Endogenous Attrition in Panels*

Laurent Davezies† Xavier D'Haultfœuille‡

First version February 2011 - This version April 2011

Abstract

We consider endogenous attrition in panels where the probability of attrition may depend on current and past outcomes, thus generalizing Rubin (1976) and Hausman & Wise (1979). We show that this probability is nonparametrically identified provided that instruments which affects the outcomes but not directly attrition, and whose distribution is identified, are available. We thus complement Hirano et al. (2001)'s framework, which does not rely on such instruments. Contrary to their approach, neither a refreshment sample nor an additive decomposition on the probability of attrition are needed. Inference is quite straightforward when the outcome and instruments are discrete, but involves a two-step procedure with a first step nonparametric estimation of an ill-posed inverse problem in the continuous case. Finally, we apply our results to the French labor force survey, and find strong evidence of attrition related to employment status. Our results also indicate that attrition depends on transitions on employment status in a way that violates the additive restriction used by Hirano et al. (2001).

Keywords: panel data, endogenous attrition, instrumental variable.

JEL classification numbers: C14, C21 and C25.

1 Introduction

Panel data are very useful to distinguish between state dependence and unobserved heterogeneity (see, e.g., Heckman, 2001), to analyze the dynamics of variables such as income

*We would like to thank Andres Santos and participants of the CREST-LMI seminar and Cemmap workshop "Recent developments in nonparametric instrumental variable methods" for their useful comments.

†CREST, laurent.davezies@ensae.fr

‡CREST-INSEE, xavier.dhaultfoeuille@ensae.fr

(see, e.g., Hall & Mishkin, 1982) or spells in duration analysis (see, e.g., Lancaster, 1990). However, these advantages may be counterbalanced by attrition, which can be especially severe when units are observed over a long period of time. Besides, attrition is often considered more problematic than standard nonresponse, because the reasons of attrition are often related to the outcomes of interest, or variations in these outcomes. Several solutions have been considered in the literature to handle these issues. A first model is to suppose that attrition is exogenous, i.e. depends on lagged values that are observed by the econometrician (see, e.g., Little & Rubin, 1987). This, however, rules out a dependence between attrition and current outcomes, and is thus likely to fail in many cases. A second model takes the opposite point of view by assuming attrition to depend on contemporaneous values only (see Hausman & Wise, 1979). Such an assumption fails to hold if attrition is related to transitions in outcomes. In the French labor force survey, for instance, households that move during the period are lost. But house moving is likely to be related to changes in the employment status. To handle more complex attrition patterns, Hirano et al. (2001) generalize the two previous models by allowing attrition to depend both on contemporaneous and lagged values. This generalization is made possible when a refreshment sample, i.e. a sample of new units surveyed at each period, is available.

In this paper, we consider still another approach, based on instruments. Contrary to Hirano et al. (2001), we do not impose any functional restrictions on the probability of attrition conditional on lagged and contemporaneous values. Refreshment sample are not needed either. On the other hand, we suppose to have in hand an instrument which is independent of attrition conditional on past and contemporaneous outcomes. A rank condition between the instrument and the contemporaneous outcome, which can be stated in terms of completeness, is also needed. Hence, the instrument is typically a lagged variable which affects the contemporaneous outcome but not directly attrition. We can use for instance past outcomes obtained from a retrospective questionnaire. We show indeed that under a nonlinear fixed effect model, such a variable is likely to meet the nonparametric rank condition, and satisfies also the conditional independence condition if attrition only depends on transitions on the outcome.

An advantage of our method is that even if no more instruments than outcomes are available, we can test for the conditional independence assumption. Another way of testing this assumption is to use refreshment samples, even though they are unnecessary in our setting. Indeed, the marginal distribution of the contemporaneous outcome is directly identified with such samples. Thus, we can compare this distribution with the one obtained under our identifying restriction.

We also conduct inference under such an attrition process. In the case of discrete outcomes and instruments, the model is parametric and a straightforward constrained maximum likelihood estimation procedure is proposed. In the continuous case, the model is semi-parametric and estimation is more involved. We first provide a necessary and sufficient condition for root-n estimability of linear functionals, and compute under this condition the asymptotic efficiency bound. Our results are close to those obtained recently by Severini & Tripathi (2011) in the case of nonparametric instrumental regressions. Second, we propose two estimation methods to estimate such linear functionals. The first is efficient but relies on rather restrictive conditions, whereas the second is not efficient but consistent under mild assumptions.

Finally, we apply our results to study transitions on the French labor market, using the labor force survey of the French national institute of statistics (INSEE). This survey is one of the most important survey conducted by INSEE, but its reliability has been much questioned inside the institute by the end of 2006 and the beginning of 2007 (i.e., during the French presidential elections campaign), as the discrepancy between the INSEE unemployment rate estimate and the one coming from administrative data started to increase. We reinvestigate this issue by studying the nature of attrition in this survey. Using the refreshment sample, we test and accept on the subsample of women the conditional independence assumption with past employment status used as an instrument. Our estimates indicate that attrition is highly related to transitions in the labor market, in a way that violated the additive restriction considered by Hirano et al. (2001). We show that this has important implications for the estimation of the probabilities of transition on the labor market.

The paper is organized as follows. In the second section, we study identification and testability under endogenous attrition, and compare our model with the existing literature. In the third section, we develop inference for both discrete or continuous outcomes. The fourth section is devoted to our application. Finally, the fifth section concludes. All proofs are gathered in the appendix.

2 Identification

2.1 The setting and main result

For simplicity, we consider a panel dataset with two dates $t = 1, 2$, and also suppose that there is no or ignorable nonresponse at date 1. We let $D = 1$ if the unit is observed at date 2, $D = 0$ otherwise. We let Y_t denote the outcome at t and $Y = (Y_1, Y_2)$. We also consider an instrument Z_1 whose role will be explained below, and let $Z = (Y_1, Z_1)$. We focus hereafter on the identification of either the joint distribution of (D, Y, Z) or on a parameter $\beta_0 = E(g(Y, Z))$. Our first assumption states the observational problem.

Assumption 1 *We observe* (D, Z) *and* Y_2 *when* $D = 1$.

To satisfy this requirement, Z_1 should be observed at the first period, or at the second period if some information on nonrespondents at the second period is available.[1] Of course, to achieve full identification of (D, Y, Z), restrictions are needed on the distribution of (D, Y, Z). If attrition directly depends on the outcome Y, the usual assumption of exogenous selection fails, and it may be difficult to find an instrument that affects the selection variable but not the outcome. On the other hand, a variable Z_1 related to Y but not directly to D may be available in this case. We thus assume the following:[2]

Assumption 2 $D \perp\!\!\!\perp Z_1 | Y$.

This assumption is identical to the one considered by D'Haultfœuille (2010) in the case of endogenous selection. This assumption was also considered by Chen (2001), Tang et al. (2003) and Ramalho & Smith (2011) in a nonresponse framework. Intuitively, this assumption states that the attrition equation depends on Y_1 and Y_2 but not on Z_1. If Y_2 was endogenous (but always observed) in this equation, we could instrument it by Z_1 to identify the causal effect of Y_2 on D. Here the problem is actually slightly different: Y_2 is observed only when $D = 1$. The identification strategy will be similar, however, as we will use the instrument to recover the conditional distribution of attrition.

Let $P(Y) = P(D = 1|Y)$. Because identification is based on inverse probability weighted moment conditions, we assume the following:

[1] Actually, our results would hold if Z_1 is observed only when $D = 1$, provided that the distribution of (Y_1, Z) is identified (through another dataset for instance).

[2] Assumption 2 could be extended to include covariates, i.e. $D \perp\!\!\!\perp Z_1 | Y, X$, provided that X is always observed. We do not include them for the sake of simplicity.

Assumption 3 *(i)* $P(Y) > 0$ *almost surely or (ii)* $P(Y) \geq c > 0$ *almost surely.*

This assumption is similar to the common support condition in the treatment effects literature. It does not hold if D is a deterministic function of Y, as in simple truncation models where $D = \mathbb{1}\{g(Y) \geq y_0\}$, y_0 denoting a fixed threshold.

