Set Identified Linear Models

Christian Bontemps[*], Thierry Magnac[†], Eric Maurin[‡]

First version, August 2006
This version, October 2007

Abstract

In this paper, we first exhibit examples of set identification where the usual moment conditions are incomplete because they depend on a unknown bounded scalar function. We show that incomplete linear moment conditions generate set identification where the identified set is bounded and convex. We characterize the identified set in both cases where the number of moment conditions is equal to, or greater than the number of parameters of interest. We derive consistent and asymptotically normal estimators of the support functions of these bounded and convex sets. We also construct procedures testing the validity of overidentifying restrictions which generalize the Sargan test. Some empirical illustrations on income data and on artificial data are provided.

[*]Toulouse School of Economics
[†]Toulouse School of Economics
[‡]PSE, Paris

1 Introduction

Point identification is often achieved by using strong and difficult to motivate restrictions on the parameters of interest. This paper contributes to the growing literature that uses weaker assumptions, under which parameters of interest are set identified only. A parameter is set identified when the identifying restrictions impose that it lies in a set that is smaller than its potential domain of variation, but larger than a single point. We exhibit a class of semi-parametric models where set identification and estimation can be achieved at low cost and using inference tools close to what is standard in applied work.

In our set-up, parameters of interest are defined by a set of restrictions that do not point-identify them and that we call incomplete linear moment restrictions. Specifically, we consider y, a dependent variable, x, a vector of p variables and assume that parameter β satisfies:

$$E(x^T(x\beta - y)) = E(x^T u(x)), \tag{1}$$

where $u(x)$ is any single-dimensional measurable function that takes its values in a given bounded interval $I(x)$ that contains zero. One leading example is the familiar linear regression model $y = x\beta + \varepsilon$, where ε is uncorrelated with x, but where the continuous dependent variable, y, is censored by interval. The issue addressed in this paper is to identify and estimate the set, B, lying in $\mathbb{R}^p$ of all values of β which satisfy Equation (1) for at least one $u(.)$. It is not difficult to show that set, B, is necessarily non-empty, convex and bounded. Convexity and boundness are the key features that we exploit to further characterize B.

A general approach to inference when a set only is identified was recently proposed by Chernozukov, Hong et Tamer (2007). They define the identified set as the set of zeroes of a functional, called the criterion, and there is no constraint on its shape. In particular, their very general procedure remains valid even when the identified set is not convex nor bounded. In this paper, we propose a novel and more direct approach to the issue of set identification when the identified set is bounded and convex. Our first contribution is a sharp characterization of the identified set using the concept of support functions which is naturally associated with convex sets (Rockafellar, 1970). In each direction of interest, which spans the unit sphere in $\mathbb{R}^p$, we show that the support function of the identified set B is the expectation of an explicit and simple random function. Second, we show that a similar characterization of the identified set also holds true when the incomplete linear moment conditions are written as a function of m instruments z:

$$E(z^T(x\beta - y)) = E(z^T u(x)), \tag{2}$$

In this *endogenous* set-up, the identified set remains convex and bounded as in the exogeneous case. Also, when there are as many instruments as explanatory variables, the identified set, B, remains necessarily non-empty. This is not the case anymore when there are supernumerary instruments. We explicit a necessary and sufficient condition, a generalization of the usual over-identifying condition *à la* Sargan, under which B is not empty. We also exhibit conditions under which the existence of supernumerary instruments restores point identification.

The next contribution of the paper is to provide a simple estimator of the support function of the identified set. This estimator is the empirical analogue of the expectation of the random function to which the support function is equal. In their closely related contribution, Beresteanu and Molinary (2006) provide an estimation procedure for a class of convex identified sets using the theory of random sets. We find it more fruitful to directly use the theory of stochastic process from which the theory of random sets is derived because the results can be obtained under simpler conditions and are easier to generalize to the endogenous case. Under standard conditions, we first show that our estimate of the support function converges almost surely to the true function, uniformly over the unit sphere of $\mathbb{R}^p$. Second, we show that the $\sqrt{n}$ inflated difference between the estimate and the true function converges in distribution to a Gaussian process whose covariance matrix is given. Interestingly enough, our approach reveals that the asymptotic results of Beresteanu and Molinary (2006) actually simplify to a quite standard linear model format for the covariance matrix. Also, our procedure provides new asymptotic results for the cases where the identified set is not strictly convex and the regressors not absolutely continuous. Given the prevalence of discrete regressors, these generalizations are worthy of attention.

