Moment Estimation with Attrition: An Application to Economic Models

John M. Abowd   Bruno Crépon   Francis Kramarz

May 2001

Author footnote

John M. Abowd is Professor of Labor Economics, Cornell University, Ithaca, NY 14853; Distinguished Senior Research Fellow, U.S. Bureau of the Census, and affiliated with the Centre de Recherche en Economie et Statistique (CREST) and the National Bureau of Economic Research (John_Abowd@cornell.edu). Bruno Crépon is Head of the Division Marchés et Stratégies d'Entreprises, Institut National de la Statistique et des Etudes Economiques (INSEE)-CREST, 15, bd Gabriel Péri, 92245 Malakoff Cedex, France (crepon@ensae.fr). Francis Kramarz is Head of the Département de la Recherche, CREST-INSEE, 15, bd Gabriel Péri, 92245 Malakoff Cedex, France and affiliated with Center for Economic Policy and Research (kramarz@ensae.fr). Abowd acknowledges the support of the National Science Foundation (SBER 96-18111 to the NBER). The authors thank Geert Ridder, Jean Marc Robin, participants of the CREST microeconometrics seminar, two editors and two referees for helpful comments. They also thank Philippe Lagarde for his assistance with the INSEE firm data. The opinions expressed in this paper are the authors' and not those of any government agency. No U.S. Title 13 confidential data were used in this paper, which was substantially completed before Abowd's appointment at the Census Bureau began. The data used in this paper are confidential but the authors' access is not exclusive. For further information contact CREST, 15, bd Gabriel Péri, 92245 Malakoff Cedex, France. All programs used in this paper are available from the authors.

Abstract

We study the effects of the attrition of firms from longitudinal samples on the estimates of dynamic labor demand models. The reasons for attrition from business-based longitudinal samples are extremely varied and are related to both the economic activity of the business and the methods of acquiring sampling frame information for those businesses. We do an exhaustive study of the available information regarding the attrition of French firms from our analysis sample. We propose flexible attrition models based on a longitudinal generalization of the missing at random assumption. We implement these models with a weighted generalized method of moments estimator that is consistent and efficient (in the class of moment estimators). Our flexible attrition models substantially alter and improve the estimation results for dynamic factor demand models. We attribute the improvement to the ability of our models to handle the very diverse reasons for attrition that our audit uncovered without requiring specific knowledge of which reason applies to a particular exiting firm.

Keywords: adjustment costs, attrition audit, business missing data, dynamic labor demand, generalized method of moments, reweighted estimation

1. INTRODUCTION

When statisticians sample fragments of the lives of businesses the effects of deaths, reorganizations, and other movements are often manifested as unexplained attrition from the sample. Model estimation and inference can be very seriously affected–most estimated coefficients are biased and inconsistent even under rather strong ancillary assumptions. The classical solution is to estimate jointly the process of interest and a model based on the economic reasons for exit. (See, for example, Heckman, 1979, and Olley and Pakes, 1996.) Current techniques may introduce serious biases and inconsistencies in estimates based upon samples of firms or establishments when they make structural assumptions about the exit process that are at odds with the real reasons for attrition.

The reasons for attrition from business-based longitudinal samples are extremely varied and are related to both the economic activity and the methods used to maintain the sampling frame. To demonstrate this heterogeneity we do an exhaustive study of the reasons for attrition of French firms from our analysis sample. Attrition from our sample is very rarely associated with the death of the economic assets used by the firm. We next show that the use of more flexible attrition models substantially changes the estimation results for dynamic factor demand models. Finally, we compare estimates with and without our attrition correction to external evidence on adjustment costs in France. Our reweighted estimates, reflecting the attrition correction, are much closer to the external evidence than the uncorrected estimates, which are implausibly low for France. Hence, modeling attrition as if it implied economic death leads to very serious inference errors that are substantially mitigated by the use of the more flexible models that we propose.

Dynamic factor adjustments are related to a set of costs that measure the firm's difficulty in reallocating its assets and employment. If adjustment costs are minimal, assets and employment move very freely among firms. In this case, the attrition of a firm from the sample probably means that its assets and employment are zero because they have moved to another economic activity, presumably represented by one of the firms already in the sample or a firm born in a subsequent period. When adjustment costs are substantial, neither assets nor employment are very mobile because the reallocation costs discourage their redeployment. In this case, the disappearance of a firm from the sample must be viewed with caution because it is unlikely that its assets or employment have been completely redeployed and, consequently, it is unlikely that another firm in the sample (or newly born) represents the missing economic activity. Thus, dynamic factor demand models are ideally suited to our application: their estimation requires the use of longitudinal data and they are very likely to be contaminated by sample attrition.

Our statistical methods are based on the missing data models of Rubin (1976), Robins *et al.* (1994 and 1995), Hirano *et al.* (1997) and Hellerstein and Imbens (1999). We make simple, flexible assumptions on the stochastic process by which firms exit from samples. Our models can handle the attrition possibilities uncovered in our audit of the French data without requiring knowledge of the actual cause of attrition for a particular firm. We show that a class of estimators based on a weighted generalized method of moments (WGMM), for which the weights are computed using the estimated exit process, allows consistent and efficient (within the class of moment estimators) estimation of any structural equation using moment conditions (Hansen 1982, Chamberlain 1987 and Newey and McFadden 1994).

In Section 2, we describe the economic problem of interest, show the impact of sample attrition on this problem, and provide statistical evidence on the variety of causes of attrition. The statistical framework that underlies our attrition models is given in Section 3. From this analysis, we derive our preferred methodology. We provide the implementation details in Section 4. In Section 5, we apply our method to the estimation of the dynamic labor demand equation for a firm. We compare various possible implementations using both formal statistical tests and external evidence. Finally, we conclude in section 6.

