Identification and Estimation of a Semiparametric Binary Response Model using Data from Repeated Cross Sections

David H. Pacini *

This Preliminary and Incomplete Version: April 1, 2011

Abstract

We investigate the identification and estimation of a semiparametric single-index dynamic binary response model when data are available from repeated cross sections instead of panels. Unlike the existing literature, we allow for time-varying regressors. Our first contribution is to provide a set of conditions sufficient for the point identification of the coefficients entering the single-index. These conditions include the existence of a continuous regressor with large support, which has been considered in the literature on identification of similar models with panel data but not in the literature on repeated cross sections. Our second contribution is to introduce a semiparametric estimator of the coefficients of interest, and give asymptotic results to justify its use in practice. The estimator is a minimizer of an U-process indexed by finite and infinite dimensional parameters. To the best of our knowledge, the statistical properties of this estimator have not been explored in the literature. We establish conditions sufficient for its consistency and asymptotic normality. Monte Carlo exercises indicate that the estimator performs well in finite sample with respect to existing estimators.

KEYWORDS: Identification; Estimation; Repeated Cross Sections.

JEL CLASSIFICATION: C25, C14

*Phd. student, Toulouse School of Economics | E-mail: davidhpacini@yahoo.com | First Version: July 15, 2010 | Please do not cite or circulate | Comments are welcomed.
1 Introduction

Dynamic binary response models have played an important role in econometrics dating back to the work of Heckman (1981). Identification and estimation of such models would ideally require panel data. In many situations however, panel data are not available but a sequence of independent samples drawn at multiple points in time, so-called repeated cross sections, are. This situation has encouraged investigations on the identification of dynamic binary response models on the basis of repeated cross sections (see the survey by Ridder and Moffit, 2007). These investigations have focused exclusively on models with time-invariant regressors. Time-varying regressors, however, are present in most applications. To rectify this situation, we investigate the identification and estimation of a semiparametric dynamic binary response model with time-varying regressors using data from repeated cross sections.

Our objective is to devise conditions under which the unknown coefficients \((\beta, \gamma)\) in the response equation \(Y_{it} = \mathbb{I}(V_{it} + X_{it}'\beta + \gamma Y_{it-1} + \alpha_i + \varepsilon_{it} \geq 0)\) can be recovered from repeated cross sections on \((Y_t, V_t, X_t', Z)\). Here \(Y_{it}\) is a scalar binary response variable for the observational unit \(i\) at period \(t\), \(\mathbb{I}()\) is the indicator function, \(V_{it}\) is a continuous regressor whose coefficient has been normalized to one, \(X_{it}'\) is a vector of (possible) time-varying regressors, \(\alpha_i\) is an observation specific ('fixed') effect, \(\varepsilon_{it}\) is an idiosyncratic error, and \(Z\) is a time-invariant variable with properties we shall spell out below. The distribution of the fixed effects \(\alpha_i\) and the distribution of the idiosyncratic errors \(\varepsilon_{it}\) are unknown. We do not restrict the dependence between the error terms \((\alpha, \varepsilon_t)\) and the regressors \(X_{it}'\), but we restrict the dependence between \((\alpha, \varepsilon_t)\) and the continuous regressor \(V_t\). In this setting, learning about the coefficients \((\beta, \gamma)\) is not straightforward because joint realizations of the time-varying variables at different points in time, say \((Y_{it}, V_{it}, X_{it}', Y_{it-1}, V_{it-1}, X_{it-1}')\), are unobserved by the sampling design.

We make two contributions. Our first contribution is to provide a set of conditions sufficient for the point identification of the coefficients \((\beta, \gamma)\) when data are available from repeated cross sections. We describe these conditions below. They extend the identification conditions given by Ridder and Moffit (2007) to the case with time-varying regressors. We believe this extension is relevant because time varying regressors are present in most applications. Our second contribution is to introduce a semiparametric estimator of the coefficients of interest \((\beta, \gamma)\), and give asymptotic results to justify its use in practice. The estimator is a minimizer of a U-process indexed by finite and infinite dimensional parameters. To the best of our knowledge, the statistical properties of this estimator have not been explored in the literature. We provide conditions sufficient for its consistency and asymptotic normality. Monte Carlo exercises indicate that the estimator performs well in finite samples.
The main insight behind our identification result is to observe that we can transform the nonlinear response equation $Y_{it} = \mathbb{I}(V_{it} + X_{it}'\beta + \gamma Y_{i,t-1} + \alpha_i + \varepsilon_{it} \geq 0)$ into the linear conditional moment $\mathbb{E}(Y_{i,t}^*|Z) - \mathbb{E}(X_{i,t}'\beta|Z) - \mathbb{E}(\gamma Y_{i,t-1}|Z) - \mathbb{E}(\alpha + \varepsilon|Z) = 0$, after assuming that: (i) The continuous regressor $V_{i,t}$ conditional on the other regressors $X_{i,t}'$ and the time-invariant variable $Z$ is independent of the individual effect $\alpha$, the idiosyncratic error $\varepsilon_{it}$, and the lagged valued of the response variable $Y_{i,t-1}$; (ii) The support of the continuous regressor $V_{i,t}$ conditional on the other regressors $X_{i,t}'$ and the time-invariant variable $Z$ is large. Here $Y_{i,t}^*$ is the random variable $Y_{i,t}^* = [Y_{i,t} - \mathbb{I}(V_{i,t} > 0)] / g_h(V_{i,t}|X_{i,t}, Z)$ with $g_h(V_{i,t}|X_{i,t}, Z)$ the density of the continuous regressor $V_{i,t}$ conditional on $(X_{i,t}, Z)$. Except for the presence of $Y_{i,t}^*$, the latter moment equation is analog to the 'cohort' level equation studied by Deaton (1985) in the context of a fixed effect linear model. Following ideas there, we can 'differentiate out' $\mathbb{E}(\alpha|Z)$ and $\mathbb{E}(\varepsilon|Z)$ by assuming that: (iii) The expectation of the idiosyncratic error $\varepsilon_{it}$ conditional on the observed time-invariant variable $Z$ is constant across time. We end up with a linear moment restriction that depends on the distribution of $(Y_{i,t}, V_{i,t}, X_{i,t}, Z)$ and of $(Y_{i,t-1}, Z)$, but it does not depend either on the distribution of $(Y_{i,t}, Y_{i,t-1})$ or of $(Y_{i,t-1}, V_{i,t}, X_{i,t})$. Hence, the lack of observations of $(Y_{i,t}, V_{i,t}, X_{i,t}', Y_{i,t-1}, V_{i,t-1}, X_{i,t-1}')$ does not longer pose an identification problem. The coefficients of interest can then be uniquely defined in terms of the available data after invoking the following 'order' and 'rank' conditions: (iv) The support of the time-invariant variable $Z$ has at least as many points as the number of unknown coefficients; and (v) The expectation of the regressors $X_{i,t}'$ conditional on the time-invariant variable $Z$ is a nonlinear function. Assumptions (i) and (ii) about the continuous regressor $V_{i,t}$ have not been imposed by previous investigations on the identification of dynamic binary response models with repeated cross sections, whereas assumptions (iii) to (v) are implied by the assumption imposed by Ridder and Moffit (2007). Assumptions (i) and (ii) are akin to the 'special regressor assumption' invoked by Honore and Lewbel (2002) when genuine panel data are available. The difference is that they require $V_{i,t}$ to be independent of $(\alpha, \varepsilon_{it})$ conditional on $(X_{i,t}', Z)$, whereas we require $V_{i,t}$ to be independent of $(\alpha, \varepsilon_{it}, Y_{i,t-1})$ conditional on $(X_{i,t}', Z)$. As a consequence, our assumptions are stronger than those by Honore and Lewbel (2002) because we rule out feedback effects from $Y_{i,t-1}$ to $V_{i,t}$.

Under the maintained identification assumptions, the coefficients $(\beta, \gamma)$ are uniquely defined by conditional moment equalities weighted by the density function of the continuous regressor $V_{i,t}$ conditional on $(X_{i,t}, Z)$. We propose to estimate $(\beta, \gamma)$ using a two-step procedure. In the first step, the density function of $V_{i,t}$ conditional on $(X_{i,t}, Z) g_h(V_{i,t}|X_{i,t}, Z)$ is estimated using kernel methods. In the second step, the finite dimensional coefficients of interest $(\beta, \gamma)$ are estimated by minimizing a criterion function that depends on the first step nonparametric estimate. The criterion function is the sample analog of the quadratic distance from zero of the conditional moments defining the coefficients of interest. It has
an U-process of order two structure, which effectively becomes a third order U-process once the first-step estimator is taken into account. Our criterion function resembles the criterion proposed by Lavergne and Patilea (2010), who study estimation and testing in generic conditional moment models indexed by finite dimensional parameters. The difference is that our conditional moments are indexed by both finite and infinite dimensional parameters.

**Child Labor Supply.** Whether the assumptions we propose to point identify and estimate the coefficients of interest \((\beta, \gamma)\) are reasonable depends on the context. To illustrate the relevance and potential strength of our approach, we discuss in the text how our assumptions apply to an example related to the study of child labor supply in Brazil. In particular, we set up a simple model of parent’s decision to make their children work, where \(\gamma\) measures a fixed job search cost. To the best of our knowledge, panel data measuring child labor force participation both at the rural and urban level are not available in Brazil, but repeated cross-sections are. With the aim of measuring \(\gamma\), we use family income net of child earnings as the special continuous regressor \(V_t\), and the years of education of the head of the households as the time-invariant variable \(Z\). More details on this example are provided in section 2.