Furthermore and more importantly, we develop a new asymptotically exact test procedure for null hypotheses such as H_0:$\beta_0 \in B$. We argue that this class of hypotheses is more attractive to economists than hypotheses about sets (such as, say, H_0:$B_0 \in B$). For example, the new Sargan condition developed above can be written this way. The convexity of support functions associated to convex sets is the key feature that simplifies our test procedure. The test statistic is constructed as the minimum value of a convex function over the compact unit sphere in a finite-dimensional space. We exploit this characteristic to derive the asymptotic distribution of the test statistic even in non-differentiable cases, that is even when the convex set B has kinks, faces and is not strictly convex. Finally, the same key feature of convexity, allows us to derive similar asymptotic properties of the estimates in the case where there are supernumerary moment restrictions. Estimates are uniformly almost surely consistent and the inflated difference between the estimated and true functions converges to a Gaussian process.

This paper belongs to the growing literature on set identification. From the very start of structural modeling, identification meant point identification. Dispersed in the literature though, there are examples of the weaker concept of set identification. Set identification can come from two broad sets of causes : information might be missing or structural models might not generate enough moment restrictions. The oldest examples of the first case corresponds to measurement errors. They were introduced by Gini (1921), Frish (1934) and further analyzed, decades later, by Klepper and Leamer (1984), Leamer (1987) or Bollinger (1996). There are many other examples of missing information generating incomplete identification (see Manski, 2003 for a survey). Seminal analysis of the incomplete information case include Fréchet (1951), Hoeffding (1940) and Manski (1989) whereas recent applications include Alvarez, Melenberg and van Soest (2001), Blundell, Gosling, Ichimura and Meghir (2007) or Honoré and Lleras-Muney (2006). Horowitz and Manski (1995) consider the case where the data are corrupted or contaminated while Moffitt and Ridder (2003) provide a survey of the results relative to two sample combination. Structural models delivering moment inequality restrictions (instead of equalities) are the second type of models leading to set identification (Andrews, Berry and Jia, 2002, Pakes, Porter, Ho and Ishii, 2005, Haile and Tamer, 2003, Ciliberto and Tamer, 2005, Galichon and Henry, 2006, among others). Set identification can also be generated by discrete exogeneous variation such as in Chesher (2003). In both cases, Chernozhukov, Hong and Tamer (2007) use a criterion approach for the definition of the identified set and subsampling techniques for estimation and inference (see also Romano and Shaikh, 2006). Rosen (2007) develops simple testing procedures.

The class of models considered in this paper belongs to both branches of the literature. Incomplete linear conditions can be interpreted as a specific set of inequality restrictions generated by some missing information. The leading examples that we propose are derived from partial observation when covariates are censored by intervals as in Manski & Tamer (2002), when the continuous regressor is observed by intervals or is discrete (Magnac & Maurin, 2007a), when outcomes are censored by intervals (Beresteanu & Molinari, 2006) or when regressors are observed in two distinct samples.

Incomplete linear moment conditions define identified sets which are convex and bounded. The approach developped in this paper relies directly on these two properties and we expect that the same procedure can be adapted to other contexts where the identified set is convex and bounded. In contrast, we believe that estimation is difficult to implement in set-ups such as those proposed by Klepper and Leamer (1984) or Erikson (1993) because the corresponding identified sets are not bounded and convex. Estimation and inference are definitely more difficult to provide in such cases although our results could also help. Finally, while our results are given in a global linear set-up, their adaptation to a local linear set-up seems to be achievable at low cost.