Before stating our main result, let us introduce some notations. For any random variable U and $p > 0$, let $L_p(U)$ denote the space of functions q satisfying $E(|q(U)|^p) < +\infty$. For any set $A \subset L_2(U)$, let also $A^\perp = \{g \in L_2(U) : \forall a \in L_2(U),\ E(g(U)a(U)|D = 1) = 0\}$. The following operator will be important for identification issues:

$$\begin{array}{rlll} T: & L_1(Y) & \to & L_1(Z) \\ & q & \mapsto & (z \mapsto E(q(Y)|D = 1, Z = z)) \, . \end{array}$$

Because Y is observed when $D = 1$, T is identified. Besids, and as indicated previously, identification hinges upon dependence conditions between Y_2 and Z, which are actually related to the null space Ker (T) of T. Finally, let $\mathcal{F} = \{q \in L_1(Y) : q(Y) \geq 1 - 1/P(Y) \text{ a.s.}\}$ and $\beta(Y) = E[g(Y,Z)|Y]$. Our main result is the following.

Theorem 2.1 *If assumptions 1-3(i) hold, then:*

1. *The distribution of* (D,Y,Z) *is identified if and only if Ker* $(T) \cap \mathcal{F} = \{0\}$.
2. β_0 *is identified if and only if* $\beta(.) \in (\textit{Ker}\,(T) \cap \mathcal{F})^\perp$.

Let us provide the intuition for the easiest result, i.e. the "if" part of the first statement. We rely on the fact that under Assumptions 2 and 3(i), it is sufficient to identify $P(Y)$ to recover the whole distribution of (D,Y,Z). Besides, we show that this function satisfies

$$T\left(\frac{1}{P}\right) = w, \tag{2.1}$$

where $w(Z) = 1/P(D = 1|Z)$. Because T and w are identified, P is identified if there is a unique solution in $(0,1]$ of this equation. This uniqueness can be established if Ker $(T) \cap \mathcal{F} = \{0\}$.

The identifying condition Ker $(T) \cap \mathcal{F} = \{0\}$ is related to various completeness conditions considered in the literature (see, e.g., Newey & Powell, 2003, Severini & Tripathi, 2006, Blundell et al., 2007, and D'Haultfœuille, 2011).[3] When Y and Z have a finite support

[3]Our condition is intermediate between the stronger "standard" completeness condition Ker $(T) = \{0\}$ and the bounded completeness condition Ker $(T) \cap \mathcal{B} = \{0\}$, where $\mathcal{B}$ is the set of bounded functions.

(respectively by $(1, ..., I)$ and $(1, ..., J)$), this assumption amounts to rank$(M) = I$, where M is the matrix of typical element $P(Y = i|D = 1, Z = j)$ (see Newey & Powell, 2003). Hence, the support of Z must be at least as rich as the one of Y ($J \geq I$) and the dependence between the two variables must be strong enough for I linearly independent conditional distributions to exist. Because the matrix M is identified, it is straightforward to test for this condition, using for instance the determinant of MM' (see Subsection 3.1 below). When Y and Z are continuous, it is far more difficult to characterize them. Conditions have been provided by Newey & Powell (2003), D'Haultfœuille (2010) and D'Haultfœuille (2011). We consider below another example, related to our panel framework, where the restriction $\text{Ker}\,(T) \cap \mathcal{F} = \{0\}$ is satisfied.

The second statement of the theorem shows that when we specialize in one parameter rather than on the full distribution of (D, Y, Z), identification is achieved under weaker restrictions. Indeed, $\text{Ker}\,(T) \cap \mathcal{F} = \{0\}$ implies that any function of Y belongs to the orthogonal of this set. This result is closely related to Lemma 2.1 of Severini & Tripathi (2011) who consider identification of linear functionals related to a nonparametric instrumental regression.

As an illustration of Theorem 2.1 with continuous outcomes, suppose that we observe at the first date a past outcome Y_0, thanks to a retrospective questionnaire. This will be the case in the application considered in Section 4. Suppose also that the outcomes satisfy the following nonlinear fixed effect model:

$$\Lambda(Y_t) = U + \varepsilon_t, \tag{2.2}$$

where $\Lambda(.)$ is a strictly increasing real function and $(U, \varepsilon_0, \varepsilon_1, \varepsilon_2)$ are independent. Such a model generalizes standard linear fixed effect model $Y_t = U + \varepsilon_t$ and is close to the accelerated failure time model in duration analysis. Note that we do not introduce covariates here for simplicity, but our result can be extended to the more realistic model considered by Evdokimov (2011), namely $\Lambda(Y_t, X_t) = \psi(U, X_t) + \varepsilon_t$ with Λ strictly increasing in Y_t, provided that the covariates X_t are always observed at each period.

We also suppose that attrition only depends on current outcomes and transitions:

$$D = g(Y_1, Y_2, \eta), \quad \eta \perp\!\!\!\perp (Y_0, Y_1, Y_2). \tag{2.3}$$

Finally, we impose the following technical restriction on U, ε_0 and ε_2. For any random variable V, we let Ψ_V denote its characteristic function.

Assumption 4 *U admits a density with respect to the Lebesgue measure, whose support is the real line. Ψ_{ε_0} vanishes only on isolated points. The distribution of ε_2 admits a* *continuous density f_{ε_2} with respect to the Lebesgue measure. Moreover, $f_{\varepsilon_2}(0) > 0$ and there exists $\alpha > 2$ such that $t \mapsto t^\alpha f_{\varepsilon_2}(t)$ is bounded. Lastly, Ψ_{ε_2} does not vanish and is infinitely often differentiable in $\mathbb{R}\backslash A$ for some finite set A.*

The assumption imposed on the characteristic function of ε_0 is very mild and satisfied by all standard distributions. The conditions on ε_2 are more restrictive but hold for many distributions such as the normal, the student with degrees of freedom greater than one[4] and the stable distributions with characteristic exponent greater than one. The following proposition shows that under these conditions, the model is fully identified using Y_0 as the instrument.

Proposition 2.2 *Let $Z = (Y_0, Y_1)$, and suppose that Assumptions 3(i), 4, Equations (2.2) and (2.3) hold. Then Assumption 2 holds and $Ker(T) \cap \mathcal{F} = \{0\}$. Thus, the distribution of (D, Y, Z) is identified.*

2.2 Comparison with the literature

We compare our approach with the most usual models of attrition.

2.2.1 Missing at random attrition

This model, which has been considered by, e.g., Rubin (1976) and Abowd et al. (1999), posits that D only depends on Y_1:

$$D \perp\!\!\!\perp Y_2 | Y_1. \tag{2.4}$$

Identification of the joint distribution of (Y_1, Y_2) follows directly from the fact that, letting f_{D,Y_1,Y_2} denote the density of (D, Y_1, Y_2) with respect to an appropriate measure,

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{f_{D,Y_1,Y_2}(1, y_1, y_2)}{P(D = 1|Y_1 = y_1)}.$$

Condition (2.4) is the equivalent, in a panel setting, of the so-called missing at random assumption (see, e.g., Little & Rubin, 1987) or the unconfoundedness assumption in the treatment effect literature (see for instance Imbens, 2004). Because it rules out any dependence between attrition and current outcomes, it is likely to fail in many cases. In a labor force survey, for instance, house moving is a common source of attrition, and is itself related to changes in employment and/or earnings.

[4]See e.g. Mattner (1992) for a proof that the conditions on the characteristic function of student distributions are indeed satisfied.

2.2.2 Dependence on current values

Compared to the first, the logic of this model is the opposite, as attrition is related to current values only:

$$D \perp\!\!\!\perp Y_1 | Y_2. \tag{2.5}$$

This assumption has been considered by Hausman & Wise (1979) in a parametric model. This assumption takes into account nonignorable attrition, but in a special way. Indeed, it rules out the possibility that transitions (i.e., functions of (Y_1, Y_2)) explain attrition. Abstracting from the parametric restrictions of Hausman & Wise (1979), identification can be proved along the same lines as previously. It suffices indeed to solve in g the functional equation

$$E\left[g(Y_2)|D = 1, Y_1\right] = 1/P(D = 1|Y_1).$$

Under completeness conditions similar to the one above, this equation admits a unique solution in g, namely $1/P(D = 1|Y_2 = .)$.