**Related Literature.** The question of whether repeated cross sections can be used to make inference on parameters of interest arising from dynamic models has attracted many researchers. Ridder and Moffit (2007) and Verbeek (2008) survey this literature. Most authors have focused on linear models (c.f., Collado, 1998; McKenzie 2004; Vella and Verbeek, 2005; Inoue, 2008). Little is known however about identification and inference on dynamic binary response models with such type of data. Ridder and Moffit (2007) discuss the identification of a dynamic binary response model with time-invariant regressors. We extend the work by Ridder and Moffit (2007) to allow for time-varying regressors. The cost of this extension is that we assume the presence of the continuous special regressor \(V_t\) with the properties discussed above.

Our estimation procedure is related to the literature on estimators defined by optimizing a criterion function with an U-process structure (e.g., Sherman, 1994; Lavergne and Patilea, 2010). Sherman (1994) presents high-level assumptions sufficient for establishing the asymptotic properties of this type of estimators. His high-level assumptions encompass both differentiable and nondifferentiable criterion functions. In our case, the criterion is differentiable. This extra piece of information enables us to avoid much of the technical detail required to verify the assumptions by Sherman (1994). In the context of models defined by conditional moments equalities, Lavergne and Patilea (2010) derive asymptotic properties for estimators defined by optimizing a criterion function with an U-process structure. As alluded before, our problem does not fit their general setup be-
cause the conditional moments here contain an unknown infinite dimensional parameter. We use however several of their insights in the construction of our estimator.

At a more general level, this paper belongs to the literature on models defined by moment equalities weighted by density functions (see e.g., Honore and Lewbel, 2002; Magnac and Maurin, 2007; Jacho-Chavez, 2009; and Khan and Tamer, 2010). Although the problem we are concerned with here is different from those consider by the aforementioned authors, our techniques are similar to theirs and we use several of their insights as well.

**Organization of the Paper.** The outline of the paper is as follows. In the next section, we set out the dynamic binary response model we focus on, and we describe the sampling process generating repeated cross sections. In section 3, we present the main result of the paper. There, we show that the coefficients of interest \((\beta, \gamma)\) are point identified under the assumptions introduced in section 2. In section 4, we introduce the estimator of \((\beta, \gamma)\), and explore its large-sample statistical properties. In Section 5, we present Monte Carlo exercises. They illustrate the implementation of the proposed estimator and its small sample properties. Section 6 concludes. A supplemental document available at [https://sites.google.com/site/davidhpacini/](https://sites.google.com/site/davidhpacini/) collects the proofs.

## 2 The Model and the Data

In this section, we first introduce a concrete example concerning child labor supply. This example motivates the model and the type of data studied in this paper. We continue by describing the semiparametric single-index dynamic binary response model we focus on. Finally, we describe the available data consisting of repeated cross sections.

### 2.1 Motivation: Child Labor Supply

Here we introduce a concrete example that motivates the model and the type of data studied in this paper.

In this example, our aim is to measure the extent of fixed search cost as a determinant of child labor supply in Brazil. We start by building up a model for the parent’s decision to make their children work. This model is based on the seminal work by Basu and Van (1998) on child labor supply. To proceed, let \(Y_{it}\) denote a random variable taking value 1 when at least one of the children in household \(i\) has worked during period \(t\) and zero otherwise. We assume that a household \(i\) at period \(t\) will make his children work if and only if household’s income from non-children labor sources, say \(V_{it}\), is lower than certain subsistence level, say \(L_{it}\), minus the term \(\gamma_0 Y_{i,t-1}\) capturing the idea of a fixed cost that parents incurs when searching a job for their children. Having incurred the search cost, the parent’s choice set for subsequent decisions changes, in the sense that the fact that
they no longer have to incur the search cost as long as the children remain in the state is taken into account in their subsequent sequential decision making. By formalizing the latter idea, we get the following rule to model parent’s decision to make their children $Y_{it} = I(V_{it} \leq L_{it} - \gamma_o Y_{it-1})$. This rule extends the "luxury axiom" by Basu and Van (1998) to consider the fixed search cost $\gamma_o Y_{it-1}$.

To control for determinants of child labor supply other than search costs, we can parametrize the subsistence level as $L_{it} = X_{it}' \beta_o + \alpha_i + \varepsilon_{it}$, where $X_{it}'$ represents a vector of time varying household characteristics, such as the number of adult-equivalent members of the household, and $(\alpha_i, \varepsilon_{it})$ represent, respectively, time-invariant and time-varying household characteristics other than $X_{it}$, such as health and preferences, affecting the subsistence level. Then, our rule to model parent’s decision to make their children becomes $Y_{it} = I(V_{it} - X_{it}' \beta_o + \gamma_o Y_{it-1} - \alpha_i - \varepsilon_{it} \leq 0)$.

Our parameter of interest is $\gamma_o$. It measures the fixed search cost. We expect $\gamma_o$ to be positive. Identification and estimation of $\gamma_o$ would ideally require panel data. To the best of our knowledge, panel data measuring child labor force participation both at the rural and urban level are not available in Brazil, but repeated cross-sections are. The Pesquisa Nacional por Amostra a Domicílio (PNAD) is a repeated survey carried out annually by the Instituto Brasileiro de Geografia e Estatística. It covers close to one hundred thousand different Brazilian households each year. This survey measures child labor force participation $Y_{it}$, household earnings $V_{it}$, household composition $X_{it}'$, and other household characteristics $Z$, such as the head of the household schooling. Nevertheless, it does not measure health and preferences. Then, $\alpha_i$ and $\varepsilon_{it}$ are unobserved.

### 2.2 The Model

Here we describe the semiparametric dynamic binary response model that will be used in the rest of the paper. This model is motivated by the example introduced above.

#### 2.2.1 Notation

To describe the model, we need first to introduce some notation. Let $(\Omega, \mathcal{A}, P_o)$ be some probability space, where $\Omega$ denotes the sample space, $\mathcal{A}$ denotes a sigma-algebra of subsets of $\Omega$, and $P_o$ is a probability measure defined on $\mathcal{A}$. Throughout, we attach the subscript (superscript) $’o’$($’o’$) to any expression to distinguish the ‘true’ value from any other value that expression can take. We use capital letters denote random variables and small letters their corresponding realizations. The notable exceptions are the random variables $\alpha_i$ and $\varepsilon_{it}$, for which the letters $a$ and $e$ are used to denote their realized values. For a generic sequence of random vectors $W_{i1}, \ldots, W_{it}$, we use the notation $W_{it}^t = (W_{i1}, \ldots, W_{it})$. The acronym a.s. is used to indicate that equality among random variables is almost sure.
Whenever convenient, we index a distribution function by the random variables it refers to. We use the notation $G_A(a)$ and $G_{A|B}(a|b)$ to denote the distribution of the random variable $A$ evaluated at $a$ and the distribution of $A$ conditional on $B$ evaluated at $(a,b)$, respectively. The absence of arguments for a function denotes the entire function rather than its value at a point. Thus, for instance, if $Y_t$ is a random variable on $(\Omega, A, P_0)$, $G_{Y_t}(y_t) \equiv P_0(Y_t \leq y_t)$ is the distribution of $Y_t$ induced by $P_0$ evaluated at $y_t$, and $G_{Y_t}$ is the whole distribution of $Y_t$. We use the capital letter $G$ to denote distributions that are 'revealed' by the available data, and $F$ for those which are not. The corresponding small letters $g$ and $f$ are used to denote the densities associated to $G$ and $F$, respectively.

Invariably, $\mathbb{R}$ denotes the real line, and $\mathbb{R}^d$ denotes $d$-dimensional euclidean space. Product space and scalar products share the symbol $\times$. The expression " stands for "defined by". The expression $E$ denotes the expectation operator.

2.2.2 Assumptions

Consider a collection of observational units $N = \{1, \ldots, i, \ldots, N\}$, i.e., individuals, households, etc., to be studied during a time interval $T = \{1, \ldots, t, \ldots, T\}$. We begin by assuming that the researcher has already postulated a theory by which:

Assumption 1 - (Single-Index Dynamic Binary Response Equation). For each observational unit $i$, a scalar binary outcome variable $Y_{it}$ is determined by the equation:

$$Y_{it} \equiv \mathbb{I}(V_{it} + X'_{it}\beta_o + \gamma_o Y_{it-1} + \alpha_i + \varepsilon_{it} \leq 0) \ a.s. \quad (A1)$$

where $\mathbb{I}(C)$ is the indicator function that takes on value one if condition $C$ is satisfied and zero otherwise, $V_{it}$ is a continuous regressor whose coefficient has been normalized to one (if necessary by multiplying it by -1), $X'_{it}$ is vector of regressors of dimension $d_X$, $(\beta_o, \gamma_o)$ are unknown parameters to be estimated from data, and $(\alpha_i, \varepsilon_{it})$ are the individual effect and idiosyncratic disturbance, respectively. We interpret $(Y_{it}, X_{it}, \alpha_i, \varepsilon_{it})$ as random variables defined on the probability space $(\Omega, A, P_0)$.