2. THE ECONOMIC AND STATISTICAL PROBLEM

2.1. Estimating Dynamic Labor Demand with Costs of Adjustment

Dynamic labor demand models been widely used in labor economics, macroeconomics, industrial organization, and other empirical fields. The extensive literature permits us to bring external evidence to bear on the reasonableness of parameter estimates (Hamermesh 1993, chapter 7). The models imply structural Euler equations that can be used as the basis for the moment conditions that allow parameter estimation without specifying a full likelihood function. Economists have been very reluctant to make the ancillary distributional assumptions that likelihood estimation requires (Hansen, 1982). Many longitudinal firm-based data sets have been developed with the specific goal of providing the information necessary to estimate dynamic factor demand models. (McGuckin and Pascoe, 1988; Hall and Mairesse, 1995; and Machin, Manning, and Meghir, 1993 provide examples from a variety of countries.)

Let L_{it} denote period t employment for firm i, K_{it} the capital stock, w_{it} the wage rate, p_{it} the price of the good produced, and $F(.,.)$ the production function. Consider the following intertemporal optimization program for the firm:

$$\max_{\{L_{it},L_{it+1},\ldots\}} E\left[\sum_{k=0}^{\infty}\delta^k(p_{it+k}F(K_{it+k},L_{it+k}) - w_{it+k}L_{it+k} - \frac{b}{2}(L_{it+k} - L_{it+k-1})^2)\right] \tag{2.1}$$

where δ is the discount factor ($0 < \delta < 1$), b is the marginal cost of adjusting one unit of employment ($b \geq 0$), and the expectation is taken over all periods $t+1, t+2, t+3, \ldots,$ conditional on information known at t

including the history of all relevant information. Solution yields the Euler equation

$$w_{it} = p_{it} \frac{\partial F(K_{it}, L_{it})}{\partial L_{it}} - b(L_{it} - L_{it-1}) + \delta b(L_{it+1} - L_{it}) + \varepsilon_{it} \tag{2.2}$$

at each date t, where ε_{it} is an *iid* shock. Stokey and Lucas (1989, chapter 9) provide a complete treatment of the conditions on $F(.,.)$ and the stochastic process governing (w_{it}, p_{it}) required to insure that equation (2.2) exists and is well-behaved. Larger values of the parameter b imply that employment will adjust more slowly to changes in the economic variables. Although there are many sources of bias in the estimation of equation (2.2), one important source, ignored in most empirical analyses, is the attrition of firms from the sample. In this article, we focus exclusively on this issue.

How could attrition affect the estimation of b in equation (2.2)? If the analysis is limited to a panel of sample survivors, then the marginal productivity of labor, $\partial F(K_{it}, L_{it})/\partial L_{it}$, may be larger for this group than in the entire population of active firms. Hence, the difference between the wage rate and marginal productivity may be smaller in the sample than in the entire population, inducing a downward bias in the estimation of b. While one could argue that firms that disappear from the sample have zero employment at the dates following their exit from the sample, there are good reasons to mistrust this assumption. First, some of the firms that leave the sample may have been acquired by other firms, either pre-existing or newly born. Such acquisitions are not restricted to the sale of less profitable firms. Second, the attrition from the sample may occur for reasons not related to end of economic entity that are difficult to quantify such as failure to file required forms, death of the owner, or errors in the treatment of the identification number. Finally, attrition may occur because of the cessation of economic activity by the enterprise, a legitimate death.

Next, we describe our efforts to understand the disappearance of firms from a sample that was constructed by the Institut National de la Statistique et des Etudes Economiques (INSEE), the French national statistical institute, which collects these data for a variety of reasons including the production of the French national income and product accounts. Our analysis shows that the assumption that employment is always zero when an entity exits from the sample is wrong. Consequently, we develop a method for modeling this attrition that does not depend upon arbitrary assumptions about the homogeneity of the causes of attrition.

2.2. The Structure of the INSEE Firm Sample

We use a longitudinal probability sample of firms called the Echantillon d'Entreprises (EE), 1978-1988, that has been routinely used by other researchers. (See Abowd, Kramarz, and Margolis, 1999, Crépon and Duguet, 1997, Hall and Mairesse, 1995). Our version, constructed by the Division des Etudes Economiques, includes 21,642 firms. The sampled entities are enterprises, which consist of one or more physical establishments engaged in related economic activity under common management. We use the terms enterprise and firm interchangeably. We describe the universe from which this sample was drawn as well as the other data sources used to investigate the reasons why firms disappear from the EE. The current EE, the 1993 version, contains data for 26,685 enterprises, sampled using the same techniques as we describe below. Table 1 shows the overall construction of the EE and the selection rules used to draw our analysis cohort.

The most inclusive source of French business data is the registry of establishments and enterprises called the Sirene. In this registry INSEE records the date of birth, date of death, and major events related to the entity's demography. There are approximately 2.2 million enterprises and 3.5 million establishments active in the registry in any given year (Chantereau and Rieu, 1995). An active entity, by definition, filed some kind of business tax return that year. The active enterprise count includes sole proprietors, with and without employees, who are issued an identifying number for tax purposes and recorded in the Sirene (*i.e.*, it is not possible to file an income tax return on business income using an individual taxpayer identification number).

The first reduction of the universe consists of selecting for the EE sampling frame only those enterprises included in the file constructed for the tax regime Bénéfice Réel Normal (BRN) by the Direction Générale des Impôts (DGI). There are approximately 600 thousand active enterprises in this file in any given year. A business is subject to the BRN tax regime if it is for-profit and meets the sales conditions stated in Table 1 (for tax years 1994 and earlier). Smaller firms may elect BRN regime taxation instead of one of the simpler tax forms but subsequent filters will eliminate those enterprises that opted into the BRN but did not meet the criteria shown in Table 1.

