Identification of the Linear Projection Model from Two-Sample Data

David H. Pacini*

This Preliminary Version:

September 28, 2011

Abstract

We investigate what can be learned about the coefficients (β, γ) in the *linear projection model* $Y = X\beta + Z'\gamma + \varepsilon$ from data consisting of two independent samples; the first sample gives information on variables (Y, Z) but not X, while the second sample gives information on (X, Z) and not Y. Here Y is a scalar response variable, X is a scalar covariate, Z is a vector of other covariates, and ε is an error term uncorrelated with the covariates (X, Z). Complications arise because none sample has joint information on the response variable and the covariates. The existing literature suggests to overcome these complications by assuming either that there exists an instrumental variable observed in both samples, or that Y and X are independent conditional on Z. Our contribution is to sharply characterize the identified set of the coefficients (β, γ) when the latter assumptions are not invoked. This set represents the limit of what can be learned about the coefficients (β, γ) given the model and the data. We show that the identified set is not a singleton, so the coefficients of interest are set not point identified by the linear projection model. This result contrasts with the existing literature, where the assumptions have the power to point identify the coefficients of interest. We employ our characterization of the identified set to construct an estimator of it. Monte Carlo experiments illustrate the implementation and the performance of the estimator.

KEYWORDS: Identification; Linear Projection Model; Two Samples.

JEL CLASIFICATION: C21, C26

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 student, Toulouse School of Economics| **E-mail:** davidhpacini@yahoo.com| **First Version:** June 15, 2010| Please do not cite or circulate| **Comments are welcomed.**

1 Introduction

Economists who use survey data for inferences often face the situation where the variables of interest are observed in different samples. In this paper we investigate what can be learned about the coefficients (β, γ) in the *linear projection model* $Y = X\beta + Z'\gamma + \varepsilon$ from data consisting of two independent samples with common variables Z; the first random sample gives information on variables (Y, Z) but not X, while the second sample gives information on variables (X, Z) but not Y. Here Y is a scalar response variable, X is a scalar covariate, Z is a vector of other covariates (possible including a constant), and ε is an scalar error term uncorrelated with the covariates (X, Z). Ridder and Moffit (2007) survey applications fitting this framework.

Within the previous context, complications arise because the coefficients (β, γ) depend on the joint distribution of the variables (Y, X), but none of the samples has joint information on these two variables. The prominent method adopted to overcome these complications is to assume that one of the covariates Z is an instrumental variable (c.f., Angrist and Krueger, 1992; Nicoletti and Ermisch, 2007; Anderson and Matsa, 2011). An alternative is to assume that the response variable Y is independent of the scalar covariate X conditional on the other covariates Z (c.f., Bostic, Gabriel and Painter, 2009), or that the error term ε is mean-independent of the covariates Z (c.f., Ichimura and Martinez-Sanchis, 2010). Here we ask what can be ascertained about the coefficients (β, γ) when the latter assumptions are not invoked. This is useful to evaluate the sensitivity of inferences about the coefficients (β, γ) to failure of the assumptions adopted by the existing literature.

Our contribution is to characterize the set of values of the coefficients (β, γ) compatible with the linear projection model and hypothetical knowledge of the distribution of (Y, Z) and of (X, Z). This set, called the identified set of (β, γ), represents the limit of what can be learned about the coefficients (β, γ) given the model and the data. We show that the identified set of (β, γ) is not a singleton, so the coefficients of interest are set not point identified by the linear projection model. This result contrasts with the existing literature, where the assumptions have the power to point identify the coefficients of interest. The size of the identified set depends on the strength of the dependence between the covariates (X, Z) and of the dependence between (Y, Z). Our characterization of the identified set of (β, γ) is based on the insight that knowledge of the distribution of (Y, Z) and of (X, Z) restricts the unknown distribution of (Y, X), and consequently the coefficients (β, γ), through the Hoeffding-Frechet distributions.

We employ our characterization of the identified set of (β, γ) to construct an estimator of it. The estimator exploits the fact that the identified set is convex and bounded. Convexity secures that the identified set is characterized by its extreme points. Boundedness secures that these extreme points are finite. In line with the literature on estimation of convex identified sets (e.g., Beresteanu an Molinari, 2008; Bontemps, Magnac, and Maurin, 2011; Kaido and Santos, 2011), we write the extreme points of the identified set in terms of moments of the available data. The estimator is the sample analog of these moments. We illustrate the performance and implementation of this estimator with Monte Carlo exercises. We also show that our identification and estimation results extend to the ecological correlation problem discussed by Robinson (1950), and to the problem of identifying the variance of the treatment effect in the potential outcome model (c.f., Heckman, Smith and Clements, 1997).

Related Literature. This paper expands upon the work by Ridder and Moffit (2007), who suggest to employ the Hoeffding-Frechet distributions to bound the joint distribution of the variables (Y, X, Z) when data are available from independent samples on (Y, X) and on (X, Z). We use their insight to characterize the identified set of the coefficients (β, γ). To the best of our knowledge, this issue has not been addressed by the existing literature. In a related paper, Cross and Manski (2002) examine identification of the function φ in the regression model $Y = \varphi(X, Z) + \varepsilon$ when data are available from independent samples on (Y, Z) and on (X, Z).[1] Our work is in the same spirit as their contribution, but applies to a different model. They impose mean-independence between the error term ε and the covariates (X, Z), whereas we impose the weaker condition that the error term ε and the covariates (X, Z) are uncorrelated. Then, when a linear parametric assumption on the function φ is imposed, the mean-independence assumption by Cross and Manski (2002) neither nest nor are nested by our assumptions.

While motivated by a different application, Fan and Zhu (2010) also employ Hoeffding-Frechet distributions. In the context of a potential outcome model, they employ Hoeffding-Frechet distributions to derive bounds on superadditive functionals of the joint distribution of potential outcomes. Our work overlaps with theirs in that each of the coefficients (β, γ) can be interpreted as a superadditive functional. We express however the identified set in terms of moment in a different way than they do. This allows us to dispense with numerical integration procedures, which is a step required to implement their estimation procedure. Our work also bears some resemblance to Hu (2006), who bounds the coefficients in a linear regression model with a mismeasured regressor when the variance of this regressor is known. The mismeasured regressor here is X. His method however is not applicable to our problem because the distribution of X, rather than its variance, is known in our setting.

At a more general level, this paper belongs to the literature on set identification of linear models with incomplete data. Papers in this literature include Gini (1921), Frisch

[1] See also Vitale (1976) and Molinari and Peski (2006).

(1934), Klepper and Leamer (1984), Krasker and Pratt (1986), Horowitz and Manski (2006), and Bontemps, Magnac and Maurin (2011). Although the problem we are concerned with here is different from those considered by the aforementioned authors, our techniques are similar to theirs and we use several of their results in our analysis.

Organization of the Paper. The outline of the paper is as follows. In the next section, we set out the linear projection model, and describe the available data. Section 3 presents the identification results. There, we characterize of the identified set of the coefficients (β, γ) delivered by the linear projection model. Section 4 discuss the estimation of the identified set and the construction of confidence intervals. Section 5 illustrates the estimation and inference procedures discussed in Section 4. Section 6 concludes, and an appendix collects all the proofs and tables.

2 The Setup

In this section, we lay out the assumptions defining the linear projection model that will be used in the rest of the paper. We also describe the available data consisting of two independent random samples with common variables.

Consider a collection $\{1, .., i, .., N\}$ of observational units (i.e., individuals, firms, etc.) to be studied at a given period in time. For each observational unit i, we begin by assuming that a scalar outcome variable Y_i is determined by a linear response equation.

Assumption 1 *(Linear Response Equation). The outcome variable Y_i is determined by the linear response equation:*

$$Y_i = X_i'\beta_o + Z_i'\gamma_o + \varepsilon_i \tag{A1}$$

where (β_o, γ_o) is a vector of unknown real numbers, and $(X_i, Z_i, \varepsilon_i)$ is a random vector defined on the probability space $(\Omega, \mathcal{A}, P_o)$ with Ω a sample space, $\mathcal{A}$ a sigma-algebra of subsets of Ω, and P_o some probability measure on $\mathcal{A}$.

In what follows, we suppress the subscript i in the notation whenever this can be done without causing confusion. We use uppercase letters to denote random variables defined on $(\Omega, \mathcal{A}, P_o)$, and lowercase letters to denote their realizations. We use the subscript (superscript) "$_o$"("^{o}") attached to any expression to distinguish the "true" value from any other value of that expression. We index a distribution functions by the random variables they refers to. The absence of arguments for a function denotes the entire function rather than its value at a point. Thus, for instance, F_X^o denote the distribution of X induced by P_o, and $F_X^o(x)$ denotes such a distribution evaluated at x. The expression $\mathbb{E}$ denotes the expectation with respect to the measure P_o.