As already said, our aim is to recover the coefficients $(\beta_o, \gamma_o)$ from repeated cross section data on $(Y_t, V_t, X_t, Z)$. In order to do so, and following common practice in econometrics, we continue by imposing restrictions on the dependence between the error terms $(\alpha, \varepsilon_t)$ and the variables $(V^T, X^T, Z)$. To motivate and justify these restrictions, we find useful to first recall the strategies available to point identify the coefficients $(\beta_o, \gamma_o)$ when genuine panel data are available. We shall only consider independent and identically distributed samples. We then drop the subscript $i$ from the random variables except when some confusion can arise. With panels, point identification of dynamic binary response models can be achieved after assuming that the idiosyncratic error $\varepsilon_t$ follows a logistic distribution (see Magnac, 2000), or by matching regressors at different time periods (see...
Honore and Kyriazidou (2000), or after assuming the presence of a special continuous regressor with large support (see Honore and Lewbel, 2002). None of these strategies directly apply to our case. The strategy by Magnac (2000) is specific to the case where the lagged outcome is the only regressor i.e., it requires $\beta_o = 0$. The strategy by Honore and Kyriazidou (2000) requires the regressors to have same support across time periods. Since we allow for time dummies among the regressors $X'_t$, which clearly have different support, this latter approach does not apply here. The strategy by Honore and Lewbel (2002) requires data to reveal the conditional density of $V_t$ given $(X_t, Z, Y_{t-1})$, say $f_{V_t|X_t, Z, Y_{t-1}}$, which is not the case when data are available from repeated cross sections. We shall show, however, that modification of Honore and Lewbel’s (2002) strategy leads to point identification of $(\beta_o, \gamma_o)$ when repeated cross sections are available.

With the aim of point identifying $(\beta_o, \gamma_o)$ when data are available from repeated cross sections, we impose the following restrictions on the dependence between the error terms $(\alpha, \varepsilon^T)$ and the variables $(V^T, X^T, Z)$:

**Assumption 2 - (Identification).** Assume that we observe replications of a time-invariant random variable $Z$ in the data. We suppose that the variables $V_t$ and $Z$ satisfy:

1. $F^o_{\alpha+\varepsilon_t, Y_{t-1}|V_t, X_t, Z} = F^o_{\alpha+\varepsilon_t, Y_{t-1}|X_t, Z}$ for all $t$ in $T$.
2. The support of $V_t$ given $(X_t, Z)$ is $[v_{tL}^i, v_{tH}^i]$ for $-\infty \leq v_{tL}^i < 0 < v_{tH}^i \leq \infty$, and the support of $X'_t \beta_o + \gamma_o Y_{t-1} + \alpha + \varepsilon_t$ is a subset of $[v_{tL}^i, v_{tH}^i]$. The conditional distribution $G^o_{V_t|X_t, Z}$ is continuous with positive density $g^o_{V_t|X_t, Z}$.
3. $\mathbb{E}(\varepsilon_t|Z) = \mathbb{E}(\varepsilon_{t-1}|Z)$ a.s. for all $t$ in $T$.
4. The support of $Z$, $Z$, has at least $d_X + 1$ points, where $d_X$ is the dimension of $X$. The function $z \mapsto \mathbb{E}(X_t|Z = z)$ is nonlinear.

We now discuss (A2). Assumption (A2.i) says that the continuous regressor $V_t$ and the random vector $(\alpha + \varepsilon_t, Y_{t-1})$ are independent conditional on the regressors $X_t$ and the time-invariant variable $Z$. This conditional independence assumption is similar to the conditional independence restriction imposed by Honore and Lewbel (2002). The difference is that they require $V_t$ to be independent of $(\alpha, \varepsilon_t)$ conditional on $(X'_t, Z)$, whereas we require $V_t$ to be independent of $(\alpha + \varepsilon_t, Y_{t-1})$ conditional on $(X'_t, Z)$. Note that (A2.i) rules out feedback effect from lagged dependent variables to current and future values of the continuous regressor $V_t$. This contrast with the model by Honore and Lewbel (2002) where feedback effects are permitted. This is part of the price we propose to pay to extend the setting proposed by Honore and Lewbel (2002) to the case where repeated cross sections instead of genuine panel data are available. In the special case of $\gamma_o = 0$ (i.e., a semiparametric static binary response model), the conditional independence assumptions (A2.i) can be replaced by independence of $V_t$ and $(\alpha, \varepsilon_t)$ conditional on $(X_t, Z)$. In such a case, (A2.i) reduces to the conditional independence assumption by Honore and Lewbel (2002). Within the child labor supply context illustrated above, assumption (A2.i)
requires household’s income net of children earnings $V_t$ to be independent, conditional on the number of members of the household $X_t$ and the educational level of the head of the household $Z_t$, of the unobserved determinants of the subsistence level $(\alpha, \varepsilon_t)$, and past child labor force participation $Y_{t-1}$. This condition may fail, for instance, if we suspect that household income and unobserved determinants of the subsistence level, such as health status, are correlated.

Assumption (A2.i) says that the continuous regressor $V_t$ has large support. It is equivalent to the large support condition imposed by Honore and Lewbel (2002). This support assumption is quite common in the literature on semiparametric binary response models. Nevertheless, Magnac and Maurin (2007) show that such an assumption represents a potential obstacle to many empirical applications because it requires the distribution of $Y_t$ conditional on $(Y_{t-1}, V_t, X_t, Z_t)$, say $F_{Y_t|Y_{t-1},V_t,X_t,Z_t}$, to vary from 0 to 1 when $V_t$ varies over its support. It is easy to imagine examples where this requirement is not met. However, the large support condition seems a plausible assumption, for instance, in our example concerning child labor supply. There, (A2.ii) says that every subsistence level $L_t$ is affordable by some level of income $V_t$.

Assumption (A2.iii) restricts the dependence between the idiosyncratic errors $\varepsilon_t$ and the time-invariant variable $Z_t$. It is the same condition on a time-invariant variable defining a 'cohort', as required by Ridder and Moffit (2007, p. 5515). In the child labor supply context introduced above, this assumption requires that the unobserved determinants of the subsistence level $(\alpha, \varepsilon_t)$ are mean independent of the educational level of the head of the household $Z_t$. This condition may fail, for instance, if we suspect that average preferences are different across time for households with the same educational level.

Assumption (A2.iv) prescribes a lower bound on the number of points in the support of time-invariant variable $Z_t$ and restricts the dependence between the regressors $X_t$ and the time-invariant variable $Z_t$. This restriction is testable. It can be interpreted as a rank condition. In the context of our example, it requires that $Z$ measures at least two levels of schooling (because the dimension of $(\beta_o, \gamma_o)$ is 2), and that the number of adult equivalent members of the household $X_t$ is nonlinearly related to the level of education of the head of the household $Z_t$. It is worth to say that assumptions (A2.iii) and (A2.iv) about the time-invariant variable $Z$ are specific to this paper, in the sense that they are stronger than necessary to get point identification in the panel data case (c.f., Honore and Lewbel, 2002). They, together with the lack of feedback effects alluded before, are the price to pay for extending the setting by Honore and Lewbel (2010) to the case at hand.

We do not impose parametric restrictions on the distribution of $(\alpha, \varepsilon_t)$, hence our model is semiparametric. We let $\alpha$ and $\varepsilon_t$ to depend on the regressors $X_{it}$ in an arbitrary way. Heteroskedasticity of general form (on $X_{it}$) is thus permitted. Our model is of the "fixed effects" type in the terminology by Arellano and Honore (2001).
2.3 The Data: Repeated Cross Sections

Here we describe data consisting in repeated cross sections. This description is useful to understand the identification problem we face.

We consider a situation in which panel data are unavailable. We assume instead that data are available from $T$ independent samples drawn from the collection of observational units $\mathcal{N}$ at multiple points in time. We begin by describing repeated cross sections made up of two independent samples. At $t = 1$, a random sample of $n_1$ individuals is drawn from $\mathcal{N}$. For each of the $n_1$ selected individuals, the values of the outcome variable $Y_t$, the regressors $(V_t, X_0)$, and additional time-invariant variables $Z$ are measured and recorded. Let $\{y_{i1}, v_{i1}, x_{i1}, z_i\}_{i=1}^{n_1}$ be the list containing the records for these $n_1$ individuals. This list is called a cross section. In order to use the statistical inference approach, we interpret this cross section as $n_1$ realizations of the independent and identically distributed (iid) random vector $(Y_1, V_1, X_0, Z)$. At $t = 2$, the $n_1$ individuals drawn at $t = 1$ are not observed again. Instead, a new independent sample of $n_2$ individuals is drawn with replacement from $\mathcal{N}$ yielding a new cross-section $\{y_{i2}, v_{i2}, x_{i2}, z_i\}_{i=n_1+1}^{n_1+n_2}$. The samples $\{Y_{i1}, V_{i1}, X_{i1}, Z_i\}_{i=1}^{n_1}, \{Y_{i2}, V_{i2}, X_{i2}, Z_i\}_{i=n_1+1}^{n_1+n_2}$ are what we call repeated cross sections. Suppose now that the process of drawing anew independent samples is performed for the remaining $t \in T$ periods. We end up thus with the sequence of $T$ repeated cross sections drawn from $\mathcal{N}$.

The following assumption sums up the sampling process generating repeated cross sections:

Assumption 3 (Repeated Cross Sections). The data are available from $T \geq 3$ independent samples:

$$\{Y_{i1}, V_{i1}, X_{i1}, Z_i\}_{i=1}^{n_1}, \ldots, \{Y_{iT}, V_{iT}, X_{iT}, Z_i\}_{i=n_{T-1}+1}^{n_T}$$

(A3)

containing independent and identically distributed (iid) replications of the random vectors $(Y_1, V_1, X_1, Z), \ldots, (Y_T, V_T, X_T, Z_t)$ generated according to assumptions (A1)-(A2). The positive integer $n \equiv n_1 + \ldots + n_T$ denotes the total number of individuals drawn from $\mathcal{N}$.