From the BRN file INSEE constructed an in-scope sampling frame, called the Universe Echantillon d'Entreprises (Universe EE) in Table 1, using all enterprises that meet *one* of the following criteria: (1) more than 20 employees, (2) more than 100 million francs of sales revenue, or (3) more than 200 million francs of total assets. The Universe EE also includes some firms that meet weaker criteria because they are included in one of the data sources that contributes information to the EE: (1) the exhaustive Bénéfices Industrials et Commerciaux-Impôts sur les Sociétés (exhaustive-BIC-IS), which is a file of all business tax returns (BRN and certain other regimes) provided to INSEE by the DGI (Chantereau and Rieu, 1995); (2) a sample from the BIC-IS (sample-BIC-IS), which is a sample of tax returns (exhaustive for larger enterprises) provided to INSEE by the DGI about nine months earlier than the exhaustive-BIC-IS in order to permit data checking and national accounts calculations to begin; or (3) the Enquête Annuelle d'Entreprises, which is an annual survey with exhaustive coverage for enterprises with at least 20 employees (INSEE, 1988). The EE is constructed from this universe using the sampling probabilities shown in Table 1. The 1988 version contained a total of 21,642 enterprises, which constitute about 11% of the active enterprises (about 11,000 firms) in any given year.

To construct a dynamically representative sample, INSEE began with a primary sample year, 1986. Every enterprise that was included in the 1986 Universe EE (*i.e.,* economically active in 1986) was at risk to be sampled with probabilities that depended upon the size of the enterprise in 1986 and the sector of economic activity. Firms with 500 or more employees were sampled with probability one. Firms with fewer than 21 employees were not sampled. Firms of intermediate sizes were sampled with probabilities between 1/30 and 1 according to the size and sector. All records in the Universe EE corresponding to a sampled firm for the years 1978-1988 were added to the EE file.

Every firm that was economically active in 1986 was then removed from the Universe EE. Complementary samples were then constructed for each of the other Universe EE files from 1978 to 1985 and from 1987 to 1988, 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 EE files in other years. In the resulting sample of firms, those with single-year gaps in their data have been retained and a missing data imputation procedure used to fill the gap. See Abowd, Crépon, Kramarz and Trognon 1995 for a more complete description of the data preparation. See Appendix Table A1 for summary statistics for our sample.

In what follows, we focus on all firms that appear in the EE sample for the first time in 1982, as shown in Table 1. These 667 firms are followed in the EE at most until 1988. None of these firms has any missing years of data. Once the firm enters the sample, economic data are available until the firm exits the sample or 1988, whichever comes first. Table 2 shows the distribution of surviving firms by year in the column labeled "Count." A firm is counted as an exit if it has no further EE activity after the exit year.

For firms in the Universe EE, more than 700 financial and operating variables are collected from a variety of administrative records and statistical surveys. As a part of its routine operations the DGI performs certain data validity checks that involve some of these variables. In addition, INSEE performs supplementary validity checks using tax and survey information that was collected from the multiple sources. The resulting INSEE file, now called the FUTE but unnamed during the period when our EE files were created, is unique in the French statistical system because of the quality and the breadth of information that it contains on the firms that satisfy at least one of the in-scope criteria. By contrast, information on smaller enterprises, those which fail to meet any of the criteria listed above, is potentially less accurate. The FUTE file, which is sometimes called the Bénéfices Industriels et Commerciaux (BIC) by the division of economic studies, is the sampling frame from which current enterprise samples are built and maintained (Fréchou and Topiol-Bensaïd, 1997).

2.3. Why Did Firms Disappear from the Sample?

Our exhaustive inquiry into the reasons why each of the 375 firms that exited from the 1982 cohort left the sample, using the sources noted in Table 1 and additional information on business restructuring is summarized in Tables 2 and 3. An economically active firm, using the French definition, would file some type of tax return as evidence of this activity. The largest tax return data base to which we have access is the BRN for each of the years from 1983 to 1994, the most recent year available at the time this was written. Our definition of "economically active" is that the enterprise was large enough so that it was required to file under the BRN tax regime or elected to do so. Table 2 summarizes the economic activity of the exiting firms by year of attrition. The column labeled "Never Appear in BRN after Exit" shows that 179 of the 375 exiting firms show no evidence of economic activity after the date of their attrition. The remaining 196 firms, show positive evidence of continued economic activity after their exit from the sample.

We divide this potentially active group into two parts. If the most recent year of economic activity (labeled "Most Recent Year to Appear in BRN") is between 1983-87, then the firm did cease economic activity before the sample ended in 1988, but the date of cessation of economic activity was misrecorded and some data are missing for the intervening years. For the 38 firms in this group there is some economic information available between the last recorded sample data and last year of the EE, 1988. In principle, a researcher might make use of these data in an economic model but because these firms are not in the Universe EE during this period, very few of the economic variables are available after attrition. The second category of economically active exits is the group for which there is evidence of continued operations after exit from the sample and after 1988. For 70 firms in this group (most recent BRN years from 1988 to 1993), the error in assuming that attrition is equivalent to cessation of economic activity is two-fold. First, the date of cessation has clearly been recorded incorrectly and is later than the date of exit from the sample. Second, the firm should have been recorded as still economically active in 1988, when our version of the EE stopped. Finally, for 88 firms (most recent BRN year 1994) the attrition error is qualitatively different. Because the most recent year of BRN data for these firms is also the most recent year available (as of the time this paper was written), the evidence strongly suggests that these firms are still economically active. The best hypothesis is that they slipped below one of the thresholds shown in Table 1 for inclusion in the Universe EE and, therefore, INSEE was unable to collect the economic data on these firms. So, for this group of firms, the attrition indicates that the data are missing but does not indicate economic death.