2.2.3 Additive restriction on the probability of attrition

Hirano et al. (2001) propose a two period framework which generalize both previous examples in the sense that D may depend on both Y_1 and Y_2. This generalization is possible when a refreshment sample, which allows one to identify directly the distribution of Y_2, is available.[5] They also suppose that

$$1/P(D = 1|Y_1, Y_2) = g(\alpha + k_1(Y_1) + k_2(Y_2)), \tag{2.6}$$

where g is a known function, but $k_1(.)$ and $k_2(.)$ are unknown. They show that $k_1(.)$ and $k_2(.)$ are identified. This allows them to recover the joint distribution of (Y_1, Y_2), since by the Bayes' rule,

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1,Y_2|D=1}(y_1, y_2)P(D = 1)g(\alpha + k_1(y_1) + k_2(y_2)),$$

Compared to our approach, Hirano et al. (2001) do not rely on any exclusion restriction. This comes at the cost of imposing the additive restriction on $P(D = 1|Y_1, Y_2)$, which may be restrictive (see below), and having a refreshment sample, which is not needed in our case.

Though the identification proof of Hirano et al. (2001) is much different from ours, the two frameworks are actually related. As shown by Bhattacharya (2008), identification in this

[5] Note that because, the distribution of Y_1 is also identified from the panel at date 1, the problem reduces to recover the copula of (Y_1, Y_2).

additive model can be directly obtained from the functional equations

$$E\left[g(k_1(Y_1) + k_2(Y_2))|D = 1, Y_i\right] = 1/P(D = 1|Y_i).$$

Thus, identification is actually achieved along similar lines as above, the instrument being equal to (Y_1, Y_2). The difference here is that only the marginal distributions of the instrument is identified. This is the reason why they have to impose Model (2.6) to the attrition process. Note that such a restriction is not innocuous. If attrition depends on transitions, then their restriction is likely fails to hold. If, as in our application, attrition occurs for individuals who move, and that moving itself occurs with a large probability when employment status changes, then $P(D = 1|Y_1, Y_2)$ depends on $\mathbb{1}\{Y_1 = Y_2\}$. Model (2.6) cannot handle such an attrition process.

2.3 Testability

An interesting feature of Assumption 2 is that it is refutable, contrary to the ignorable attrition assumption considered above. First, Equation (2.1) may have no solution. This is especially clear when (Y, Z) has a finite support. If indeed Y and Z take respectively I and J distinct values, with $J > I$, then (2.1) can be written as a system of J equations with I unknown parameters, so that the model is overidentified. But even when $I = J$, the model is testable since under Assumption 2 (and maintaining Assumption 3-(i)), Equation (2.1) admits at least one solution in $(0, 1]$, namely $P(Y)$. We derive a formal statistical test of this condition in Subsection 3.1 below.

A stronger test of the conditional independence assumption can be derived if a refreshment sample is available, as in Hirano et al. (2001). In this case, the marginal distribution of Y_2 is identified. But then we can reject the conditional independence assumption if for all Q satisfying $T(1/Q) = w$, there exists t such that

$$E\left[\frac{D\mathbb{1}\{Y_2 \le t\}}{Q(Y)}\right] \ne P(Y_2 \le t).$$

3 Estimation

We now turn to inference within our framework of endogenous attrition. In line with our application, we focus here on the case where (Y, Z) is discrete. In the continuous case, estimation is more involved because we have to estimate nonparametrically $P(.)$ using Equation (2.1), that is by solving an ill-posed inverse problem. As in nonparametric IV models (see Severini & Tripathi, 2011), root-n estimation of the parameter $\beta_0 = E(g(Y, Z))$ may not be possible. We leave this issue for future research.[6]

As previously, we focus on the estimation of the distribution of (D, Y, Z), but also on the parameter $\beta_0 = E(g(Y, Z))$. We first posit an i.i.d. sample of n observations.

Assumption 5 *We observe an iid sample* $((D_1, D_1Y_{21}, Z_1), ..., (D_n, D_nY_{2n}, Z_n))$.

We denote the support of Y_t and Z_1 by respectively $\{1, ..., I\}$ and $\{1, ..., J\}$, with $I \leq J$.[7] In this case, the data (D, DY_2, Z) are distributed according to a multinomial distribution. To get asymptotic efficient estimators, we consider constrained maximum likelihood estimation hereafter.

For a fixed y, let $p_{1ij} = P(D = 1, Y_2 = i, Z_1 = j|Y_1 = y)$ and $p_{0.j} = P(D = 0, Z_1 = j|Y_1 = y)$ denote the probabilities corresponding to the observations, and define $p_1 = (p_{111}, ..., p_{1IJ})$, $p_{0.} = (p_{0.1}, ..., p_{0.J})$ and $p = (p_1, p_{0.})$. Note that we let the dependence in y implicit hereafter. p is the natural parameter of the statistical model here, as it fully describes the distribution of (D, DY_2, Z_1) (conditional on Y_1).[8] However, it does not directly allow us to recover the whole distribution of (D, Y_2, Z_1). This is why we also introduce $p_{0ij} = P(D = 0, Y_2 = i, Z_1 = j|Y_1 = y)$, and define p_0 as p_1. Then any parameter θ_0 of the distribution of (D, Y_2, Z_1) is a function of (p_0, p_1), and we write $\theta_0 = g(p_0, p_1)$.[9]

Finally, we adopt the same notations for the constrained maximum likelihood estimator $\widehat{p}$ as for p, and we let $n_{1ij} = \sum_{k:Y_{1k}=y} D_k \mathbb{1}\{Y_{2k} = i\}\mathbb{1}\{Z_{1k} = j\}$ and define $n_{0.j}$ accordingly. The following proposition shows how to compute $\widehat{p}$ and an efficient estimator of θ_0 in our attrition model.

Proposition 3.1 *Suppose that Assumptions 1-3(i) hold. Then the maximum likelihood*

[6]The intermediate cases where Y is discrete but Z is continuous is addressed, in a setting close to ours, by Ramalho & Smith (2011).

[7]Of course, our setting readily extends to a case where discrete covariates X are available, and Assumption 2 holds conditionally (i.e., $D \perp\!\!\!\perp Z|Y, X$).

[8]This parametrization is also convenient for the unconstrained model where Assumption 2 does not necessarily hold.

[9]We thus consider here implicitly parameters that depend on the distribution of (D, Y_2, Z_1) conditional on Y_1. To estimate unconditional parameters, it suffices to integrate conditional parameters over the empirical distribution of Y_1. Because Assumption 2 does not impose any restriction on the distribution of Y_1, this results in asymptotically efficient estimators.

estimator $\widehat{p}$ satisfies

$$\widehat{p} = \arg \max_{(q,b)\in[0,1]^{(I+1)J}\times\mathbb{R}^I} \sum_{j=1}^{J} \left[n_{0.j} \ln q_{0.j} + \sum_{i=1}^{I} n_{1ij} \ln q_{1ij} \right]$$

$$s.t. \left| \begin{array}{cc} \sum_{j=1}^{J} \left[q_{0.j} + \sum_{i=1}^{I} q_{1ij} \right] = 1, & \\ b_i \geq 0 & i = 1, ..., I, \\ \sum_{i=1}^{I} q_{1ij} b_i = q_{0.j} & j = 1, ..., J. \end{array} \right. \qquad (C)$$

Suppose moreover that the matrix P_1 of typical element p_{1ij} has rank I and g is differentiable. Then θ_0 is identifiable and can be estimated efficiently by

$$\widehat{\theta} = g(\widehat{p}_0, \widehat{p}_1),$$

where $\widehat{p}_{0ij} = \widehat{b}_i \widehat{p}_{1ij}$, and $\widehat{b} = (\widehat{b}_1, ..., \widehat{b}_I)$ is a solution of constraints (C) taken at $q = \widehat{p}$.