Notice that joint realizations of the time varying variables for the same observational unit, say $(Y_{it}, V_{it}, X_{it}, Y_{is}, V_{is}, X_{is})$ for $t \neq s$, are never observed. Such absence of observations is not related to the sample size $n$, but to the sampling scheme generating the available data.
3 Identification Results

Here we show that the assumptions imposed in the previous section on the continuous regressor $V_t$ and the time-invariant variable $Z$ have the power to uniquely define the coefficients $(\beta_0, \gamma_0)$ when data are available from repeated cross sections. This is the main result of the paper.

The main insight behind our identification strategy is to observe that assumptions (A2.i) and (A2.ii) about the continuous regressor $V_t$ enable us to construct a linear moment condition from the response equation (A1). To get point identification of the coefficients of interest, we ‘differentiate out’ from this linear moment equation the terms related to the individual effect $\alpha$ and the idiosyncratic error $\varepsilon_t$. To do so, we invoke assumptions (A2.iii) and (A2.iv) about the time-invariant variable $Z$.

To proceed, define the random variable $Y_{it}^*$ by:

$$Y_{it}^* = \frac{[Y_{it} - \mathbb{I}(V_{it} > 0)]}{g_{V_{it}|X_t,Z}(V_{it}|X_{it},Z_i)} \text{ a.s.}$$

where $g_{V_{it}|X_t,Z}$ denotes the density of $V_t$ conditional on $(X_t, Z)$. The variable $Y_{it}^*$ is similar to the one constructed by Honore and Lewbel (2002). The difference is that here $g_{V_{it}|X_t,Z}$, instead of $f_{Y_{it}|Y_{t-1},X_t,Z}$, is required in the definition of $Y_{it}^*$. The following lemma shows that the single-index restriction embodied in the response equation (A1), and the restrictions (A2.i) and (A2.ii) on the continuous regressor $V_t$, allow us to write the expectation of $Y_{it}^*$ conditional on $Z$ as an additive separable function of $(X_t, Z), (Y_{t-1}, Z)$, and $(\alpha, \varepsilon_t, Z)$.

**Lemma 1** Let assumptions (A1), (A2.i) and (A2.ii) hold. Define $Y_{it}^*$ as in (1). Then,

$$E(Y_{it}^*|Z) = \beta_0 E(X_t|Z) + \gamma_0 E(Y_{t-1}|Z) + E(\alpha|Z) + E(\varepsilon_t|Z) \text{ a.s.}$$

for all $t$ in $\{2, \ldots, T\}$.

**Proof.** See the supplemental document. ■

Except for the presence of $Y_{it}^*$ in the right hand side, equation (1) is analog to the 'cohort' level equation studied by Deaton (1985) in the context of a fixed effect linear model. Following ideas there, we can 'differentiate out' $E(\alpha|Z)$ and $E(\varepsilon_t|Z)$ by invoking assumptions (A2.iii)-(A2.iv). This in turn yields the following proposition:

**Proposition 1** (Point Identification of the Coefficients $\beta_0, \gamma_0$) Let assumptions (A1), (A2) and (A3) hold. Then, the coefficients $(\beta_0, \gamma_0)$ are uniquely defined by:

$$E[Y_{it}^* - Y_{i,t-1}^* - (X_t - X_{t-1})\beta_0 - \gamma_0(Y_{t-1} - Y_{t-2})|Z] = 0 \text{ a.s.}$$

(CMR)

for all $t$ in $\{2, \ldots, T\}$. 

11
Proof. See the supplemental document.

According to Proposition 1, assumptions (A1)-(A2) and three independent cross sections on \((Y_t, V_t, X_t, Z)\) yield point identification of the coefficients \((\beta_o, \gamma_o)\). Proposition 1 is constructive in that the conditional moments (CMR) can be employed, by way of the analog principle, to construct a consistent estimator of \((\beta_o, \gamma_o)\). We discuss estimation of \((\beta_o, \gamma_o)\) but in the section 4.

The identification result in Proposition 1 can be extended to the case where \(q\) lags of the response variable enter, in additive form, in the response equation (A1). In such a case, the support of \(Z\) needs to have at least \(d_X + q\) points. It is also worth saying that there is an issue related to the robustness of the identification result in Proposition 1.

Identification of parameters by means of weighted moment conditions, such as (CMR), are known to be sensitive to conditions imposed on the support of the random variables (see Magnac and Maurin, 2007). Indeed, point identification of the coefficients \((\beta_o, \gamma_o)\) is lost for the case at hand if the support of \(V_t\) given \((X_t, Z_t)\) is not a superset of the support of \(X_t^o + Y_{t-1} + \alpha + \varepsilon_t\), i.e. if the large support assumption (A2.ii) is violated. In such case, the model and the available data contain no information about the coefficients of interest, in the sense that the identified set of \((\beta_o, \gamma_o)\) is the parameter space.

4 Inference Procedures

In this section, we propose a semiparametric two-step estimator of the coefficients of interest and explore its statistical properties. Since the section is long, an outline of the topics covered is in order. We introduce the two-step estimator, next. Then, we provide a number of sufficient conditions under which the proposed estimator is consistent and asymptotically normal. We also establish the semiparametric efficiency bound for the coefficients of interest.

4.1 Motivation

We aim at estimating the coefficients \((\beta_o, \gamma_o)\) defined by the conditional moment restrictions (CMR) in proposition 1 with at hand independent replications of \((Y_1, V_1, X_1, Z)\), \((Y_T, V_T, X_T, Z)\). There are three sort of unknown parameters in (CMR): the finite dimensional coefficients of interest \((\beta_o, \gamma_o)\); the density \(g_{Y_t|X_t,Z}^o\) in the definition of \(Y_t^*\); and the densities \(g_{Y_t,V_t,X_t|Z}^o\), \(g_{X_t|Z}^o\), and \(g_{Y_{t-1}|Z}^o\) defining the moment restriction (CMR). Since we do not impose parametric restrictions on \(\left(g_{Y_t,V_t,X_t|Z}^o, g_{Y_{t-1},V_{t-1},X_{t-1}|Z}^o\right)\), our estimation problem is semiparametric in nature.

General methods for estimating parameters of interest defined by conditional moment restriction containing unknown nuisance functions are already available in Ai and
Chen (2003) and Otsu (2011). Applying Ai and Chen (2003) approach to our setting involves to approximate the conditional density function $g_{V_t|X_t,Z}$, the conditional expectations $E[Y_t|Z]$, $E[X_t|Z]$, $E[Y_t|Z]$ using sieve methods, and then to estimate the coefficients $(\beta_o, \gamma_o)$ and the sieve parameters jointly by applying the method of minimum distance subject to the restriction that the sieve approximation of the density $g_{V_t|X_t,Z}$ integrate to one. Applying Otsu (2011) approach to our setting involves to approximate the conditional density function $g_{V_t|X_t,Z}$ also by sieve methods, and then to estimate the the coefficients $(\beta_o, \gamma_o)$ and the sieve parameters jointly by applying the method of empirical likelihood subject to the restriction that the sieve approximation of the density $g_{V_t|X_t,Z}$ integrate to one. Implementing these approaches can be costly from a computational point of view because it requires solving constrained nonlinear optimization problems of increasing complexity with the sample size.

We propose to estimate the coefficients $(\beta_o, \gamma_o)$ in two steps. In the first step, the nonparametric component of the variable $Y_t$ (i.e., the conditional density function $g_{V_t|X_t,Z}$) is estimated using kernel methods. In the second step, the coefficients of interest $(\beta_o, \gamma_o)$ are estimated by minimizing a criterion function that depends on the first-step estimates. This two-step procedure enables us to overcome part of the practical disadvantages associated to the one-step approaches. This is because we do not have to solve constrained nonlinear optimization problems of increasing complexity with the sample size. We provide the details of each step, next.

4.2 Description of the Estimator

Here we introduce a semiparametric two-step estimator for the unknown coefficients of interest $(\beta_o, \gamma_o)$. The estimator, which is based on the analog principle, is, as far as we know, novel.

4.2.1 Preliminaries

To streamline the description of our estimator, we impose two additional restrictions. First, we assume that $X_{t}'$ is a continuous random variable, and $Z$ is a discrete random variable. The case of a random vector $X_{t}'$ with discrete/continuous components and/or continuous time-invariant variables $Z$ can be accommodated, as we discuss below, at the cost of a more cumbersome notation. Second, we strengthen the assumption that the expectation of the sum of the individual effects $\alpha$ and the idiosyncratic error $\varepsilon_t$ conditional on time-invariant variables $Z$ is constant, i.e., $E(\alpha + \varepsilon_t|Z) = E(\alpha + \varepsilon_s|Z)$ for $t \neq s$, by assuming that the sum of $\alpha$ and $\varepsilon_t$ is mean independent of the time-invariant variable $Z$, i.e. $E(\alpha + \varepsilon_t|Z) = 0$. The following assumption collects these restrictions:

**Assumption 4 (Continuous $X_t$, Discrete $Z$).** For ease of exposition, we assume that:
(A.4.i) For all \( t \) in \( T \), the support of the regressor \( X_t, X_t, \) is a convex subset of \( \mathbb{R} \). The support of the time-invariant variable \( Z \) is \( Z = \{0, ..., \bar{z}\} \).

