Least Square Linear Prediction with Two-Sample Data

David H. Pacini

Discussion Paper No. 12/631

November 19, 2012

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN

Least Square Linear Prediction with Two-Sample Data

David H. Pacini*

University of Bristol

November 19, 2012

Abstract

This paper investigates the identification and estimation of the least square linear predictor for the conditional expectation of an outcome variable Y given covariates (X, Z') from data consisting of two independent random samples; the first sample contains replications of the variables (Y, Z') but not X, while the second sample contains replications of (X, Z') but not Y. The contribution is to characterize the identified set of the least square linear predictor when no assumption on the joint distribution of (Y, X, Z'), except for the existence of second order moments, is imposed. We show that the identified set is not a singleton, so the least square linear predictor of interest is set identified. The characterization is used to construct a sample analog estimator of the identified set. The asymptotic properties of the estimator are established and its implementation is illustrated via Monte Carlo exercises.

KEYWORDS: Identification; Least Square Linear Prediction; Two samples.

JEL CLASSIFICATION: C21, C26

*Author address: Department of Economics, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK; Email: David.Pacini@bristol.ac.uk; Tel.: +44 (0) 11 79 28 84 37

1. Introduction

Least square linear predictors are employed to approximate conditional expectations while guarding against misspecification and the curse of dimensionality (see Goldberger, 1991 or Wooldridge, 2002 for a textbook exposition). Economists who use survey data for inferences about least square linear predictors often face the situation where variables of interest are observed in different samples (c.f., Meghir and Palme, 1999; Bover, 2005; Fang, Keane and Silverman, 2008; Flavin and Nakagawa, 2008; Bostic, Stuart and Painter, 2009; Brzozowski, Gervais, Klein, and Suzuki, 2010). Consumption and wealth, for example, are seldom measured together for a single sample of households. Separate measurements however are more often available. In the US, for instance, the Consumer Expenditure Survey measures consumption and other household socioeconomic characteristics but wealth for a sample of households. The Survey of Consumer Finances measures wealth and socioeconomic characteristics but consumption for a different sample of households. In this context, complications arise because the least square linear predictor depends on moments of variables observed in different samples. The prominent method adopted to overcome these complications is to impose additional assumptions on the distribution of the variables of interest. Assumptions include restricting the dependence between the variables observed in different samples (c.f., Rassler, 2002) or requiring the presence of an instrumental variable observed in all samples (c.f., Angrist and Krueger, 1992; Arellano and Meghir, 1992). In general, these assumptions are not testable. Then, it is worthy of consideration to analyze the sensitivity of inference to a failure of them. Little is known however about what can be ascertained about least square linear predictors in such a case, except that the underlying conditional expectation is unidentified (Cross and Manski, 2002; Ridder and Moffit, 2007).

Motivated by the previous situation, we investigate what can be learned about the least square linear predictor for the conditional expectation of an outcome variable Y given covariates (X, Z') 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 variables (X, Z') but not Y. Here Y is a scalar outcome variable (such as household expenditures), X is a scalar covariate (such as household wealth) and Z is a vector of other covariates possibly including a constant. Complications arise because the unknown least square linear predictor depends on the joint distribution of the

variables (Y, X, Z') but none of the samples has joint information on these variables. The main contribution of this paper is to characterize the *identified set* of the least square linear predictor, that is, the set of least square linear predictors compatible with knowledge of the distribution of (Y, Z') and of (X, Z'). This characterization is useful to evaluate the sensitivity of inferences to a failure of the assumptions commonly invoked to achive point identification of the least square linear predictor of interest. It is based on the insight by Ridder and Moffit (2007) that Hoeffding-Frechet distributions bound the joint distribution of (Y, X, Z') from knowledge of their marginals. We employ these bounds to show that the *identified set* of the least square linear predictor is a line. We derive an analytical expression for the endpoints of this line, and employ the sample analog method to estimate them. We establish the asymptotic properties of the resulting sample analog estimator following existing results on estimation of convex identified sets (c.f., Beresteanu and Molinari, 2008; Bontemps, Magnac and Maurin, 2012). We illustrate the implementation of the estimator via Monte Carlo experiments.

The problem of learning about the least square linear predictor when data are available from two independent samples has been studied in several strands of literature under different concerns and methodologies.[1] The first strand of related literature focuses on matching-based estimation of linear regression coefficients (see the survey by Ridder and Moffit, 2007). In this strand of literature, complications arising from the lack of observations on (Y, X) are overcome by imputing the values of Y in the second sample (or the values of X in the first sample). The imputation procedures employed are valid under the assumption that Y and X are independent conditional on Z. This assumption however often finds very little justification in practice. We do not assume either that Y and X are independent conditional on Z or that the conditional expectation of Y given (X, Z') is linear. Our results are thus useful to see what is lost when the conditional independence assumption and/or the linearity assumption not valid. It turns out that what is lost is point identification. The second strand of related literature studies estimation when instrumental variables are available (e.g., Angrist and Krueger, 1992: Arellano and Meghir, 1992). In this strand of literature, complications arising from the lack of observations on (Y, X) are overcome by assuming that one component in the list of common variables Z is an instrument. Although this assumption does guarantee

[1]The fact that the two samples are independent distinguishes our problem from the problem with samples with common units (e.g., Devereux and Triphati, 2009; Komarova, Nekipelov, Yakovlev, 2012, and references therein).

point identification, it is often the case that instrumental variables are not available. Our results are useful to estimate the least square linear predictor when there are no instruments common to both samples. The last strand of related literature focuses on nonparametric identification of the conditional expectation of Y given (X, Z') when the common variables Z are discrete (e.g., Vitale, 1979; Cross and Manski, 2002; Molinari and Peski, 2006). In this strand of literature, identification analysis is carried out without imposing additional assumption delivering point identification of the conditional expectation of Y given (X, Z'). Our work is in the same spirit, but applies to a different setting. First, we do not restrict the common variables Z to be discrete. Second, our focus is on the least square linear predictor rather than the conditional expectation itself.

The outline of the paper is as follows. In the next section, we define the least square linear predictor, describe the data, and discuss an example fitting our setup. In Section 3, we present the main result of the paper, namely the characterization of the identified set. We also discuss the force of additional assumptions to shrink the identified set to a singleton. In Section 4, we describe the sample analog estimator of the identified set and establish its asymptotic properties. In Section 4, we illustrate via Monte Carlo exercises the finite-sample performance of the estimator. In Section 5, we discuss how our identification results can be extended to the problem of measuring the variance of treatment effect and to the ecological correlation problem. Section 6 concludes.

2. The Setup

In this Section, we first set out the assumptions defining the least square linear predictor. We then describe the available data consisting of two independent random samples with common variables. Finally, we discuss an example fitting this setup.

The parameter of interest in this paper is a vector of coefficients representing the least square linear predictor under square loss of a conditional expectation. To define this vector, we need to introduce some notation. We 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 define the random vector (Y_i, X_i, Z_i') on a probability space endowed with probability measure P_o. We suppress the subscript i in the notation whenever this can be done without causing confusion. We use the expression $\mathbb{E}$ to denote the expectation associated to P_o. We employ uppercase letters

to denote random variables and lowercase letters to denote their realizations.

We define the vector of coefficients of interest by the following assumption:

Assumption 1 *(Parameter of Interest) Knowledge is sought about the coefficients* $\beta_o = (\beta_{1o}, \beta_{2o}')'$ *defined by:*

(A1.i) $$\beta_o := \arg\min_{b_1, b_2} \mathbb{E}[(Y - Xb_1 - Z'b_2)^2]$$

where the joint distribution of (Y, X, Z')*, say* $F^o_{Y,X,Z}$*, is such that:*

(A1.ii) The random vector (Y, X, Z') *has finite variance;*

(A1.iii) The variance of (X, Z') *has rank* $1 + d_Z$*, where* d_Z *is the dimension of* Z*.*

An equivalent way of writing (A1.i) is $\beta_o := \arg\min_{b_1, b_2} \mathbb{E}[(\mathbb{E}(Y|X, Z) - Xb_1 - Z'b_2)^2]$, which shows that β_o can be interpreted as the least square linear predictor of the conditional expectation of Y given (X, Z') under quadratic loss. The moment restriction (A1.ii) and the rank restriction (A1.iii) secure that β_o is finite. The first order condition associated to the quadratic programming problem (A1.i) is $\mathbb{E}[(Y - X\beta_{1o} - Z'\beta_{2o})(X, Z')] = 0$. This uncorrelatedness restriction is weaker than the mean-independence restriction $\mathbb{E}[(Y - X\beta_{1o} - Z'\beta_{2o})|X, Z'] = 0$, which defines β_o as the vector of partial derivatives of the conditional expectation of Y given (X, Z'). Indeed, Assumption (A1) does not restrict the conditional expectation of Y given (X, Z') to be a linear function. This distinction is relevant in our context because uncorrelatedness and mean-independence shall deliver different identification results for the conditional expectation Y given (X, Z').

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

Assumption 2 *(Data) Let* $G^o_{Y,Z}$ *denote the* (Y, Z')*-marginal distribution of* $F^o_{Y,X,Z}$*. A similar notation is adopted for* $G^o_{X,Z}$*. Data are available from two independent samples. 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* $\{X_i, Z_i'\}_{i=n_1+1}^{n}$*, 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.*

To close this Section, we discuss a concrete example fitting our setup. This example comes from the work by Bostic, Gabriel, and Painter (2009, BGP from now on). They employ two-sample data to measure the dependence between household consumption and housing wealth after controlling for households characteristics. We concentrate only on one of the specifications considered by BGP. Let Y_i denote the log consumption of household i living in the US in 2001. The standard specification for Y_i proposed by BGP is $Y_i = X_i\beta_{1o} + Z_i'\beta_{2o} + \varepsilon_i$, where X_i is the log of household's house value, and Z_i is a vector of household characteristics including income, number of members, and like controls. As in BGP, we assume that the error term ε and the covariates (X, Z') are uncorrelated, which is equivalent to say that the vector (β_{1o}, β_{2o}) is the least square linear predictor of the conditional expectation of log consumption given the covariates (X, Z'). Interest is on the coefficient β_{1o}, which measures the degree of linear association between log consumption Y and log house value X after controlling for household's characteristics Z. Learning about β_{1o} would ideally require measurements on (Y, X, Z) for a single sample of households. Since such data are not available, BGP employ data from two samples; the Consumer Expenditure Survey (CEX) and the Survey of Consumer Finances (SCF).[2] The CEX provides information on households' consumption and characteristics, that is on (Y, Z), but not on households' house value X. The SCF in turn provides information on households' house value and characteristics, that is on (X, Z'), but not on household consumption Y. The CEX and CSF do not survey the same households because they are independent samples. To overcome the complications arising from the lack of joint realizations on (Y, X), BGP employ an imputation procedure. This procedure is valid under the assumption that log consumption Y and households' house value X are independent conditional on household characteristics Z, which in turns implies $\mathbb{E}(Y|X, Z) = \mathbb{E}(Y|Z)$. We are concerned with the situation where this conditional independence assumption might not hold. In the next section, we derive result permitting to evaluate the sensitivity of inferences to a failure of the assumption that Y and X are independent conditional on Z.