Proposition 3.1 establishes that the maximum likelihood of p can be obtained by a constrained maximization with quite simple (although nonlinear) constraints. It also shows how to compute an efficient estimator of θ_0. The idea behind the introduction of the $(b_i)_i$ is that, by Bayes' rule and Assumption 2,

$$p_{0ij} = \frac{P(D = 0|Y_1 = y, Y_2 = i)}{P(D = 1|Y_1 = y, Y_2 = i)} p_{1ij},$$

and b_i represents the odds $P(D = 0|Y_1 = y, Y_2 = i)/P(D = 1|Y_1 = y, Y_2 = i)$. The inequality constraints $b_i \geq 0$ then ensure that $P(D = 1|Y = i)$ is indeed a probability, while the equality constraints are a rewriting of Equation (2.1) in the discrete context.

The identifying condition $\text{rank}(P_1) = I$ is the equivalent of $\text{Ker}\,(T) \cap \mathcal{F} = \{0\}$ here. It can be easily tested in the data because under the null hypothesis that $\text{rank}(P_1) < I$, we have $\mu_0 \equiv \det(P_1 P_1') = 0$. Then, letting $\widehat{\mu} = \det(\widehat{P}_1 \widehat{P}_1')$, $\sqrt{n}\widehat{\mu}$ tends to a zero mean normal variable under the null by the delta method. We use this result to test for the rank condition in our application Section 4.

Finally, as noted before, we can test for Assumption 2 by two ways. The first and standard one is that the equality constraints in (C) may not hold when $J > I$, because there is no $(b_i)_{1\leq i\leq I}$ such that $\sum_{i=1}^{I} b_i p_{1ij} = p_{0.j}$. Basically, this arises when the different values of Z are not "compatible", as with the Sargan test in linear IV models. The second is that the $(b_i)_{1\leq i\leq I}$ which satisfy these equality constraints must be nonnegative. This may not hold in general, even when $I = J$. To test for both conditions simultaneously, we consider the Wald test developed by Kodde & Palm (1986). In our framework, the unconstrained model where Assumption 2 does not necessarily hold is simply the multinomial model on (D, DY_2, Z) parameterized by p, and the maximum likelihood estimator $\widehat{p}_U$ simply corresponds to the sample proportions. The contraints (C) corresponding to Assumption 2 hold if and only if there exists $b \geq 0$ (understood componentwise) such that $P_1'b = p_{0.}$. This condition is equivalent to[10]

$$\left[P_1'(P_1P_1')^{-1}P_1 - I\right] p_{0.} = 0, \quad (P_1P_1')^{-1}P_1p_{0.} \geq 0.$$

Let us rewrite these constraints as $h_1(p) = 0$ and $h_2(p) \geq 0$, and let $h(p) = (h_1(p), h_2(p))$. Let also $\mathcal{H}_0 = 0^J \times \mathbb{R}^{+I}$ denote the set of $h = (h_1, h_2)$ satisfying these constraints. Denote by Σ_{ii} (resp. Σ_{12}) the asymptotic variance of $\widehat{h}_i \equiv h_i(\widehat{p}_U)$ (resp. covariance of $h_1(\widehat{p}_U)$ and $h_2(\widehat{p}_U)$), and by Σ the asymptotic variance of $\widehat{h} \equiv h(\widehat{p}_U)$. Finally, let $\widehat{\Sigma}$ denote a consistent estimator of Σ. The test statistic W_n is then defined as

$$W_n = n \min_{h \in \mathcal{H}_0} \left(h - \widehat{h}\right)' \widehat{\Sigma}^- \left(h - \widehat{h}\right),$$

where $\widehat{\Sigma}^-$ denotes the Moore-Penrose inverse of $\widehat{\Sigma}$.[11] Computing W_n is straightforward as it corresponds to a quadratic programming problem. Proposition 3.2 provides the asymptotic distribution of W_n.

Proposition 3.2 *We have, for all $c \geq 0$,*

$$\sup_{h \in \mathcal{H}_{0\Sigma}} \lim_{n \to \infty} \Pr(W_n \geq c) = P\left(\min_{h \in \mathcal{H}_0} (h - U)' \Sigma^- (h - U) \geq c\right). \tag{3.1}$$

where $\mathcal{H}_{0\Sigma}$ is the subset of $\mathcal{H}_0$ such that the asymptotic variance of $\widehat{h}$ is equal to Σ and $U \sim \mathcal{N}(0, \Sigma)$.

Following the analysis of Kodde & Palm (1986), it is also possible to express this asymptotic distribution as a mixture of chi-square. The corresponding weights, however, do not have a closed form in general, so that it is actually easier to approximate the asymptotic distribution using (3.1) rather than their expression.[12] We use such simulations to compute our p-values in the application below.

[10]Indeed, the existence of $b \in \mathbb{R}^{+I}$ such that $P_1'b = p_{0.}$ is equivalent to the fact that the least square solution $(P_1P_1')^{-1}P_1p_{0.}$ to $\min_{b \in \mathbb{R}^I} \left\|P_1b - p_{0.}\right\|$ satisfies the equation exactly and belongs to $\mathbb{R}^{+I}$.

[11]$\widehat{\Sigma}$ is not full rank in general, because the rank of Σ_{11} is $J - I$, while $h_1(p) \in \mathbb{R}^J$. This is logical, since we only have $J - I$ overidentifying equality constraints here.

[12]It is also possible to derive lower and upper bounds on the critical values of this test, see, e.g., Kodde & Palm (1986).

4 Application

4.1 Introduction

In this section, we apply the previous results to estimate transitions on employment status in the French labor market. Beyond the unemployment rate, measuring such transitions is important to assess, for instance, the importance of short and long-term unemployment. We use for that purpose the Labor Force Survey (LFS) conducted by the French national institute of statistics (INSEE). This survey is probably the best tool to measure such transitions in France. Indeed, and compared to administrative data or other surveys, it properly measures unemployment with respect to the standard ILO definition, has a comprehensive coverage of the population and its sample size is large. Since 2003, it is a rotating panel where approximately 5,900 new households are surveyed each quarter. These new households are then questioned the five following quarters. On the first and sixth wave, interrogations are face to face, while on the others they are conducted by telephone. It has been argued that the use of phone may introduce specific measurement errors (see, e.g., Biemer, 2001), so we focus on the first and last interrogations hereafter. We also restrict ourselves to people between 15 and 65 and pool together all labor force surveys on the period 2003-2005.

Table 1: Summary statistics on the French LFS (first waves during 2003-2005).

Statistics	All	Men	Women
Main sample:			
Number of individuals	107,031	52,245	54,786
Attrition rate on last waves	21.78%	22.26%	21.31%
Participation rate on first waves	68.17%	73.91%	62.69%
(Uncorrected) participation rate on last waves	67.38%	72.75%	62.32%
Unemployment rate on first waves	9.68%	9.05%	10.39%
(Uncorrected) unemployment rate on last waves	8.02%	7.22%	8.90%
Refreshment sample for last waves:			
Number of observations	109,404	53,337	56,067
Participation rate on the refreshment sample	67.92%	73.31%	62.78%
Unemployment rate on the refreshment sample	9.97%	9.43%	10.57%

Table 1 provides some summary statistics on our dataset, which emphasize that attrition may be problematic in the LFS survey. This is especially striking when we compare the (uncorrected) participation and unemployment rate on last waves and the one on the refreshment sample (i.e., entrants at the same time). We observe differences around 1.5 percent points on participation rates, and around 2 percent points on unemployment rates. This does not correspond to the so-called rotation bias, which concern for the same period the variation of measured outcome between the different waves for a rotating panel (see, e.g., Bailar, 1975, and more recently on the French LFS, Goux, 2005). Here, we consider a panel of household (and not of house) with a specific attrition scheme du to the collect in the French LFS: moving households are not followed. This is likely to affect transition estimates on the labor market, because moving is likely to be related to such transitions.

As suggested in Section 2, we propose to correct for potentially endogenous attrition by using past employment status, measured by a retrospective question asked on the first waves. The underlying assumption is that attrition depends on the current transition on this outcome, but not on previous ones. This assumption is plausible if most of the endogeneity in attrition stems from the moving of households. The instrument Z we use is employment status 6 months before the first wave. We choose to divide this variable in three categories (unemployed, employed, and out of labour force) as our outcome which is contemporary employment status. To assess the strength of our instrument we implement rank test between Y_2 and Z conditionally to Y. We also test our instrument following the approach developed in Proposition 3.2.We also assess the plausibility of our instrument by comparing the corrected participation and unemployment rates we obtain on last waves with the one on the refreshment sample. Finally, as a matter of comparison, we also correct for attrition under ignorability or using the method of Hirano et al. (2001) with a logistic cumulative distribution for the link function.