The following restrictions complete the description of the linear projection model:

Assumption 2 *(Regularity). The distribution of (X, Z, ε),* say $F_{X,Z,\varepsilon}^o$, *is such that:*

(A2.i) The error ε and the covariates (X, Z) are uncorrelated i.e., $\mathbb{E}[(X, Z)'\varepsilon] = 0$.
(A2.ii) The $1+d_Z$ vector of covariates (X, Z) satisfies rank $(\mathbb{E}[(X, Z) \cdot (X, Z)']) = 1+d_Z$, where rank() stands for the rank of the matrix within the parenthesis.
(A2.iii) The random vector (Y, X, Z) from Ω to $\mathcal{Y} \times \mathcal{X} \times \mathcal{Z}$ has finite variance.

We call the triplet $(\beta_o, \gamma_o, F^o_{X,Z,\varepsilon})$ the *true structure.* All the elements in the true structure are unknown. Since in most application only the coefficients (β_o, γ_o) are of interest, we focus on them and not on the distribution $F^o_{X,Z,\varepsilon}$. We interpret the coefficients (β_o, γ_o) as partial correlations, or as the coefficients of the best linear predictor under quadratic loss of Y based on (X, Z).

We now describe the available data. If a common sample of (Y, X, Z) were available, identification and inference on the coefficients (β_o, γ_o) would be straightforward. Here a common sample of (Y, X, Z) is unavailable. Instead, we assume that data are available from two independent random samples with common variables:

Assumption 3 *(Data are Available from Independent iid Samples). Let $F^o_{Y,X,Z}$ denote the distribution of (Y, X, Z) generated by the true structure $(\beta_o, \gamma_o, F^o_{X,Z,\varepsilon})$. We denote by $G^o_{Y,Z}$ the (Y, Z)-marginal distribution of $F^o_{Y,X,Z}$. A similar notation is adopted for $G^o_{X,Z}$. The data are available from independent samples drawn from $G^o_{Y,Z}$ and $G^o_{X,Z}$. The first sample, say $\{Y_i, Z_i\}_{i=1}^{n_1}$, contains independent and identically distributed (iid) replications of the variables (Y, Z) generated from $G^o_{Y,Z}$ for a group of n_1 observational units. The second sample, say $\{X_j, Z_j\}_{j=n_1+1}^{n+1}$, contains iid replications of the variables (X, Z) generated from $G^o_{X,Z}$ for a group of different $n_2 = n - n_1$ observational units.*

Since we are working with iid samples, the distribution of the sample $\{Y_i, Z_i\}_{i=1}^{n_1}$ and $\{X_j, Z_j\}_{j=n_1+1}^{n+1}$ are fully characterized by the distributions $G^o_{Y,Z}$, $G^o_{X,Z}$, respectively. We say thus that $G^o_{Y,Z}$ and $G^o_{X,Z}$ represent the available data free of sample variation. For identification purposes, we assume that the distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$ are known, even if their exact knowledge can not be derived from any finite number of observations. This is useful to separate the identification problem from the statistical inference one.

2.1 Example

Here we discuss a concrete example that fits the previous setup. This example comes from the work by Bostic, Gabriel, and Painter (2009), who employ two samples to measure the partial correlation between housing wealth and consumption at the household level in the United States.

We concentrate only in one of the regressions considered by Bostic et. al. (2009). Let Y_i denote the log of the total expenses of a household i living in the US in 2001. The standard specification for Y_i proposed by Bostic et. al. (2009) is $Y_i = X_i\beta_o + Z_i'\gamma_o + \varepsilon_i$, where X_i is the log of household's house value, and Z_i is a vector of household characteristics such as the household current income, human capital and like controls including number of members in the household and education, age and marital status of the head of the household. As in Bostic et. al. (2009), we assume that the error term ε and the vector of covariates (X, Z') are uncorrelated. This is a strong assumption which may fail, for instance, if the income, or some of the components of Z, are chosen by the household after observing ε. Here however we follow Bostic et. al. (2009) and we abstract from such type of situations. The interest is on the coefficient β_o, which represents the partial correlation between log consumption Y and log household's house value X after controlling for the variables included in Z'.

Learning about the coefficient β_o would ideally require a sample of household with measurements on (Y, X, Z). Such sort of data however is not available. Bostic et. al. (2009) employ two samples in order to learn about β_o; the Consumer Expenditure Survey (CEX), and the Survey of Consumer Finances (SCF). The CEX provides information on household expenses and income, that is on (Y, Z), but not on household's house value X. The SCF provides information household house value and income, that is on (X, Z) but not on household expenses Y. The CEX and CSF do not survey the same households because they are independent samples. To overcome the lack of joint realizations on (Y, X), Bostic et. al. (2009) employ an imputation procedure under the assumption that Y and X are independent conditional on Z. By contrast, we do not impose such a conditional independence assumption. The results we present below can thus be employed to evaluate the sensitivity of their inferences to a failure of the latter conditional independence assumption.

3 Identification

In this section, we ask whether the linear projection model and the two independent samples described in the previous section provide information about the unknown coefficients of interest (β_o, γ_o). To answer this question, we characterize the identified set of the coefficients (β_o, γ_o). This set contains all values of the coefficients of interest compatible with the linear projection model and the available data. This characterization is the main result of the paper. We employ it to show that the coefficients (β_o, γ_o) are set not point identified.

Characterizing the identified set of the coefficients (β_o, γ_o) involves to characterize first the set of values of the expectation $\mathbb{E}(YX)$ compatible with the distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$. To see why, note that under Assumptions (A1) and (A2.i), the coefficients (β_o, γ_o) are defined by the equations:

$$\beta_o \equiv \left[\mathbb{E}(YX) - \mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZY)\right] \times \left[\mathbb{E}(X^2) - \mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)\right]^{-1} \quad (1)$$

$$\gamma_o \equiv \mathbb{E}(ZZ')^{-1} \cdot \left[\mathbb{E}(ZY) - \mathbb{E}(ZX)\beta_o\right] \quad (2)$$

where the expression " $\equiv$ " stands for "*defined by*". The rank restriction (A2.ii) secures that the denominators in the latter display are different from zero, so the coefficients of interest (β_o, γ_o) are well-defined. The consequence of having observations on (Y, Z) and (X, Z) but not on (Y, X) is that all the expectations in the right hand side of (1) are known except for $\mathbb{E}(YX)$. However, the values that the expectation $\mathbb{E}(YX)$ can take are not completely free; they are restricted by the marginal distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$.

Our strategy to characterize the identified set of (β_o, γ_o) has two parts. In the first part, we characterize the identified set of the expectation $\mathbb{E}(YX)$. In the second part, we characterize the identified set of the coefficients (β_o, γ_o) by employing the characterization of the identified set of the expectation $\mathbb{E}(YX)$.

3.1 The Identified Set of the Expectation $\mathbb{E}(YX)$

Here we show that the identified set of the expectation $\mathbb{E}(YX)$ is a segment of the real line, and we express the extreme points of this segment in terms of moments of the available data. This characterization of the identified set of $\mathbb{E}(YX)$ is key to derive the main result of the paper, namely, a characterization of the identified set of the coefficients (β_o, γ_o).

We begin by defining the identified set of $\mathbb{E}(YX)$. Heuristically, this set contains all the values of the expectation of the product between Y and X compatible with the available data free of sample variation. Since in our case the available data free of sample variation are represented by the marginal distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$, we formally define the identified set of $\mathbb{E}(YX)$ by:

$$\Theta_I \equiv \left\{ \begin{array}{l} \theta \in \mathbb{R} : \theta = \int\int yx f_{Y,X}(y,x)dydx \\ f_{Y,X}(y,x) = \int_{\mathcal{Z}} f_{Y,X|Z}(y,x|z) g^o_Z(z)dz \\ g^o_{Y|Z}(y|z) = \int_{\mathcal{X}} f_{Y,X|Z}(y,x|z)dy \quad \forall x \in \mathcal{X}, z \in \mathcal{Z} \\ g^o_{X|Z}(x|z) = \int_{\mathcal{Y}} f_{Y,X|Z}(y,x|z)dy \quad \forall y \in \mathcal{Y}, z \in \mathcal{Z} \end{array} \right\}, \quad (3)$$

where $f_{Y,X}$ denotes the density of (Y, X), and similarly for $f_{Y,X|Z}$, g^o_Z, $g^o_{Y|Z}$ and $g^o_{X|Z}$. Heuristically, θ belongs to Θ_I if and only if there exists a collection of conditional densities $\left\{y, x \mapsto f_{Y,X|Z}(y,x|z)\right\}_{z\in\mathcal{Z}}$ matching the density of (Y, X) induced by the model, that is $f_{Y,X}$, with the available data free of sample variation, that is the densities $g^o_{Y|Z}$ and $g^o_{X|Z}$.

The above definition of the identified set of $\mathbb{E}(YX)$ is not an operational one, in the sense that it does not allow the computation of Θ_I. We look for an operational characterization, next.