[2]One important caveat to this example should be kept in mind. A key presumption underlying our analysis is that data are obtained by simple random sampling (see Assumption 2). The SCF does not use a simple random but a dual frame sampling design. The identification results below however still applies.

3. Identification

In this Section, we characterize the identified set of the vector of coefficients β_o. This is the main result of the paper. We also discuss the force of additional assumptions on the joint distribution of (Y, X, Z) to shrink the identified set to a singleton.

For identification purposes, we assume that the distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$ characterizing the two samples are known. We begin the identification analysis by describing the identification problem. Let $\lambda_o := \mathbb{E}(YX)$ denote the value of the expectation of the product of Y and X. Using the the first order condition of the programming problem (A1.i) the coefficients of interest β_o can be written

$$\begin{pmatrix} \beta_{1o} \\ \beta_{2o} \end{pmatrix} := \begin{pmatrix} \frac{[\lambda_o - \mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZY)]}{[\mathbb{E}(X^2) - \mathbb{E}(XZ')\mathbb{E}(ZZ')^{-1}\mathbb{E}(ZX)]} \\ \mathbb{E}(ZZ')^{-1} \cdot [\mathbb{E}(ZY) - \mathbb{E}(ZX)\beta_{1o}] \end{pmatrix}$$

Since $G^o_{Y,Z}$ and $G^o_{X,Z}$ are known, all the expectations in the latter display are known, except for λ_o. Let $m(\lambda_o)$ denote the right hand side in the latter display. The identification problem is to derive an operational characterization of the set B_S of vectors β in $\mathbb{R}^{d_Z+1}$ such that $\beta = m(\lambda)$ and λ is compatible with knowledge of the marginal distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$.

The next Theorem shows that knowledge of the distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$ restricts the expectation λ_o to lie in an interval, whose extreme points are moments of the available data.

Theorem 1 *(Bounds on the Expectation of the Product of Y and X). Let Assumptions (A1) and (A2) hold. Let $Q^o_{X|Z}$ denote the quantile function of X given Z, and define the quantities:*

$$\lambda_L := \mathbb{E}\big[YQ^o_{X|Z}\big(1 - G^o_{Y|Z}(Y|Z)\big)|Z\big] \quad ; \quad \lambda_U := \mathbb{E}[YQ^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z]$$

where $G^o_{Y|Z}$ is the distribution of Y given Z. Then, the expectation $\lambda_o := \mathbb{E}(YX)$ lies in the interval $[\lambda_L, \lambda_U]$.

The interval $[\lambda_L, \lambda_U]$ contains values of the expectation of the product of Y and X compatible with the marginal distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$. A few remarks about this interval are in order:

Remark 1.1 Our intuition for the result in Theorem 1 is the following. When the distributions of (Y, Z) and of (X, Z) are given, the maximal possible value λ_U of the expectation λ_o of the product of Y and X occurs when Y is commonotic with X given Z, that is, when $X = Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)$ or the joint distribution Y and X conditional on Z is the upper Hoeffing-Frechet bound. The upper bound λ_U is then the expectation $E(YX)$ evaluated at $X = Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)$. Similarly, the minimal possible λ_L of the expectation λ_o of the product of Y and X occurs when Y is anti-commonotic with X given Z , that is, when $X = Q^o_{X|Z}(1 - G^o_{Y|Z}(Y|Z)|Z)$ or the joint distribution Y and X conditional on Z is the lower Hoeffing-Frechet bound.

Remark 1.2 The bounds $\lambda_L \leq \mathbb{E}(YX) \leq \lambda_U$ are sharp. Note that the lower bound λ_L does not coincide with upper bound λ_U because $1 - G^o_{Y|Z}(Y|Z)|Z)$ is different from $G^o_{Y|Z}(Y|Z)$. There is therefore more than one value of the expectation of the product of Y and X compatible with the available data free of sample variation. We could also bound the expectation $\mathbb{E}(YX)$ from the fact that the variance matrix of (Y, X, Z) is positive semidefinite (c.f., Ridder and Moffit, 2007). The resulting bounds however are not sharp. We appeal to the following intuition to illustrate this point. Positive semidefinitiveness of the variance matrix of (Y, X, Z) implies that the conditional correlation between Y and X given Z is in the interval $[-1, 1]$. From these bounds, we can derive bounds on the expectation of the product of Y and X. The latter bounds are not sensitive to the functional form of the marginal distributions $G^o_{Y|Z}$ and $G^o_{X|Z}$. By contrast, for some forms of $G^o_{Y|Z}$ and $G^o_{X|Z}$ (i.e, for $G^o_{Y|Z}$ and $G^o_{X|Z}$ both lognormal), the bounds in Theorem 1 restrict the correlation between Y and X conditional on Z to lie in the interior of the interval $[-1, 1]$. Hence, bounds based on the restriction that the variance matrix of (Y, X, Z) is positive semidefinite are generally wider than those in Theorem 1.

Remark 1.3 Theorem 1 is new in the form stated, but its intuition was anticipated by numerous previous authors (Heckman, Smith and Clements, 1997; Fan and Zhu, 2010). For our purposes, it serves as a backdrop to the next result in the paper, namely a sharp and operational characterization of the identified set of β_o.

With Theorem 1 in hand, it is now straightforward to establish that the identified set B_S is a bounded convex set. From the definition of the the coefficients of interest, notice that these parameters equal the value of a linear mapping $\lambda \mapsto m(\lambda)$ evaluated at λ_o. The identified set B_S

is therefore nonempty (set e.g., $\lambda = \mathbb{E}(YX)$), bounded (since it is a bounded transformation of the bounded interval $[\lambda_L, \lambda_U]$), and convex (since it is a linear transformation of the interval $[\lambda_L, \lambda_U]$). From the mapping $m(\lambda)$ defining the coefficients of interest notice that the identified set B_S can be further characterized as the line in $\mathbb{R}^{d_Z+1}$ joining the points $(\beta_{1L}, \beta_{2l}')$ and $(\beta_{1U}, \beta_{2u}')$ with

$$\begin{aligned} \beta_{1L} &:= \frac{[\lambda_L - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZY)]}{[\mathbb{E}(X^2) - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZX)]} \qquad \beta_{1U} := \frac{[\lambda_U - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZY)]}{[\mathbb{E}(X^2) - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZX)]} \\ \beta_{2l} &:= \mathbb{E}(\Sigma ZY) - \mathbb{E}(\Sigma ZX)\beta_{1L} \qquad \beta_{2u} := \mathbb{E}(\Sigma ZY) - \mathbb{E}(\Sigma ZX)\beta_{1U} \end{aligned}$$

These points can be estimated from data using the sample analog principle.

In a given application, such as the one discussed in Section 2, one may be interested only ine one component of the vector β_o. In such a case, the one-dimensional projections of the identified set would be of interest. The one-dimensional projection of the identified set in the β_1-axis is the segment $[\beta_{1L}, \beta_{1U}]$. If Z has one component, the one-dimensional projection of the identified set on the β_2-axis is the segment $[\beta_{2L}, \beta_{2U}] := [\inf_\lambda (0,1)'m(\lambda), \sup_\lambda (0,1)'m(\lambda)]$, so

$$\begin{aligned} \beta_{2L} &= \mathbb{E}(\Sigma ZY) - \mathbb{E}(\Sigma ZX)\big[\beta_{1L}\mathbf{1}(\mathbb{E}(ZX) < 0) + \beta_{1U}\mathbf{1}(\mathbb{E}(ZX) \geq 0)\big] \\ \beta_{2U} &= \mathbb{E}(\Sigma ZY) - \mathbb{E}(\Sigma ZX)\big[\beta_{1U}\mathbf{1}(\mathbb{E}(ZX) < 0) + \beta_{1L}\mathbf{1}(\mathbb{E}(ZX) \geq 0)\big], \end{aligned}$$

where $\mathbf{1}()$ is the indicator function, and Σ is the inverse of the variance of Z. To extend this idea to the case that Z is a vector and relate our results with the literature on identification of convex sets, we next employ the concept of *support function*. For some $d_Z + 1$-vector belonging to the unit sphere $\mathbb{S}^{1+d_Z}$ in $\mathbb{R}^{1+d_Z}$, the support function $q \mapsto s(q)$ of the identified set B_S is:

$$s(q) := \sup_{\lambda \in [\lambda_L, \lambda_U]} q' \cdot m(\lambda)$$

where the bounds λ_L and λ_U were introduced in Theorem 1. To each direction q, the support function $s(q)$ equals the signed distance between zero and the orthogonal hyperplane that is tangent to the identified set. Put differently, the support function characterizes the boundary of the identified set. The fact that the identified set is convex does guarantee that its support function $q \mapsto s(q)$ fully characterizes it (see Hiriart-Urruty and Lemarechal, 2004). By evaluating the support function

in a given direction, we can calculate the lower-dimensional projections of the identified set. To complete the operational characterization of the identified set, in the next theorem we write the support function in terms of the distributions $G^o_{Y,Z}$ and $G^o_{X,Z}$.

Theorem 2 *(Operational Characterization of the Identified Set). Let Assumptions (A1)-(A2) hold. Define the scalars λ_L and λ_U as in Theorem 1, and*

$$\beta_{1L} := \frac{[\lambda_L - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZY)]}{[\mathbb{E}(X^2) - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZX)]} \qquad \beta_{1U} := \frac{[\lambda_U - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZY)]}{[\mathbb{E}(X^2) - \mathbb{E}(XZ')\Sigma\mathbb{E}(ZX)]}$$

where Σ is the inverse of the variance matrix of Z. Let q denote a vector belonging to the unit sphere in $\mathbb{R}^{1+d_Z}$. Split the vector q into $q = (q_1, q_2')$, where q_1 is a scalar and q_2 is a vector with the remaining d_Z components. Let $\mathbf{1}()$ denote the indicator function. Then, the identified set of the coefficients β_o is characterized by

$$B_S = \left\{ \beta \in \mathbb{R}^{d_Z+1} : q'\beta \leq s(q) \ ; \textit{for all } q \in \mathbb{S}^{d_Z} \right\}$$

where the support function $s(q)$ equals:

$$\begin{aligned} s(q) &= \mathbf{1}(q_1 \neq 0) \times [\mathbf{1}(q_1 < 0)\beta_{1L} + \mathbf{1}(q_1 > 0)\beta_{1U}] + \mathbb{E}(q_2'\Sigma ZY) \\ &- \mathbb{E}(q_2'\Sigma ZX)\big[\beta_{1L}\,\mathbf{1}(\mathbb{E}(q_2'\Sigma ZX) \geq 0) + \beta_{1U}\,\mathbf{1}(\mathbb{E}(q_2'\Sigma ZX) < 0)\big] \end{aligned}$$

Some comments to Theorem 2 follow:

Remark 2.1 The expression for the support function in the Theorem characterizes the boundary of the identified set. As already said, this expression can be employed to calculate the lower-dimensional projections of the identified set. The projections of the identified set can also be calculated by first evaluating the identifying mapping $\lambda \mapsto m(\lambda)$ at λ_L and λ_U, and then arranging the resulting values into segments. The endpoints of each segment will depend on the sign of the elements of the vector $\mathbb{E}(XZ)$. The support function does the latter arrangement for us. In more general settings, Beresteanu and Molinari (2008), Bontemps, Magnac and Maurin (2012) and Kaido and Santos (2012) also employ the concept of support function to characterize convex identified sets.