4.2 The results

We first check the rank condition between Z_1 and Y_2 conditional on gender and Y_1, relying on the determinant test proposed in Subsection 3.1. Results are displayed in Table 2. The p-value of the rank test associated to any state Y_1 are always smaller than 10% for both men and women. We also implement the test developed in the Proposition 3.2. Though some inequality constraints are binding in our estimates, we do not reject the independence assumption $Z \perp\!\!\!\perp D|Y_1, Y_2$ here (see Table 3).

Table 2: Rank test between Z and Y_2 conditional on gender and Y_1.

	P-value (Men)	P-value (Women)
Y_1 = Empl.	0.0047	0.0031
Y_1 = Unempl.	0.0695	0.0539
Y_1 = Out L.F.	0.0498	0.0867

Table 3: Test of $Z \perp\!\!\!\perp D|Y_1, Y_2$ by gender.

	P-value (Men)	P-value (Women)
Y_1 = Empl.	0.7430	0.8818
Y_1 = Unempl.	0.5724	0.5418
Y_1 = Out L.F.	0.7186	0.7088

Second, we estimate the probabilities of attrition (or non-attrition) conditional on (Y_1, Y_2). Our results, displayed in Table 4, confirm that (under the validity of our instrument), attrition is related to transitions on employment status. People who remain stable on the labor market have always a significant larger probability to respond in the second wave than people who change. In particular, we observe a large attrition for those who move from employment to unemployment or inactivity whereas attrition seems negligible for those who remain unemployed at both periods. As suggested above, such transitions are likely to be related to house movings. For instance, transitions from inactivity to employment or unemployment mostly correspond to students who enter the labor market and move at the same time. Such features cannot be captured under the missing at random (MAR) scheme $D \perp\!\!\!\perp Y_2|Y_1$, or the additive model of Hirano et al. (2001). In particular, they tend to underestimate the probability of attrition for people whose status change on the labor market, and to overestimate them for stable trajectories (see Table 7 in appendix for tests on the difference between our IV models and the two others).

Table 4: $\widehat{P}(D=1|Y_1,Y_2)$ for women under various assumption

	Men			Women		
	Y_2 = Empl.	Y_2 = Unempl.	Y_2 = Out L.F.	Y_2 = Empl.	Y_2 = Unempl.	Y_2 = Out L.F.
IV						
Y_1 = Empl.	84.33 (0.85)	34.33 (3.92)	46.31 (7.04)	85.72 (0.92)	46.45 (6.42)	44.57 (4.46)
Y_1 = Unempl.	55.56 (4.33)	100.00 (3.91)	51.01 (9.27)	52.46 (4.25)	100.00 (2.06)	76.38 (13.43)
Y_1 = Out L.F.	54.83 (11.35)	55.85 (11.35)	85.72 (1.99)	56.43 (11.16)	67.10 (16.77)	84.34 (1.10)
MAR						
Y_1 = Empl.	78.22 (0.22)	78.22 (0.22)	78.22 (0.22)	79.00 (0.24)	79.00 (0.24)	79.00 (0.24)
Y_1 = Unempl.	65.90 (0.79)	65.90 (0.79)	65.90 (0.79)	69.77 (0.77)	69.77 (0.77)	69.77 (0.77)
Y_1 = Out L.F.	79.52 (0.34)	79.52 (0.34)	79.52 (0.34)	79.77 (0.28)	79.77 (0.28)	79.77 (0.28)
HIRR						
Y_1 = Empl.	79.01 (0.28)	59.98 (2.07)	76.44 (2.25)	79.57 (0.32)	66.59 (2.21)	77.57 (2.02)
Y_1 = Unempl.	75.84 (1.43)	55.55 (1.24)	73.01 (2.04)	76.04 (1.46)	61.89 (1.40)	73.81 (1.77)
Y_1 = Out L.F.	82.41 (1.69)	65.09 (2.33)	80.15 (0.47)	82.01 (1.67)	69.99 (2.07)	80.18 (0.39)

Note: Standard error in brackets computed with 1000 bootstrap samples.

Before presenting our results on transitions, we estimate the distribution of Y_2 with our IV method and compare it with the one of the refreshment sample. We also estimate this distribution supposing that data are missing at random (MAR), i.e. $D \perp\!\!\!\perp Y_2|Y_1$. Table 5 shows that on the five statistics related to the distribution of Y_2, our estimator is close, and not statistically significant at usual levels, to the one based on the refreshment sample. Those based on the MAR assumptions, on the other hand, do differ significantly for several features of Y_2. In other words, we can reject, using the refreshment sample, the hypothesis that attrition only depends on past outcomes, while our independence condition is not rejected in the data. Note that we cannot use the refreshment sample to properly compare our method with the one of Hirano et al. (2001) because by construction, their estimator exactly matches the distribution of Y_2 on the refreshment sample.

Table 5: Comparison of the methods with the refreshment sample

	Men			Women		
	REF.	MAR	IV	REF.	MAR	IV
$P(Y_2 = \text{Empl.})$	66.40	67.47 (<0.0001)	64.59 (0.0545)	56.15	56.81 (0.0055)	55.07 (0.2718)
$P(Y_2 = \text{Unempl.})$	6.92	5.62 (<0.0001)	7.53 (0.2851)	6.63	5.78 (<0.0001)	6.51 (0.8529)
$P(Y_2 = \text{Out L.F.})$	26.69	26.92 (0.2825)	27.88 (0.2033)	37.22	37.40 (0.4300)	38.42 (0.1373)
Participation rate	73.31	73.08 (0.2825)	72.12 (0.2033)	62.78	62.60 (0.4300)	61.58 (0.1373)
Unemployment rate	9.43	7.68 (<0.0001)	10.44 (0.1884)	10.57	9.24 (<0.0001)	10.58 (0.9913)

Note: we indicate the p-values of the difference with the refreshment sample under parentheses. Computation based on 1000 bootstrap samples.

Finally, we compute transitions on the labor market using our IV method, the MAR assumption and the additive method of Hirano et al. (2001) (see Table 6). Not surprisingly given the discrepancies on the probabilities of attrition, our results differ significantly from those obtained by the other methods (see Table **??** for the tests of differences). Other methods tend in particular to overestimate stability on the labor market. If it is impossible to discriminate between our IV method and the one of Hirano et al. (2001) without extra information on these transitions, some patterns on unemployment seem to support our model over theirs in this application. In particular, the estimated probability of staying unemployed after 15 months are respectively equal to 25% with our IV method and around 44% with the one of Hirano et al. (2001) (41% for women and 47% for the men). These latter figures seem particularly high, compared to the rate we observe between the first wave and 11 months before, namely 44,9% for women and 49.6% for men. It is also notably at odds with the ??% rate of long term (i.e., one year or more) unemployment directly observed on the LFS.

Table 6: Estimated probability of transitions for women under various assumption

	Men			Women		
	Y_2 = Empl.	Y_2 = Unempl.	Y_2 = Out L.F.	Y_2 = Empl.	Y_2 = Unempl.	Y_2 = Out L.F.
IV						
Y_1 = Empl.	85.86 (0.85)	6.12 (0.70)	8.02 (1.13)	83.46 (0.89)	4.73 (0.63)	11.81 (1.15)
Y_1 = Unempl.	49.00 (3.71)	25.85 (1.46)	25.15 (3.74)	52.61 (3.92)	25.30 (0.95)	22.08 (3.91)
Y_1 = Out L.F.	13.81 (2.63)	6.47 (1.16)	79.72 (1.84)	12.74 (2.12)	5.92 (1.41)	81.34 (1.05)
MAR						
Y_1 = Empl.	92.56 (0.16)	2.69 (0.10)	4.75 (0.13)	90.56 (0.19)	2.78 (0.11)	6.66 (0.16)
Y_1 = Unempl.	41.31 (1.06)	39.23 (1.01)	19.46 (0.82)	39.56 (0.98)	36.27 (0.95)	24.18 (0.84)
Y_1 = Out L.F.	9.52 (0.27)	4.55 (0.20)	85.93 (0.33)	9.02 (0.23)	4.98 (0.17)	86.00 (0.28)
HIRR						
Y_1 = Empl.	91.64 (0.20)	3.50 (0.12)	4.86 (0.15)	89.92 (0.23)	3.30 (0.12)	6.79 (0.18)
Y_1 = Unempl.	35.90 (0.96)	46.54 (0.91)	17.57 (0.75)	36.29 (0.97)	40.87 (0.90)	22.85 (0.83)
Y_1 = Out L.F.	9.19 (0.27)	5.56 (0.24)	85.26 (0.36)	8.77 (0.23)	5.68 (0.17)	85.55 (0.28)

Note: Standard error in brackets computed with 1000 bootstrap samples.