The exiting firms could be registered as "dead" in the Sirene, independent of the evidence that we found in Table 2. Table 3 addresses this issue. We consider the evidence from the Sirene, which records the death or transfer of assets of a firm (date de cessation d'activité), and from the file Modification de Structure (MDST; Chantereau and Rieu, 1995), which records the mandatory asset transfer declarations filed with the Centre de Formalités des Entreprises (CFE). Still, we are unable to characterize the status of the assets of many exiting firms. The end of economic activity can be recorded as a result of one of the following events: (1) a legally mandated filing to the CFE that declares an end to taxable activity and/or transfer of all assets to another entity (also recorded with the successor entity); or (2) an investigation triggered by the failure to file a tax return, employee declaration, or other legally mandated report for two years. When the firm files the required form, the information is recorded in the Sirene. Otherwise, INSEE tries to estimate the date of cessation of economic activity by using the documents that it has. Based on a follow-up of a subset of these firms, INSEE believes that 99% of the deaths recorded through the follow-up investigation are true cessations of economic activity (Francoz, 1996). Prior to 1989, the investigation for determining the death of the entity was conducted using much less information than is currently available (Francoz, 1996). Even when the determination of death is accurate for the firms so recorded, there remains the problem of determining the disposition of assets for firms whose status is unknown. To address this problem we use the MDST.

Table 3 excludes those firms whose last reported economic activity is in 1993 or 1994 because they are very likely still active, given that 1994 was the most recent year of the BRN available to us. In our discussion of Table 3 we consider two groups of exiting firms, based on the information shown in Table 2. Of the 217 firms that we believe ended their economic activity before 1988, 112 have a death record in the Sirene–105 with no transfer of assets and 7 with transfer through merger. For 103 of these firms there is no record of BRN filing. They may be inactive but the information required to certify death is not available. For 2 of the remaining firms in the first group a merger was recorded. For 40 of the 54 exiting firms for which there is evidence of continued economic activity in the BRN between 1988 and 1992 the Sirene and MDST provide no additional death or asset transfer information. Twelve of these 54 firms are recorded as dead and, among these, one transferred its assets to another entity.

Our analysis of the cohort of 667 EE firms born in 1982 and followed by the EE until 1988 shows that attrition has a multiplicity of causes, most of which are not captured by any meaningful economic model. The information that we used to classify the attrition by reason is not usually available to researchers using similarly constructed samples. Such supplementary information on exiting firms can only be used at INSEE (or, in general, at national statistical agencies under appropriate confidentiality controls) at a great additional cost since tracking down the attrition involves using many other data sources within the agency. Therefore, our strategy for the remainder of this paper is to develop statistical procedures that do not rely on modeling the reason for the attrition of the firm. Our procedures allow for the multiplicity of true reasons for attrition by using the available information on exiting firms and survivors in new ways.

3. ATTRITION PROCESSES

We consider a population of N entities following the process $\underline{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$, with y_{it} a $K \times 1$ vector, where $i = 1$ to N, denotes firms, t denotes time, and $t = 1$ is the date of birth, common to all entities and known. Here, y_{it} is a vector that includes both dependent and explanatory variables of the process. Denote the conditions that relate these variables as $Eg(\underline{y}_i, \theta) = 0$. Denoting $\underline{y}_{it} = (y_{i1}, \ldots, y_{it})$, we have $Eg_t(\underline{y}_{it}, \theta) = 0$ for $t = 1, \ldots, T$, 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 directly.

Entities and all of their variables are observed from date $t = 1$ to 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. 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

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

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

The initial conditions are $s_{i1} = \pi_{i1} = 1$ for all i. Notice that the moment conditions may have no direct empirical counterpart since data are observed at date t if and only if entity i survived until that date. Direct estimation of equations (3.1) and (3.2), based only on observed data, may lead to biased estimates because of the sample attrition.

Following Rubin (1976) 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 (for notational simplicity we do not show the parameters except in our moment conditions) missing at random means

$$l(\underline{y}_i, s_{it} \mid \underline{y}_{it-1}) = l(\underline{y}_i \mid \underline{y}_{it-1}) l(s_{it} \mid \underline{y}_{it-1}) \tag{3.3}$$

or equivalently

$$P(s_{it} = 1 \mid \underline{y}_i) = P(s_{it} = 1 \mid \underline{y}_{it-1}), \tag{3.4}$$

which implies that $l(y_{it} \mid \underline{y}_{it-1}, s_{it} = 1) = l(y_{it} \mid \underline{y}_{it-1})$, $l(\underline{y}_i) = l(y_{iT} \mid \underline{y}_{iT-1}, s_{iT} = 1) l(y_{iT-1} \mid \underline{y}_{iT-2}, s_{iT-1} = 1) \cdots l(y_{i1}, s_{i1} = 1)$, and $l(\underline{y}_i) = l(y_{iT} \mid \underline{y}_{iT-1}) l(y_{iT-1} \mid \underline{y}_{iT-2}) \cdots l(y_{i1})$.We apply this result directly to our moment conditions rather than to the densities themselves because economists usually do not specify the full probability model for $\underline{y}_i$ and because our empirical analysis of the causes of attrition indicates that there is heterogeneity in the reasons for attrition. Moment conditions provide tools that do not require extensive parametric modeling of the attrition process.