Remark 2.2 The characterization of the identified set of β_o in Theorem 2 is sharp, that is, it contains

the values of the coefficients of interest compatible with Assumptions (A1)-(A2) and no others. This means that all the elements in B_S are observationally equivalent to β_o. No amount of data generated according to Assumption (A2) can distinguish the elements in B_S from β_o.

Remark 2.3 According to Theorem 2, the identified set B_S has more than one element because the bounds β_{1L} and β_{1U} on β_{1o} do not coincide. Therefore, the vector of coefficients β_o is set identified when data are available from independent samples on (Y, Z) and (X, Z). This result contrast with the existing literature on two-sample combination with instrumental variables or statistical matching techniques (see the survey by Ridder and Moffit, 2007), where additional assumptions on the joint distribution of (Y, X, Z) deliver point identification of the coefficients of interest. In the next subsection, we discuss these additional assumptions.

Remark 2.4 The characterization in Theorem 2 is operational in the sense that it can be employed, by the way of the analog principle, to construct a sample analog estimator of the identified set. This estimator is obtained after replacing in the sample analog of the support function the unknown functions $G^o_{Y|Z}$ and $Q^o_{X|Z}$ by nonparametric estimates. We discuss the asymptotic properties of such an estimator in the next section.

3.1 Obtaining Point Identification

We have emphasized the fact that Assumptions (A1)-(A2) deliver set identification of the vector of coefficients β_o. We now discuss the force of additional assumptions on the joint distribution of the variables (Y, X, Z) to achieve point identification.

If the covariates Z and X are uncorrelated, i.e. $\mathbb{E}(ZX) = 0$, the vector of coefficients β_{2o} are point identified and β_{1o} is not. This result corresponds to the equality between the "short regression" and the "long regression" as discussed by Goldberger (1991) or the absence of omitted variables bias. This suggests that the two samples may be informative about β_{2o} when the correlation between the covariates is small.

If at least one of the elements in the vector β_{2o} is zero, then the vector of coefficients β_o is point identified. This is equivalent to assume that one of the common variables Z is an instrument (c.f., Angrist and Krueger, 1992). To see why, fix Z to be a scalar. This is without loss of generality. Then, when $\beta_{2o} = 0$ it follows from the identifying mapping $m(\lambda)$ that the coefficient β_{1o} is equal

to $\beta_{1o} = \mathbb{E}(YZ)/\mathbb{E}(XZ)$. In such a case, the coefficient β_{1o} is point identified because $\mathbb{E}(ZY)$ and $\mathbb{E}(XZ)$ so are (and the denominator $\mathbb{E}(XZ)$ is different from zero because the rank condition A1.iii).

If the uncorrelatedness condition implied by (A1.i) is replaced by the mean-independence condition $\mathbb{E}[(Y - X\beta_{1o} - Z'\beta_{2o})|X, Z'] = 0$, then the vector of coefficients β_o is point identified. To see why, notice that mean-independence implies that any measurable function of Z, such as Z_k^2, is uncorrelated with ε. In such a case, any of these functions can be used as an instrument to point identify the coefficients of interest. This situation corresponds to the linear regression model studied by Ichimura and Martinez-Sanchis (2010). After restricting Z to be discrete, it also fits in the setting studied by Cross and Manski (2002) and Molinari and Peski (2006).

Finally, if Y is independent of X conditional on Z, then the vector of coefficients β_o is also point identified. To see why, note that under this conditional independence assumption the expectation $\mathbb{E}(YX)$ is equal to $\mathbb{E}[\mathbb{E}(Y|Z)\mathbb{E}(X|Z)]$. Point identification follows after evaluating the identifying mapping $\lambda \mapsto m(\lambda)$ at $\mathbb{E}[\mathbb{E}(Y|Z)\mathbb{E}(X|Z)]$. This conditional independence restriction is behind the validity of the matching procedures reviewed by Ridder and Moffit (2007). When Z is a discrete variable, this assumption also justifies the procedure of interpreting the least square linear predictor of the conditional expectation of $\mathbb{E}(Y|Z)$ given $\mathbb{E}(X|Z)$ as the coefficient β_{1o}.

4. Estimation and Inference

The need to reflect sampling variability makes it desirable to extend the previous identification results to develop estimation and inference procedures for the coefficients of interest. In the previous Section, we have assumed that the distributions of the two samples were known. In this Section, we drop this assumption. We estimate these distributions from data and employ the characterization derived in Theorem 2 to construct a consistent sample analog estimator for the identified set. We also approximate the sampling distribution of such an estimator, whereby confidence intervals for the coefficients of interest follow.

We begin by describing the estimator of the identified set. Recall that the identified set is a line with support function described in Theorem 2. We estimate this support function in three steps. In the first step, we estimate the functions $y, z \mapsto G^o_{Y|Z}(y, z)$ and $x, z \mapsto Q^o_{X|Z}(x|z)$ by a nonparametric method. In a second step we estimate the bounds λ_L and λ_U on the expectation of

the product of Y and X by:

$$\hat{\lambda}_L := n_1^{-1} \sum_{i=1}^{n_1} Y_i \hat{Q}_{X|Z}(1 - \hat{G}_{Y|Z}(Y_i|Z_i) \quad ; \quad \hat{\lambda}_U := n_1^{-1} \sum_{i=1}^{n_1} Y_i \hat{Q}_{X|Z}(\hat{G}_{Y|Z}(Y_i|Z_i),$$

where $\hat{G}_{Y|Z}$ and $\hat{Q}_{X|Z}$ are the first-step nonparametric estimators. We then, estimate the bounds β_{1L} and β_{1U} on the coefficient β_{1o} by their sample analogs, once λ has been replaced by $\hat{\lambda}_L$ and $\hat{\lambda}_U$, respectively. In the last step, we estimate the support function by its sample analog, once the bounds β_{1L} and β_{1U} have been replaced by its estimators:

$$\begin{aligned} \hat{s}(q) \quad &:= \quad \mathbf{1}(q_1 \neq 0) \left[\hat{\beta}_{1L} \times \mathbf{1}(q_1 \leq 0) + \hat{\beta}_{1U} \times \mathbf{1}(q_1 > 0) \right] + q_2' \hat{\Sigma} n_1^{-1} \sum_{i=1}^{n_1} Z_i Y_i \\ &- \quad q_2' \hat{\Sigma} n_2^{-1} \sum_{i=n_1+1}^{n} Z_i X_i \big[\beta_{1U} \mathbf{1}(q_2' \hat{\Sigma} n_2^{-1} \sum_{i=n_1+1}^{n} Z_i X_i < 0) + \beta_{1L} \mathbf{1}(q_2' \hat{\Sigma} n_2^{-1} \sum_{i=n_1+1}^{n} Z_i X_i \geq 0) \big] \end{aligned}$$

where $\hat{\Sigma}$ is an estimate of the inverse variance matrix of Z.

To establish uniform consistency of the estimator $\hat{s}(q)$, we employ standard results on empirical processes (see van der Vaart, 1998, Chapter 19).

Proposition 1 *(Uniform Consistency) Let Assumptions A1-A2 hold. Let $\mathcal{G}$ denote the space of functions from the support of (Y, Z) into the unit interval $[0, 1]$ which are cadlag in the first argument. Let $\mathcal{Q}$ denote the space of functions from the Cartesian product of $[0, 1]$ and the support of Z into the support of X which are non decreasing in the first argument. Define the functions:*

$$\begin{aligned} f_{L,G,Q}(Y, Z) \quad &:= \quad Y Q_{X|Z}(1 - G_{Y|Z}(Y|Z)|Z) \\ f_{U,G,Q}(Y, Z) \quad &:= \quad Y Q_{X|Z}(G_{Y|Z}(Y|Z)|Z) \end{aligned}$$

for any (G, Q) in the product space $\mathcal{G} \times \mathcal{Q}$. Assume further that:

A3.i) There exist positive constants M and η such that $|||\Sigma|||^{1+\eta} \leq M$, $\mathbb{E}(||ZY||^{1+\eta}) \leq M$ and $\mathbb{E}(||ZX||^{1+\eta})$, where $|||\Sigma|||$ denotes the trace of $\Sigma := \mathbb{E}(ZZ')^{-1}$.

A3.ii) The classes $\mathcal{F}_L := \{f_{L,G,Q}, (G, Q) \in \mathcal{G} \times \mathcal{Q}\}$ and $\mathcal{F}_U := \{f_{U,G,Q}, (G, Q) \in \mathcal{G} \times \mathcal{Q}\}$ are Glivenko-Cantelli.

A3.iii) The functions $(G^o_{Y|Z}, Q^o_{X|Z})$ and their estimates $(\hat{G}_{Y|Z}, \hat{Q}_{X|Z})$ belong to the space $\mathcal{G} \times \mathcal{Q}$.

Then, the estimator of the support function $\hat{s}(q)$ converges in probability to the support function $s(q)$ uniformly over q in the unit sphere in $\mathbb{R}^{d_Z+1}$.

In the proof, we use restriction (A3.i) to apply the Law of Large Numbers to different averages of interest. Restrictions (A3.ii) and (A3.iii) are employed to make the Law of Large Numbers to hold uniformly over different sets.

We use similar tools to derive sufficient conditions securing the weak convergence of the scaled difference $S_n(q) := n^{1/2}\big[\hat{s}(q) - s(q)\big]$.

Proposition 2 *(Uniform Asymptotic Normality) Let Assumptions A1-A3 hold. Assume further that:*

A4.i) There exist positive constants M and η such that $|||\Sigma|||^{2+\eta} \leq M$, $\mathbb{E}(||ZY||^{2+\eta}) \leq M$ and $\mathbb{E}(||ZX||^{2+\eta})$.

A4.ii) The classes $\mathcal{F}_L := \{f_{L,G,Q}, (G,Q) \in \mathcal{G} \times \mathcal{Q}\}$ and $\mathcal{F}_U := \{f_{U,G,Q}, (G,Q) \in \mathcal{G} \times \mathcal{Q}\}$ are Donsker.

A4.iii) The support of the covariates (X,Z) is a convex compact set. The estimator $|||\hat{\Sigma}|||$ for the trace of the matrix Σ is unbiased, i.e., $\mathbb{E}(|||\hat{\Sigma}|||) = |||\Sigma|||$.

Then, the stochastic process $S_n(q) := n^{1/2}\big[\hat{s}(q) - s(q)\big]$ weakly converges to a Gaussian process over q in the unit sphere in $\mathbb{R}^{d_Z+1}$.