(A.4.ii) For all \( t \) in \( T \), \( \mathbb{E}(\alpha + \epsilon_t | Z) = 0 \).

As a consequence of the restriction (A.4.ii), the coefficients of interest \((\beta_o, \gamma_o)\) are uniquely defined by the conditional moment restriction:

\[
\mathbb{E}(Y_t^* - X_t'\beta_o - \gamma_oY_{t-1} | Z) = 0 \quad a.s.,
\]

(2)

The restriction (A.4.ii) simplifies thus the notation because we do not need to work with the conditional expectations in first differences (CMR).

### 4.2.2 First Step: Kernel Density Estimation

The first step involves nonparametrically estimating the conditional density function \( g_{v_{i|X_t,Z}}^{o} \). We propose to estimate this function using kernel methods. There are alternative methods for estimating a conditional density nonparametrically. For instance, \( g_{v_{i|X_t,Z}}^{o} \) could be estimated using sieve methods. Our choice of a kernel estimate for the conditional density \( g_{v_{i|X_t,Z}}^{o} \) is due to the fact that its statistical properties are relatively well established, whereas the statistical properties of sieve estimates for conditional density functions appear to be less well known (see Ichimura and Todd, 2007).

To describe this first step estimator, we need to introduce some additional notation. Let \( v \mapsto K(v) \) denote a bounded symmetric univariate kernel function and \( \{a_{ni}\} \) a sequence of non-negative real numbers (bandwidth). For each \( v \) in \( \mathbb{R} \), write \( K_a(v) \) for \( a_n^{-1} \times K(v/a_n) \). Do likewise with \( K_b(x) \), for the bandwidth \( \{b_{ni}\} \) and \( x \) in \( \mathbb{R} \). For a sequence of non-negative real numbers \( \{c_{ni}\} \) in \([0, (\bar{z} - 1)/\bar{z}]\), let \( z \mapsto L_c(z, Z_l) \) denote the discrete univariate kernel function \( L_c(z, Z_l) = (c_{ni}/(\bar{z} - 1))^{1-\delta(Z_l=z)} \times (1 - c_{ni})^{\delta(Z_l=z)} \), where \( \bar{z} \) is the largest value in the support of \( Z \).

We propose to estimate the conditional density \( g_{v_{i|X_t,Z}}^{o} \) by:

\[
\hat{g}_{v_{i|X_t,Z}}(v, x, z) \equiv \hat{g}_{v_{i|X_t,Z}}(v, x, z) = \frac{\hat{g}_{v_{i|X_t,Z}}(v, x, z)}{\hat{g}_{X_t,Z}(x, z)}
\]

where

\[
\hat{g}_{v_{i|X_t,Z}}(v, x, z) \equiv (n_t - 2)^{-1} \sum_{l=1, l\neq i, l\neq j}^{n_t} K_a(V_{it}, v) \times K_b(X_{lt}, x) \times L_c(z, Z_l)
\]

is the leave-two-out kernel estimator of the density \( g_{v_{i|X_t,Z}}^{o}(v, x, z) \), and

\[
\hat{g}_{X_t,Z}(x, z) \equiv (n_t - 2)^{-1} \sum_{l=1, l\neq i, l\neq j}^{n_t} K_b(X_{lt}, x) \times L_c(z, Z_l)
\]
is the leave-two-out kernel estimator of the density $g_{X_t,Z}^o(x,z)$. Leaving two observations out is convenient for our theoretical developments below. Note that the same kernel $K_v(X_t,x) \times L_o(z,Z_t)$ is used to estimate $g_{V_t,X_t,Z}^o$ and $g_{X_t,Z}^o$. This is done for convenience.

4.2.3 Second Step: Smooth Minimum Distance Estimation

The second step involves estimating the coefficients of interest $(\beta_o, \gamma_o)$ by minimizing a criterion function. Following insights in Lavergne and Patilea (2010), we employ as a criterion function the sample analog of the quadratic distance from zero of the conditional moments (2), where the conditional density $g_{V_t|X_t,Z}^o$ is replaced by its first step estimate. Alternative criterion functions can be employed. For instance, it is possible to construct first unconditional moments using some particular functions of the conditioning variables, and then setting the criterion function to be a quadratic form of the sample analog of these unconditional moments. This is the so-called generalized method of moment criterion. With this criterion, the selection of the number of unconditional moments to be used can be a vexing problem. The Monte Carlo exercises below borne out this problem. Our choice of the criterion function is intended to avoid this drawback.

To describe the second step, abbreviate $\hat{g}_{V_t,X_t,Z}$ to $\hat{g}_t$ and $g_{V_t|X_t,Z}^o$ to $g_t^o$. Define the vectors $\hat{g} \equiv (\hat{g}_2, ..., \hat{g}_T, \hat{g}_T)$ and $g^o \equiv (g^o_2, ..., g^o_T, g^o_T)$. Let $\binom{n_t}{2} \equiv n_t \times (n_t - 1)$ be the number of permutations of $n_t$ taken two elements at time; $\hat{Y}^*_it \equiv \frac{Y_{it} - \hat{E}(V_{it}|X_{it},Z_t)}{\hat{g}_i(X_{it},Z_t)}$; and $\varpi_{ij}$ be the scalar weight $\varpi_{ij} \equiv \varpi(Z_i)^{-1/2} \times K_{ij} \times \varpi(Z_j)^{-1/2}$ with $z \mapsto \varpi(z)$ the value of a known positive transformation of $Z$, such as $\varpi(Z) = 1$, and $K_{ij} \equiv \mathbb{I}(Z_i = Z_j)$. Define also the function $\hat{m}_t(\beta, \gamma, \hat{g}_t)$ by:

$$\hat{m}_t(\beta, \gamma, \hat{g}_t) = \frac{1}{2} \left( \frac{n_t}{2} \right)^{-1} \sum_{i=n_{t-1}+1}^{n_t} \sum_{i<j} \left( \hat{Y}^*_it - X_{it}^t \beta \right) \times \varpi_{ij} \times \left( \hat{Y}^*_jt - X_{jt}^t \beta \right)$$

$$+ \frac{1}{2} \left( \frac{n_t-1}{2} \right)^{-1} \sum_{i=n_{t-2}+1}^{n_{t-1}} \sum_{i<j} \gamma Y_{it-1} \times \varpi_{ij} \times Y_{jt-1} \gamma$$

$$- \frac{1}{n_t \times (n_t - 1)} \sum_{i=n_{t-1}+1}^{n_t} \sum_{j=n_{t-1}+1}^{n_{t-1}} \left( \hat{Y}^*_it - X_{it}^t \beta \right) \times \varpi_{ij} \times Y_{jt-1} \gamma,$$

Let $\hat{m}(\beta, \gamma, \hat{g}) \equiv (\hat{m}_2(\beta, \gamma, \hat{g}_2), ..., \hat{m}_T(\beta, \gamma, \hat{g}_T))^{1/2}$ be a $(T - 1)$ dimensional vector containing the $\hat{m}_t(\beta, \gamma, \hat{g}_t)$ functions defined above.

The estimator $(\hat{\beta}, \hat{\gamma})$ of the coefficients of interest $(\beta_o, \gamma_o)$ we propose is:

$$(\hat{\beta}, \hat{\gamma}) = \arg \min_{\beta, \gamma} \frac{1}{2} \hat{m}(\beta, \gamma, \hat{g})^T \hat{m}(\beta, \gamma, \hat{g})$$ (2SSMMD)

where the vector $\hat{m}(\beta, \gamma, \hat{g})$ has been defined in the previous paragraph, and the scaling factor $1/2$ is added to facilitate the theoretical results below. This is the second step of
the estimator. We denote by \( \hat{C}(\beta, \gamma, \hat{g}) \equiv \hat{m}(\beta, \gamma, \hat{g})' \tilde{m}(\beta, \gamma, \hat{g}) \) the criterion function in (2SSMD).

### 4.2.4 Discussion

The guiding principle for choosing the criterion \( \hat{C}(\beta, \gamma, \hat{g}) \) in (2SSMD) is the fact that the true value of the coefficients of interest \((\beta_0, \gamma_0)\) minimize the limit in probability of \( \hat{C}(\beta, \gamma, \hat{g}) \). This is because the \( t \)-th element of the vector \( \hat{m}(\beta, \gamma, \hat{g}) \) converges in probability to the expectation of the product between \( \varphi(Z) \) and the square of the conditional restrictions (2), that is, \( \hat{m}(\beta, \gamma, \hat{g}) \rightarrow P \varphi(Z) E(Y_t^* - X_t\beta - \gamma Y_{t-1}|Z)^2 \). The criterion \( \hat{C}(\beta, \gamma, \hat{g}) \) estimates thus a sum of weighted distance of the sample analog of the conditional moment restrictions (2) from zero. Since \((\hat{\beta}, \hat{\gamma})\) renders the vector \( \hat{m}(\beta, \gamma, \hat{g}) \) closest to zero in the metric associated with the scalar product defined by the identity matrix, and \( \hat{g} \) is a first step estimate, we say that \((\hat{\beta}, \hat{\gamma})\) is a two-step smooth minimum distance estimator (2SSMD). Unlike the corresponding GMM criterion, \( \hat{C}(\beta, \gamma, \hat{g}) \) does not depend on a user’s choice parameter selecting unconditional moments to be employed in the estimator of the coefficients of interest.