Proposition 3.1. *Under the Rubin attrition rule (3.3), the following two equations hold*

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

$$\pi_{it} = q_{it} \pi_{it-1} \tag{3.6}$$

Proof: Given the independence assumption of equation (3.3), we have $E(g_t(\underline{y}_{it}, \theta) s_{it} \mid \underline{y}_{it-1}) = E(g_t(\underline{y}_{it}, \theta) \mid \underline{y}_{it-1}) E(s_{it} \mid \underline{y}_{it-1})$. Notice that $E(g_t(\underline{y}_{it}, \theta) s_{it} \mid y_{it-1})$ has a sample analog, since $s_{it} = 0$ implies that 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 directly on s_{it} because some of the y_{it-1} are not observed. But, this is possible conditional on $s_{it-1} = 1$, which permits observation of q_{it}. More precisely, we have the relation $E(s_{it} \mid \underline{y}_{it-1}) = E(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 1) P(s_{it-1} = 1 \mid \underline{y}_{it-1}) + E(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 0) P(s_{it-1} = 0 \mid \underline{y}_{it-1})$ Given that $E(s_{it} \mid \underline{y}_{it-1}, s_{it-1} = 0) = 0$, $P(s_{it-1} = 1 \mid \underline{y}_{it-1}) = P(s_{it-1} = 1 \mid \underline{y}_{it-2}) = \pi_{t-1}$, and equation (3.3), we have $\pi_{it} = q_{it} \pi_{it-1}$. Thus, given

$\pi_{i1} = q_{i1}$, we have $\pi_{it} = q_{it} \cdots q_{i1}$ and it follows that π_{it} is identifiable. Q.E.D.

Equations (3.5) and (3.6) have several implications. Consider the case of GMM estimation. There is a set of orthogonality conditions, satisfied on the whole population, given by the moment equations. Decompose the function g into $g = (g_1, g_2, \ldots, g_T)$ with g_t a function of $(y_{i1}, \ldots, y_{it})$. The equation (3.5) can be applied to each g_t for $t = 1, ..., T$. The orthogonality conditions $E\left(g_t(\underline{y}_{it}, \theta) s_{it}/\pi_{it}\right) = 0$ have direct empirical counterparts and can be estimated as a function of the sampled data in the presence of attrition. For extensions of this framework to other cases, see Abowd, Crépon, Kramarz, and Trognon (1995) and Abowd, Crépon, and Kramarz (1997).

4. IMPLEMENTATION OF THE WGMM METHOD

We now discuss our procedure for implementing the attrition rule (3.4). Assume that the nuisance parameters, q_{it}, have a logistic regression form. (For the non-parametric case, see Newey, 1994.) We can estimate a set of parameters $\{\beta_t\}_{t=1,...,T}$ defined by the orthogonality conditions on $(\psi_t)_{t=1,...,T}$, the derivatives of the logistic regression functions with respect to β_t, $E\left(\psi_t(\beta_t, \underline{y}_{it})\right) = 0$. The complete set of conditions is

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

Standard errors were computed using the corresponding formulas for the GMM estimation of the equations (4.1). For the computation of the parameters themselves, it is simpler to solve the equations that define the β parameters first and then to replace them by their estimated values for the solution of the parameter of interest θ (Crépon, Kramarz and Trognon, 1998). There is 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 coefficients are defined by maximization of the logistic regression likelihood, (Crépon *et al.*, 1998; Newey, 1994; Newey and McFadden, 1994).

5. APPLICATION TO THE LABOR DEMAND PROBLEM

The weights for the WGMM estimator were modeled using logistic regressions. We used several sets of attrition-correcting weights based on variables in the labor demand equation, their lagged values, and variables that are meant to better capture the factors affecting the attrition of the firms. We describe these models and the results but we do not present the logistic regression coefficients in tabular form since our main focus is the labor demand equation.

The first logistic regression model predicted attrition based upon the wage rate, employment, the ratio of production to employment (Q/L), and the interactions $(Q/L)\log L$ and $Q/L\log K$ and all possible lags. The last three variables arise from the choice of a translog production function (see below) to specify the labor demand equation. The second attrition model used the same variables as the first but also included an indicator variable for presence in the BRN following attrition. The third model was identical to the first but allowed for different slope coefficients on (Q/L), $(Q/L)\log L$, and $Q/L\log K$ when these variables were below the first quartile or above the third quartile. The fourth model used the same variables as the first with lags of order 4 and 5 excluded. Finally, the fifth model added three financial variables–return on assets, financial return on assets, and debt ratio–to the variables used in the first regression. Summary statistics and definitions for all variables are shown in Table A.1.

We conclude that reweighting to account for attrition is necessary. Statistical tests confirm that y_{t-1} and z_{t-1} must always be included in the logistic regressions, for each of the z variables considered in the attrition models outlined above. The estimated attrition probabilities also show that all variables matter at some date and for some lag back to 1982, the date of birth of all our firms. Thus, attrition models that ignore information since the date of birth may be misspecified. Both economic and financial information are important–when both types of variables are included, at each lag, at least one real and one financial variable are significantly different from zero. When the indicator functions for very dispersed variables are introduced, the results show that firms at both ends of the spectrum display similar attrition probabilities. Hence, firms may exit for opposite reasons, as we showed in our audit. Finally, when we include an indicator for presence in the BRN after attrition, strong evidence that the firm is still economically active after attrition, our results are essentially the same as when this variable is excluded. This is important since analysts would not usually have access to such information.