Some comments follows. To verify the high level assumptions (A4.ii) and (A3.ii), we can combine existing results on empirical processes (see van der Vaart, 1998) and on sieve estimation (see Chen, 2007). From a result in van der Vaart (1998, Example 19.9), we know that the class $\mathcal{F}_L$ is Donsker (and Glivenko-Cantelli) if the function $y, z \mapsto f_{L,G,Q}(y,z)$ has bounded continuous second partial derivatives for any (G,Q) in the product space $\mathcal{G} \times \mathcal{Q}$. Under the compact support restriction (A4.iii), this latter happens whenever the functions $y, z \mapsto G(y,z)$ and $\tau, z \mapsto Q(\tau,z)$ have bounded continuous second partial derivatives. A similar reasoning follows for the class $\mathcal{F}_L$. Then, in order to meet restriction (A3.iii), we need to employ estimators $y, z \mapsto \hat{G}_{Y|Z}(y,z)$ and $\tau, z \mapsto \hat{Q}_{X|Z}(\tau,z)$ with continuous second order partial derivatives. This in turn can be achieved by employing a cubic spline estimator (see Chen, 2007). There is a kind of tension between assumptions (A4.ii) and (A3.iii): imposing restrictions on the space $\mathcal{G} \times \mathcal{Q}$, such as differentiability of its elements, facilitates the verification of the Donsker condition (A4.ii) but at the same time it requires to

verify extra conditions about the estimators $y, z \mapsto \hat{G}_{Y|Z}(y,z)$ and $\tau, z \mapsto \hat{Q}_{X|Z}(\tau, z)$ in order to meet restriction (A3.iii). We now turn to the discussion of Assumption (A4.iii). This assumption restricts the variables X or Z to have both continuous bounded supports. When X or Z has mass points (i.e., when A4.iii is violated), the normal asymptotic approximation in Proposition 2 is not longer valid. This is because the scaled difference $S_n(q)$ is not longer continuous in the estimator of the inverse of variance of Z. A similar issue arises in more general settings employing the sample analog of a support function to estimate a convex identified set (c.f., Bontemps, Magnac and Maurin, 2012). Approximating the sampling distribution of the proposed estimator for this latter case is left for future research.

To use Proposition 2 to construct confidence regions requires consistent estimation of the asymptotic variance of the limiting process $q \mapsto S_n(q)$. A nonparametric bootstrap procedure can be shown to give asymptotically valid approximation to this variance. Then, confidence regions on the true value of the variance of interest can be constructed by employing the general results in Bontemps, Magnac and Maurin (2012).

5. Monte Carlo Experiments

In this section, we employ simulated data to illustrate the implementation and performance of the estimator of the support function described in the previous Section.

The data generation process is as follows. For computational simplicity, we let Z_i to be a random variable. For the true coefficients $(\beta_{1o}, \beta_{2o}) = (.1, .5)$, we then generate:

$$Y_i = X_i\beta_{1o} + Z_i\beta_{2o} + \varepsilon_i \quad i = 1, .., n$$

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 samples, we split the n draws of the vector (Y, X, Z) into two samples of size n_1 and n_2, respectively. In the first sample, we drop the realized values of X, while in the second sample we drop the realized values of Y. Notice that the distribution $F^o_{X,Z,\varepsilon}$ in the true structure $(\beta_o, F^o_{X,Z,\varepsilon})$ is completely determined by the covariance matrix of (X, Z, ε). For simplicity sake, we fix the variances of X and Z to one. The design variable in this experiment is the correlation between X and Z, say ρ_{XZ}. We choose values for ρ_{XZ} in the

set $\{-.25, 0, .25\}$ to evaluate the sensitivity of the identified set B_S to changes in the correlation between X and Z. Note that the coefficient γ is point identified when the covariates X and Z are uncorrelated, i.e., when $\rho_{XZ} = 0$. Table 1 reports the values of the smallest and largest values of β_1 and β_2 compatible with the simulated data (i.e., the one-dimensional projections of the identified set) for different values of the correlation between the covariates X and Z.

TABLE 1. *One-dimensional Projections of the Identified Set*

Bound	Correlation Between X and Z (ρ_{XZ})										
	-.9	-.75	-.5	-.25	-.01	0	.01	.25	.5	.75	.9
β_{1L}	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1	-.1
β_{1U}	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1
β_{2L}	.32	.35	4	.45	.49	.5	.5	.5	.5	.5	.5
β_{2U}	.5	.5	.5	.5	.5	.5	.51	.55	.60	.65	.68

The smallest and largest value of β_{1o} do not change with the correlation between X and Z. As expected, the difference between β_{2U} and β_{2L} decreases as the correlation ρ_{XZ} between the covariates approaches to zero. The identified set is informative about the sign of the coefficient β_{2o} whatever the correlation between the covariates is. The true value of the coefficients of interest is an extreme point of the identified set.

We estimate the smallest and largest value of each of the coefficients (β_{1o}, β_{2o}) compatible with the available data. These values are equal to:

$$\begin{aligned} \beta_{1L} - s((-1,0)) &= 0 \quad ; \quad \beta_{1U} - s((1,0)) = 0 \\ \beta_{2L} - s((0,-1)) &= 0 \quad ; \quad \beta_{2U} - s((0,1)) = 0 \end{aligned}$$

The interval $[\beta_{1L}, \beta_{1U}]$ is the projection of the identified set onto the β_1-axis. A similar interpretation follows for the interval $[\beta_{2L}, \beta_{2U}]$. Recall that the sample analog estimator of the points $(\beta_{1L}, \beta_{1U}, \beta_{2L}, \beta_{2U})$ is obtained after replacing the unknown functions $G^o_{Y|Z}$ and $Q^o_{X|Z}$ by nonparametric estimates. 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 estimate the quantile function $Q^o_{X|Z}$). We choose a sample size of $n_1 = n_2 = 250$.[3] The number of replications in each experiment is equal to 100.

Table 2 below reports the mean squared error (MSE) of the estimated smallest and largest values of the coefficients (β_{1o}, β_{2o}), together with their Monte Carlo average (labeled Mean), for different values of 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 is minimized for a choice of the number of knots between 90 and 110. There, the variance term is the main component of the mean square error. When the number of knots is too small (i.e., 10) or too big (i.e., 130), the estimator exhibits a significant mean square error. When the number of knots is small, the bias is negative for the estimated values of the lower bounds and positive for the estimated values of the upper bounds, and the bias renders the estimator more likely to be outside the identified set (see the column labeled Cove.). In such cases, we can expect confidence intervals with coverage probabilities above the pre-specified nominal value. By contrast, when the number of knots is too big, the bias renders the estimated values more likely to be inside the identified set, so we can expect confidence intervals with coverage probabilities below the nominal value. Since the choice of the number of knots has an important effect on the performance of the estimator, it would be useful to have a way to choose them in practice. Obtaining such a result is however beyond the scope of this paper, but remain an important topic for future research.

[3]Result for larger sample sizes are available upon request.

TABLE 2. *Monte Carlo Experiments: Sensitivity to First-Step Nonparametric Estimator*

Sample Sizes = 250

		$\rho_{XZ} = -.25$				$\rho_{XZ} = 0$				$\rho_{XZ} = .25$			
Knots	Bound	Mean	MSE	Bias	Cove.	Mean	MSE	Bias	Cove.	Mean	MSE	Bias	Cove.
10	β_{1L}	-.821	.525	98%	100%	-.769	.452	99%	100%	-.717	.385	99%	100%
	β_{1U}	.724	.395	98%		.772	.458	98%		.815	.517	98%	
	β_{2L}	.254	.047	81%	96%	.451	.008	30%	100%	.323	.038	80%	99%
	β_{2U}	.657	.031	81%		.536	.006	18%		.689	.026	73%	
30	β_{1L}	-.612	.268	97%	100%	-.572	.226	98%	100%	-.510	.173	96%	100%
	β_{1U}	.527	.187	97%		.559	.215	97%		.595	.250	97%	
	β_{2L}	.308	.027	73%	92%	.461	.006	21%	34%	.376	.021	71%	88%
	β_{2U}	.605	.016	69%		.524	.005	10%		.641	.014	58%	
50	β_{1L}	-.451	.131	94%	100%	-.414	.104	95%	100%	-.373	.081	91%	100%
	β_{1U}	.375	.081	94%		.411	.103	93%		.458	.137	93%	
	β_{2L}	.351	.016	61%	92%	.471	.005	14%	27%	.408	.013	60%	84%
	β_{2U}	.566	.009	48%		.516	.005	5%		.607	.008	36%	
70	β_{1L}	-.336	.064	86%	99%	-.301	.051	77%	96%	-.281	.069	65%	100%
	β_{1U}	.276	.041	74%		.298	.051	76%		.342	.084	69%	
	β_{2L}	.381	.011	44%	68%	.476	.005	10%	22%	.436	.009	42%	72%
	β_{2U}	.541	.006	23%		.509	.004	1%		.584	.007	15%	
90	β_{1L}	-.248	.031	70%	84%	-.193	.037	23%	92%	-.151	.061	4%	93%
	β_{1U}	.192	.018	45%		.226	.031	52%		.257	.049	50%	
	β_{2L}	.404	.007	25%	55%	.481	.005	6%	18%	.455	.007	26%	54%
	β_{2U}	.519	.004	7%		.506	.004	1%		.555	.006	1%	
110	β_{1L}	-.188	.021	36%	52%	-.153	.044	6%	59%	-.024	.361	1%	67%
	β_{1U}	.097	.017	1%		.144	.025	7%		.224	.414	3%	
	β_{2L}	.421	.006	14%	36%	.485	.004	4%	13%	.462	.027	4%	37%
	β_{2U}	.495	.004	1%		.501	.004	1%		.524	.023	2%	
130	β_{1L}	-.105	.084	1%	47%	-.131	.195	1%	35%	-.014	.086	8%	45%
	β_{1U}	.066	.031	3%		.073	.032	2%		.147	.076	2%	
	β_{2L}	.444	.009	3%	55%	.488	.004	3%	7%	.484	.007	3%	29%
	β_{2U}	.488	.007	1%		.497	.004	1%		.522	.008	9%	

This table presents different measures describing the finite sample performance of the estimator of the support function evaluated at the canonical directions. The label "$\rho_{X,Z}$" indicates value of the correlation between the covariates in the exercises. "Knots" stands for the number of knots employed in the nonparametric estimation of the conditional quantile function of X given Z and the conditional distribution function of Y given Z. "Mean" is the Monte Carlo average of the estimates. "MSE" stands for the mean square error and "Bias" the percentage of the mean square error corresponding to the bias. "Cove." is the percentage of times that the true coefficient of interest, say β_{jo}, lies in the estimated interval $[\hat{\beta}_{jL}, \hat{\beta}_{jU}]$ for $j \in \{1, 2\}$. The number of Monte Carlo replications is 100.

6. Further Applications

In this section, we discuss two other applications of the identification results in Section 3. These applications are the measurement of the variance of the treatment effect (e.g., Heckman, Smith and Clements, 1997), and the measurement of the correlation coefficient from aggregate data (Robinson, 1950).

6.1 Variance of the Treatment Effect

To discuss how our identification results apply to the measurement of the variance of the treatment effect, we need to introduce some notation. 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. The variance of the treatment effect is defined by:

$$\sigma^2(\lambda_o) := \mathbb{V}(Y) + \mathbb{V}(X) - 2\lambda_o + 2\mathbb{E}(Y)\mathbb{E}(X)$$

where $\lambda_o := \mathbb{E}(YX)$ denotes the expectation of the product of Y and X, and $\mathbb{V}$ denotes the variance operator.

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

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

where λ_U and λ_L are defined in Theorem 1. The identified set Σ_I is a segment of the real line. The end points of this set correspond to the value of the function $\lambda \mapsto \sigma^2(\lambda)$ evaluated at the quantities λ_U and λ_L, respectively. Unlike the case for the coefficients in a linear projection, calculating the support function does not help in the construction of an estimator of the identified because the identified set is just an interval.

