-1}, \ldots, y_{it-m})$$

then π_{it}^{m+1} is identified; (ii) the g_t are restricted to those in which only the observed ys i.e. $(y_{i1}, \ldots, y_{it})$ enter so that the expectation $E(g_t(\underline{y}_{it}) \mid s_{im} = 1)$ be identified; and that (iii) $E(g_t(\underline{y}_{it}) \mid s_{im} = 1) = E(g_t(\underline{y}_{it})) = 0$; we can use Robin's *et al.*'s attrition rule for moment estimation.

Notice that all our other propositions can be extended in this setting. This discussion also demonstrates that the two attrition rules, Rubin's missing at random, on the one hand, and Robins, Rotnitzky and Zhao's rule on the other, differ in their premises, consequences, and applicability.

4. Implementation of the Method

Consider now Rubin's data missing at random attrition rule. Assume for simplicity that the nuisance parameters, q_t, have a logit form (for the case of first-stage non parametric estimates, see Newey (1994)). Hence, we can estimate a set of parameters $(\beta_t)_{t=1,\dots,T}$ defined by the following orthogonality conditions on $(\psi_t)_{t=1,\dots,T}$, the derivatives of the logit functions with respect to β_t,

$$E(\psi_t(\beta_t, \underline{y}_{it})) = 0$$

Thus, the complete set of conditions for the GMM is

$$\begin{cases} E(\frac{g_t(\underline{y}_{it}, \theta) s_{it}}{q_{it}(\beta_t, \underline{y}_{it-1}) \cdots q_{i2}(\beta_2, \underline{y}_{i1})}) = 0 \\ E(\psi_t(\beta_t, \underline{y}_{it-1}) \mid s_{it-1} = 1) = 0 \end{cases} \quad t = 1, \cdots, T \tag{4.1}$$

The standard errors can be computed using the corresponding formulas for the GMM estimation of the equations (4.1). However, for the computation of the parameters themselves, it is simpler to solve first for the equations that define the β parameters and to replace them by their value in the estimating equations for the estimation of the parameter of interest θ (see Crépon, Kramarz and Trognon forthcoming). There will be no efficiency loss as long as there are as many independent orthogonality conditions, ψ, as there are parameters β as is the case if the β's are defined by maximization of the logit likelihood, for example (see, again, Crépon, Kramarz and Trognon forthcoming).

4.1. Computation of the Standard Errors

To show how to compute the standard errors of our parameters, consider first the restatement of the preceding GMM program (4.1) as

$$\begin{cases} E(f(\underline{y}_i, \theta, \beta)) = 0 \\ E(\psi_t(\beta_t, \underline{y}_{it-1}) \mid s_{it-1} = 1) = 0 \end{cases} \quad t = 1, \cdots, T$$

where

$$f(\underline{y}_i, \theta, \beta) \equiv \frac{g_t(\underline{y}_{it}, \theta) s_{it}}{q_{it}(\beta_t, \underline{y}_{it-1}) \cdots q_{i2}(\beta_2, \underline{y}_{i1})}.$$

The estimated parameter $\widehat{\theta}$ is defined by

$$\overline{F_\theta(\underline{y}_i, \widehat{\theta}, \widehat{\beta})}' S \overline{f(\underline{y}_i, \widehat{\theta}, \widehat{\beta})} = 0$$

where

$$\overline{F_\theta(\underline{y}_i,\widehat{\theta},\widehat{\beta})} \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{\partial f(\underline{y}_i,\widehat{\theta},\widehat{\beta})}{\partial\theta'}$$

$$\overline{f(\underline{y}_i,\widehat{\theta},\widehat{\beta})} \equiv \frac{1}{N}\sum_{i=1}^{N} f(\underline{y}_i,\widehat{\theta},\widehat{\beta})$$

and S is the optimal weighting matrix, derived below. Expanding around the true value θ_0, we get

$$\overline{F_\theta(\underline{y}_i,\widehat{\theta},\widehat{\beta})}' S \left[\overline{f_{i0}} + \overline{F_\theta(\underline{y}_i,\widehat{\theta},\widehat{\beta})}(\widehat{\theta}-\theta_0) + \overline{F_\beta(\underline{y}_i,\widehat{\theta},\widehat{\beta})}(\widehat{\beta}-\beta_0)\right] = 0$$

where the obvious definitions apply. Using the following approximation of $\widehat{\beta}$