Our approach in some sense resembles the two-step estimator based on genuine panel data by Honore and Lewbel (2002). In their setting, the coefficients of interest are defined by unconditional moment restrictions weighted by a conditional density function. They also nonparametrically estimate a conditional density function by using kernel smoothing in the first step. Their second step minimizes a weighted distance of the sample analog of the unconditional moment restrictions defining the parameters of interest. Since our coefficients of interest are defined by conditional moment restrictions, the second step of the estimator by Honore and Lewbel (2000) does not directly apply here. This motivates us to introduce the criterion function \( \hat{C}(\beta, \gamma, \hat{g}) \) in (2SSMD).

The estimator presented above can be extended to cover the case with discrete regressors \( X_t \) and/or continuous conditioning variable \( Z \). The case with discrete regressors \( X_t \) can be accommodated by replacing \( K_b \) in the first step with a discrete kernel, say \( L_b \). The case with continuous conditioning variable \( Z \) requires two modifications. First, the kernel \( L_c \) in the first step needs to be replaced with a continuous kernel, say \( K_c \). Second, the factor \( K_{ij} \) in the function \( \hat{m}_t(\beta, \gamma, \hat{g}_t) \) should be set equal to \( K_c(Z_i - Z_j) \) instead of \( I(Z_i = Z_j) \). In this case, the function \( \hat{m}_t(\beta, \gamma, \hat{g}_t) \) becomes a kernel weighted U-process.

### 4.3 Asymptotic Properties

Having introduced an estimator for the coefficients of interest, it is desirable to determine its exact distribution. Here we establish conditions under which our two-step smooth minimum distance estimator \((\hat{\beta}, \hat{\gamma})\) of \((\beta_0, \gamma_0)\) is consistent, ‘root-n’ consistent, and its
exact distribution can be approximated by a normal distribution.

Before going to the results, it is important to clarify the type of asymptotic approximation we consider. In the applications we have in mind, the number of repeated cross sections is small with respect to the number of observations in each of them. Then, our asymptotic approximation shall refer to the case where the number of observational units \( n \) goes to infinity while the total number of time periods \( T \) remain constant. In particular, the limit situation we shall consider suppose that each of the sample sizes \( n_t \) tends to infinity, all at the same rate. We consider, therefore, sequences of sample sizes \( \{n_{t,k}\}_{k=1}^{\infty} \) with total sample size \( n_k = \sum_{t=1}^{T} n_{t,k} \) such that \( \lim_{k \to \infty} \frac{n_{t,k}}{n_k} = \delta_t \), where \( \{\delta_t\}_{t=1}^{T} \) is a sequence of positive numbers satisfying \( \sum_{t=1}^{T} \delta_t = 1 \). For ease of exposition, in what follows we drop the subscript \( k \) from the sequences of sample sizes.

### 4.3.1 Consistency

To establish consistency, our strategy is to take the usual route and show that the criterion function \( \hat{C}(\beta, \gamma, \hat{g}) \) converges uniformly to the nonrandom function \( C(\beta, \gamma; g_o) \equiv \mathbb{E}[\hat{m}(\beta, \gamma, g_o)] \mathbb{E}[\hat{m}(\beta, \gamma, g_o)] \). The fact that this nonrandom function attains a unique maximum at \((\beta_o, \gamma_o)\) will implies in turns that \((\beta, \hat{\gamma})\) is consistent.

Generally speaking, to prove that \( \hat{C}(\beta, \gamma, \hat{g}) \) converges uniformly to \( C(\beta, \gamma, g_o) \) is enough to show that \( \beta, \gamma, g \mapsto \hat{m}(\beta, \gamma, g) \) is bounded and that the first-step estimator \( \hat{g}_t \) converges in probability to the conditional density \( g^o_{V_t|X_t,Z} \) in the suitable way. The following restrictions are convenient to fulfill these requirements:

**Assumption A5 (Sufficient Conditions for the Consistency of \((\beta, \hat{\gamma})\)).**

(A5.i) Let \( \delta \) be some positive number. Let \( \mathcal{V}_t \) and \( \mathcal{X}_t \) denote, respectively, the support of \( V_t \) and \( X_t \). Let \( \mathcal{G}_t^\delta \) denote the class of bounded away from zero functions \( g_t : \mathcal{V}_t \times \mathcal{X}_t \mapsto [\delta, 1] \), whose partial derivatives up to order \( \lambda > (d_X + 1)/2 \) exist and are uniformly bounded by some constant. For every \( z \) in \( Z \), the function \( v, x \mapsto g^o_{V_t|X_t,Z}(v|x,z) \) belongs to \( \mathcal{G}_t^\delta \). The coefficients \( (\beta_o, \gamma_o) \) belong to the compact space \( \mathcal{B} \times \Gamma \) included in \( \mathbb{R}^2 \). The random vector \((V^T, X^T, Z, \varepsilon^T, \alpha)\) has finite second moments.

(A5.ii) For all \( t \), \( \sum_{y=0}^{1} \int_{X_t} \int_{Z_t} \frac{g^o_{V_t|X_t,Z}(y,v,x,z)}{g^o_{V_t|X_t,Z}(v,v,x,z)} g^o_{X_t,Z}(x,z) \delta dxdzdv < \infty \)

(A5.iii) The weighting function \( z \mapsto \omega(z) \) belongs to the space \( \mathcal{W} \) of bounded positive functions with finite bracketing number \( N_{11}^{(1)}(\varepsilon, \mathcal{W}, P) \).

(A5.iv) The kernel functions \( v \mapsto K_a(v) \) and \( x \mapsto K_b(x) \) satisfy \( \int_v K_a(v) dv = \int_x K_b(x) dx = 1 \), \( \int_v v K_a(v) dv = \int_x x K_b(x) dx = 0 \), and \( K_a(v) = K_a(-v) \), \( K_b(x) = K_b(-x) \). The bandwidths \( a = a_n \), \( b = b_n \) and \( c = c_n \) satisfy \( a_n = o(1) \), \( b_n = o(1) \), \( c_n = o(1) \), \( \lim_{n \to \infty} na_n b_n c_n = \infty \).

We now discuss this assumption. Assumption (A5.i) defines the parameter space for the conditional density \( g^o_{V_t|X_t,Z} \), and for the coefficients \((\beta_o, \gamma_o)\). Notice that we restrict
$g_{V_t|X_t,Z}^o$ to be bounded away from zero. Since $g_{V_t|X_t,Z}^o$ is in the denominator of the moment restrictions defining the coefficients of interest, this assumption is useful to guard against difficulties that arise when $g_{V_t|X_t,Z}^o$ is close to zero. This assumption is standard in the literature on estimation of parameters defined by moments weighted by conditional density functions (see Honore and Lewbel, 2002, assumption B.3). Assumption (A5.ii) says that the second derivative of the characteristic function of the random variable $Y_t^*$ evaluated at zero is finite. This secures the existence of the integrable envelope for $Y_t^*$.

Assumption (A5.iii) secures that the weighting function $z \mapsto \omega(z)$ is well defined. Assumption (A5.iv) restricts the kernel functions employed in the first step to estimate the conditional density $g_{V_t|X_t,Z}^o$. The restriction on the kernels there ensures that $\hat{g}_t$ converges uniformly in probability to the conditional density $g_{V_t|X_t,Z}^o$. This in turn implies that the vector of kernel estimates $\hat{g} = (\hat{g}_2, ..., \hat{g}_T)$ consistently estimates the vector of conditional density functions $g_o = (g_{V_2|X_2,Z}^o, ..., g_{V_T|X_T,Z}^o)$ elementwise.

The following proposition establishes that assumption (A5) is sufficient for consistency,

**Proposition 2 (Consistency of the Estimator $\hat{\beta}, \hat{\gamma}$)** Let assumptions (A1)-(A6) hold. Define the smooth minimum distance estimator $(\hat{\beta}, \hat{\gamma})$ as in display (2SSMD). Then, $(\hat{\beta}, \hat{\gamma})$ is consistent for $(\beta_o, \gamma_o)$.

**Proof.** See the supplemental document.

### 4.3.2 ‘Root-n’ Consistency and Asymptotic Normality

We now turn to specifying sufficient conditions for ‘root-n’ consistency and asymptotic normality of the two-step smooth minimum distance estimator $(\hat{\beta}, \hat{\gamma})$.

To do so, our strategy is to first demonstrate that $n^{1/2}\hat{m}(\beta, \gamma, \hat{g}_t)$ is asymptotically normal, and then show that the sequence $n^{1/2}\left( (\hat{\beta}, \hat{\gamma}) - (\beta_o, \gamma_o) \right)$ inherits the asymptotic normality of $n^{1/2}\hat{m}(\beta, \gamma, \hat{g}_t)$ by a first order expansion of the criterion function $\hat{C}(\beta, \gamma, \hat{g})$ about $(\beta_o, \gamma_o)$. This strategy is similar to the one employed in the literature on the asymptotic behavior of semiparametric two-step estimators that minimizes a quadratic form (c.f., Andrews, 1994; Newey, 1994). The difference is that here $\hat{m}_t(\beta, \gamma, \hat{g}_t)$ is an U-process of order three, whereas the analog expression in the previous literature is an U-process of order one, i.e. an empirical process.

There is a potential caveat when approximating the distribution of estimators of parameters defined by moment restrictions weighted by density functions. Since the density function may be arbitrarily close to zero, the variances of certain variables may not exist. This can lead to estimators that do not converge to the usual parametric rate. This issue has been raised by Khan and Tamer (2010). The assumption that the density function $g_{V_t|X_t,Z}^o$ is bounded away from zero (see assumption A5.ii) comes in handy to circumvent this problem.