5 Conclusion

In this paper, we develop an alternative method to correct for endogenous attrition in panel. We allow for both dependence on current and past outcomes and, thanks to the availability of an instrument, do not need to impose functional restrictions on the probability of attrition, contrary to Hirano et al. (2001). The application suggests that our method may do a good job for handling attrition processes which mostly depend on transitions.

The paper raises two challenging issues, related to our main conditional independence assumption. The first is whether the refreshment sample could be used to weaken this assumption, rather than to test for it. This may be useful in settings where this condition is considered too stringent. The second is whether one can build bounds on parameters of interest if the conditional independence assumption is replaced by weaker conditions such as monotonicity ones. Although not considered here, these questions are clearly at the top of our research agenda.

References

Abowd, J. M., Crépon, B. & Kramarz, F. (1999), 'Moment estimation with attrition: An application to economic models', *Journal of the American Statistical Association* **96**, 1223–1231.

Bailar, B. A. (1975), 'The effects of rotation group bias on estimates from panel surveys', *Journal of the American Statistical Association* **70**, 23–30.

Ben-Israel, A. & Greville, T. N. E. (2003), *Generalized inverses:Theory and applications*, Springer.

Bhattacharya, D. (2008), 'Inference in panel data models under attrition caused by unobservables', *Journal of Econometrics* **144**, 430–446.

Biemer, P. P. (2001), 'Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing', *Journal of Official Statistics* **17**, 295–320.

Bierens, H. J. (1982), 'Consistent model specification tests', *Journal of Econometrics* **20**, 105–134.

Blundell, R., Chen, X. & Kristensen, D. (2007), 'Nonparametric iv estimation of shape-invariant engel curves', *Econometrica* **75**, 1613–1669.

Chen, C. (2001), 'Parametric models for response-biased sampling', *Journal of the Royal Statistical Society, Series B* **63**, 775–789.

D'Haultfœuille, X. (2010), 'A new instrumental method for dealing with endogenous selection', *Journal of Econometrics* **154**, 1–15.

D'Haultfœuille, X. (2011), 'On the completeness condition in nonparametric instrumental regression', *Econometric Theory, forthcoming* .

Evdokimov, K. (2011), Nonparametric identification of a nonlinear panel model with application to duration analysis with multiple spells. Working paper.

Goux, D. (2005), 'L'impact des changements intervenus dans l'enquête emploi en 2003 sur la qualité de ses résultats', *Actes des Journées de Méthodologie Statistique 2005, http://jms.insee.fr* .

Hall, R. & Mishkin, F. S. (1982), 'The sensitivity of consumption to transitory income: estimates from panel data on households', *Econometrica* **50**, 461–481.

Hausman, J. A. & Wise, D. A. (1979), 'Attrition bias in experimental and panel data: the Gary income maintenance experiment', *Econometrica* **47**, 455–473.

Heckman, J. (2001), 'Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture', *Journal of Political Economy* **109**, 673–748.

Hirano, K., Imbens, G. W., Ridder, G. & Rubin, D. B. (2001), 'Combining panel data sets with attrition and refreshment samples', *Econometrica* **69**, 1645–1659.

Imbens, G. (2004), 'Nonparametric estimation of average treatment effects under exogeneity: a review', *The Review of Economics and Statistics* **86**, 4–29.

Kodde, D. A. & Palm, F. C. (1986), 'Wald criteria for jointly testing equality and inequality restrictions', *Econometrica* **54**, 1243–1248.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press.

Lee, G. M., Tam, N. N. & Yen, N. D. (2005), *Quadratic programming and affine variational inequalities: a qualitative study*, Springer.

Little, R. & Rubin, D. B. (1987), *Statistical analysis with Missing Data*, John Wiley & Sons, New York.

Mattner, L. (1992), 'Completeness of location families, translated moments, and uniqueness of charges', *Probability Theory and Related Fields* **92**, 137–149.

Newey, W. & Powell, J. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Perlman, M. D. (1969), 'One sided problems in multivariate analysis', *Annals of Mathematical Statistics* **40**, 549–567.

Ramalho, E. A. & Smith, R. J. (2011), 'Discrete choice nonresponse', *Review of Economic Studies, forthcoming* .

Rubin, D. B. (1976), 'Inference and missing data', *Biometrika* **63**, 581–592.

Severini, T. & Tripathi, G. (2006), 'Some identification issues in nonparametric linear models with endogenous regressors', *Econometric Theory* **22**, 258–278.

Severini, T. & Tripathi, G. (2011), 'Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors', *Journal of Econometrics, forthcoming* .

Tang, G., Little, R. J. A. & Raghunathan, T. E. (2003), 'Analysis of multivariate missing data with nonignorable nonresponse', *Biometrika* **90**, 747–764.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Appendix A: proofs

Theorem 2.1

Consider first point 1. By definition of T and Assumption 2,

$$\begin{aligned} T\left(\frac{1}{P}\right) &= E\left(\frac{D}{P(Y)}|Z\right)w(Z) \\ &= E\left(\frac{E(D|Y,Z)}{P(Y)}|Z\right)w(Z) \\ &= w(Z). \end{aligned} \tag{5.1}$$

Now, suppose that there exists $Q(Y) \in (0,1]$ that also satisfies $T(1/Q(Y)) = w(Z)$. Then

$$T\left(\frac{1}{Q(Y)} - \frac{1}{P(Y)}\right) = 0.$$

Besides, $1/Q(Y) - 1/P(Y) \geq 1 - 1/P(Y)$, so that $1/Q(Y) - 1/P(Y) \in \mathcal{F}$. Hence, $Q(Y) = P(Y)$ and $P(.)$ is identified as the unique solution of (5.1), since T and $w(Z)$ are identified. To see that the distribution of (D,Y,Z) is identified in this case, denote by $f_{D,Y,Z}$ the density of (D,Y,Z) with respect to an appropriate measure, and observe that by Bayes' rule,

$$\begin{aligned} f_{D,Y,Z}(0,y,z) &= P(D=0|Y=y,Z=z)f_{Y,Z}(y,z) \\ &= \frac{P(D=0|Y=y,Z=z)}{P(D=1|Y=y,Z=z)}f_{Y,Z|D=1}(y,z) \\ &= \frac{1-P(y)}{P(y)}f_{Y,Z|D=1}(y,z) \end{aligned}$$

Thus, $f_{D,Y,Z}(0,y,z)$ is identified in the data. $f_{D,Y,Z}(1,y,z)$ is also identified. The "if" part of statement 1 follows. To get the "only if" part, suppose that there exists $h \in \text{Ker}\,(T) \cap \mathcal{F}$, $h \neq 0$ such that $T(h) = 0$. Then define $Q = P/(1+hP)$. By definition of $\mathcal{F}$, $Q \in (0,1]$ and by definition of $\text{Ker}\,(T)$, $T(1/Q) = w$. By Theorem 2.4 of D'Haultfœuille (2010), one can rationalize a distribution of (D,Y,Z) satisfying Assumption 2 and $P(D=1|Y) = Q(Y)$. Because $Q \neq P$, this implies that the joint distribution of (D,Y,Z) is not identified (because the joint distribution of (D,Y) is not).