$$\widehat{\beta} - \beta_0 \simeq \overline{\phi_{i0}}$$

one gets that $\widehat{\theta}$ is given by:

$$\widehat{\theta} - \theta_0 \simeq -(E(F_{\theta i})' S E(F_{\theta i}))^{-1} E(F_{\theta i})' S \left[\overline{f_{i0} + E(F_{\beta i})\phi_{i0}}\right]$$

where $E(F_{\theta i})$ and $E(F_{\beta i})$ stand for the expectation of the corresponding derivatives of f evaluated at the true value of the parameters and $\overline{\phi_{i0}}$ is the first order term of the expansion of $E\psi_t$ around θ_0. In this framework, all the usual results from the GMM apply for $\widehat{\theta}$. However, the variance matrix of the estimated parameters and the optimal weight matrix have to be modified because we must account for the estimated β, hence we use $f_{i0} + E(F_{\beta i})\phi_{i0}$ instead of f_{i0}. Hence, the optimal weighting matrix is

$$S = \mathrm{Var}[f_{i0} + E(F_{\beta i})\phi_{i0}]^{-1}.$$

Given the optimal weighting matrix S, evaluated at the second-stage estimates of $\widehat{\theta}$ and $\widehat{\beta}$, we have

$$\mathrm{Var}\left[\widehat{\theta}\right] = (E(F_{\theta i})' S E(F_{\theta i}))^{-1}$$

4.1.1. Computation of $E(F_{\beta_i})$ and $E(F_{\theta i})$

The vector f is defined by $f' = (f_1', \ldots, f_T')$ with

$$f_t = \frac{g_t(\underline{y}_{it},\theta)s_{it}}{q_{it}(\beta_t,\underline{y}_{it-1})\cdots q_{i2}(\beta_2,\underline{y}_{i1})}$$

Thus, we have

$$\frac{\partial f_t}{\partial \beta_s'} = \begin{cases} 0 \text{ if } s > t \\ -f_t \frac{\partial \log(q_{is})}{\partial \beta_s'} \end{cases} = -f_t \frac{\partial \log(q_{is})}{\partial \beta_s'}$$

It turns out that $E(F_{\beta i})$ is a block inferior triangular matrix with block (t, s) equal to

$$E(F_{\beta i})_{t,s} = E(\frac{\partial f_t}{\partial \beta_s'}) = -E(f_t \frac{\partial \log(q_{is})}{\partial \beta_s'})$$

Denote $\nabla \log q_i = \left[\frac{\partial \log(q_{i2})}{\partial \beta_2'}, \ldots, \frac{\partial \log(q_{iT})}{\partial \beta_T'}\right]$, then we have

$$E(F_{\theta i}) = -E(f \nabla \log q_i)$$

4.1.2. Computation of $\widehat{\beta}$

Let β denote the vector of β_t, $t = 1$ to T. Each β_t is defined by

$$E(\psi_t(\beta_t, \underline{y}_{it-1}) \mid s_{it-1} = 1) = 0$$

Assume for simplicity that the dimension of ψ_t is equal to that of β_t and that β_t is defined by a maximum likelihood estimate. An estimate $\widehat{\beta}_t$ of β_t is defined by

$$\sum_i \psi_t(\widehat{\beta}_t, \underline{y}_{it-1}) s_{it-1} = 0$$

Expanding around the true value, the estimated parameter can be obtained using

$$\sum_i \psi_t(\beta_t, \underline{y}_{it-1}) s_{it-1} + (\sum_i \frac{\partial \psi_t}{\partial \beta_t}(\beta_t, \underline{y}_{it-1}) s_{it-1})(\widehat{\beta}_t - \beta_t) \simeq 0$$

Thus, we have

$$(\widehat{\beta}_t - \beta_t) \simeq -\left(\frac{1}{N} \sum_i \frac{\partial \psi_t}{\partial \beta_t}(\beta_t, \underline{y}_{it-1}) s_{it-1}\right)^{-1} \frac{1}{N} \sum_i \psi_t(\beta_t, \underline{y}_{it-1}) s_{it-1}$$

Defining $\phi_i' = (\phi_{1i}', \cdots, \phi_{Ti}')$ with ϕ_{it} defined by

$$\phi_{it} = -\left(\frac{1}{N} \sum_i \frac{\partial \psi_t}{\partial \beta_t}(\beta_t, \underline{y}_{it-1}) s_{it-1}\right)^{-1} \psi_t(\beta_t, \underline{y}_{it-1}) s_{it-1}$$

one gets

$$\widehat{\beta} - \beta \simeq \overline{\phi_i}.$$

5. Labor Demand

In this section, we use our methodology to estimate the parameters of interest in a system of dynamic labor demand equations. First, we describe the data. Then, we present the estimation results. In particular, we compare the estimates for the speed of adjustment with those obtained in the usual framework that does not take attrition into account.