18
To establish 'root-n' consistency and asymptotic normality, we invoke the following restrictions:

Assumption A6 (Sufficient Conditions for the "root-n" Asymptotic Normality of \( \hat{\beta}, \hat{\gamma} \)).

(A6.i) \((\beta_o, \gamma_o)\) is an interior point of \( B \times \Gamma \).

(A6.ii) For all \( t \), the bandwidths \( a_n, b_n \) and \( c_n \) satisfy \( n_t^{1/2} a_n = o(1) \), \( n_t^{1/2} b_n = o(1) \), and \( n_t^{1/2} b_n = o(1) \). The kernel functions \( v \mapsto K_n(v) \) and \( x \mapsto K_b(x) \) are of order \( o \).

(A6.iii) For each \( z \) in the support of \( Z \), for every \( t \) in \( T \), and some numbers \((u, u')\) in an open neighborhood of zero, there exists some functions \( v_t, x_t \mapsto h(v_t, x_t) \) and \( x_t \mapsto h'(x_t) \) such that:

\[
\begin{align*}
\| \mathbb{E}(Y_t^* | v_t + u, x_t + u', z) - \mathbb{E}(Y_t^* | v_t, x_t, z) \| & \leq h(v_t, x_t) ||(u, u')|| \\
\| \mathbb{E}(Y_t^* | x_t + u', z) - \mathbb{E}(Y_t^* | v_t, x_t, z) \| & \leq h'(x_t) ||u'||
\end{align*}
\]

(A6.iv) The variance of the random variable \( \mathbb{E}(Y_t^* (Y_t^* - X_t \beta - \gamma Y_{t-1}) | V_{it}, X_{it}, Z_{it}) \) exists.

We now discuss this assumption. Assumptions (A6.i)-(A6.iii) are convenient to show that \( n_t^{1/2} \left( (\hat{\beta}, \hat{\gamma}) - (\beta_o, \gamma_o) \right) \) has an asymptotic linear representation which does not depend on the kernels. Assumption (A6.iv) enables us to apply the central limit theorem to this linear asymptotic representation. In particular, assumption (A6.i) enables us to Taylor expand the vector \( \hat{\nu}(\beta, \gamma, \hat{g}) \) about the true values of the coefficients of interest. It is standard in the literature on asymptotic theory of optimization estimators. Assumption (A6.ii) on the bandwidth enables us to eliminate the second order terms in such a Taylor expansion. Higher order kernels are needed to insure that the asymptotic distribution of \( n_t^{1/2} \left( (\hat{\beta}, \hat{\gamma}) - (\beta_o, \gamma_o) \right) \) is centered around zero. Assumption (A6.ii) implies that the kernel estimators \( \hat{g}_{V_t, X_t, Z_t}(v, x, z) \) and \( \hat{g}_{X_t, Z_t}(x, z) \) converges uniformly (in the sup norm) to their true counterparts at a rate faster than \( n_t^{1/4} \). This requirement is common in the literature on semiparametric two-step estimators (c.f., Andrews, 1994). The Lipchitz condition (A6.iii) allows us to get rid off the kernels in the first order terms of the Taylor expansion. This type of continuity is standard in the literature on estimation of expectations of kernel estimators (c.f., Powell, Stock and Stoker, 1989). Assumption (A6.iv) secures that the variance of \( n_t^{1/2} \hat{\nu}(\beta, \gamma, \hat{g}_t) \) exists.

The following proposition establishes that assumptions (A5) and (A6) are sufficient for 'root-n' consistency and asymptotic normality:

**Proposition 3** (Asymptotic Normality of the Estimator \( \hat{\beta}, \hat{\gamma} \)) Let assumptions (A1)-(A6) hold. Define the smooth minimum distance estimator \( (\hat{\beta}, \hat{\gamma}) \) as in display (2SSMD). Then, \( n_t^{1/2} \left( (\hat{\beta}, \hat{\gamma}) - (\beta_o, \gamma_o) \right) \) converges in distribution to a random vector with normal distribution \( \mathcal{N}(0, (J'J)^{-1} J \Sigma J' (J'J)^{-1}) \), where \( J \) is a \((T-1) \times (d_X + 1)\) matrix with characteristic
\[ J_t = -2\mathbb{E}[\varpi(Z) \times (Y_t^* - X_t^*\beta_o - \gamma_o Y_{t-1}) \cdot (X_t, Y_{t-1})], \]

and \( \Sigma \) is a square \((T-1)\) diagonal matrix with characteristic element

\[
\sigma_t = \begin{bmatrix}
\varpi(Z)Y_t^*\mathbb{E}[Y_t^* - X_t^*\beta_o - \gamma_o Y_{t-1}|X_t, Z] \\
-\varpi(Z)Y_t^*\mathbb{E}[Y_t^* - X_t^*\beta_o - \gamma_o Y_{t-1}|V_t, X_t, Z] \\
+\varpi(Z)\mathbb{E}[Y_t^* (Y_t^* - X_t^*\beta_o - \gamma_o Y_{t-1})|V_t, X_t, Z] \\
-\varpi(Z)\mathbb{E}[Y_t^* (Y_t^* - X_t^*\beta_o - \gamma_o Y_{t-1})|X_t, Z]
\end{bmatrix}
\]

**Proof.** See the supplemental document.

To conduct inference, one can either estimate the variance \((J'J)^{-1}J\Sigma J'J^{-1}\), or adopt the subsampling procedure discussed by Politis, Romano and Wolf (1999). Regarding the first approach, a consistent estimator of the variance \((J'J)^{-1}J\Sigma J'J^{-1}\) can be formed by plugging in estimators of the different pieces. This requires estimating conditional expectations, which involves choosing further smoothing parameters. This makes the subsampling procedure a more desirable approach, albeit implementing such a procedure in our case can be costly from a computational point of view because of the optimization nature of the estimator.

### 4.3.3 Semiparametric Efficiency

Our development of the asymptotic properties of the estimator \((\hat{\beta}, \hat{\gamma})\) naturally leads to the question about the precision of such estimator. We establish now the smallest possible variance associated to a regular estimator of the coefficients of interest \((\beta_o, \gamma_o)\). Even if our estimator is not semiparametric efficient, such a variance is useful to characterize the departures of our estimator from asymptotic efficiency.

Before going on, it should be noted that the setting here is different to the one considered by Jacho-Chavez (2009). Specifically, he derives efficiency bounds for parameters of interest defined by unconditional moment restrictions weighted by a density function when data are available from a single sample. As our parameters of interest are defined by conditional moment restrictions and data are available from multiple samples, his results do not directly apply to our case. In the same line, the relationship between the non-parametric component of \(Y_t^*\) (i.e., \(g_{V}\{X_t, Z\}\)) and the density function \(g_{Y_t, V, X_t, Z}\) defining the conditional moment \(\mathbb{E}(Y_t^*|Z)\) prevents us to employ the general methodologies developed by Ai and Chen (2003) to calculate efficiency bounds for parameters of interest defined by conditional moment restrictions containing unknown functions.

We tailor the techniques by Severini and Tripathi (2001) and insights from Magnac and Maurin (2007) to get the following result:

**Proposition 4 (Efficiency Bound)** Let assumptions \((A1)-(A4)\) hold. Then, the semi-
parametric efficiency bound for estimating \((\beta_o, \gamma_o)\) is

\[ I^{-1} = \mathbb{E}[D(Z) \times \Omega(Z) \times D(Z)']^{-1}, \]

where \(D(Z) \equiv (\mathbb{E}(X_t|Z), \mathbb{E}(Y_{t-1}|Z))' \) a.s., and

\[ \Omega(Z) = V [Y^*_t + \mathbb{E}(Y^*_t|X_t, Z) - \mathbb{E}(Y^*_t|V_t, X_t, Z) - X_t' \beta_o - \gamma_o \mathbb{E}(Y_{t-1})|Z] \]

**Proof.** See the supplemental document. ■

Notice that the semiparametric information matrix \(I^{-1}\) is different from zero whenever \(\mathbb{E}(Y^*_t^2)\) is finite. Consequently, the parameters \((\beta_o, \gamma_o)\) can be estimated at the parametric rate only if \(\mathbb{E}(Y^*_t^2)\) is finite. Assumption (A5.i)-(A5.ii) are imposed to overcome this difficulty.

## 5 Monte Carlo Experiments

The previous asymptotic results give conditions under which our proposed estimator of the coefficients of interest will be well behaved in large samples. In this section, we explore the finite sample performance of the estimator by reporting results of Monte Carlo experiments. These experiments also illustrate the implementation of the estimator.

**Data Generating Process.** We now describe the data generating process employed in the Monte Carlo experiments. We want to evaluate the finite-sample behavior of the two-step minimum distance estimator in the following situations: 

- **(A)** All the assumptions imposed so far hold, and none of the ‘cohorts’ defined by \(Z\) have ‘few’ observations; 
- **(B)** All the assumptions imposed so far hold, but one of the ‘cohorts’ defined by \(Z\) has ‘few’ observations; 
- **(C)** The distribution of the continuous regressor \(V_t\) conditional on the other variables is not bounded away from zero (i.e., when condition (A5.i) fails); 
- **(D)** The support of the continuous regressor is not large (i.e., when condition (A2.ii) fails).