Now let us turn to point 2 of the theorem. Let $\beta(Y) = E(g(Y,Z)|Y)$, so that $\beta_0 = E(\beta(Y))$. Let $\mathcal{Q}$ the set of function $\{Q \in (0,1] | T(1/Q) = w\}$. $\mathcal{Q}$ is identified in the data and let Q

any element of $\mathcal{Q}$. Let $h = 1/Q - 1/P$, so that by construction, $h \in \text{Ker}\,(T) \cap \mathcal{F}$. Besides,

$$\begin{aligned} E\left(\frac{D\beta(Y)}{Q(Y)}\right) &= E\left(\frac{D\beta(Y)}{P(Y)}\right) + E\left(Dh(Y)\beta(Y)\right) \\ &= E\left(\beta(Y)\right) + E\left(h(Y)\beta(Y)|D=1\right)P(D=1) \\ &= \beta_0 + E\left(h(Y)\beta(Y)|D=1\right)P(D=1). \end{aligned}$$

By assumption, $\beta(Y) \in (\text{Ker}\,(T) \cap \mathcal{F})^{\perp}$. This means that the second term in the right-hand side is zero. As a result, β_0 is identified, and the "if" part is proved. To show the "only if", we reason similarly to previously. Let $h \in \text{Ker}\,(T) \cap \mathcal{F}$ be such that $E(\beta(Y)h(Y)|D=1) \neq 0$. Let $Q = P/(1+hP)$. By construction, $Q \in \mathcal{Q}$. By Theorem 2.4 of D'Haultfœuille (2010) once more, one can rationalize a distribution of (D, Y, Z) satisfying Assumption 2 and $E(D|Y) = Q(Y)$. With this distribution,

$$\begin{aligned} E(\beta(Y)) &= E\left(\frac{D\beta(Y)}{Q(Y)}\right) \\ &= E\left(\frac{D\beta(Y)}{P(Y)}\right) + E\left(Dh(Y)\beta(Y)\right) \\ &= \beta_0 + E\left(h(Y)\beta(Y)|D=1\right)P(D=1). \end{aligned}$$

The second term in the right-hand side is different from zero by assumption. Hence, we have built a distribution consistent with the data and such that $E(\beta(Y)) \neq \beta_0$. This implies that β_0 is not identified, and the result follows.

Proposition 2.2

First, remark that if $\eta \perp\!\!\!\perp (Y_0, Y_1, Y_2)$, then $\eta \perp\!\!\!\perp Y_0|Y_1, Y_2$. As a result, $D \perp\!\!\!\perp Y_0, Y_1|Y_1, Y_2$, and Assumption 2 holds with $Z = (Y_0, Y_1)$. Now, suppose that $T(h) = 0$, for $h \in \mathcal{F}$, and let us prove that $h = 0$. First, $T(h) = 0$ rewrites as

$$0 = E(Dh(Y_1, Y_2)|Y_0, Y_1) = E(\widetilde{h}(Y_1, \widetilde{Y}_2)|Y_0, Y_1),$$

with $\widetilde{Y}_t = \Lambda(Y_t)$ and $\widetilde{h}(y_1, y_2) = h(y_1, \Lambda(y_2)) \times P(y_1, y_2)$. As a result, for all $t \in \mathbb{R}$,

$$E(\widetilde{h}(Y_1, \widetilde{Y}_2)e^{it\widetilde{Y}_0}|Y_1) = 0$$

Because $\varepsilon_0 \perp\!\!\!\perp (U, Y_1, Y_2)$,

$$E(\widetilde{h}(Y_1, \widetilde{Y}_2)e^{itU}|Y_1)\Psi_{\varepsilon_0}(t) = 0.$$

Thus, by assumption, $t \mapsto E(\widetilde{h}(Y_1, \widetilde{Y}_2)e^{itU}|Y_1)$ is equal to zero except perhaps on a set of isolated points. Because this function is continuous by dominated convergence, it is actually equal to zero on the whole line. This implies (see e. g. Bierens, 1982, Theorem 1)

$$E(\widetilde{h}(Y_1, \widetilde{Y}_2)|Y_1, U) = 0.$$

Now, ε_2 is independent of (Y_1, U) and U admits a density with respect to the Lebesgue measure. Thus, for almost all y_1 almost every u,

$$\int \widetilde{h}(y_1, u - v) f_{-\varepsilon_2}(v) dv = 0. \tag{5.2}$$

Fix y_1 so that Equation (5.2) holds for almost every u. Because $h \geq 1 - 1/P$ by assumption, $\widetilde{h}$ is bounded below by -1. Letting $g = \widetilde{h}(y_1, .)$ and $\star$ denote the convolution product, we have $(g+1) \star f_{-\varepsilon_2} = 1$ almost everywhere. Besides, by the first step of the proof of Proposition 2.2 of D'Haultfœuille (2010), there exist positive c_1, c_2 and $0 < \alpha' < \alpha - 2$ such that

$$c_1 \leq (f_{\varepsilon_2} \star f_{\alpha'})(x) \times (1 + |x|)^{\alpha' + 1} \leq c_2, \tag{5.3}$$

where $f_{\alpha'}$ denotes the density of an α'-stable distribution of characteristic function $\exp(-|t|^{\alpha'})$. Moreover, because $g + 1, f_{-\varepsilon_2}$ and $f_{\alpha'}$ are nonnegative, we can apply Fubini's theorem, so that

$$(g+1) \star (f_{-\varepsilon_2} \star f_{\alpha'}) = ((g+1) \star f_{-\varepsilon_2}) \star f_{\alpha'} = 1 \star f_{\alpha'} = 1.$$

Thus, $g \star \phi = 0$, with $\phi = f_{-\varepsilon_2} \star f_{\alpha'}$. In other words, we have a similar result as (A.5) in the proof of D'Haultfœuille (2010). Applying the third step of this proof shows that the location family generated by ϕ is complete. Thus $g = 0$ almost everywhere. Because y_1 was arbitrary, $h(Y_1, Y_2) = 0$ almost surely, and $\text{Ker}\,(T) \cap \mathcal{F} = \{0\}$ $\square$

Proposition 3.1

We first prove that $D \perp\!\!\!\perp Z|Y$ is equivalent to the existence of $b_i \geq 0$, for $i = 1, ..., I$ such that $\forall (i,j), p_{0ij} = b_i p_{1ij}$.

For $i \in [1, I]$, let $A(i)$ the set of j such that $P(Y = i, Z = j) > 0$. The "only if" is obtained by:

$$\begin{aligned}
Z \perp\!\!\!\perp D|Y \quad &\Leftrightarrow \quad \forall (i,j) \in A, \tfrac{1 - P(D=1|Y=i,Z=j)}{P(D=1|Y=i,Z=j)} \text{ does not depend on } j \\
&\Leftrightarrow \quad \forall i, \exists b_i \geq 0 \text{ s.t. } \forall j \in A(i), P(D = 0|Y = i, Z = j) = b_i P(D = 1|Y = i, Z = j) \\
&\Rightarrow \quad \forall i, \exists b_i \geq 0 \text{ s.t. } \forall j, P(D = 0, Y = i, Z = j) = b_i P(D = 1, Y = i, Z = j) \\
&\Rightarrow \quad \forall i, \exists b_i \geq 0 \text{ s.t. } \forall j, p_{0ij} = b_i p_{1ij}
\end{aligned}$$

Reciprocally, if B denotes the set of i such that $P(D = 0, Y = i) = 0$, we have $D \perp\!\!\!\perp Z|Y = i$ for every $i \in B$. For $i \notin B$, the result follows from:

$$
\begin{aligned}
\forall(i,j), \exists b_i \geq 0, p_{0ij} = b_i p_{1ij} \quad &\Rightarrow \quad \forall(i,j), p_{0ij} = \frac{\sum_i p_{0ij}}{\sum_i p_{1ij}} p_{1ij} \\
&\Rightarrow \quad \forall(i,j), p_{0ij} = \frac{P(D=0,Y=i)}{P(D=1,Y=i)} p_{1ij} \\
&\Rightarrow \quad \forall i \notin B, \forall j, P(Z=j|D=0,Y=i) = P(Z=j|D=1,Y=i) \\
&\Rightarrow \quad \forall i \notin B, Z \perp\!\!\!\perp D|Y=i
\end{aligned}
$$