5.1. The Data

We start with a representative unbalanced longitudinal sample of French firms from the INSEE Echantillon d'entreprises 1978-1988. As constructed by the division of economic studies, the 1988 version of this sample includes 21,642 firms. The sample is cross-sectionally representative in each of the eleven years; that is, the age structure of the sampled firms is representative of the age structure in the population for each year. The enterprise universe consists of every enterprise that responded to the BIC (Benefices industriels et commerciaux) in any year between 1978 and 1988. In the primary sample year, 1986, every entreprise that completed the BIC in 1986 was at risk to be sampled with probabilities that depended upon the size of the entreprise in 1986 and the sector of economic activity. Firms with 500 or more employees were sampled with probability one. Firms with less than 20 employees were not sampled.[2] Firms of intermediate sizes were sampled with probabilities between $\frac{1}{30}$ and 1 according to the size and sector. Every firm that existed in 1986 was then removed from the universe. Complementary samples were then constructed for each of the other years using the same sampling probabilities as in 1986. After a complementary sample was drawn for a particular year all firms at risk to be sampled that year were eliminated from the universe.

In principle, the BIC is exhaustive, however many enterprises have at least one year of missing data between the first and last years they appear in the BIC. From the initial sample of 21,642 firms, 8.2% were discarded for having two missing years between the first and last year they appear. An additional 14.9% of the firms, having one year of missing data in the BIC, were used in the dynamic sample with the missing year imputed (using the procedure described below). Finally, 16,645 enterprises had complete data and were used without any imputations. The enterprises that appear in the dynamic sample represent 91.8%

[2] Firms with less than 20 employees (full- or part-time) were not required to complete the BIC until 1984.

(unweighted) of the firms in the universe and 90.4% (weighted[3]). The number of observations (firm-years) is 129,447 of which 2.5% contain some imputed data. We eliminated all data from large governmental entreprises (the SNCF, RATP, la Poste, Air France, Electricité de France, Gaz de France, governmental savings banks, etc.) and all data from a group of extremely anomalous entreprises.[4] All told we eliminated data for 53 entreprises, including the governmental firms.

To apply our method. as we need to know the complete history of the firms, we focus on all firms that were born in 1982 (i.e. within our sample period). These 667 firms are followed at most until 1988. Of these 667 firms, 292 survived to this last date. Furthermore, 139 lived only one year, 76 lived exactly two years, 54 lived exactly three years. 28 lived exactly four years, 28 lived exactly five years, 50 lived exactly six years. and finally, 292 lived at least seven years, *i.e.* they were still alive in 1988, the end of the sample period.

5.2. Construction of Variables

From the BIC, we used total employment at december 31^{st}, real value-added, real total employer labor cost per employee (frais de personnel y.c. charges patronales divided by total employment), gross operating sales, real total assets, and financial ratios. The financial ratios are defined as follows. Investment ratio (investment divided by value-added), margin ratio (operating profit divided by value-added), return on assets (operating profit divided by total assets), return on fixed assets (operating profit divided by fixed assets), financial return on assets (operating profit plus earnings on financial assets minus interest charges, *i.e.* profit brut courant avant impôts, divided by total assets), average interest rate (interest charges divided by total long-term plus mid-term debt), debt ratio (total long-term plus mid-term debt divided by total assets), ratio of long-term assets (equity plus total long-term and mid-term debt divided by total assets), and solvency ratio (interest charges divided by operating profit). All such ratios are used either by banks or financial analysts to evaluate the future of firms. To compute the variables defined in real terms, we used the value-added price index at the industry level with no change in definition from the BIC. Summary statistics of the variables are presented in Table A.1.

[3]The weights come from the sampling plan of the echantillon d'entreprises, described above.
[4]These firms had negative assets or some other gross accounting anomaly.