With the previous motivation in mind, we generate repeated cross sections using the dynamic binary response model:

\[ Y_{it} \equiv \mathbb{I} [V_{it} + \gamma_o Y_{it-1} + \alpha_i + \varepsilon_{it} \leq 0] \text{ a.s., } t = 0, \ldots, 3 \tag{3} \]

with initial condition \(Y_{i0} \equiv \mathbb{I} (V_{i0} + \alpha_i + \varepsilon_{i0} \leq 0) \) a.s.. The parameter of interest is \(\gamma_o.\) For the sake of computational simplicity, we do not consider regressors \(X_t.\) The distribution
of the random vector \((\alpha, \varepsilon^T, V^T, Z)\) is specified as follows:

\[
F_{\alpha}^o(a; s_\alpha, p_\alpha) \prod_{t=0}^{3} F_{\varepsilon t}^o(e_t; s_\varepsilon, p_\varepsilon) \prod_{t=0}^{3} G_{Vt|Z}^o(v_t; s_t, p_t, \bar{v}_t) G_{Z}^o(z; p_z, \bar{z})
\]  

(4)

with \(F_{\alpha}^o(a; s_\alpha, p_\alpha)\) the beta distribution with shape parameters \((s_\alpha, p_\alpha)\); \(F_{\varepsilon t}^o(e_t; s_\varepsilon, p_\varepsilon)\) the beta distribution with shape parameters \((s_\varepsilon, p_\varepsilon)\); \(G_{Vt|Z}^o(v_t; s_t, p_t, \bar{v}_t)\) the distribution of the random variable \(U_t \times \bar{v}_t \times 2 - \bar{v}_t\), where \(\bar{v}_t\) is a positive scalar and \(U_t\) follows a beta distribution with shape parameters \((s_t, p_t)\); and \(G_{Z}^o(z; p_z, \bar{z})\) the binomial distribution with success probability \(p_z\) and \(\bar{z}\) the number of trials. The success probability \(p_z\) governs the number of observation in each of the \(\bar{z} + 1\) 'cohorts' defined by \(Z\). To govern the lower bound of the density \(g_{Vt|Z}^o\), we set the shape parameter \(p_t\) at \(p_t = [z/(\bar{z} + 2)] + q_t\), where \(z\) is the realized value of \(Z\), and \(q_t\) is positive number. The parameter \(\bar{v}_t\) governs the support of \(V_t\) conditional on \(Z\), which is indeed \([-\bar{v}_t, \bar{v}_t]\). We dismiss the realizations for the initial period \(t = 0\). Data consist then in the independent samples \(\{Y_{i1}, V_{i1}, Z_{i1}\}_{i=1}^{n_1}, \ldots, \{Y_{i3}, V_{i3}, Z_{i5}\}_{i=n_2+1}^{n_3}\).

The design variables are: the parameters \((s_\alpha, p_\alpha)\) indexing the beta distribution of the individual effect; the parameters \((s_\varepsilon, p_\varepsilon)\) indexing the beta distribution of the idiosyncratic disturbances; the parameters \(\{s_t, q_t, \bar{v}_t\}_{t=0}^{3}\) indexing the distribution of \(V_t\) conditional on \(Z\); and the parameters \((p_z, \bar{z})\) indexing the distribution of \(Z\). In all the experiments, we keep fixed the values of the parameters indexing the distributions of the individual effects and idiosyncratic error at \(s_\alpha = 2\), \(p_\alpha = 2\), \(s_\varepsilon = 1\) and \(p_\varepsilon = 1\). We set \(s_t = .75\). We set \(\bar{z} = 2\), so the support of \(Z\) is \(\{0, 1, 2\}\). To evaluate the sensitivity of the estimator to the size of the 'cohorts' defined by \(Z\), we make vary the success probability \(p_z\) in \(\{.5, .9\}\). To evaluate the behavior of the estimator when the density \(g_{Vt|Z}^o\) is not bounded away from zero, we make vary the shape \(\{q_t\}_{t=0}^{3}\) of the distribution of \(V_t\) conditional on \(Z\) between \(\{q_t = .25\}_{t=0}^{3}\) and \(\{q_t = 4\}_{t=0}^{3}\). When \(q_t = .25\), the density \(g_{Vt|Z}^o\) is bounded away from zero. When To evaluate the behavior of the estimator when the support of the continuous regressor is not large, we make vary \(\bar{v}_t\) in \(\{5, 1\}\). We have four designs: \(A\) \(\{q_t = .25\}_{t=0}^{3}\); \(p_z = .5\), and \(\bar{v}_t = 5\); \(B\) \(\{q_t = .25\}_{t=0}^{3}\); \(p_z = .95\), and \(\bar{v}_t = 5\); \(C\) \(\{q_t = 4\}_{t=0}^{3}\); \(p_z = .5\), and \(\bar{v}_t = 5\); and \(D\) \(\{q_t = .25\}_{t=0}^{3}\); \(p_z = .5\), and \(\bar{v}_t = 1\). They correspond to the situations \(A\), \(B\), \(C\) and \(D\) described above.

**Implementation.** In order to implement our estimator, we have to specify a weighting factor \(\varpi(Z)\), a kernel function \(K_\alpha(v)\), and bandwidth sequences \((a_n, c_n)\), compatible with the assumptions made in the previous section. For the weighting factor, we set \(\varpi(Z) = 1\). For the kernel function, we let \(K_\alpha(v)\) to be the fourth order Epanechnikov kernel.

Regarding the bandwidth sequences, we choose the 'naive' rule \(a_n = \sqrt{V(V)} \times n^{-1/h}\).

---

1Specifically, we let \(K(v) = \frac{15}{8} \left( 1 - \frac{4}{3} v^2 \right) \frac{3}{4} \left( 1 - v^2 \right) I(|v| < 1)\).
\[ c_n = [\bar{z}/(\bar{z} + 1)] \times n_t^{-1/2}, \] where \( h \) is a positive integer. In order to evaluate the sensitivity of the estimator, we make vary \( h \) in the grid \{4, 5, ..., 10\}. For each design, we make vary the size of the cross sections in \( n_t \in \{250, 500, 1000\} \). The number of simulation in each experiment is equal to 500. All the experiments were carried out in \( R \).

**Results.** As competitors to the two-step smooth minimum distance estimator (2SSMD), we consider two generalized method of moment estimators with identity weighting matrix. The first generalized method of moment estimator employs the true density \( g^o_{V_{i|Z}} \) instead of an estimator of it. This estimator is infeasible. We call it GMM. It is our benchmark. The GMM can be interpreted as an infeasible version of the sieve minimum distance estimator proposed by Ai and Chen (2003). The second generalized method of moment estimator employs the kernel estimates of density \( g^o_{V_{i|Z}} \) instead of the density itself. This is a feasible estimator. We call it 2SGMM.

We now summarize the main findings. So far, we have only run the designs \( A \) and \( B \). In all the experiments in these designs, the 2SSMD estimator outperforms in terms of RMSE and MAE both the infeasible GMM and the feasible 2SGMM. The feasible 2SGMM never beats the infeasible GMM. For the 2SSMD and the 2SGMM, we compute the ratios of the root mean squared error (RMSE) and mean absolute deviation (MAE) with respect to the ones of GMM. When all the assumptions hold and none of the cohorts have few observations (design A), the RMSE of the 2SSMD is 10%-30% that of the GMM, whereas the RMSE of the 2SGMM is 140%-200% that of the GMM. When all the assumptions hold and one of the cohorts has few observations (design B), the RMSE of the 2SSMD is 30%-50% that of the GMM, whereas the RMSE of the 2SGMM is 120%-170% that of the GMM. Similar results are obtained for the MAE. Tables with the complete results of the experiments are in the supplemental document to this paper.

6 Summary and Conclusions

We investigate the identification and estimation of a semiparametric single-index dynamic binary response model when data are available from repeated cross sections instead of panels. Unlike the existing literature, we allow for time varying variables among the regressors. Our first contribution is to provide a set conditions sufficient for the point identification of the coefficients entering the single-index. Under the maintained identification assumptions, the coefficients of interest are uniquely defined by a conditional moment equality weighted by a conditional density function. Our second contribution is to propose a consistent and asymptotically normal estimator estimator of the coefficients of interest. Monte Carlo experiments show that our estimator behaves well in finite samples with respect to existing procedures.

Our focus is on identification and estimation of the coefficients entering the single-
index. In nonlinear models, such as the one studied here, it is often of interest to learn not only about these coefficients but also about the marginal effects. We left the identification and estimation of marginal effects for future research. It is clear that the approach discussed here will suffer from many of the problems that make identification from multiple independent random sample difficult. These include problems associated to the high sensitivity of point identification to a failure of the maintained assumptions. Indeed, one of the main limitations of this paper is that we handle only the case when all the conditions defining our model hold. Exploring identification when some of our assumptions fails might be appropriate. Our approach will suffer as well from the problems that make inference on parameters defined by moment restrictions weighted by the inverse of a density function a delicate issue. These include problems associated to slow speed of convergence of the estimators when some densities are not bounded away from zero. An analysis similar to that in Khan and Tamer (2010) might also be appropriate.

Acknowledgments. I wish to thank my advisor Thierry Magnac for invaluable help during this project. Pascal Lavergne, Fabiana Gomez, and seminar participants at Toulouse School of Economics have offered useful comments and suggestions for which I am very grateful. All remaining errors are my responsibility.

Corresponding Address. David H. Pacini, Toulouse School of Economics, Manufacture des Tabacs, 21 Allée de Brienne, Bureau MF003, (31000) Toulouse, France; E-mail: davidhpacini@yahoo.com; Website: https://sites.google.com/site/davidhpacini/

References