The labor demand relation, equation (2.2), uses a translog production function, $F(L, K)$. The equation is estimated in first differences to allow for the possibility of a firm effect in the error term. Instrumental variables include all lags of $L, K, Q/L, (Q/L) - w, (Q/L)\log L$, and $(Q/L)\log K$. Table 4 contains the estimation results. Column (1) reports the estimates using the first logistic regression discussed above to estimate the attrition probabilities. Column (2) contains the estimates when we model attrition using the additional information from the BRN. Columns (3)-(5) use the other variations discussed above. Column (6) reports unweighted estimates that result from the direct estimation of equation (2.2) without any correction for attrition. Finally, column (7) presents estimates for the balanced panel of firms that are born in 1982 and remain in the sample through 1988.

The row corresponding to the variable $[(L_{t+1} - L_t)\delta - (L_t - L_{t-1})]$ contains the estimates of b, the adjustment cost parameter. The following rows give estimates of the components of the translog production function. The Sargan statistics (Hansen 1982) for the overidentifying restrictions, the number of degrees of freedom, and the probability value are given in the following rows. Rows labeled "Test: H_w vs. $H_{nw} = H_0$ $(\chi^2(1))$" and "Test: H_{nw} vs. $H_w = H_0$ $(\chi^2(1))$" give the results of the non-nested tests (Smith, 1992) for the weighted model against the null hypothesis of a non-weighted model (resp., the non-weighted model against the null hypothesis of a weighted model). The last rows provide descriptive statistics for the distribution of the estimated elasticities of production for labor, $\partial \log Q/\partial \log L$.

Attrition corrections have an important effect on the estimation of the key parameters of the labor demand equation. In columns (1)-(5), where we have corrected for attrition biases, the estimated adjustment cost is between 80,000FF (\$16,000 US) and 100,000FF (\$20,000 US) per worker, while in columns (6) and (7), where there is no attrition bias correction, the estimated adjustment costs are much lower (20,000FF $\approx$ \$4,000 US). The results in column (2), which are based upon the post-sample BRN data, are fully consistent with those in columns (1) and (3)-(5). Having access to evidence of post-sample economic activity does not substantially change the attrition correction. Use of the BRN post-attrition status fully supports the conclusion that attrition correction based on the pre-attrition sample data is appropriate.

Adjustment costs include hiring and firing costs, required severance payments, and reorganization costs. Using data for 1992, six years after our analysis period ends, Abowd and Kramarz (2001) provide estimates of components of these costs based on direct reports from French firms. The reported costs in that paper do not include reorganization costs whereas the estimates of equation (2.2) do include such costs. Table 5 summarizes the adjustment costs directly measured by that survey and supports our conclusion that the estimated adjustment costs in Table 4 columns (6) and (7) are too low to be credible given French separation laws (Abowd and Kramarz, 2001).

The estimated elasticities of production are implausibly high in the uncorrected estimates with a large number of firms having elasticities greater than 1, a theoretical impossibility for profit-maximizing firms. In particular, one expects the elasticity of production for labor to be close to 0.6, the approximate share of labor in total costs (Hamermesh, 1993, chapter 7). The distribution of elasticities of production is very concentrated around the median, showing that the attrition bias corrected parameter estimates imply relatively constant estimated elasticities.

The test of overidentifying restrictions is a common specification check that measures the quality of the instruments. In all columns, the overidentifying restrictions are accepted at conventional significance levels. Hence, we do not learn anything about the relative quality of the instruments from these tests. While these tests can sometimes be informative for the choice of instruments, they do not test the quality of the attrition model. As shown in Hirano *et al.* (1997), there is no test of the restrictions associated with the Rubin attrition rule because it is a just-identified model.

Since the overidentifying restriction tests were not conclusive, we turned to non-nested tests of the weighted models against the non-weighted model (Smith, 1992) that are designed for two competing models estimated by GMM, each model being based on a set of orthogonality conditions. Our non-nested tests are also inconclusive: we cannot reject the non-weighted model under the null of a weighted model and, simultaneously, we cannot reject the weighted model under the null of the non-weighted model.

Finally, we examine the Verbeek and Nijman (1992 and 1996) suggestion that if balanced estimates are close to the unbalanced estimates, then attrition should not matter. This suggestion is refuted by the results for our attrition models. Inspection of the estimated costs of adjustment and elasticities of production in the last two columns, which are not corrected for attrition, shows that although the balanced and unbalanced results are close to each other, neither resembles those that have been corrected for attrition. We have already shown that the attrition-corrected results make better economic sense.

6. CONCLUSION

For a longitudinal sample of French firms, sample attrition is a very complex and heterogeneous process that does not imply that the firm has become economically inactive. For at least half of the exiting firms an audit strongly suggests that output and employment were positive for at least some years after the firm left the sample through attrition. These audit results, based on an elaborate follow-up of every exiting firm, lead us to propose a simple way to account for attrition without a formal economic model. We use a moment-based approach, rather than a likelihood approach, because we are reluctant to put additional structure on the attrition process. Our approach relies on Rubin's missing at random assumption adapted to a longitudinal data framework. We show how to implement this method and demonstrate empirically that attrition matters, using a model of dynamic labor demand. In our application we used firm data but our methodology could be applied to correct for heterogeneous attrition within a panel of individuals or in dynamic samples of other economic entities. The methods that we propose allow researchers to use longitudinal data with substantial attrition without directly modeling all of the economic and statistical reasons for that attrition. Our methods also encourage analysts to make use of all available data, rather than just using certain subsets of the variables or observations. The combination of these two recommendations should improve the quality of the research based upon such samples.

References

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

[2] Abowd, J.M., B. Crépon, and F. Kramarz (1997), "Moment Exstimation with Attrition, " NBER technical working paper 214.

[3] Abowd, J.M., F. Kramarz, and D.N. Margolis (1999), "High Wage Workers and High Wage Firms," *Econometrica,* 67, 251-333.