5.3. Estimation results

The estimated equation is a classical dynamic labor demand relation. For comparison purposes, we used the same equation as the one estimated in Arellano and Bond (1991):

$$L_{it} = \alpha_1 L_{it-1} + \alpha_2 L_{it-2} + \sum_{j=0}^{1} \beta_{wj} w_{it-j} + \sum_{j=0}^{1} \beta_{kj} k_{it-j} + \sum_{j=0}^{1} \beta_{qj} q_{it-j} + \sum_{k=1985}^{1988} \delta_k I(t=k) + \varepsilon_{it} \tag{5.1}$$

for $t = 1985$ to $t = 1988$. This equation is estimated in first differences. Instrumental variables are the lags of our exogenous variables $(t-2, \ldots, t-5)$ and endogenous variables $(t-4$ and $t-5)$.

Table 1 contains all of our estimation results. Column (1) reports the estimates when we use all of the y variables–employment, wage, sales, and capital, which enter the moment equations g–as well as all eight financial ratios, which do not enter these moment equations, to model death.[5] Column (2) contains the estimates when we model death using all of the y variables and value-added, operating profit, and total debts, which do not enter the moment equations g. Column (3) relies on a model of death restricted to the y variables and excluding both the financial ratios and the other economic variables used in columns (1) and (2). Column (4) reports unweighted estimates which result from the estimation of equation (5.1) without any correction for the missing data caused by firms that exit before 1988. Finally, column (5) presents estimates for the balanced panel i.e. firms that are born in 1982 and survive at least until the end of the period, 1988.

Columns (1) to (3) show that our weighting framework has in important effect upon the estimation of the key parameters of the labor demand equation. In these columns, all coefficients have the correct sign and reasonable magnitude. According to the results presented in column (1), the cofficient on the lag of order 2 for employment is not significantly different from zero and of trivial magnitude. Hence, one would conclude that it is not necessary to disaggregate employment in multiple skill categories (see Nickell 1984). However, columns (2) and (3) show that both coefficients on lagged employment are significantly different from zero and of nontrivial magnitude; therefore, some disaggregation by skill level may be appropriate for equation (5.1). Notice here that the coefficients imply

[5]The logit results are not reported for the ψ moment conditions. Each logit equation and, therefore, its associated ψ_ts contain the past (until 1982, the first year for all our firms) of all the variables used to predict death.

that the two AR(1) processes for the different skill levels have coefficients of about 0.84 and 0.08, respectively, on lagged employment. Hence, the employment skill level with the first coefficient–the slower to adjust–summarizes most of the dynamics (again, see Nickell 1984). The estimated logit coefficients of the death models tend to support the results presented in columns (2) and (3); because the financial variables do not predict well the death of the firm. The χ^2 statistic (test of overidentifying restrictions related to the instruments) is lower in column (3). While this test is informative for the choice of instruments, it is not a test of the quality of the death model. As shown in Hirano *et al.* (1996), Rubin's missing at random assumption exactly identifies the density $l(\underline{y}_i)$. Hence, unless some additional information is introduced, there exists no test of overidentifying restrictions associated with the attrition rule.

Inspection of the last two columns, in which estimates are not corrected for attrition, demonstrates the size of the biases especially since in the absence of missing data effects, these columns estimated the same model as columns (1)-(3). The biases are large for all coefficients. The uncorrect estimates imply that employment adjustments are quick, because the autoregressive coefficients on employment are relatively small, and that the wage elasticity of employment is determined primarily by the first lag whereas in the corrected results both lags of the wage rate are important. With respect to the speed of adjustment the unadjusted results are close to those obtained on the balanced sample, which suggest that Nijman and Verbeek's (1996 and forthcoming) suggestion to compare balanced versus unbalanced estimates is not necessarily a good way to assess the importance of attrition in the estimates. For the speed of adjustment coefficients, our column (4) and (5) estimates are close but both are far from our corrected ones in columns (1)-(3). It is quite clear from our estimates that attrition matters and correction is crucial.

6. Conclusion

In this article, we propose a simple way to take attrition into account within a GMM framework. Our approach relies on Rubin's data missing at random assumption adapted to a panel data framework. We show how to implement this method and demonstrate empirically that attrition matters in simple economic problems such as labor demand equations. Our application was performed on firm data but our methodology could have been applied to attrition within a panel of individuals (the PSID, NLSY, German SOEP, French Enquête Emploi) to estimate, for instance, wage equations or to attrition that occcurs in dynamic samples of other entities.