The latter characterization of the identified set of the variance of the treatment effect is new. Concurrent work by Fan and Zhu (2010) studies identification on *superadditive integral functionals* of the distribution of (Y, X). Since $Y, X \mapsto \sigma_o^2(F_{Y,X})$ 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 $\sigma_o^2(\lambda_U)$, $\sigma_o^2(\lambda_L)$ in terms of moments in a different way than they do. We shall show that this allows us to dispense with numerical integration procedures, a step required by Fan and Zhu (2010). A plug-in estimator of the bounds $\sigma_o^2(\lambda_U) \leq \sigma^2(\lambda_o) \leq \sigma_o^2(\lambda_L)$ can be obtained after replacing in their sample analogs the unknown functions $Q_{X|Z}^o$ and $G_{Y|Z}^o$

by nonparametric estimates. To facilitate the description of this estimator, and without loss of generality, suppose that Y and X have both zero (known) mean and unit (known) variance. Then, the plug-in estimator of the lower bound $\sigma_o^2(\lambda_U)$ is:

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

where $\hat{Q}_{X|Z}$ and $\hat{G}_{Y|Z}$ are nonparametric estimator of $Q^o_{X|Z}$ and $G^o_{Y|Z}$, respectively. A similar expression follows for the estimator of the upper bound. We now compare the estimator $\hat{\sigma}_L$ with the one proposed by Fan and Zhu (2010). Their estimator for the lower bound $\sigma_o^2(\lambda_U)$ is:

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

where $I_{ib} := \mathbf{1}(|Z_i| \leq b)$ is a trimming sequence with $\mathbf{1}(\cdot)$ the indicator function, and $\hat{Q}_{Y|Z}$ is a nonparametric estimator of the quantile functions $Q^o_{Y|Z}$. Implementing the estimator $\hat{\sigma}_L^{FZ}$ requires to employ a numerical integration procedure to compute the integral in (1). By contrast, no numerical integration procedure is required to implement $\hat{\sigma}_L$. To establish the connection between these two estimators, it suffices to perform the change-of-variable $u = G^o_{Y|Z}(Y|Z)$ in (1) and replace $G^o_{Y|Z}$ by its nonparametric estimate. We can say then that our estimator $\hat{\sigma}_L$ thus replace the numerical integration procedure by a sum.

6.2 Measuring the Correlation Coefficient from Aggregate Data

Measuring the correlation between two random variables from aggregate data is another application where our identification results apply. To be more precise, suppose that knowledge is sought about the correlation coefficient between two discrete random variables, say Y and X:

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

but data provide only *estimates* of the distribution of (Y, Z) and of (X, Z). Difficulties arise because joint realizations of Y and X are not observed. This is the so-called ecological correlation problem (c.f., Robinson, 1950). A leading example of this problem 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 $\rho(\lambda_o)$ 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 $G^o_{X|Z}$.

Under hypothetical knowledge of the distributions $G^o_{Y|Z}$ and $G^o_{X|Z}$, all the expectations in $\rho(\lambda_o)$ are known except for λ_o. The ecological correlation problem consists in determining what we can learn about $\rho(\lambda_o)$. One approach to solve the ecological correlation problem 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 $\lambda \mapsto \rho(\lambda)$ is linear and increasing, Theorem 1 can be employed to extend the insight by Robinson (1950) to provide a sharp characterization of the identified set of $\rho(\lambda_o)$.

7. Summary and Conclusions

Applied researchers interested in making inference about least square linear predictors are often confronted to the case where the relevant variables are measured in two independent samples, neither of which contains information on all the variables of interest. The existing literature suggests to overcome the difficulties associated to such lack of data by imposing additional assumptions, which may be difficult to justify in a given application. This paper shows that the least square linear predictor is set identified when data are available from two samples and no additional assumptions on the data generating process are invoked. We characterize the identified set of the least square linear predictor and show that this set can be estimated by the sample analog principle. We evaluate

then the force of additional assumptions to remove set identification. These results highlight the trade-off between imposing restrictions and achieving point identification when data are incomplete. As it stressed by the literature on set identification (c.f., Manski, 2003), analyzing this trade-off is worthy of consideration in applications where the assumptions leading to point identification are under suspicion.

There are at least two topics which deserve further research. The first topic relates to the generalization of our identification results to the case where the covariate observed in only one sample is a vector rather than an scalar. The second topic relates to the choice of the smoothing parameters for the estimator of the identified set.

Acknowledgments. *I wish to thank my advisor Thierry Magnac for invaluable help during this project. Stephan Bonhomme, Christian Bontemps, Andrew Chesher, Fabiana Gomez, Gregory Jolivet, Pascal Lavergne, Nour Meddahi, Jorge Ponce, Christoph Rothe, Senay Sokullu, Frank Windmeijer and seminar participants at University of Bristol, the European Winter Meeting of the Econometric Society '11, the EC^2 '10 meeting, Carlos III/Madrid, the NESG '10 meeting, the ENTER Jamboree '10 meeting, and Toulouse School of Economics have offered useful comments and suggestions for which I am very grateful. All remaining errors are my responsibility.*

References

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. 328-36.

ARELLANO, Manuel and Costas MEGHIR (1992): "Female Labor Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Data Sets", *Review of Economic Studies*, 59(3) pp. 537-59.

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", *Econometrica*, forthcoming.

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.

BOVER, Olympia (2005): "Wealth Effects on Consumption: Microeconomic Estimates from the Spanish Survey of Household Finances", *Documentos the Trabajo No 0522*, Banco de Espana.

BRZOZOWSKI, Matthew, GERVAIS, Martin, KLEIN, Paul and SUZUKI, Michio (2010): "Consumption, Income, and Wealth Inequality in Canada", *Review of Economic Dynamics*, 13 (1), pp. 52-75.

CROSS, Phillip and Charles MANSKI (2002): "Regressions, Short and Long", *Econometrica*, 70(2), pp. 357-68.

CHEN, Xiaohong (2007): "Large Sample Sieve Estimation of Semi-nonparametric Models", in J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Volume 6B, Elsevier.

DEVEREUX, Paul and Gautam TRIPATHI (2009):"Optimally Combining Censored and Uncensored Datasets", *Journal of Econometrics*, 151(1) pp. 17-32.
FANG, Hanming, Michael KEANE and Dan SILVERMAN (2008): "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market", *Journal of Political Economy*, 116(2) pp. 303-50.
FAN, Yanqin and Dongming ZHU (2010): "Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes", *unpublished manuscript.*
FLAVIN, Marjorie and Shinobu NAKAGAWA (2008): "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence", *American Economic Review*, 98(1) pp. 474-95.
GOLDBERG, Arthur (1991): *A Course in Econometrics*, Harvard University Press.
HECKMAN, James, Jeffrey 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.
HIRIART-URRUTUY, Jean-Baptiste and Claude LEMARECHAL (2004): *Fundamentals of Convex Analysis*, Springer.
ICHIMURA, Hidehiko and Elena MARTINEZ-SANCHIS (2010): "Identification and Estimation of GMM Models by Combining Two Data Sets", *unpublished manuscript.*
KAIDO, Hiroaki and Andres SANTOS (2011): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities", *unpublished manuscript.*
KOMAROVA, Tatiana, Denis NEKIPELOV and Evgeny YAKOVLEV (2012): "Identification, Data Combination and the Risk of Disclosure", unpublished manuscript.
MANSKI, Charles (2003): *Partial Identification of Probability Distributions*, Springer.
MEGHIR, Costas and Marten PALME (1999): "Assesing the Effect of Schooling on Earnings Using a Social Experiment", *The Institue for Fiscal Studies*, Working Paper No W99/10.
MOLINARI, Francesca and Marcin PESKI (2006): "Generalization of a Result on Regressions, Short and Long", *Econometric Theory*, 22(1) pp. 159-63.
RASSLER, Susanne (2002): *Statistical Matching*, Springer.
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 Sociolo- gical Review,* 15(3), pp. 351-357.
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.
van der VAART, A. (1998): *Asymptotic Statistics*, Cambridge University Press.
VITALE, Richard (1979): "Regression with Given Marginals", *The Annals of Statistics*, 7(3), pp. 653-658.
WOOLDRIDGE, Jeffrey (2002): *Econometric Analysis of Cross Sections and Panel Data*, MIT Press.

Appendix: Proofs

Proof of Theorem 1 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. By construction, λ_o belongs to the range Λ_I of the mapping $F_{Y,X,Z} \mapsto \int yx dF_{Y,X,Z}(y,x,z)$ from $\mathcal{F}_{Y,X,Z}$ into the real line. We shall show that $\Lambda_I = [\lambda_L, \lambda_U]$.

To proceed, notice that the class $\mathcal{F}_{Y,X,Z}$ is non-empty, convex. To see why $\mathcal{F}_{Y,X,Z}$ is non-empty, it suffices to note that 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 verify 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:

$$\int \int_{-\infty}^{x} \int_{-\infty}^{z} f(y,a,b)\, dydadb \quad = \quad G^{o}_{X,Z}(x,z) \quad ; \quad \int \int_{-\infty}^{x} \int_{-\infty}^{z} \tilde{f}(y,a,b)\, dydadb = G^{o}_{X,Z}(x,z)$$

Let ψ be a number between zero and one. Multiply both sides of the first equality by ψ. Multiply both sides of the second equality by $(1-\psi)$. Summing the resulting expressions yields:

$$\psi \int_{-\infty}^{y} \int \int_{-\infty}^{z} f(y,a,b)\, dydadb + (1-\psi) \int_{-\infty}^{y} \int \int_{-\infty}^{z} \tilde{f}(y,a,b)\, dydadb = G^{o}_{X,Z}(x,z)$$

Then, the convex combination of $F, \tilde{F}$ has *(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 $\psi F(y,x,z) + (1-\psi)\tilde{F}(y,x,z)$ belongs to the class $\mathcal{F}_{Y,X,Z}$. Therefore, $\mathcal{F}_{Y,X,Z}$ is convex. Since the mapping $F_{Y,X,Z} \mapsto \int yx dF_{Y,X,Z}(y,x,z)$ is linear, and convexity is preserved under linear transformations (see Rockafellar, 1970), we have that Λ_I is an interval, say $\Lambda_I = [\lambda_L, \lambda_U]$. We now characterize the endpoints λ_L and λ_U. We can define

$$\begin{aligned} &\lambda_L := \min_{F_{Y,X,Z}} \textstyle\int yx \; dF_{Y,X,Z}(y,x,z) \\ &\text{s.t. } G^{o}_{Y,Z}(y,z) = \lim_{x\to\infty} F_{Y,X,Z}(y,x,z) \;\; \forall y \in \mathcal{Y}, z \in \mathcal{Z} \\ &\quad G^{o}_{X,Z}(x,z) = \lim_{y\to\infty} F_{Y,X,Z}(y,x,z) \;\; \forall x \in \mathcal{X}, z \in \mathcal{Z} \end{aligned}$$

for the endpoint λ_L, and the corresponding maximization problem for the endpoint λ_U. These programming problems have linear objective functions with linear constraints. Since the function $y, x \mapsto yx$ in the objective function is a strictly superadditive function, it follows from a result in Ruschendorf (1991, Proposition 7) that the function $y, x, z \mapsto G^{L}_{Y,X,Z}(y,x,z)$ with

$$G^{L}_{Y,X,Z}(y,x,z) = \int_{-\infty}^{z} \max\{0, G^{o}_{Y|Z}(y|s) + G^{o}_{X|Z}(x|s) - 1\} dG^{o}_{Z}(s)$$

is the unique argument of the minimum in the minimization problem we have described above. In turn, the function $y, x, z \mapsto G^{U}_{Y,X,Z}(y,x,z)$ with

$$G^{U}_{Y,X,Z}(y,x,z) = \int_{-\infty}^{z} \min\{G^{o}_{Y|Z}(y|s), G^{o}_{X|Z}(x|s)\} dG^{o}_{Z}(s)$$

is the unique argument of the maximum in 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 claim in the Theorem follows after evaluating the mapping $F_{Y,X,Z} \mapsto \int yx \; dF_{Y,X,Z}(y,x,z)$ at the Hoeffding-Frechet distributions. In particular, replace the lower Hoeffding-Frechet bound $G^{L}_{Y,X,Z}$ in the objective function of the programming problem defining the extreme point λ_L:

$$\lambda_L = \int_{\mathcal{Z}} \int_{\mathcal{Y}\times\mathcal{X}} y, x dG^{L}_{Y,X,Z}(y,x,z) dG^{o}_{Z}(z)$$