[4] Abowd, J.M. and F. Kramarz (2001), "The Costs of Hiring and Separation," Cornell University working paper.

[5] Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

[6] Chantereau, J. and C. Rieu (1995), "Guide d'Utilisation des Fichiers SUSE III," INSEE internal publication, 262/E220.

[7] Crépon, B. and E. Duguet (1997), "Research and Development, Competition and Innovation: Pseudo Maximum Likelihood and Simulated Maximum Likelihood Methods Applied to Count Data Models with Heterogeneity," *Journal of Econometrics,* 79, 355-378.

[8] Crépon, B., Kramarz, F., and A. Trognon (1998), "Parameters of Interest and Nuisance Parameters: an Application to Autoregressive Error Component Models," *Journal of Econometrics,* 82, 135-156.

[9] Francoz, D. (1996), "Les cessations d'entreprises depuis 1989," *INSEE Première* (June) 1-4.

[10] Fréchou, H. and A. Topiol-Bensaïd, A. (1997), "Panorama des Fichiers de la Division Marchés et Stratégies d'Entreprises," INSEE internal note, 75 (1995-Revised 1997)/G231.

[11] Hall, B.H. and J. Mairesse (1995), "Exploring the Relationship between R&D and Productivity in French Manufacturing Firms," *Journal of Econometrics,* 65, 263-293.

[12] Hamermesh, D (1993), *Labor Demand,* Princeton, The Princeton University Press.

[13] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029-1054.

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

[15] Hellerstein, J.K. and G.W. Imbens (1999), "Imposing Moment Restrictions from Auxiliary Data by Weighting," *Review of Economics and Statistics*, 81, 1-14.

[16] Hirano, K., Imbens, G. W., Ridder, G. and D. Rubin (1997), "Combining Panel Data Sets with Attrition and Refreshment Samples," NBER technical working paper no. 230.

[17] INSEE, "Dictionaire des Sources Statistiques INSEE Paris" (1988), *Collection: Archives et Documents*, 3 volumes, 0249-6186, 258.

[18] Machin, S., A. Manning, and C. Meghir (1993), "Dynamic Models of Employment based on Firm-Level Panel Data," in J.C. van Ours, G.A. Pfann and G. Ridder (eds.), *Labor Demand and Equilibrium Wage Formation*, Amsterdam: Elsevier Science Publishers.

[19] McGuckin, R. and G. Pascoe (1998), "The Longitudinal Research Database (LRD): Status and Research Possibilities," *Survey of Current Business*, 30-37.

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

[21] Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, 4, Amsterdam: North Holland Press, 2111-2245.

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

[23] 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.

[24] 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.

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

[26] Smith, R.J. (1992), "Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica*, 60, 973-980.

[27] Stokey, N. and R.E. Lucas, Jr. (1989), *Recursive Methods in Economics Dynamics*, Cambridge: Harvard University Press.

[28] Verbeek, M. and T.E. Nijman (1992) "Testing for Selectivity Bias in Panel Data Models," *International Economic Review*, 33, 681-703.

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

Table 1
Construction of the Echantillon d'Enterprise from Underlying Firm Data Sources

Source	***Number of Enterprises***	***Definition or Criterion***
Sirene	2,200,000	Registry, estimate of the number active in a given year
Régime du Bénéfice Réel Normal (BRN)	600,000	For-profit economic entities, sales>3.5M FF, nonservice, or sales>1M FF, services; estimate of the number active in a given year
Universe Echantillon d'Entreprises (Universe EE)	100,000	BRN + (employees>20 or sales>100M FF or assets>200M FF)[exhaustive BIC-IS] or (employees>10 or sales>3.5M or assets>5M FF)[sample-BIC-IS] or (20<employees)[EAE], in a given year
Echantillion d'Entreprises (EE)	21,642	Sampling rates: employees>499 1/1; 20<employees<500 between 1/30 and 1/1; employees<21 not sampled
EE born 1982	667	Never appeared in Universe EE before 1982
Exit 1982-1987	375	Last year in Universe EE < 1988
Still present 1988	292	Still present in Universe EE in 1988

Sources: Chantereau and Rieu (1995) Frechou and Topiol Bensaid (1997).
Notes: BIC-IS means Bénéfices Industriels et Commerciaux-Impôts sur les Sociétés; EAE means Enquête Annuelle d'Entreprises; M=million.

Table 2
Analysis of the Most Recent Economic Activity of Enterprises
That Appear in the Echantillon d'Entreprise 1982 Cohort

Exit Year	***Count***	***Never Appear in BRN after Exit***	***Appear in BRN at Least Once after***	***Most Recent Year to Appear in BRN***											
				1983	***1984***	***1985***	***1986***	***1987***	***1988***	***1989***	***1990***	***1991***	***1992***	***1993***	***1994***
1982	139	62	77	7	6	4	1	5	2	4	4	5	7	5	27
1983	76	44	32	0	0	1	4	1	3	1	4	1	2	3	12
1984	54	20	34	0	0	5	1	2	0	2	0	1	2	4	17
1985	28	20	8	0	0	0	1	0	0	1	1	1	1	0	3
1986	28	18	10	0	0	0	0	0	0	0	0	1	1	2	6
1987	50	15	35	0	0	0	0	0	0	4	0	2	4	2	23
1988	292	na	na	0	0	0	0	0	na	na	na	na	na	na	na
Total	667	179	196	7	6	10	7	8	5	12	9	11	17	16	88

Sources: Authors' calculations using Echantillon d'Entreprises and Bénéfice Réel Normal (BRN). Note: The last available year of data in the BRN was 1994 at the time this table was created. Firms listed as exits in 1988 were censored by the end of the sample.