3.1.1 Conditional Hoeffding-Frechet Distributions

To find an operational characterization of the identified set of $\mathbb{E}(YX)$, we find useful to interpret this set as the range of a mapping. To describe this mapping, let $\mathcal{F}_{Y,X,Z}$ denote the class of distribution functions with support on $\mathcal{Y}\times\mathcal{X}\times\mathcal{Z}$ for which the (Y,Z)-marginal and the (X,Z)-marginal are given by $G^o_{Y,Z}$ and $G^o_{X,Z}$, respectively. We view the identified set of $\mathbb{E}(YX)$ as the range Θ_I of the mapping $F_{Y,X,Z} \mapsto \int yxdF_{Y,X,Z}(y,x,z)$ from the class $\mathcal{F}_{Y,X,Z}$ into the real line.

Our strategy to characterize the identified set of $\mathbb{E}(YX)$ has two steps. In the first step, we establish the convexity of domain class $\mathcal{F}_{Y,X,Z}$, and then show that the mapping $F_{Y,X,Z} \mapsto \int yxdF_{Y,X,Z}(y,x,z)$ carries this property over its range Θ_I. Convexity enables to characterize Θ_I by its extreme points. The second step of our strategy is to write the extreme points of Θ_I in terms of moments of the available data. The fact that the class $\mathcal{F}_{Y,X,Z}$ has been previously studied in the literature on distribution with given marginals (c.f., Ruschendorf, 1991) facilitates the actual implementation of this strategy.

To proceed, we establish the following properties of the class $\mathcal{F}_{Y,X,Z}$:

Proposition 1 *(Geometric Properties of $\mathcal{F}_{Y,X,Z}$) Let $\mathcal{F}_{Y,X,Z}$ denote the class of distribution functions with support on $\mathcal{Y}\times\mathcal{X}\times\mathcal{Z}$ for which the (Y,Z)-marginal and the (X,Z)-marginal are given by $G^o_{Y,Z}$ and $G^o_{X,Z}$, respectively. The class $\mathcal{F}_{Y,X,Z}$ is non-empty and convex.*

Proof. See the Appendix A. ■

Since $F_{Y,X,Z} \mapsto \int yxdF_{Y,X,Z}(y,x,z)$ is a linear map, and convexity is preserved under linear transformations (see Rockafellar, 1970) , we have the following corollary to Proposition 1:

Corollary 1 *(Geometric Properties of the Identified Set of $\mathbb{E}(YX)$) Let assumptions (A1) and (A2) hold. Define the identified set Θ_I as in (1). Then, Θ_I is a non-empty convex subset -i.e., a segment- of the real line.*

As a consequence of Corollary 1, the identified set Θ_I can be characterized by the smallest θ_L and largest θ_U value that the mapping $F_{Y,X,Z} \mapsto \int yx\ dF_{Y,X,Z}(y,x,z)$ can take over the space of functions $\mathcal{F}_{Y,X,Z}$. These values are formally defined by the programming problems:

$$\begin{aligned}
&\theta_L \equiv \min\nolimits_{F_{Y,X,Z}} \int yx\ dF_{Y,X,Z}(y,x,z) \\
&\text{s.t. } G^o_{Y,Z}(y,z) = \lim\nolimits_{x\to\infty} F_{Y,X,Z}(y,x,z) \quad \forall y\in\mathcal{Y}, z\in\mathcal{Z} \\
&\qquad G^o_{X,Z}(x,z) = \lim\nolimits_{y\to\infty} F_{Y,X,Z}(y,x,z) \quad \forall x\in\mathcal{X}, z\in\mathcal{Z}
\end{aligned}$$

for the smallest value θ_L, and the corresponding maximization problem for the largest value θ_U. These programming problems have linear objective functions with linear constraints. They are the object of study in the literature on mass transportation. Solving this type of programming problem is a potentially delicate issue, that have attracted considerable attention. However, since here $y, x \mapsto yx$ in the objective function is a strictly superadditive function, our problem is much less complicated and has a well known unique closed form solution (see Ruschendorf, 1991). In particular, the function

$$G^L_{Y,X,Z}(y,x,z) = \int_{-\infty}^{z} \max\{0, G^o_{Y|Z}(y|s) + G^o_{X|Z}(x|s) - 1\} dG^o_Z(s) \tag{4}$$

solve the minimization problem we have described above. In turn, the function

$$G^U_{Y,X,Z}(y,x,z) = \int_{-\infty}^{z} \min\{G^o_{Y|Z}(y|s), G^o_{X|Z}(x|s)\} dG^o_Z(s) \tag{5}$$

solves the corresponding maximization problem. The functions $G^L_{Y,X,Z}$ and $G^U_{Y,X,Z}$ are referred to as the conditional Hoeffding-Frechet distributions. The lower conditional Hoeffding-Frechet distribution $G^L_{Y,X,Z}$ can be interpreted as the copula on (Y, X, Z) with given $G^o_{Y,Z}$ and $G^o_{X,Z}$ marginals minimizing the correlation between Y and X. A similar interpretation follows for the upper conditional Hoeffding-Frechet distribution $G^U_{Y,X,Z}$.

3.1.2 Characterization of the Identified Set of the Expectation $\mathbb{E}(YX)$

The next theorem provides an operational characterization of the identified set of $\mathbb{E}(YX)$. This characterization arises after evaluating the mapping $F_{Y,X,Z} \mapsto \int yx \; dF_{Y,X,Z}(y,x,z)$ at the Hoffding-Frechet distributions (4) and (5).

Theorem 1 *(Characterization of the Identified Set of $\mathbb{E}(YX)$). Suppose that data are available from independent random samples of (Y, Z) and of (X, Z) as described in Assumption (A3). Let $Q^o_{X|Z}$ denote the quantile function of X given Z, and define the quantities:*

$$\theta_L = \mathbb{E}[Y \cdot Q^o_{X|Z}(1 - G^o_{Y|Z}(Y|Z)|Z)] \;\; ; \;\; \theta_U = \mathbb{E}[Y \cdot Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)]$$

where $G^o_{Y|Z}$ is the distribution of Y given Z. Then, the identified set of $\mathbb{E}(YX)$ delivered by the linear projection model is the interval $\Theta_I = [\theta_L, \theta_U]$.

Proof. See the Appendix A. ■

The characterization of the identified set of $\mathbb{E}(YX)$ in Theorem 1 is operational because it express Θ_I in terms of moments of the available data. This result seems to be novel. It is key in the characterization we derive below of the identified set of the coefficients

(β_o, γ_o). Theorem 1 prescribes the bounds $\theta_L \leq \mathbb{E}(YX) \leq \theta_U$ on the expectation $\mathbb{E}(YX)$. These bounds are sharp in the sense that they can not be improved unless additional assumptions are imposed. Note that the lower bound θ_L does not coincide with upper bound θ_U because $1-G^o_{Y|Z}(Y|Z)|Z)$ and $G^o_{Y|Z}(Y|Z)$ neither does. Then, the identified set of the expectation θ_o delivered by the linear projection model has more than one element. This means that different structures admitted by assumptions (A1)-(A2), and compatible with the available data, can deliver an identical value of θ. No amount of data generated according to assumption (A3) can distinguish the elements in the identified set Θ_I from $\mathbb{E}(YX)$. As a result, we say that the expectation $\mathbb{E}(YX)$ is set identified, and all the values of θ in the segment $\Theta_I = [\theta_L, \theta_U]$ are observational equivalent to $\mathbb{E}(YX)$.

We can interpret the bounds $\theta_L \leq \mathbb{E}(YX) \leq \theta_U$ as follows. Observe that the expression defining the upper bound θ_U replaces X_i in the expectation $\mathbb{E}(Y_i X_i)$ by the quantity $Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)$, that is, the quantile of X conditional on Z evaluated at the distribution of Y conditional on Z. Given the data, the quantity $Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)$ is the least upper bound for the value that the unobserved variable X_i can take. Similarly, $Q^o_{X|Z}(1-G^o_{Y|Z}(Y|Z)|Z)$ is the greatest lower bound on X_i. Since joint realization of (Y_i, X_i) are not observed, we can think of X_i as an unobserved real random variable bracketed by the observed random variables $Q^o_{X|Z}(1-G^o_{Y|Z}(Y|Z)|Z)$ and $Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)$. The bounds $\theta_L \leq \mathbb{E}(YX) \leq \theta_U$ correspond to the expectation between Y_i and the variables bracketing X_i.

Before going on, it is important to say that there is a way to bound the expectation $\mathbb{E}(YX)$ alternative to Theorem 1. This alternative way involves to invoke the Cauchy-Schwarz inequality (c.f., Ridder and Moffit, 2007). The bounds on the expectation $\mathbb{E}(YX)$ delivered by the Cauchy-Schwarz inequality however are not sharp in general. The intuition behind such a difference is that the Cauchy-Schwarz inequality restricts the correlation between Y and X conditional on Z to lie in the interval $[-1, 1]$ regardless of the functional form of the marginal distribution functions $G^o_{Y|Z}$ and $G^o_{X|Z}$, whereas for some specific functional forms of $G^o_{Y|Z}$ and $G^o_{X|Z}$ the bounds in Theorem 1 restrict such conditional correlation to lie in the interior of the interval $[-1, 1]$.