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,

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

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

$$\lambda_L = \int_{\mathcal{Z}} \int_{[0,1]} Q^{o}_{Y|Z}(\tau,z) \times Q^{o}_{X|Z}(1-\tau,z) d\tau dG^{o}_{Z}(z)$$

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

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

where the expectation is with respect to the joint distribution of (Y, X), $G^o_{Y,Z}$. The expression for the upper bound $\lambda_U = \mathbb{E}[Y \cdot Q^o_{X|Z}(G^o_{Y|Z}(Y|Z)|Z)]$ follows from a similar reasoning. ■

Proof of Theorem 2. Let $\beta_1(\lambda)$ and $\beta_2(\lambda)$ denote the components of $m(\lambda)$. We can re-express the support function as

$$s(q) = \sup_{\lambda\in[\lambda_L,\lambda_U]} q_1\beta_1(\lambda) + \sup_{\lambda\in[\lambda_L,\lambda_U]} q_2'\beta_2(\lambda)$$

Since the function $\lambda \mapsto \beta_1(\lambda)$ is linear and increasing, the value function of the linear programming problem $\sup_{\lambda\in[\lambda_L,\lambda_U]} q_1\beta_1(\lambda)$ is:

$$\mathbf{1}(q_1 \neq 0)\left[\beta_{1L} \times \mathbf{1}(q_1 < 0) + \beta_{1U} \times \mathbf{1}(q_1 > 0)\right]$$

Similarly, the value function of the linear programming problem $\sup_{\lambda\in[\lambda_L,\lambda_U]} q_2'\beta_2(\lambda)$ is equal to:

$$\mathbb{E}(q_2'\Sigma ZY) - \mathbb{E}(q_2'\Sigma ZX)\big[\beta_{1U}\mathbf{1}(\mathbb{E}(q_2'\Sigma ZX) < 0) + \beta_{1L}\mathbf{1}(\mathbb{E}(q_2'\Sigma ZX) \geq 0)\big]\Big]$$

where Σ is the inverse of the variance matrix of Z. ■

Notation. To simplify the arguments in the proofs below and avoid clutter, we introduce the following notation. Recall that the support function of the identified set is:

$$\begin{aligned}s(q) &= \mathbf{1}(q_1 \neq 0) \times [\mathbf{1}(q_1 < 0)\beta_{1L} + \mathbf{1}(q_1 > 0)\beta_{1U}] \\ &+ \mathbb{E}(q_2'\Sigma_o ZY) - \mathbb{E}(q_2'\Sigma_o ZX)\big[\beta_{1U}\mathbf{1}(q_2'\Sigma_o\mu_o < 0) + \beta_{1L}\mathbf{1}(q_2'\Sigma_o\mu_o \geq 0)\big]\Big]\end{aligned}$$

where $\Sigma_o := \mathbb{E}(ZZ')^{-1}$ and $\mu_o := \mathbb{E}(ZX)$. The estimator we consider is

$$\begin{aligned}\hat{s}(q) &:= \mathbf{1}(q_1 \neq 0)\left[\hat{\beta}_{1L} \times \mathbf{1}(q_1 \leq 0) + \hat{\beta}_{1U} \times \mathbf{1}(q_1 > 0)\right] \\ &+ n_1^{-1}\sum_{i=1}^{n_1} q_2'\hat{\Sigma}Z_iY_i - n_2^{-1}\sum_{i=n_1+1}^{n} q_2'\hat{\Sigma}Z_iX_i\big[\beta_{1U}\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0) + \beta_{1L}\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} \geq 0)\big]\end{aligned}$$

where $\hat{\Sigma}$, $\hat{\beta}_{1L}$, $\hat{\beta}_{1U}$ and $\hat{\mu}$ are estimates of Σ_o, β_{1L}, β_{1U} and μ_o, respectively. We denote by M a generic majorization constant. Let Σ denote an estimate of the inverse of the variance of Z, and let μ denote an estimate of the expectation of the product of Z and X. Let $||.||$ denote a generic norm and define $|||\Sigma||| := Tr(\Sigma)$. Let define the space:

$$\Theta = \mathbb{S}^{d_Z} \times \{\Sigma \in \mathbb{R}^{d_Z} \times \mathbb{R}^{d_Z} : |||\Sigma||| < M\} \times \{\mu \in R^{d_Z} : ||\mu|| < M\} \times \{\beta_{1L} : |\beta_{1L}| < M\} \times \{\beta_{1U} : |\beta_{1U}| < M\}$$

Consider the functions $f_{1,\theta}$, $f_{2,\theta}$ and $f_{3,\theta}$ indexed by $\theta = (q, \Sigma, \beta_{1L}, \beta_{1U}, \mu) \in \Theta$ such that:

$$\begin{aligned}f_{1,\theta}(Y,Z) &:= q_2'\Sigma ZY \\ f_{2,\theta}(X,Z) &:= q_2'\Sigma ZX\beta_{1U}\mathbf{1}(q_2'\Sigma\mu < 0) \\ f_{3,\theta}(X,Z) &:= q_2'\Sigma ZX\beta_{1L}\mathbf{1}(q_2'\Sigma\mu \geq 0),\end{aligned}$$

With this notation in hand, the difference $\hat{s}(q) - s(q)$ can be expressed as:

$$\begin{aligned}
\hat{s}(q) - s(q) &= \mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 < 0)(\hat{\beta}_{1L} - \beta_{1L}) \\
&+ \mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 > 0)(\hat{\beta}_{1U} - \beta_{1U}) \\
&+ n_1^{-1} \sum_{i=1}^{n_1} f_{1,\hat{\theta}}(Y_i, Z_i) - \mathbb{E}\big(f_{1,\theta_o}(Y,Z)\big) \\
&+ n_2^{-1} \sum_{i=n_1+1}^{n} f_{2,\hat{\theta}}(X_i, Z_i) - \mathbb{E}\big(f_{2,\theta_o}(X,Z)\big) \\
&+ n_2^{-1} \sum_{i=n_1+1}^{n} f_{3,\hat{\theta}}(X_i, Z_i) - \mathbb{E}\big(f_{3,\theta_o}(X,Z)\big)
\end{aligned}$$

Proof of Proposition 1 We want to prove $\sup_q \big|\hat{s}(q) - s(q)\big| = o_P(1)$. The proof is similar to other results establishing consistency of empirical analogs of support functions (c.f., Molinari and Beresteanu, 2008; Bontemps, Magnac and Maurin, 2012, Proof of Proposition 9). The main difference is in the presence of the unknown functions $Q_{X|Z}$ and $G_{Y|Z}$. Our starting point is the difference $\hat{s}(q) - s(q)$. Apply the triangle inequality, add-and-subtract $\mathbb{E}\big(f_{j,\hat{\theta}}\big)$ for $j \in \{1,2,3\}$, apply the triangle inequality again to obtain that $\sup_q \big|\hat{s}(q) - s(q)\big|$ is bounded by:

$$\begin{aligned}
\sup_q \big|\hat{s}(q) - s(q)\big| &\leq \sup_q \left|\mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 < 0)(\hat{\beta}_{1L} - \beta_{1L})\right| && (2) \\
&+ \sup_q \left|\mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 > 0)(\hat{\beta}_{1U} - \beta_{1U})\right| && (3) \\
&+ \sup_q \left|n_1^{-1} \sum_{i=1}^{n_1} f_{1,\hat{\theta}}(Y_i, Z_i) - \mathbb{E}\big(f_{1,\hat{\theta}}(Y,Z)\big)\right| && (4) \\
&+ \sup_q \sum_{j=1}^{2} \left|n_2^{-1} \sum_{i=n_1+1}^{n} f_{j,\hat{\theta}}(X_i, Z_i) - \mathbb{E}\big(f_{j,\hat{\theta}}(X,Z)\big)\right| && (5) \\
&+ \sup_q \left|\mathbb{E}\big(f_{1,\hat{\theta}}(Y,Z)\big) - \mathbb{E}\big(f_{1,\theta_o}(Y,Z)\big)\right| && (6) \\
&+ \sup_q \sum_{j=1}^{2} \left|\mathbb{E}\big(f_{j,\hat{\theta}}(X,Z)\big) - \mathbb{E}\big(f_{j,\theta_o}(X,Z)\big)\right| && (7)
\end{aligned}$$

We shall show that (2) to (7) in the latter display are $o_P(1)$.

To show that (2) is $o_P(1)$, start by noticing that (2) $= |(\hat{\beta}_{1L} - \beta_{1L})|$ because $\mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 < 0)$ is nonnegative. Then, estabilishing uniform convergence in probability of (2) boilds down into establishing pointwise convergence in probability of $(\hat{\beta}_{1L} - \beta_{1L})$. Recall that the estimator $\hat{\beta}_{1L}$ of the lower bound β_{1L} is:

$$\hat{\beta}_{1L} := \left[\hat{\lambda}_L - \Big(n_2^{-1} \sum_{i=n_1+1}^{n} X_i Z_i'\Big)\hat{\Sigma}\Big(n_1^{-1} \sum_{i=1}^{n_1} Z_i Y_i\Big)\right] \times \left[n_2^{-1} \sum_{i=n_1+1}^{n} X_i^2 - \Big(n_2^{-1} \sum_{i=n_1+1}^{n} X_i Z_i'\Big)\hat{\Sigma}\Big(n^{-1} \sum_{i=n_1+1}^{n_2} X_i Z_i\Big)\right]^{-1}$$

We are working with iid samples (see Assumption 2) with finite second order moments (see Assumption A1.ii). Hence, the Law of Large Numbers implies that $n_2^{-1}\sum_{i=n_1+1}^{n} X_i Z_i'$, $\sum_{i=1}^{n_1} Z_i Y_i$, and $n_2^{-1}\sum_{i=n_1+1}^{n} X_i^2$ in the latter display converges in probability to their population counterparts. We show below (see point v) that $\hat{\Sigma}$ is also consistent for Σ_o. To conclude that $\big(\hat{\beta}_U - \beta_U\big)$ is $o_P(1)$, we need to show that $\big(\hat{\lambda}_U - \lambda_U\big)$ is $o_P(1)$. This later requirements follows directly from conditions (A3.ii) and (A3.iii) in the Proposition. Verifying that (3) is $o_P(1)$ is similar and thus ommited.