Table 1
Labor Demand Model (GMM estimation)

Dependent Variable: Employment at t	(1) (StE)	(2) (StE)	(3) (StE)	(4) (StE)	(5) (StE)
Employment at $t-1$	0.6329	0.9124	0.9174	0.3455	0.3300
	(0.1022)	(0.1513)	(0.1575)	(0.0620)	(0.1055)
Employment at $t-2$	-0.0202	-0.0736	-0.0667	-0.0237	-0.0205
	(0.0241)	(0.0251)	(0.0185)	(0.0220)	(0.0148)
Wage at t	-0.7326	-1.0666	-1.0270	-1.3856	-0.5402
	(0.2896)	(0.2772)	(0.2761)	(0.1648)	(0.3102)
Wage at $t-1$	-0.4003	-0.7265	-0.7013	0.1602	-0.1283
	(0.2597)	(0.1800)	(0.1834)	(0.1281)	(0.1052)
Capital at t	0.0012	-0.0239	-0.0215	-0.0152	-0.0013
($\times$ 100)	(0.0091)	(0.0041)	(0.0029)	(0.0052)	(0.0033)
Capital at $t-1$	0.0183	0.0402	0.0395	0.0570	-0.0024
($\times$ 100)	(0.0112)	(0.0056)	(0.0053)	(0.0048)	(0.0082)
Output at t	0.0006	0.0008	0.0008	0.0012	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Output at $t-1$	0.0001	-0.0004	-0.0004	-0.0003	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
χ^2	24.62	23.59	15.99	36.07	21.76
Degrees of Freedom	22	22	22	22	22
p-value	0.317	0.369	0.816	0.030	0.484
Number of Observations	667	667	667	667	292

Sources: BIC.
Notes: All regressions include time dummies. All regressions use lagged exogenous and endogenous variables as instruments. In column (1), the first-stage logit regressions use employment, wage, output, capital, and 8 financial ratios. In column (2), the logit regression uses employment, wage, output, capital, value-added, total debts, operating profit. In column (3), the logit regression uses employment, wage, output, capital. The regression of column (4) is not corrected for attrition. In column (5), we use a balanced sample

Table A.1
Summary Statistics
(firms born in 1982, sample period 1982-1988)

Variable:	Mean	Standard Deviation
Employment (Full-Time Equivalent)	184.5	843.6
Total Labor Cost (per employee, log, FF 1980)	4.340	0.419
Sales (Millions FF 1980)	88.44	329.14
Value-Added (Millions FF 1980)	18.63	90.29
Operating Profit / Total Assets	1.755	53.65
Operating Profit / Fixed Assets	-0.836	39.70
Financial Charges/ Long and Mid-Term Debt	0.502	8.848
Investment/ Value-Added	0.137	3.040
Operating Profit/ Value-Added	0.126	3.944
(Operating Profit + Financial Products -Financial Charges) / Total Assets	-0.464	21.67
Long and Mid-Term Debt / Total Assets	2.939	50.61
Long-Term Assets / Total Assets	2.224	17.43
Financial Charges/ Total Assets	0.030	10.19
Number of Observations	3,049	3,049

Source: BIC.

Appendix: Practical Implementation Issues

In this appendix, we list the different steps that have been followed in the empirical section when estimating a labor demand relationship based on firm data. At each date t, one must:

- Compute (logit) estimates of the attrition process for that date. This provides the probability of dying at each date given survival at the preceding date: q_{it}.
- Stack one above the other these probabilities. This gives a $N \times 1$ matrix q_t;
- Compute $\frac{\partial \log(q_{it})}{\partial \beta_t'} = (1 - q_{it})\underline{y}_{it-1}$
- Compute

$$\phi_{it} = -\left(\frac{1}{N}\sum_i \frac{\partial \psi_t}{\partial \beta_t}(\beta_t, \underline{y}_{it-1})s_{it-1}\right)^{-1} s_{it-1}\left[s_{it}(1 - q_{it}) - (1 - s_{it})q_{it}\right]\underline{y}_{it-1}$$