3.2 The Identified Set of the Coefficients (β_o, γ_o)

Here we characterize the identified set of the coefficients (β_o, γ_o) delivered by the linear projection model. To do so, we employ the characterization of the identified set of the expectation $\mathbb{E}(YX)$ derived in Theorem 1.

To proceed, we define the identified set of (β_o, γ_o). The linear projection model (A1)-

(A2) implies that the equations:

$$\beta = \frac{[\theta - \mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZY)]}{[\mathbb{E}(X^2) - \mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)]} \tag{6}$$

$$\gamma = \mathbb{E}(ZZ')^{-1}\mathbb{E}(ZY) - \mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)\beta(\theta) \tag{7}$$

have one and only one solution (β, γ) for any value of θ in the the identified set of $\mathbb{E}(YX)$. The identified set of (β_o, γ_o), say $\mathcal{I}_S$, is the collection of such solutions:

$$\mathcal{I}_S \equiv \left\{(\beta, \gamma) \in \mathbb{R}^{1+d_Z} : (\beta, \gamma')' = m(\theta, P_o) \text{ for all } \theta \in \Theta_I\right\}$$

where $m(\theta, P_o)$ is a $1+d_Z$-dimensional vector collecting the equations (6)-(7). The function $\theta \mapsto m(\theta, P_o)$ relates the coefficients (β, γ) with the expectation θ of the product between Y and X. Note that this function is linear. The identified set of (β_o, γ_o) is therefore nonempty (set e.g., $\theta = \mathbb{E}(YX)$), convex and bounded (since it is a linear transformation of the convex and bounded set Θ_I).

The above definition of the identified set of (β_o, γ_o) is not an operational one, in the sense that it does not allow the computation of $\mathcal{I}_S$ based on hypothetical knowledge of $G^o_{Y|Z}$ and $G^o_{X|Z}$. We look for an operational characterization of the set $\mathcal{I}_S$, next

3.2.1 The Support Function

The key object that we exploit to characterize the identified set of (β_o, γ_o) is its support function. To formally define this function, we need to introduce some notation. Let q be a vector of dimension $1 + d_Z$ belonging to the unit sphere $\mathbb{S}^{1+d_Z}$ in $\mathbb{R}^{1+d_Z}$. The support function $q \mapsto s(q, P_o)$ of the set $\mathcal{I}_S$ is defined by:

$$s(q, P_o|\mathcal{I}_S) \equiv \sup_{\theta \in [\theta_L, \theta_U]} q' \cdot m(\theta, P_o) \text{ for some } q \in \mathbb{S}^{1+d_Z}$$

To each direction q, the support function $s(q, P_o)$ equals the signed distance between zero and the orthogonal hyperplane that is tangent to $\mathcal{I}_S$.

The fact that the identified set $\mathcal{I}_S$ is convex guarantees that its support function $q \mapsto s(q, P_o)$ fully characterizes it (see Rockafellar, 1970). The support function represents the identified set $\mathcal{I}_S$ as an element of a functional space. To gain some intuition about this concept, note that the support function of the identified set of the expectation $\mathbb{E}(YX)$ is $s(q, P_o|\Theta_I) = \mathbb{I}(q = 1)\theta_U - \mathbb{I}(q = -1)\theta_L$ for q in $\{-1, 1\}$. Beresteanu and Molinari (2008), Bontemps, Magnac and Maurin (2011), and Kaido and Santos (2011) also employ the support function to characterize convex identified sets but in different applications.

3.2.2 Characterization of the Identified Set of the Coefficients (β_o, γ_o)

The next theorem provides an operational characterization of the identified set of (β_o, γ_o). This characterization arises after writing the support function $s(q, P_o|\mathcal{I}_S)$ in terms of expectations of the available data.

Theorem 2 *(Characterization of the Identified Set of (β_o, γ_o)). Let Assumptions (A1)-(A2) hold and suppose that data are available from independent random samples of (Y, Z) and of (X, Z) as described in Assumption (A3). Define the scalars θ_L and θ_U as in Theorem 1, and the function $\theta \mapsto \beta(\theta)$ as in equation (6). Let $\beta_L = \beta(\theta_L)$ and $\beta_U = \beta(\theta_U)$ denote the values of the function $\theta \mapsto \beta(\theta)$ evaluated at the scalars θ^L and θ^U, respectively. Let q denote a vector belonging to the unit sphere in in $\mathbb{R}^{1+d_Z}$. Split the vector q into $q = (q_\beta, q_\gamma')'$, where q_β is a scalar and q_γ is a vector with the remaining d_Z components. Then, the identified set of the coefficients (β_o, γ_o) delivered by the linear projection model is characterized by*

$$\mathcal{I}_S = \left\{ (\beta, \gamma) \in \mathbb{R}^{d_Z+1} : q' \cdot (\beta, \gamma) \leq s(q, G^o_{Y|Z}, Q^o_{X|Z}) \text{ ; for all } q \in \mathbb{S}^{d_Z} \right\}$$

where the support function $s(q, G^o_{Y|Z}, Q^o_{X|Z})$ equals to:

$$\begin{aligned} s(q, G^o_{Y|Z}, Q^o_{X|Z}) \;\; = \;\; & \mathbb{I}(q_\beta \neq 0) \left[\beta_L \times \mathbb{I}(q_\beta < 0) + \beta_U \times \mathbb{I}(q_\beta > 0) \right] + \mathbb{E}(q_\gamma' \Sigma_Z^{-1} Z Y) \\ & - \mathbb{E}\left[q_\gamma' \Sigma_Z^{-1} Z X \times \beta_U \times \mathbb{I}(q_\gamma' \Sigma_Z^{-1} Z < 0) \right] \\ & - \mathbb{E}\left[q_\gamma' \Sigma_Z^{-1} Z X \times \beta_L \times \mathbb{I}(q_\gamma' \Sigma_Z^{-1} Z \geq 0) \right] \end{aligned}$$

with $\Sigma_Z^{-1} = \mathbb{E}(ZZ')^{-1}$ the inverse of the variance matrix of Z.

Proof. See the Appendix A. ■

Putting differently, Theorem 2 characterizes the extreme points of the identified set of (β_o, γ_o) in terms of moments of the available data. According to Theorem 2, the identified set $\mathcal{I}_S$ has more than one element because the bounds $\beta_L = \beta(\theta_L, P_o)$ and $\beta_U = \beta(\theta_U, P_o)$ do not coincide. Therefore, the coefficients (β_o, γ_o) are set not point identified when data are available from independent samples on (Y, Z) and (X, Z). We stress the fact that the characterization of the identified set of (β_o, γ_o) in Theorem 2 is sharp, that is, it contains the values of the coefficients of interest compatible with the linear projection model (A1)-(A2) and the data described by assumption (A3) and no others. This means that all the elements in $\mathcal{I}_S$ are observational equivalent to (β_o, γ_o). The characterization in Theorem 2 is operational in that it can be employed, by the way of the analog principle, to construct an estimator of the identified set. We discuss the construction of such an estimator in Section 4.

Underpinning the characterization of the identified set $\mathcal{I}_S$ are the conditional Hoeffding-Frechet distributions introduced above. Ridder and Moffit (2007) employ these distributions to bound the joint distribution of (Y, X, Z). In the literature on treatment effects, Fan and Zhu (2010) also employ them to bound functionals of the joint distribution of potential outcomes. Theorem 2 seems the first result employing conditional Hoeffding-Frechet distributions to bound the coefficients of a linear projection model. Since Hoeffding-Frechet distributions are usually not informative about the distributions they bound, it is tempting to think that $\mathcal{I}_S$ would not very informative about (β_o, γ_o). However this is not generally the case. If the quantity $\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)$ is "small", data can be informative about γ_o. For instance in the extreme case where Z and X are uncorrelated, i.e. $\mathbb{E}(ZX) = 0$, the partial correlations γ_o are point identified. On the other hand, if the partial correlation between the response variable Y and one of the common covariates Z is "small", data can be informative about β_o. To see why, consider the extreme case where Z is an scalar whose partial correlation with Y is zero, that is, $\gamma_o = 0$. This case corresponds to the instrumental variable assumption on Z putting forward by many authors. When Z is an scalar instrument, it follows from the equation in (6) linking γ and β that the coefficient β_o is equal to $\beta_o = \mathbb{E}(YZ)/\mathbb{E}(XZ)$. Then, β_o is point identified because $\mathbb{E}(ZY)$ and $\mathbb{E}(XZ)$ so do (and the denominator $\mathbb{E}(XZ)$ is different from zero because the rank condition A2.ii).