To show that (4) and (5) are $o_P(1)$, we exploit the following argument. We know from a result by van der Vaart (1998, Theorem 19.4, p. 270) that if: (i) the true values $(\Sigma_o, \mu_o, \beta_{1L}, \beta_{1U})$ and their estimates $(\hat{\Sigma}, \mu, \hat{\beta}_{1L}, \hat{\beta}_{1U})$ belong

to the space Θ; and (ii) the classes $\mathcal{F}_j := \{f_{j,\theta} : \theta \in \Theta\}$ for $j \in \{1,2,3\}$ are Glivenko-Cantelli; then (2) and (3) are $o_P(1)$. To verify condition (i) above, we note that the true values $(\Sigma_o, \mu_o, \beta_{1L}, \beta_{1U})$ belong to Θ by the moment restriction (A.ii) and the rank condition (A1.iii) for a sufficiently large M. We show below (see point v) that we can force the estimates $(\hat{\mu}, \hat{\beta}_L, \hat{\beta}_U, \hat{\Sigma})$ to belong to the space Θ, if necessary, by trimming. To verify the Glivenko-Cantelli condition (ii), note that the classes $\mathcal{F}_j := \{f_{j,\theta} : \theta \in \Theta\}$ for $j \in \{1,2,3\}$ are parametric. The index set Θ is bounded by construction. Then, it follows from a result in van der Vaart (1998, Example 19.7, p. 271) that $\mathcal{F}_j$ for $j \in \{1,2,3\}$ are Glivenko-Cantelli whenever: (iii) there exist measurable functions m_j for $j \in \{1,2,3\}$ such that

$$|f_{j,\theta_1}(y,x,z) - f_{j,\theta_2}(y,x,z)| \leq m_j(y,x,z)||\theta_1 - \theta_2||$$

for every θ_1, θ_2; and (iv) $\mathbb{E}(|m_j(Y,X,Z)|^r) < \infty$ for some $r \geq 1$. We now verify the existence of the measurable function m_1 satisfying (iii) and (iv) above for the class $\mathcal{F}_1$. From the definition of $f_{1,\theta}(Y,Z)$, notice that the function $q_2\Sigma \mapsto f_{1,\theta}(y,z)$ is linear and it is defined on a bounded set. Hence, a result about continuity of convex functions (see e.g., Hiriart-Urruty and Lemarechal, 2004, Theorem 3.1.2, p. 103) implies that there exists $M \geq 0$ such that such that for any θ_1, θ_2:

$$\begin{aligned} |f_{1,\theta_1}(y,z) - f_{1,\theta_2}(y,z)| &\leq \sup\{||zy||, M\}||q_{2,1}'\Sigma_1 - q_{2,2}'\Sigma_2|| \\ &\leq m_1(y,z) \times ||\theta_1 - \theta_2|| \end{aligned} \tag{8}$$

where $m_1(y,z) = \sup\{||zy||, M\}$, and the term $\sup\{||zy||, M\}$ controls for those points in Θ that are not in the relative interior of the domain of $f_{1,\theta}(X,Z)$. Under restriction (A3.i), we have $\mathbb{E}(\sup\{||ZY||, M\}) < \infty$. In condition (iv), set $r = 1$ to conclude that $\mathcal{F}_1$ is Glivenko-Cantelli. We now verify the existence of a measurable function m_2 satisfying (iii) and (iv) above for the class $\mathcal{F}_2$. From the definition of $f_{2,\theta}(X,Z)$, notice that $q_2\Sigma \mapsto f_{2,\theta}(X,Z)$ is also a linear function defined on a bounded set. Hence, for any θ_1, θ_2 there exists $M \geq 0$ such that

$$\begin{aligned} |f_{2,\theta_1}(y,z) - f_{1,\theta_2}(y,z)| &\leq M||q_{2,1}'\Sigma_1 - q_{2,2}'\Sigma_2|| \times ||\mu_1 - \mu_2|| \sup\{||zx\beta_{1U,1}||, ||zx\beta_{1U,2}||\} \\ &\leq m_2(y,z) \times ||\theta_1 - \theta_2|| \end{aligned} \tag{9}$$

where $m_2(y,z) = M \sup\{|\beta_{1U,1}| \times ||zx||, |\beta_{1U,2}| \times ||zx||\}$, and the term $\sup\{|\beta_{1U,1}| \times ||zx||, |\beta_{1U,2}| \times ||zx||\}$ controls for the points in Θ that are not in the relative interior of the domain of $f_{2,\theta}(X,Z)$. Under restrictions (A3.iii), we have that the expectation of $m_2(Y,Z)$ is finite. In condition (iv), set now $r = 1$ to conclude that $\mathcal{F}_2$ is Glivenko-Cantelli. Verifying that $\mathcal{F}_3$ is Glivenko-Cantelli is similar and thus ommited.

To show that (6) and (7) are $o_P(1)$, we combine inequalities (8) and (9) with the Dominated Convergence Theorem (see Pollard, 2002, page 32). We begin with (6). In inequality (8), set $\theta_1 = \hat{\theta}$ and $\theta_2 = \theta_o$. Since $\hat{\theta}$ converges in probability to θ_o (see point v), it follows from inequality (6) that:

$$|f_{1,\hat{\theta}}(y,z) - f_{1,\theta_o}(Y,Z)| \leq m_1(y,z) \times o_P(1) \tag{10}$$

If the function $\theta \mapsto f_{1,\theta}(y,z)$ is uniformly bounded, then (10) and the Dominated Convergence Theorem imply that (6) is $o_P(1)$. We now verify that $\theta \mapsto f_{1,\theta}(y,z)$ is uniformly bounded. By the Cauchy-Schwarz inequality:

$$f_{1,\theta}(y,z) \quad := \quad q_2'\Sigma zy \leq ||q_2'||^{1/2}|||\Sigma|||^{1/2}||zy||^{1/2}$$

Since $||q_2'||^{1/2} < 1$ and $|||\Sigma|||^{1/2} < M$ (see Assumption A3.i), we have:

$$\sup_{q_2} |f_{1,\theta}(Y,Z)| \leq M \times ||zy||^{1/2}$$

Since the vector (Z,Y) has finite variance matrix (see condition A1.ii), we further obtain:

$$\mathbb{E}\left(\sup_{q_2} |f_{1,\theta}(Y,Z)| \right) \leq M$$

Then, we can conclude from (10) and the Dominated Convergence Theorem that (6) is $o_P(1)$. Verifying that (7) is $o_P(1)$ is similar thus omitted.

To conclude the proof, we need to show that: (v) $\hat{\Sigma}$ is consistent for Σ and it belongs to $\{\Sigma : |||\Sigma||| < M\}$. To establish (v), we follow Bontemps, Magnac and Maurin (2012, Proof of Proposition 9). Let $\bar{\Sigma} := (n^{-1}\sum_{i=1}^{n} Z_i Z_i')^{-1}$ denote the sample analog of Σ. We define the bounded sample analog estimator $\hat{\Sigma}$ as:

$$\hat{\Sigma} := \begin{cases} \bar{\Sigma} & \text{if } |||\bar{\Sigma}||| < M \\ (M/|||\bar{\Sigma}|||) \times \bar{\Sigma} & \text{otherwise} \end{cases}$$

The estimator $\hat{\Sigma}$ belongs to $\{\Sigma : |||\Sigma||| < M\}$ by construction. Under (A3.i), the estimator $\hat{\Sigma}$ is consistent for Σ. ■

Proof of Proposition 2 To prove this proposition, we decompose the empirical process $S_n(q) := n^{1/2}\big[\hat{s}(q) - s(q)\big]$ into a sum of several pieces and show that each piece converges weakly to a Gaussian random process. The fact that a sum of Gaussian processes is a Gaussian process will guarantee that the combination of the pieces delivers the desired result.

The empirical process S_n is equal to:

$$\begin{aligned} S_n(q) &= \mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 < 0) n^{1/2}\big(\hat{\beta}_L - \beta_L\big) && (11) \\ &+ \mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 > 0) n^{1/2}\big(\hat{\beta}_U - \beta_U\big) && (12) \\ &+ n^{1/2}\Big[n_1^{-1}\sum_{i=1}^{n_1} f_{1,\hat{\theta}}(Y_i, Z_i) - \mathbb{E}\big(f_{1,\theta}(Y,Z)\big)\Big] && (13) \\ &+ n^{1/2}\Big[n_2^{-1}\sum_{i=n_1+1}^{n} f_{2,\hat{\theta}}(X_i, Z_i) - \mathbb{E}\big(f_{2,\theta}(X,Z)\big)\Big] && (14) \\ &+ n^{1/2}\Big[n_2^{-1}\sum_{i=n_1+1}^{n} f_{3,\hat{\theta}}(X_i, Z_i) - \mathbb{E}\big(f_{3,\theta}(X,Z)\big)\Big] && (15) \end{aligned}$$

As anticipated above, we shall show that each of the terms in the right hand side of the latter display weakly converges to a Gaussian process. To prove that (11) converges to a Gaussian process, we combine the Slutzky Lemma, the Central Limit Theorem and the Uniform Central Limit Theorem by van der Vaart (1998, Theorem 18.4). The proof for (12) is similar and thus omitted. We begin by establishing convergence in distribution of the expression $n^{1/2}\big(\hat{\beta}_L - \beta_L\big)$ in (13). For convenience sake, define the quantities

$$\begin{aligned} \hat{A} &:= \Big(n_2^{-1}\sum_{i=n_1+1}^{n} X_i Z_i'\Big)\hat{\Sigma}\Big(n_1^{-1}\sum_{i=1}^{n_1} Z_i Y_i\Big) \\ \hat{B} &:= \left[n_2^{-1}\sum_{i=n_1+1}^{n} X_i^2 - \Big(n_2^{-1}\sum_{i=n_1+1}^{n} X_i Z_i'\Big)\hat{\Sigma}\Big(n_2^{-1}\sum_{i=n_1+1}^{n} X_i Z_i\Big)\right]^{-1} \end{aligned}$$