- For each of the two last steps, stack one above the other the transposed vectors. This gives two $N \times \dim(\underline{y}_{it-1})$ matrix $\tilde{\phi}_t$ and $\nabla \widetilde{\log}(q_t)$
- Compute the $N \times T$ weight matrix π and stack one next to the other the $\widetilde{\phi_t}$ matrices and $\nabla \widetilde{\log}(q_t)$ matrices. This gives two matrices of the same size $N \times \dim(\beta)$: $\tilde{\phi}$ and $\nabla \widetilde{\log}(q)$
- Perform a first estimate using any weight matrix S;
- Denoting $\tilde{f}$ be the matrix of orthogonality conditions $N \times \dim f$, compute an estimate $\widehat{E(F_{\theta i})}$ of $E(F_{\theta i})$ as $-\frac{1}{N}\tilde{f}'\nabla \widetilde{\log} q$
- Compute the residual of estimated orthogonality conditions $\tilde{\epsilon}' = \tilde{f}' + \widehat{E(F_{\theta i})}\tilde{\phi}'$
- Compute an estimate of its variance matrix $\widehat{W} = \frac{1}{N}\tilde{\epsilon}'\tilde{\epsilon}$
- Compute an estimate of the variance of the first estimate using $\widehat{W}$ or perform the best estimate using $\widehat{W}^{-1}$ as weight matrix.

References

[1] Abowd. J.M., Crépon, B., Kramarz, F., and A. Trognon (1995), "A la Recherche des Moments Perdus: Covariance Models for Panel Data with Endogenous Death", NBER technical working paper 180.

[2] Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.

[3] Baden-Fuller, C. (1989), "Exits from Declining Industries and the Case of Steel Castings", Economic Journal, 99, 949-961.

[4] Baltagi. B.H. and Y.J. Chang (1994), "Incomplete Panels, a Comparative Study of Alternative Estimation for the Unbalanced One-Way Error Component Regression Model," *Journal of Econometrics*, 62, 67-89.

[5] Crépon. B., Kramarz, F., and A. Trognon (forthcoming), "Parameters of Interest and Nuisance Parameters: an Application to Autoregressive Error Component Models", *Journal of Econometrics.*

[6] Dionne. G., Gagné, R., and C. Vanasse (1996), "Inferring Technological Parameters from Incomplete Panel Data", Université de Montréal working paper.

[7] Hausman, J.A. and D.A. Wise (1979)"Attrition Bias in Experimental Data: the Gary Income Maintenance Hypothesis", *Econometrica*, 47, 455-473.

[8] Heckman, J.J (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-161.

[9] Hirano. K., Imbens, G. W., Ridder, G. and D. Rubin (1996) "Combining Panel Data Sets with Attrition and Refreshment Samples," Harvard working paper.

[10] Imbens. G. W. and J. K. Hellerstein (1996) "Imposing Moment Restrictions from Auxiliary Data by Weighting," NBER technical working paper 202.

[11] Meghir. C. and Saunders (1988), "Attrition in Company Panels and the Estimation of Investment Equations" UCL working paper.

[12] Newey, W.K. (1994), "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, 62, 1349-1382.

[13] Nickell, S. (1984) "An Investigation of the Determinants of Manufacturing Employment in the United Kingdom" *Review of Economic Studies* 51(4): 529-558.

[14] Olley, G.S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica*, 64, 1263-1298.

[15] Ridder, G. (1990), "Attrition in Multi-Wave Panel Data", in Hartog, Ridder, and Theeuwes (eds.), in *Panel Data and Labor Market Studies*, North-Holland.

[16] Robins, J.M., Rotnitzky, A., and L.P.Zhao (1994), "Estimation of Regression Coefficients When Some Regressors are not Always Observed", *Journal of the American Statistical Association*, 89, 846-866.

[17] Robins, J.M., Rotnitzky, A., and L.P.Zhao (1995), "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data", *Journal of the American Statistical Association*, 90, 106-121.

[18] Rubin, D.B. (1976), "Inference and Missing Data", Biometrika, 63, 581-592.

[19] Schary, M. (1991), "The Probability of Exit", *Rand Journal of Economics*, 22, 339-353.

[20] Sutton, J. (1996), "Gibrat's Legacy", LSE working paper, The Economics of Industry Group.

[21] Verbeek, M and T. E. Nijman (1996) "Incomplete Panels and Selectivity Bias," in the *Econometrics of Panel Data*, L. Matyas and P. Sevestre eds., Kluwer Academic Publishers, pp. 449-490.

[22] Verbeek, M and T. E. Nijman (forthcoming) "Testing for Selectivity Bias in Panel Data Models," *International Economic Review*.