Theorem 2 is useful to characterize the smallest and largest values of the coefficients (β_o, γ_o) compatible with the data. This is useful for practical purposes because these values, rather than the whole identified set, are usually of interest in applications. In particular, the smallest and largest values of that the coefficients (β_o, γ_o) can take equal the support function evaluated at the *canonical directions*. To be more precise, define q_L^k in $\mathbb{S}^{d_Z}$ as a vector whose $k'th$ entry is equal to minus one. Since q_L^k lives in the unit sphere $\mathbb{S}^{1+d_Z}$ all the other entries are necessarily equal to zero. We call q_L^k the *negative canonical* k*-direction*. Similarly, define q_U^k in $\mathbb{S}^{d_Z}$ as a vector whose $k'th$ entry is equal to one. We call q_U^k the *positive canonical* k*-direction*. Notice that the value of the support function in the negative canonical 1-direction is $s(q_L^1, G_{Y|Z}^o, Q_{X|Z}^o) = \beta_L$, that is, the smallest value of the coefficient β_o compatible with the available data. In the same sense, the value of the support function in the positive canonical 2-direction $s(q_U^2, G_{Y|Z}^o, Q_{X|Z}^o)$ is equal to the largest value of the first component of the vector γ_o compatible with the available data.

3.3 Discussion of Other Applications

Here we discuss two other potential applications of our identification results. These applications are the measurement of the variance of the treatment effect (e.g., Djebbari and Smith, 2008), and the measurement of the correlation coefficient from aggregate data (Robinson, 1950).

3.3.1 Variance of the Treatment Effect

The evaluation of social programs is a first application where our identification results apply. To be more precise, let Y, X and Z represent, respectively, the potential outcome from receiving a treatment, the potential outcome from not receiving the treatment, and some background variables not affected by the treatment. In some investigations in this literature (e.g. Heckman, Smith and Clements, 1997; Djebbari and Smith, 2008) knowledge is sought about the variance $\sigma^2(\theta_o)$ of the difference $Y - X$:

$$\sigma^2(\theta_o) \equiv \mathbb{V}(Y) + \mathbb{V}(X) - \theta_o + \mathbb{E}(Y)\mathbb{E}(X)$$

where $\theta_o \equiv \mathbb{E}(YX)$ denotes the expectation of the product between Y and X, and $\mathbb{V}$ denotes the variance operator. The parameter $\sigma_o^2(\theta_o)$ is the so-called variance of the treatment effect.

In randomize experiments, data are available from independent samples on (Y, Z) and on (X, Z). This implies that all the expectations in the definition of $\sigma_o^2(\theta_o)$ are point identified except θ_o. Theorem 1, and the fact that the function $\theta \mapsto \sigma^2(\theta)$ is decreasing, imply that the identified set the variance of the treatment effect is:

$$\Sigma_I = \left\{\sigma^2 \in \mathbb{R}_+ : \sigma_o^2(\theta_U) \leq \sigma^2 \leq \sigma_o^2(\theta_L)\right\}$$

The support function of the identified set Σ_I is:

$$s(q, G^o_{Y|Z}, Q^o_{X|Z}) = \mathbb{I}(q = 1)\sigma_o^2(\theta_L) - \mathbb{I}(q = -1)\sigma_o^2(\theta_U) \text{ for } q \in \{-1, 1\}$$

The bounds $\sigma_o^2(\theta_U) \leq \sigma^2(\theta_o) \leq \sigma_o^2(\theta_L)$ are sharp, unless additional assumptions are invoked.

Discussion. The latter characterization of the identified set of the variance of the treatment effect is new in the formed stated. Concurrent work by Fan and Zhu (2010) studies identification on *superadditive integral functionals* of the joint distribution of (Y, X) when the distributions of (Y, Z) and of (X, Z) are given. Since $F_{Y,X} \mapsto -\sigma_o^2(F_{Y,X})$ is a superadditive functional, the variance of the treatment effect fits their setting. Our characterization of the identified set Σ_I is however different from theirs. We express the bounds $\sigma_o^2(\theta_U)$, $\sigma_o^2(\theta_L)$ in terms of moments in a different way than they do. In the next paragraph, we show that this allows us to dispense with numerical integration procedures while estimating the bounds $\sigma_o^2(\theta_U)$ and $\sigma_o^2(\theta_L)$, a step required by the latter authors.

A plug-in sample analog estimator of these bounds can be obtained after replacing in their expressions the integral by sums and the unknown functions $Q^o_{X|Z}$ and $G^o_{Y|Z}$ by nonparametric estimates. To facilitate the description of such estimator, and without loss of generality, suppose that Y and X have both zero (known) mean and unit (known) variance. Then, the estimator of the lower bound $\sigma_o^2(\theta_U)$ is:

$$\hat{\sigma}_L = 2 - n_1^{-1} \sum_{i=1}^{n_1} Y_i \hat{Q}_{X|Z}(\hat{G}_{Y|Z}(Y_i, Z_i)|Z_i)$$

where $\hat{G}_{Y|Z}$ is a nonparametric estimator of the distribution function $G^o_{Y|Z}$. A similar expression follows for the estimator of the upper bound. We now compare the previous estimator with the one proposed by Fan and Zhu (2010). Fan and Zhu's (2010) estimator of the lower bound $\sigma_o^2(\theta_U)$ is:

$$\hat{\sigma}_L^{FZ} = 2 - n^{-1} \sum_{i=1}^{n} I_{ib} \int_0^1 \hat{Q}_{Y|Z}(u|Z_i)\hat{Q}_{X|Z}(u|Z_i)du \tag{8}$$

where $I_{ib} \equiv \mathbb{I}(|Z_i| \leq b)$ is a trimming sequence with $\mathbb{I}(\cdot)$ the indicator function, and $\hat{Q}_{Y|Z}$, $\hat{Q}_{X|Z}$ are, respectively, nonparametric estimates of the quantile functions $Q^o_{X|Z}$, $Q^o_{X|Z}$.Implementing the estimator $\hat{\sigma}_L^{FZ}$ requires to employ a numerical integration procedure to compute the integral in (8). Unlike $\hat{\sigma}_L^{FZ}$, no numerical integration procedure is required to implement $\hat{\sigma}_L$. To establish the connection between $\hat{\sigma}_L^{FZ}$ and $\hat{\sigma}_L$, it suffices to perform the change-of-variable $u = G^o_{Y|Z}(Y|Z)$ in (8) and replace $G^o_{Y|Z}$ by its nonparametric estimate. The estimator $\hat{\sigma}_L$ thus replace the numerical integration required to implement $\hat{\sigma}_L^{FZ}$ by the use of an empirical sum.

3.3.2 Ecological Correlation

The so-called ecological correlation problem is another application where our identification results apply. There knowledge is sought about the correlation coefficient between two discrete random variables, say Y and X,:

$$\rho(\theta_o) = \frac{\theta_o - \mathbb{E}(Y)\mathbb{E}(X)}{\mathbb{V}(Y)^{1/2}\mathbb{V}(X)^{1/2}}$$

but data provide only with *estimates* of the distribution of (Y, Z) and of (X, Z). A leading example of this situation arises in the study of voting behavior in elections with secret ballot. Let Y_i denote the vote of individual i, let X_i denote the educational level of voter i, and let Z_i denote the precinct where i votes. Suppose we are interested in the correlation $\rho(\theta_o)$ between voting behavior Y and educational level X in a presidential election with secret ballot. Since votes are secret, it is impossible to jointly observe voting behavior Y_i and educational level X_i. Election returns, however, allow us to estimate the distribution $G^o_{Y|Z}$ of the voting behavior by electoral precinct. Moreover, from census data we can estimate the distribution $G^o_{X|Z}$ of educational level by electoral precinct.

Hence the available data free of sample variation consist of the distributions $G^o_{Y|Z}$ and $G^o_{X|Z}$ as in Assumption A3. Under hypothetical knowledge of these distributions, all the expectations in $\rho(\theta_o)$ are known except for θ_o.

The current approach to solve the ecological correlation problem (e.g., Gentzkow, 2006; Snyder and Stromberg, 2010; Da Silveria and De Mello, 2011) is to aggregate the discrete variables Y and X by Z into shares, and then calculate the correlation between these shares. Robinson (1950) criticizes the tacit interpretation of the correlation between the shares, the so-called ecological correlation, as the correlation between Y and X. He points out the fact that there are many values of the correlation between Y and X compatible with knowledge of the estimates of the distribution of (Y, Z) and of (X, Z). Nevertheless, he neither characterizes such feasible values nor proposes inference procedures. Since $\theta \mapsto \rho(\theta)$ is linear and increasing, Theorem 1 can be employed to extend the insight by Robinson (1950) to provide a sharp characterization of the feasible values of the correlation between Y and X.

4 Inference Procedures

In this section, we discuss estimation and inference procedures for the coefficient of interest (β_o, γ_o). These procedures are based on the identification results derived in the previous section.

4.1 Estimation

We focus on the estimation of the smallest and largest values of the coefficients (β_o, γ_o) compatible with the available data. Our focus on these values, rather than in the whole identified set of (β_o, γ_o), is motivated by two reason. First, in most applications only the smallest and largest values of (β_o, γ_o) are of interest. Second, estimates of the whole identified set of (β_o, γ_o) can be difficult to display and communicate when the dimension of the covariate Z is greater than two (which is usually the case).