With this notation in hand, rewrite the scaled difference $n^{1/2}\big(\hat{\beta}_L - \beta_L\big)$ as:

$$n^{1/2}\big(\hat{\beta}_L - \beta_L\big) = \hat{B} n^{1/2}(\hat{\lambda}_L - \lambda_L) - \hat{B} n^{1/2}(\hat{A} - A) + (\lambda_L - A) n^{1/2}(\hat{B} - B)$$

where A and B are, respectively, the population analogs of $\hat{A}$ and $\hat{B}$. We now analyze convergence of each of the terms in the right hand side of the latter display. Consider the first term. Under conditions (A1.ii), (A1.iii) and (A2), we have that the average $\hat{B}$ converges in probability to B by the law of large numbers. Under conditions (A4.ii) and (A3.iii), we have that the scaled difference $n^{1/2}(\hat{\lambda}_L - \lambda_L)$ converges in distribution to random variable with zero mean normal distribution by the Donsker Theorem in van der Vaart (1998, Theorem 19.5). The product $\hat{B} n^{1/2}(\hat{\lambda}_L - \lambda_L)$ converges then to random variable with zero mean normal distribution by Slutzky Lemma. Consider now the second term. Under conditions (A1.ii)(A3.iii), (A2) and (A4.i), the scaled difference $n^{1/2}(\hat{A} - A)$ converges in distribution to

a random variable with zero mean normal distribution by the Central Limit Theorem. The product $\hat{B}n^{1/2}(\hat{A}-A)$ then converges then to random variable with zero mean normal distribution by Slutzky Lemma. Asymptotic normality of the last term follows from a similar reasoning. Hence, $n^{1/2}\big(\hat{\beta}_L - \beta_L\big)$ converges in distribution to a random variable with zero mean normal distribution because equals a sum of terms converging in probability to normal random variables. Since $\mathbf{1}(q_1 \neq 0)\mathbf{1}(q_1 < 0)$ is not random, Slutzky lemma does guarantee that (11) converges in distribution to a normal distribution for a given q_1. To extend the latter pointwise convergence in distribution result to uniform convergence over q, we need to verify that $q \mapsto (11)(q)$ is asymptotically tight (see van der Vaart, 1998, Theorem 18.4). For any $q_{1,1}$ and $q_{1,2}$ in the unit interval, consider the difference:

$$\begin{aligned}(11)(q_{1,1}) - (11)(q_{1,2}) &= \mathbf{1}(q_{1,1} \neq 0)\mathbf{1}(q_{1,1} < 0)\big(\hat{\beta}_L - \beta_L\big) - \mathbf{1}(q_{1,2} \neq 0)\mathbf{1}(q_{1,2} < 0)\big(\hat{\beta}_L - \beta_L\big) \\ &= (\mathbf{1}(q_{1,1} \neq 0)\mathbf{1}(q_{1,1} < 0) - \mathbf{1}(q_{1,2} \neq 0)\mathbf{1}(q_{1,2} < 0))\big(\hat{\beta}_L - \beta_L\big)\end{aligned}$$

For any $q_{2,2}$, $sup_{q_{1,1}}(\mathbf{1}(q_{1,1} \neq 0)\mathbf{1}(q_{1,1} < 0) - \mathbf{1}(q_{1,2} \neq 0)\mathbf{1}(q_{1,2} < 0)) \leq 1$. Hence, for every any partition of the unit interval into finitely many intervals I_1,..,I_k of size $\delta > 0$ we have :

$$\sup_k \sup_{q_{1,1},q_{1,2} \in I_k} n^{1/2}(11)(q_{1,1}) - (11)(q_{1,2}) \quad \leq \quad n^{1/2}(\hat{\beta}_L - \beta_L)$$

To conclude that $q \mapsto (11)(q)$ is asymptotically tight, note that $n^{1/2}(\hat{\beta}_L - \beta_L) = O_P(1)$ by our discussion above.

We now establish the weak convergence of (13). Add-and-subtract $\mathbb{E}(f_{1,\hat{\theta}}(Y,Z))$ within the sum in (13) to obtain:

$$\begin{aligned}n^{1/2}\Big[n_1^{-1}\sum_{i=1}^{n_1} f_{1,\hat{\theta}}(Y,Z) - \mathbb{E}\big(f_{1,\theta}(Y,Z)\big)\Big] &= n^{1/2}\Big[n_1^{-1}\sum_{i=1}^{n_1} f_{1,\hat{\theta}}(Y,Z) - \mathbb{E}\big(f_{1,\hat{\theta}}(Y,Z)\big)\Big] &(16)\\ &+ n^{1/2}\Big[\mathbb{E}\big(f_{1,\hat{\theta}}(Y,Z)\big) - \mathbb{E}\big(f_{1,\theta}(Y,Z)\big)\Big] &(17)\end{aligned}$$

We shall show that (16) is asymptotically equivalent to another random element which weakly converges to a Gaussian process, and that (17) is asymptotically negligible. Consider first (16). Under assumption (A4.i), the expression (16) is asymptotically equivalent to:

$$n^{1/2}\Big[n_1^{-1}\sum_{i=1}^{n_1} f_{1,\theta}(Y,Z) - \mathbb{E}\big(f_{1,\theta}(Y,Z)\big)\Big] \tag{18}$$

To see why, recall from inequality (8) in the proof of Proposition 1 that:

$$|f_{1,\hat{\theta}}(Y,Z) - f_{1,\theta}(Y,Z)| \leq m_1(Y,Z) \times o_P(1)$$

Under (A4.i), we have that the expectation of the square of $m_1(Y,Z)$ is finite so that $\mathbb{E}\big(|f_{1,\hat{\theta}}(Y,Z) - f_{1,\theta}(Y,Z)|^2\big) = o_P(1)$. It follows then from a result in van der Vaart (1998, Lemma 19.4) that (16) and (18) have the same asymptotic distribution. To establish the weak convergence of (18), it suffices to verfiy that the class $\mathcal{F}_1$ is Donsker. This in turns follows from inequality (8) and Assumption (A4.i) (see van der Vaart, 1998, p. 271). We now show that (17) is $o_P(1)$. Use $f_{1,\theta}(Y,Z) := q_2'\Sigma ZY$ to write:

$$\begin{aligned}(17) &= n^{1/2}\Big[\mathbb{E}\big(q_2'\hat{\Sigma}ZY\big) - \mathbb{E}\big(q_2'\Sigma ZY\big)\Big] \\ &\leq n^{1/2}\mathbb{E}\big(||q_2'(\hat{\Sigma}-\Sigma)||\big)^{1/2}\mathbb{E}\big(||ZY||\big)^{1/2} \\ &\leq n^{1/2}\big(\mathbb{E}(||q_2'\hat{\Sigma}||) - ||q_2'\Sigma||\big)^{1/2}M \\ &\leq n^{1/2}0^{1/2}M \leq o(1) \leq o_P(1)\end{aligned}$$

where the first inequality follows from the Cauchy-Schwartz inequality, the second one from the triangle inequality and assumption (A1.iii), and the third inequality from Assumption (A4.iii)

We now establish the weak convergence of (14). Add-and-substract $\mathbb{E}(f_{2,\hat{\theta}}(X,Z))$ within the sum in (14) to

obtain:

$$n^{1/2}\Big[n_2^{-1}\sum_{i=n_1+1}^{n_2} f_{2,\hat{\theta}}(X_i,Z_i) - \mathbb{E}\big(f_{2,\theta}(X,Z)\big)\Big] = n^{1/2}\Big[n_2^{-1}\sum_{i=n_1+1}^{n_2} f_{2,\hat{\theta}}(X_i,Z_i) - \mathbb{E}\big(f_{2,\hat{\theta}}(X,Z)\big)\Big] \quad (19)$$

$$+ \quad n^{1/2}\Big[\mathbb{E}\big(f_{2,\hat{\theta}}(X,Z)\big) - \mathbb{E}\big(f_{2,\theta}(X,Z)\big)\Big] \quad (20)$$

As for (13), we shall show that (19) is asymptotically equivalent to another random element which weakly converges to a Gaussian process, and that (20) is $o_P(1)$. Consider first (19). Under assumption (A4.i), the term (21) in the latter display is asymptotically equivalent to:

$$n^{1/2}\Big[n_2^{-1}\sum_{i=n_1+1}^{n_2} f_{2,\theta}(X_i,Z_i) - \mathbb{E}\big(f_{2,\theta}(X,Z)\big)\Big] \quad (21)$$

To see why, from inequality (9) in the proof of Proposition 1 notice that:

$$|f_{2,\hat{\theta}}(X,Z) - f_{2,\theta}(X,Z)| \leq o_P(1)$$

so that $\mathbb{E}\big(|f_{2,\hat{\theta}}(X,Z) - f_{2,\theta}(X,Z)|^2\big) = o_P(1)$. It follows then again from a result in van der Vaart (1998, Lemma 19.4) that (19) and (21) have the same asymptotic distribution. To establish the the weak convergence of (21), from the inequality (9) and assumption (A4.i) we note that the family $\mathcal{F}_2$ is Donsker. We now show that, under the support restriction (A4.iii), the term (20) is $o_P(1)$. Use $f_{2,\theta}(X,Z) := q_2'\Sigma ZX\mathbf{1}(q_2'\Sigma\mu < 0)\beta_U$ to write:

$$\begin{aligned}
(20) &= n^{1/2}\big[\mathbb{E}(q_2^{\hat{\beta}_{1U}\prime}\hat{\Sigma}ZX\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0)) - \mathbb{E}(\beta_{1U}q_2'\Sigma_o ZX\mathbf{1}(q_2'\Sigma_o\mu_o < 0))\big] \\
&= n^{1/2}\mathbb{E}\Big[\hat{\beta}_{1U}\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0)q_2'(\hat{\Sigma} - \Sigma_o)ZX\Big] \quad (22) \\
&+ n^{1/2}\mathbb{E}\Big[\beta_{1U}q_2'\Sigma_o ZX\big(\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0) - \mathbf{1}(q_2'\Sigma_o\mu_o < 0)\big)\Big] \quad (23) \\
&+ n^{1/2}\mathbb{E}\Big[(\hat{\beta}_{1U} - \beta_{1U})\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0)q_2'\Sigma_o ZX\Big] \quad (24)
\end{aligned}$$

Considering (22) first, use the Cauchy-Schwartz to write

$$\begin{aligned}
(22) &\leq n^{1/2}\mathbb{E}(\hat{\beta}_{1U}^2)^{1/2}\mathbb{E}(\mathbf{1}(q_2'\hat{\Sigma}ZX < 0))^{1/2}\mathbb{E}(||q_2'(\hat{\Sigma} - \Sigma_o)||)^{1/2}\mathbb{E}(||ZX||)^{1/2} \\
&\leq O_P(n^{1/2}) \times 1 \times 0^{1/2} \times M \leq o(1) \leq o_P(1)
\end{aligned}$$

where the second inequality follows from assumption (A1.iii), assumption (A4.iii), and assumption (A3.i). Consider now the term (23). Use again the Cauchy-Schwartz inequality to write:

$$\begin{aligned}
(23) &\leq n^{1/2}\mathbb{E}(\hat{\beta}_{1U}^2)^{1/2}\mathbb{E}(||q_2'\Sigma_o ZX||)^{1/2}\mathbb{E}\Big(\big(\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0) - \mathbf{1}(q_2'\Sigma_o\mu_o < 0)\big)^2\Big)^{1/2} \\
&\leq O_P(n^{1/2})\mathbb{E}\Big(\big|\mathbf{1}(q_2'\hat{\Sigma}\hat{\mu} < 0) - \mathbf{1}(q_2'\Sigma_o\mu_o < 0)\big|\Big)^{1/2}
\end{aligned}$$

where the second inequality follows from assumption (A4.i) and because the square of the difference of two binary variables is equal to the absolute value of the difference. Under assumption (A4.i) and (A4.iii), it follows from a result in Bontemps, Magnac and Maurin (2012, proof of Lemma 13) that the expectation in the right hand side of the latter display is $o_P(n^{-1/2})$, hence (23) $= o_P(1)$. By similar arguments we have (24) $= o_P(1)$. Verifying the weak convergence of (15) is similar and thus ommited. ■