We deduce that the distribution (D, DY, Z) is compatible with $D \perp\!\!\!\perp Z|Y$ if and only if $\forall(i,j), \exists p_{0ij} \geq 0, \exists b_i \geq 0$ such that $\sum_i p_{0ij} = p_{0.j}$ and $p_{0ij} = b_i p_{1ij}$.
The previous condition is also equivalent to $\forall i, \exists b_i \geq 0$ such that $\forall j, p_{0.j} = \sum_i b_i p_{1ij}$, because if such $(b_i)_{i=1,...,I}$ exist, p_{0ij} can always be defined by $p_{0ij} = b_i p_{1ij}$.
Then the maximum likelihood estimator $\widehat{p}$ is defined by:

$$
\widehat{p} = \arg \max_{q \in [0,1]^{(I+1)J}} \sum_{j=1}^{J} \left[n_{0.j} \ln q_{0.j} + \sum_{i=1}^{I} n_{1ij} \ln q_{1ij} \right]
$$
$$
\text{s.t.} \left| \begin{array}{l} \sum_{j=1}^{J} \left[q_{0.j} + \sum_{i=1}^{I} q_{1ij} \right] = 1, \\ \forall i, \exists b_i \geq 0 \ \text{ s.t. } \ \forall j, \sum_{i=1}^{I} q_{1ij} b_i = q_{0.j}. \end{array} \right.
$$

Or equivalently:

$$
\widehat{p} = \arg \max_{(q,b) \in [0,1]^{(I+1)J} \times \mathbb{R}^I} \sum_{j=1}^{J} \left[n_{0.j} \ln q_{0.j} + \sum_{i=1}^{I} n_{1ij} \ln q_{1ij} \right]
$$
$$
\text{s.t.} \left| \begin{array}{ll} \sum_{j=1}^{J} \left[q_{0.j} + \sum_{i=1}^{I} q_{1ij} \right] = 1, & \\ \quad b_i \geq 0 & i = 1, ..., I, \\ \sum_{i=1}^{I} q_{1ij} b_i = q_{0.j} & j = 1, ..., J. \end{array} \right. \qquad (C)
$$

To complete the proposition, remark that $\widehat{p}$ is an efficient estimator of p. The Fisher information matrix of p is singular because $\sum_{j,i} p_{1ij} + \sum_j p_{0.j} = 1$. However the parameter u defined by the $IJ + J - 1$ first components of p has a nonsingular Fisher information matrix. If matrix P_1 has rank I then $p_0 = P_1'(P_1 P_1')^{-1} P_1 p_{0.}$, and then $p_0 = l(u)$ and $\theta = g(p_0, p_1) = k(u)$ with l and k being differentiable at u. Because $\widehat{P}_1$ has rank i with probability tending to one, $\widehat{b}$ is equal to $(\widehat{P}_1 \widehat{P}_1')^{-1} \widehat{P}_1 \widehat{p}_{0.}$ with probability tending to one and then $\widehat{b} = m(\widehat{u})$ with m differentiable with probability tending to one. It follows that $\widehat{b}$, $\widehat{p}_0$ and $\widehat{\theta}$ are efficient estimators of $(P_1 P_1')^{-1} P_1 p_{0.}$, p_0 and θ (see for instance van der Vaart (2000), section 8.9).

Proposition 3.2

First, define the function g as

$$g(h, \Sigma_{11}, \Sigma_{12}, V) = h_1' \Sigma_{11}^- h_1 + \min_{x \geq 0} \left(x - h_2 + \Sigma_{12}' \Sigma_{11}^- h_1 \right)' V^{-1} \left(x - h_2 + \Sigma_{12}' \Sigma_{11}^- h_1 \right), \quad (5.4)$$

where $h = (h_1, h_2)$. Some algebra shows that W_n satisfies $W_n = g(\sqrt{n}\widehat{h}, \widehat{\Sigma}_{11}, \widehat{\Sigma}_{12}, \widehat{V})$, where $\widehat{V} = \widehat{\Sigma}_{22} - \widehat{\Sigma}_{12}' \widehat{\Sigma}_{11}^- \widehat{\Sigma}_{12}$. Then, because Σ is fixed on $\mathcal{H}_{0\Sigma}$, the asymptotic distribution of W_n only depends on the second term in the right-hand side of (5.4). Because $h_2 - \Sigma_{12}' \Sigma_{11}^- h_1 = h_2 \geq 0$ under the null hypothesis, $\lim_{n\to\infty} \Pr(W_n \geq c)$ is maximal over $\mathcal{H}_{0\Sigma} \cap \{0\}^{J-I} \times \mathbb{R}^{+I}$ at $h = 0$ (see Theorem 8.3 of Perlman, 1969). In this case, $U_n \equiv \sqrt{n}\widehat{h} \xrightarrow{\mathcal{L}} U$, with $U \sim \mathcal{N}(0, \Sigma)$. Besides, $\widehat{\Sigma} \xrightarrow{P} \Sigma$. Hence (see, e.g., van der Vaart, 2000, Theorem 2.7) $(U_n, \widehat{\Sigma}) \xrightarrow{\mathcal{L}} (U, \Sigma)$. To obtain our results, it thus suffices to show that g is continuous, and apply the continuous mapping theorem (see, e.g., van der Vaart, 2000, Theorem 2.3).

Let $\mathcal{M}_1$ (resp. $\mathcal{M}_2$) denote the set of positive (resp. positive definite positive) matrices of dimension J and rank $J - I$ (resp. I and I), so that $\Sigma_{11} \in \mathcal{M}_1$, $V \in \mathcal{M}_2$ and with probability approaching one, $\widehat{\Sigma}_{11} \in \mathcal{M}_1$ and $\widehat{V} \in \mathcal{M}_2$. Similarly, let $\mathcal{M}_3$ denote the set of $J \times I$ matrices, so that $\Sigma_{12} \in \mathcal{M}_3$. The generalized inverse function $\Sigma_{11} \mapsto \Sigma_{11}^-$ is continuous on $\mathcal{M}_1$ (see, e.g., Ben-Israel & Greville, 2003). The quadratic programming in (5.4) is also continuous (see, e.g., Lee et al., 2005, Chapter 13). By composition of continuous function, g is thus continuous on $\mathbb{R}^{I+J} \times \mathcal{M}_1 \times \mathcal{M}_3 \times \mathcal{M}_2$, and the result follows $\square$

Appendix B: additional tables

Table 7: Comparison between our method and others on $\widehat{P}(D = 1|Y_1, Y_2)$

	Men			Women		
	Y_2 = Empl.	Y_2 = Unempl.	Y_2 = Out L.F.	Y_2 = Empl.	Y_2 = Unempl.	Y_2 = Out L.F.
IV-MAR						
Y_1 = Empl.	6.11 (<0.0001)	−43.89 (<0.0001)	−31.91 (<0.0001)	6.72 (<0.0001)	−32.55 (<0.0001)	−34.43 (<0.0001)
Y_1 = Unempl.	−10.34 (<0.0001)	34.07 (<0.0001)	−14.86 (0.1038)	−17.31 (<0.0001)	30.23 (<0.0001)	6.61 (0.6227)
Y_1 = Out L.F.	−24.69 (0.0325)	−23.67 (0.0527)	6.20 (0.0021)	−23.33 (0.0364)	−12.66 (0.4504)	4.58 (<0.0001)
IV-HIRR						
Y_1 = Empl.	5.32 (<0.0001)	−25.64 (<0.0001)	−30.13 (0.0001)	6.15 (<0.0001)	−20.14 (0.0025)	−33.00 (<0.0001)
Y_1 = Unempl.	−20.28 (<0.0001)	44.45 (<0.0001)	−22.01 (0.0178)	−23.58 (<0.0001)	38.11 (<0.0001)	2.57 (0.8504)
Y_1 = Out L.F.	−27.58 (0.0175)	−9.24 (0.4595)	5.57 (0.0058)	−25.57 (0.0213)	−2.89 (0.8641)	4.16 (0.0002)

Note: p-value in brackets computed with 1000 bootstrap samples.