According to Theorem 2, the smallest (β_L, γ_L) and largest (β_U, γ_U) values of the coefficients (β_o, γ_o) are defined by in terms of the support function $s(q, G^o_{Y|Z}, Q^o_{X|Z})$ by the unconditional moment conditions:

$$\beta_L - s(q^1_L, G^o_{Y|Z}, Q^o_{X|Z}) = 0 \quad ; \quad \beta_U - s(q^1_U, G^o_{Y|Z}, Q^o_{X|Z}) = 0 \tag{9}$$

$$\gamma_{k,L} - s(q^k_L, G^o_{Y|Z}, Q^o_{X|Z}) = 0 \quad ; \quad \gamma_{k,U} - s(q^k_U, G^o_{Y|Z}, Q^o_{X|Z}) = 0 \tag{10}$$

for $k = 1, .., d_Z$. Moment conditions (9)-(10) contain the unknown functions $G^o_{Y|Z}$ and $Q^o_{X|Z}$. The values $(\beta_L, \gamma_L, \beta_U, \gamma_U)$ can be estimated by the method-of-moments after replacing the unknown functions $G^o_{Y|Z}$ and $Q^o_{X|Z}$ by nonparametric estimates. We denote the corresponding estimator by $(\hat{\beta}_L, \hat{\gamma}_L, \hat{\beta}_U, \hat{\gamma}_U)$. Different nonparametric estimators for the functions $G^o_{Y|Z}$ and $Q^o_{X|Z}$ are available in the literature. These include kernel-type estimators, local-linear estimators, and series-based estimators. In the illustration below, we estimate $G^o_{Y|Z}$ and $Q^o_{X|Z}$ by series of cubic splines.

The estimator $(\hat{\beta}_L, \hat{\gamma}_L, \hat{\beta}_U, \hat{\gamma}_U)$ belongs to the class of two-step semiparametric estimators studied, among others, by Chen, Linton and van Keilegom (2003). Consistency, root-n consistency, asymptotic normality, and validity of the bootstrap can be established after verifying the conditions proposed by these authors. In what follows, we assume that our setting meets the conditions proposed by these authors for root-n asymptotic normality and validity of the bootstrap.

It is important to note that the plug-in estimator $(\hat{\beta}_L, \hat{\gamma}_L, \hat{\beta}_U, \hat{\gamma}_U)$ is related to the general estimation procedure studied by Bontemps, Magnac and Maurin (2011) for incomplete linear models. The difference is that here the support function $q \mapsto s(q, G^o_{Y|Z}, Q^o_{X|Z})$ involved in the definition of the estimator depends on the infinite-dimensional nuisance parameters $G^o_{Y|Z}$, $Q^o_{X|Z}$. This difference prevents us to directly apply their asymptotic results the case at hand.

4.2 Confidence Intervals and Hypothesis Testing

The construction of confidence intervals for set identified parameters, such as (β_o, γ_o), raises a number of conceptual and technical issues that are subject of a currently active literature. We discuss two of these issues, next.

The first issue relates to the object to be covered by the confidence interval. In particular, should a confidence interval cover a set (such as the interval $[\beta_L, \beta_U]$ in our case), or should it cover the true value of the parameter (such as β_o)? The answer to this question depends on what is conceived as the quantity of interest. Romano and Shaikh (2010) provide methods to construct confidence intervals and testing hypothesis when the quantity of interest is a set. Their methods apply to our case whenever the estimator $(\hat{\beta}_L, \hat{\gamma}_L, \hat{\beta}_U, \hat{\gamma}_U)$ converges in distribution to a limit continuous distribution (such as the normal). As alluded before, the asymptotic normal approximation can be obtained in our case after verifying the conditions by Chen, Linton and van Keilegom (2003). Imbens and Manski (2004) propose confidence intervals for the true value of the coefficient of interest. Their methods apply to our case whenever the estimator $(\hat{\beta}_L, \hat{\gamma}_L, \hat{\beta}_U, \hat{\gamma}_U)$ converges in distribution to a normal limit distribution.

The second issue relates to the coverage probabilities of different confidence intervals. Some confidence intervals display undesirable behavior because they fail to capture the possibility that parameters of interest might be point identified. This undesirable behavior is reflected in the lack of coverage probabilities close to the nominal values for some possible distributions generating the data. Romano and Shaikh (2010) provide sufficient conditions under which their confidence intervals exhibit uniform coverage probabilities. Stoye (2009) discusses the case for confidence intervals covering the true value of the parameter of interest. Applied to our case, the conditions by Stoye (2009) include the existence of an asymptotic normal approximation for the estimator $(\hat{\beta}_L, \hat{\gamma}_L, \hat{\beta}_U, \hat{\gamma}_U)$, and the inequality restriction $(\hat{\beta}_L, \hat{\gamma}_L) \leq (\hat{\beta}_U, \hat{\gamma}_U)$ elementwise Here, this inequality restriction is satisfied by construction.

5 Illustration

In this section, we employ simulated data to illustrate the implementation and performance of the estimation and inference procedures discussed above.

Data Generating Process. For computational simplicity, we let Z_i to be a univariate random variable. For a true parameter $(\beta_o, \gamma_o) = (.5.5)$, we then generate:

$$Y_i = X_i\beta_o + Z_i\gamma_o + \varepsilon_i \quad i = 1, .., n$$

where ε_i is a standard normal random variable independent of (X_i, Z_i). The joint distribution of (X_i, Z_i) is bivariate normal. In order to create two independent sample as described in Assumption (A3), we drop the realizations of ε and split the n drawn of the vector (Y, X, Z) into two samples of size n_1 and n_2, respectively. In the first sample, we also drop the realized values of X, while in the second sample we drop the realized values of Y. We end up with two independent random samples $\{Y_i, Z_i\}_{l=1}^{n_1}$ and $\{X_i, Z_i\}_{i=1}^{n_2=n-n_1}$.

It should be notice that the distribution $F^o_{X,Z,\varepsilon}$ in the true structure $(\beta_o, \gamma_o, F^o_{X,Z,\varepsilon})$ is completely determined by covariance matrix of (X, Z, ε). We fix the variances of X and Z to one. The design variable of the experiment is the correlation between X and Z, say $\rho_{X,Z}$. We choose values for $\rho_{X,Z}$ in the set $\{-.25, .25\}$ to contrast the sensitivity of the identified set $\mathcal{I}_S$ to changes in the correlation between X and Z. We choose a sample size of $n_1 = n_2 = 250$. The number of replications in each experiment is equal to 1000.

Implementation. We estimate the distribution function $G^o_{Y|Z}$ and the quantile function $Q^o_{X|Z}$ by series of cubic splines. Implementing such estimators requires to choose the location and the numbers of knots. We place the knots at the quantiles of Z. We choose different numbers of knots and evaluate the sensitivity of the results to these different choices. All the experiment were carried out in the program R using the libraries "mvtnorm" (to generate bivariate normal random numbers), "splines" (to generate cubic spline basis) and "quantreg" (to solve estimate the quantile function $Q^o_{X|Z}$).

Results. Table I reports the smallest and largest values of the coefficients (β_o, γ_o) compatible with the available data for different values of the correlation between the covariates X and Z. As expected, the difference between the largest γ_U and smallest γ_L value of

γ_o decreases as the correlation ρ_{XZ} between the covariates X and Z approaches to zero. The smallest β_L and largest β_U value of β_o compatible with the data do not change with the correlation between X and Z.

Table II reports the mean square error (MSE) of the estimated smallest and largest values of the coefficients (β_o, γ_o), together with their Monte Carlo average (labeled Mean), for different values for the correlation between the covariates (ρ_{XZ}), and different choices of the number of knots. The results suggest that the choice of the number of knots has an important effect on the mean square error of the estimator of the lower and upper bounds on the coefficients of interest. In the experiments, the mean square error seems to be minimized for a choice of the number of knots between 40 and 50. There, the variance term is the main component of the mean square error. When the number of knots is too small (ie, 10) or too big (i.e. 90), the estimator exhibits a significant mean square error. There, the variance term accounts only for a half of the mean square error. When the number of knots is small, the bias is negative for the estimators of the lower bounds and positive for upper bounds. Then, the bias renders the estimator more likely to be outside the identified set. In such a case, we can expect confidence intervals with coverage probabilities above the prespecified nominal value. By contrast, when the number of knots is to big, the bias renders the estimator more likely to be inside the identified set (because the bias is positive for the estimators of the lower bounds and negative for upper bounds). Then, when the number of knots is too big we can expect confidence intervals with coverage probabilities below the nominal value.