Table 3
Detailed Resolution of Reasons for Attrition for Enterprises That Exit the Echantillon d'Entreprises 1982-1987

Description of resolution for the attrition of these enterprises			***Number of enterprises***
Never appear in the BRN from 1988-1994	No death code found in Sirene	No recorded merger	103
		Merger recorded	2
	Officially listed as "dead" in Sirene	No recorded merger	105
		Merger recorded	7
Appear at least once in BRN from 1988-1992 (see note)	No death code found in Sirene	No recorded merger	40
		Merger recorded	2
	Officially listed as "dead" in Sirene	No recorded merger	11
		Merger recorded	1
Atttritions from 1982 cohort in Echantillon d'Entreprises			271

Sources: Authors' calculations using the Echantillon d'Entreprises, Bénéfice Réel Normal (BRN), Sirene, and Modifications de Structure (for mergers). Note: 104 cases that appear in the BRN in 1993 and 1994 are not counted because additional merger activity could have occured in subsequent years.

Table 4
Labor Demand Model (Weighted GMM Estimation)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Independent Variables:								
$[(L_{t+1} - L_t)\,\delta - (L_t - L_{t-1})]$		78.840	82.854	116.416	121.101	76.573	19.912	16.452
		(2.254)	(3.428)	(6.482)	(6.393)	(2.273)	(7.625)	(9.750)
Q_t/L_t		0.832	0.654	0.871	0.799	1.101	1.797	2.443
		(0.058)	(0.072)	(0.127)	(0.117)	(0.064)	(0.157)	(0.205)
(Q_t/L_t) log L_t		0.077	0.064	0.077	0.068	0.093	0.168	0.220
		(0.004)	(0.005)	(0.009)	(0.009)	(0.004)	(0.012)	(0.017)
(Q_t/L_t) log K_t		-0.079	-0.059	-0.096	-0.092	-0.120	-0.168	-0.235
		(0.008)	(0.009)	(0.016)	(0.015)	(0.009)	(0.020)	(0.024)
Test of overidentifying restrictions: χ^2		19.70	24.58	16.71	17.88	19.05	11.72	15.18
Degrees of freedom		16	16	16	16	16	16	16
P-value		0.234	0.078	0.404	0.331	0.266	0.763	0.511
Test: H_{nw} vs $H_w = H_0$ ($\chi^2(1)$)		12.12	8.70	11.15	11.97	11.23	-	10.58
P-value		0.001	0.003	0.001	0.001	0.001	-	0.001
Test: H_w vs $H_{nw} = H_0$ ($\chi^2(1)$)		8.50	9.02	6.40	6.32	8.07	-	6.51
P-value		0.004	0.003	0.011	0.012	0.004	-	0.011
Elasticity of Production:	Median	0.45	0.351	0.457	0.42	0.595	0.963	1.32
	99th percentile	0.64	0.496	0.692	0.644	0.888	1.379	1.906
	75th percentile	0.497	0.387	0.519	0.48	0.673	1.076	1.48
	25th percentile	0.393	0.311	0.383	0.349	0.503	0.858	1.17
	1st percentile	0.19	0.161	0.123	0.094	0.171	0.429	0.569
Number of firms		667	667	667	667	667	667	292

Source: Authors' calculations based on the Echantillon d'Entreprises, cohort entering 1982.

Notes: The dependent variable in all columns is the wage rate at date t: w_t. All regressions use lagged endogenous and exogenous variables as instruments. In column (1), the logistic regression of the attrition weights uses employment, wage, output/employment (Q/L), (Q/L) log L, (Q/L) log K, and all possilble lags of order 1 to 5. In column (2), the logistic regression uses the same variables as in (1) plus an indicator for being in the BRN data set after attrition from the original sample. In column (3), the logistic regression includes the same variables as in (1) plus specific slope coeficients for (Q/L), (Q/L) log L, and (Q/L) log K when these variables had values below the first quartile or above the third quartile. Essentially identical results were obtained using the first and ninth deciles. In column (4), we use the same variables as in column (1) with the lag of order 4 and 5 excluded. In column (5), the logistic regression uses employment, wage, output, capital, value-added, total debt, and operating profit. The model in column (6) is not corrected for attrition. In column (7), we use a balanced sample. Standard errors are in parentheses.

Table 5
Distribution of Separation Costs per Worker

Statistic	***Retirement or Pre-retirement***	***Layoff or Firing***	***Total Separations***
Mean	84.102	101.952	79.776
90th percentile	169.418	218.237	168.591
75th percentile	78.034	98.614	82.124
Median	38.249	37.019	36.833
25th percentile	19.647	13.477	14.722
10th percentile	10.740	5.707	6.244
Number of firms	1,418	1,976	2,697

Source: Enquête sur la Structure des Salaires et des Coûts, 1992.
Notes: All costs are shown in thousands of 1992 French francs.
The original sample contains 7,903 enterprises. Statistics in
this table are based on firms with positive separations.

Table A.1
Summary Statistics for Firms born in 1982

Variable	***Mean***	***Standard Deviation***
Full-time equivalent employment	184.5	843.6
Log total labor cost per employee (log FF 1980)	4.340	0.419
Sales (millions FF 1980)	88.440	329.14
Value-added (millions FF 1980)	18.630	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 returns - financial charges) / total assets	-0.464	21.670
Long and mid-term debt / total assets	2.939	50.610
Long-term assets / total assets	2.224	17.430
Financial charges/ total assets	0.030	10.190
Number of observations	3,049	

Source: Authors' calculations based on the 667 firms in Echantillon d'Entreprises.