6 Summary and Conclusions

Applied researchers employing the linear projection model are often confronted to the case where there is no single database that contains all the relevant variables. The existing literature suggests to overcome the difficulties associated to such lack of data by imposing ancillary assumptions. Researchers invoking such ancillary assumptions however rarely warn the readers about the consequences on the results of a failure of them. We conjecture that this omission is due to the lack of appropriate methods to determine the reliability of the results. This paper supplies such methods and given examples of their implications. The most important finding is that the coefficients in the linear projection model are set not point identified when no ancillary assumptions are invoked. The extent of the identified set depends on the dependence between the covariates.

There are at least two issues which deserves further research. The first one relates to the generalization of our results to the case where the covariate X observed in only one sample is a random vector rather than an scalar. The second issue relates the choice of the tuning parameters in the estimator we propose for the identified set.

Acknowledgments. *I wish to thank my advisor Thierry Magnac for invaluable help during this project. Stephan Bonhomme, Christian Bontemps, Niklas Buehren, Andrew Chesher, Fabiana Gomez, Nour Meddahi, Jorge Ponce, Christoph Rothe and seminar participants at Banco Central del Uruguay, EC*^{2} *meeting in Toulouse, Carlos III/Madrid, the NESG '10 meeting in Leuven, the ENTER Jamboree '10 meeting in Toulouse, and 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/*

Appendix A: Proofs and Tables

Proof of Proposition 1 The class $\mathcal{F}_{Y,X,Z}$ is non-empty because it contains the multivariate distribution which is such that Y and X are conditionally independent given Z - i.e., the distribution $F(y,x,z) = \int G^o_{Y|Z}(y,s) G^o_{X|Z}(x,s) dG_Z(s)$ is in $\mathcal{F}_{Y,X,Z}$. To verfy that $\mathcal{F}_{Y,X,Z}$ is convex, consider the elements $F, \tilde{F}$ both in $\mathcal{F}_{Y,X,Z}$ with associated densities f, $\tilde{f}$. These two elements satisfy:

$$\begin{aligned} \int \int_{-\infty}^{x} \int_{-\infty}^{z} f\left(y,a,b\right) dydadb &= G^o_{X,Z}(x,z); \\ \int \int_{-\infty}^{x} \int_{-\infty}^{z} \tilde{f}\left(y,a,b\right) dydadb &= G^o_{X,Z}(x,z) \end{aligned}$$

Let λ be a number between zero and one. Multiply both sides of the first equality by λ. Multiply both sides of the second equality by $(1-\lambda)$. Sum the resulting expressions. After taking common factor $G^o_{X,Z}(x,z)$, we have:

$$\lambda \int_{-\infty}^{y} \int \int_{-\infty}^{z} f\left(y,a,b\right) dydadb + (1-\lambda) \int_{-\infty}^{y} \int \int_{-\infty}^{z} \tilde{f}\left(y,a,b\right) dydadb = G^o_{X,Z}(x,z)$$

Then, the convex combination of $F, \tilde{F}$ has marginal *(X,Z)*-marginal distribution $G^o_{X,Z}$. By a similar argument, it is possible to show that the convex combination of $F, \tilde{F}$ has marginal (Y,Z)-marginal distribution $G^o_{Y,Z}$. It follows then that $\lambda F\left(y,x,z\right) + (1-\lambda)\tilde{F}\left(y,x,z\right)$ belongs to $\mathcal{F}_{Y,X,Z}$. Therefore, $\mathcal{F}_{Y,X,Z}$ is convex. ■

Proof of Theorem 1 *(Characterization of the Identified Set of θ_o).* We start by replacing the lower Hoeffding-Frechet bound $G^L_{Y,X,Z}$ in the objective function of the programming problem defining the extreme point θ^L:

$$\theta_L = \int_{\mathcal{Z}} \int_{\mathcal{Y}\times\mathcal{X}} y, x dG^L_{Y,X,Z}(y,x,z) dG^o_Z(z)$$

Let $Q^o_{Y|Z}(\tau,z)$ and $Q^o_{X|Z}(\upsilon,z)$ denote, respectively, the τ-quantile of Y given $Z=z$ and the υ-quantile of X given $Z=z$. By using the quantile substitution $y=Q^o_{Y|Z}(\tau,z)$ and $x=Q^o_{X|Z}(\upsilon,z)$ we get,

$$\theta^L=\int_{\mathcal{Z}}\int_{[0,1]\times[0,1]}Q^o_{Y|Z}(\tau,z)\times Q_{X|Z}(\upsilon,z)d\max\{0,\tau+\upsilon-1\}dG^o_Z(z)$$

Since $d\max\{0,\tau+\upsilon-1\}$ is different from zero only at $\tau+\upsilon-1=0$, we have the following analytical expression for θ^L:

$$\theta^L=\int_{\mathcal{Z}}\int_{[0,1]}Q^o_{Y|Z}(\tau,z)\times Q^o_{X|Z}(1-\tau,z)d\tau dG^o_Z(z)$$

By the change-of-variable $\tau=G^o_{Y|Z}(y|z)$:

$$\begin{aligned}\theta^L &= \int_{\mathcal{Z}}\int_{\mathcal{Y}}y\times Q^o_{X|Z}(1-G^o_{Y|Z}(y|z),z)dG^o_{Y,Z}(y,z)\\ &= \mathbb{E}[Y\cdot Q^o_{X|Z}(1-G^o_{Y|Z}(Y|Z)|Z)]\end{aligned}$$

where the expectation is with respect to the joint distribution of (Y,X), $G^o_{Y,Z}$. A similar reasoning leads to $\theta_U=\mathbb{E}[Y\cdot Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)]$. ■

Proof of Theorem 2 *(Characterization of the Identified Set of* (β_o,γ_o)*).* The support function of the set $\mathcal{I}_S$ is defined by:

$$s(q,P_o)\equiv\sup_{\theta\in[\theta_L,\theta_U]}q'\cdot m(\theta,P_o)\text{ for some }q\in\mathbb{S}^{1+d_Z}$$

where $m(\theta,P_o)\equiv(\beta(\theta,P_o),\gamma(\theta,P_o)')'$ is the function that relates the coefficients (β,γ) with the expectation θ of the product between Y and X. Recall that $\beta(\theta,P_o)$ $\gamma(\theta,P_o)$ are equal to

$$\begin{aligned}&\beta(\theta,P_o)=[\theta-\mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZY)]\times[\mathbb{E}(X^2)-\mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)]^{-1}\\ &\gamma(\theta,P_o)=\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZY)-\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)\beta(\theta)\end{aligned}$$

Split the vector of directions q into $q=(q_\beta,q_\gamma')$, where q_β is a scalar, and q_γ is a vector with d_Z components. Using this notation, we can reexpress the support function as

$$s(q,P_o)\equiv\sup_{\theta\in[\theta_L,\theta_U]}\left[q_\beta\beta(\theta,P_o)+q_\gamma'\gamma(\theta,P_o)\right]\text{ for some }(q_\beta,q_\gamma')\in\mathbb{S}^{1+d_Z}$$

Since $\theta\mapsto\beta(\theta,P_o)$ and $\theta\mapsto\gamma(\theta,P_o)$ are linear functions, the support function $s(q,P_o)$

evaluated at q is equal to:

$$s(q, P_o) \equiv \sup_{\theta \in [\theta_L, \theta_U]} q_\beta \beta(\theta, P_o) + \sup_{\theta \in [\theta_L, \theta_U]} \gamma(\theta, P_o)$$

Since the function $\theta \mapsto \beta(\theta, P_o)$ is increasing, the value function of the linear programming problem $\sup_{\theta \in [\theta_L, \theta_U]} q_\beta \beta(\theta, P_o)$ is:

$$\mathbb{I}(q_\beta \neq 0) \left[\beta_L \times \mathbb{I}(q_\beta < 0) + \beta_U \times \mathbb{I}(q_\beta > 0)\right]$$

where $\beta_L = \beta(\theta_L, P_o)$ is equal to the function $\theta \mapsto \beta(\theta, P_o)$ evaluated at θ^L, and similarly for β_U. In turns, the value function of the linear programming problem $\sup_{\theta \in [\theta_L, \theta_U]} \gamma(\theta, P_o)$ is:

$$\begin{aligned} &\mathbb{E}(q_\gamma' \Sigma_Z^{-1} Z Y) - \mathbb{E}\left[q_\gamma' \Sigma_Z^{-1} Z X \times \beta_U \times \mathbb{I}(q_\gamma' \Sigma_Z^{-1} Z < 0)\right] \\ &-\mathbb{E}\left[q_\gamma' \Sigma_Z^{-1} Z X \times \beta_L \times \mathbb{I}(q_\gamma' \Sigma_Z^{-1} Z \geq 0)\right] \end{aligned}$$

Therefore, the support function is equal to:

$$\begin{aligned} s(q, G^o_{Y|Z}, Q^o_{X|Z}) \quad = \quad & \mathbb{I}(q_\beta \neq 0) \left[\beta_L \times \mathbb{I}(q_\beta < 0) + \beta_U \times \mathbb{I}(q_\beta > 0)\right] + \mathbb{E}(q_\gamma' \Sigma_Z^{-1} Z Y) \\ & -\mathbb{E}\left[q_\gamma' \Sigma_Z^{-1} Z X \times \beta_U \times \mathbb{I}(q_\gamma' \Sigma_Z^{-1} Z < 0)\right] \\ & -\mathbb{E}\left[q_\gamma' \Sigma_Z^{-1} Z X \times \beta_L \times \mathbb{I}(q_\gamma' \Sigma_Z^{-1} Z \geq 0)\right] \end{aligned}$$

Since β_L and β_U depend on the distribution function $G^o_{Y|Z}$ and the quantile function $Q^o_{X|Z}$ so does the support function $s(q, G^o_{Y|Z}, Q^o_{X|Z})$. ■

Tables

Table I. Smallest and Largest Values of the Coefficients Compatible with the Data

Bound	Correlation Between X and Z (ρ_{XZ})										
	$-.9$	$-.75$	$-.5$	$-.25$	$-.1$	0	$.1$	$.25$	$.5$	$.75$	$.9$
β_L	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5
β_U	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
γ_L	-.4	-.25	0	.25	.25	.5	.5	.5	.5	.5	.5
γ_U	.5	.5	.5	.5	.5	.5	.6	.75	1	1.25	1.4

Table II. Monte Carlo Exercises. Mean Square Error and Bias of the Estimator
Sample Sizes: $n_1 = n_2 = 250$. *Value of the coefficients:* $(\beta_o, \gamma_o) = (.5, .5)$

		$\rho_{XZ} = -.25$			$\rho_{XZ} = 0$			$\rho_{XZ} = .25$		
Knots	Bound	MSE	Mean	%Bias	MSE	Mean	% Bias	MSE	Mean	% Bias
10	$\hat{\beta}_L$	.295	-.881	48 %	.252	-.851	48 %	.168	-.784	47 %
	$\hat{\beta}_U$	.201	.812	48 %	.251	.848	48 %	.352	.915	49 %
	$\hat{\gamma}_L$	.028	.162	26 %	.012	.488	21 %	.034	.395	31 %
	$\hat{\gamma}_U$	.021	.581	30 %	.009	.542	18 %	.017	.813	23 %
40	$\hat{\beta}_L$	.021	-.576	29 %	.009-	.536	14 %	.021	-.488	1 %
	$\hat{\beta}_U$	.006	.509	1 %	.009	.535	13 %	.006	.594	34 %
	$\hat{\gamma}_L$	.008	.232	3 %	.007	.474	8 %	.008	.482	2 %
	$\hat{\gamma}_U$	.005	.497	1 %	.007	.529	10 %	.005	.758	1 %
50	$\hat{\beta}_L$	.007	-.495	1 %	.010	-.467	10 %	.028	-.400	34 %
	$\hat{\beta}_U$	.013	.443	22 %	.011	.457	15 %	.008	.512	2 %
	$\hat{\gamma}_L$	.007	.254	1 %	.009	.476	5 %	.000	.500	1 %
	$\hat{\gamma}_U$	.005	.485	3 %	.008	.521	5 %	.006	.728	7 %
60	$\hat{\beta}_L$	.025	-.419	25 %	.028	-.396	38 %	.052	-.352	41 %
	$\hat{\beta}_U$	.044	.361	44 %	.028	.395	37 %	.021	.431	21 %
	$\hat{\gamma}_L$	.009	.271	4 %	.008	.471	9 %	.009	.527	7 %
	$\hat{\gamma}_U$	.006	.479	6 %	.006	.512	2 %	.008	.725	6 %
90	$\hat{\beta}_L$	.112	-.271	46 %	.139	-.244	46 %	.226	-.178	45 %
	$\hat{\beta}_U$	.193	.199	46 %	.171	.251	36 %	.129	.266	41 %
	$\hat{\gamma}_L$	.012	.306	26 %	.006	.487	2 %	.015	.561	23 %
	$\hat{\gamma}_U$	.017	.419	36 %	.006	.512	2 %	.018	.671	33 %

References

ANDERSON, Michael, and David MATSA (2011): "Are Restaurants Really Supersizing America", *American Economic Journal: Applied Economics*, 3(1), pp. 152-188.

ANGRIST, Joshua and Alan KRUEGER (1992): “The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples”, *Journal of the American Statistical Association* 87, pp. 328–336.

BERESTEANU, Arie and Francesca MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, 76(4) pp. 763-814.

BONTEMPS, Christian, Thierry MAGNAC and Eric MAURIN (2011): "Set Identified Linear Models", *TSE Working Paper Series 09-090.*

BOSTIC, Raphael, Gabriel STUART and Gary PAINTER (2009): “Housing Wealth, Financial Wealth, and Consumption: New Evidence from Micro Data”, *Regional Science and Urban Economics,* 39 (1) pp. 79-89.

CHEN, Xiaohong, Olivier LINTON and Ingrid van KEILEGOM (2003): "Estimation of Semiparametric Models when the Criterion Function Is not Smooth", *Econometrica*, 71(5), 1591-1608.

CROSS, Phillip and Charles MANSKI (2002): "Regressions, Short and Long", *Econometrica*, 70(2), pp. 357-368.

Da SILVEIRA, Bernardo, and Joao De MELLO (2011): "Campaign Adverstising and Elections Outcome: Quasi Natural Experiment Evidence from Gubernatorial Elections in Brazil", *Review of Economic Studies*, 78(2), pp. 590-612.

FAN, Yanquin and Dongming ZHU (2010): "Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes", *unpublished manuscript.* Available at: http://www.econ.yale.edu/seminars/ecm/ecm10/fan-100303.pdf

FRISCH, Ragnar (1934): *Statistical Confluence Analysis by Means of Complete Regression Systems*, University Institute for Economics, Norway.

GENTZKOW, Matthew (2006): "Television and Voter Turnout", *Quarterly Journal of Economics*, 121(3), pp. 931-972.

GINI, Conrado (1921): "Sull'interpolazione di una Retta quando i Valori della Variabile Independenti Soni Affetti da Errori Accidentali", *Metroeconomica*, 1, pp. 63-82.

HECKMAN, James, Jeffry SMITH and Nancy CLEMENTS (1997): "Making the Most Out of Programme Evaluation and Social Experiments: Accounting for Heterogeneity in Program Impacts", *Review of Economic Studies*, 64(4), pp. 487-535.

HOROWITZ, Joel and Charles MANSKI (2006): "Identification and Estimation of Statistical Functionals Using Incomplete Data", *Journal of Econometrics*, 132, pp. 445-459.

HU, Yingyao (2006): "Bounding Parameters in a Linear Regression Model with a Mismeasured Regressor Using Additional Information", *Journal of Econometrics*, 133, pp. 51-70.

ICHIMURA, Hidehiko and Elena MARTINEZ-SANCHIS (2010): "Identification and Estimation of GMM Models by Combining Two Data Sets", *unpublished manuscript.*

IMBENS, Guido and Charles MANSKI (2004): "Confidence Intervals for Partially Identified Parameters", *Econometrica*, 74 pp. 1845-1857

KAIDO, Hiroaki and Andres SANTOS (2011): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities", *unpublished manuscript.*

KLEPPER, Steven and Edward LEAMER (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables ", *Econometrica*, 52:1, pp. 163-184.

KRASKER, William and John PRATT (1986): "Bounding the Effects of Proxy Variables on Regression Coefficients", *Econometrica*, 54, pp. 641-655.

MOLINARI, Francesca and Marcin PESKI (2006): "Generalization of a Result on Regressions, Short and Long", *Econometric Theory*, 22(1) pp. 159-163.

NICOLETTI, Chetti and John ERMISCH: "Intergenerational Earnings Mobility: Changes Across Cohorts in Britain", *The BE Journal of Economic Analysis and Policy*, 7(2).pp. 1-36.

ROCKAFELLAR, Tyrrel (1970): *Convex Analysis*, Princeton University Press.

ROMANO, Joseph, and Azeem SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models", Econometrica, 78(1), pp. 169-211.

RUSCHENDORF, Ludger (1991): "Bounds for Distributions with Multivariate Marginals", in K. Mosler and M. Scarsini (eds.), *Stochastic Orders and Decision under Risk*, IMS Lecture Notes-Monograph Series, Vol. 19, pp. 285-310.

RIDDER, Geert and Robert MOFFIT (2007): "The Econometrics of Data Combination", in J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Volume 6B, Elsevier.

ROBINSON, W. (1950): "Ecological Correlation and the Behavior of Individuals", *American Sociological Review,* 15(3), pp. 351-357.

SNYDER, James and David STROMBERG (2010): "Press Coverage and Political Accountability", *Journal of Political Economy*, 118(2), pp. 355-81.

STOYE, Jörg (2009): "More on Confidence Intervals for Partially Identified Parameters", *Econometrica*, 77(4), pp. 1299-1315.

VITALE, Richard (1979): "Regression with Given Marginals", *The Annals of Statistics*, 7(3), pp. 653 -658.