Section 2 develops three examples that are of interest for applied econometricians and generate incomplete linear moment conditions. Section 3 characterizes the identified set using these moment restrictions. We analyze the case where the number of parameters is equal to the number of restrictions as well as the case where the number of restrictions is larger than the number of parameters and we provide the extension of the Sargan hypothesis. For the sake of simplicity, Section 4 specializes to the case of outcomes measured by intervals and under general conditions, we derive asymptotic properties of estimates in the case of no moment restrictions in surplus, exact test procedures and asymptotic properties of estimates using supernumerary restrictions. Section 5 is devoted to Monte Carlo experiments about the testing procedures and Section 6 presents the results of an empirical illustration using censored income data. Section 7 proposes an extension to the case of ecological inference. Section 8 concludes.

2 Set Identification in Linear Models: Examples and a General Framework

Let us consider the familiar linear regression model $y = x\beta + \epsilon$, where y is a continuous dependent variable, x a vector of independent variables of dimension K (where $E(x^T x)$ full rank) and ϵ a random variable uncorrelated with x. When y and x are perfectly observed, the set B of parameters observationally equivalent to the true parameter boils down to a singleton defined by the usual moment condition,

$$E(x^T(x\beta - y)) = 0.$$

The regression coefficient $\beta = E(x^T x)^{-1}E(x^T y)$ is the only parameter such that there is an ϵ uncorrelated with x satisfying $x\beta + \epsilon \equiv y$. In the remainder of this section, we show that there is a wide variety of contexts such that the identification set B of the linear regression model is not defined by the previous moment condition anymore, but by the following generalisation,

$$E(x^T(x\beta - y)) = E(x^T u(x)),$$

where $u(x)$ is a measurable function that takes its values in a given uniformly bounded interval $I(x)$ that contains zero. In such a case, it is shown in Section 3 that B remains a non-empty convex and bounded set, but it is not necessarily a singleton anymore. The familiar moment condition above defines only one specific admissible value of the parameter (i.e., the one which corresponds to $u(x) = 0$). We first exhibit examples that lead to such a framework.

2.1 Example 1 : Linear Regressions with Interval Data on the Dependent Variable

The first interesting set of examples corresponds to the case where the dependent variable y is observed by interval only (see *e.g.* Manski and Tamer, 2002). Household income, individual wages, hours worked or time spent at school represent continuous outcomes that are often reported by interval only in survey or administrative data[1]. For example, the long standing (and still growing) literature on the long run variations in the distribution of income relies on tax data reporting the number of tax payers for a finite number of income brackets only (see e.g., Piketty, 2005). Researchers typically use parametric extrapolation techniques to estimate the fractiles of the latent income distributions and to analyse variations across periods and countries. The robustness of these analyses to alternative extrapolation assumptions remains unclear, however.

In these examples, the data are given by the distribution of a random variable $w = (y, y^*, x)$ where $y^* \in [y_0, y_K)$ is a bounded outcome, x a vector of L covariates.[2] A derivative variable is y, the result of censoring y^* by intervals. If $y_0 < y_1 < ... < y_{K-1} < y_K$ denote the bounds of

[1] Also, for anonymity reasons, it is more and more often the case that only interval information is made available to researchers even though the information collected was actually continuous.

[2] Without bounds on y^*, parameter β is not identified in the strong sense, *i.e.* any value of β rationalizes the data. It stems from the well known argument that there is no robust estimator fro the mean (see Magnac and Maurin, 2007a, for an example).

the K intervals, y can be re-defined as the center of the observed intervals, i.e.

$$y = \sum_{k=0}^{K-1} (\frac{y_k + y_{k+1}}{2}) 1(y^* \in [y_k, y_{k+1})). \tag{3}$$

The choice of the mid-point is just a normalization as proved in the Appendix after the proof of Proposition 1 below. Any choice which preserves the ordering is possible and does not affect further developments. Variable y is discrete and only realizations of (y, x) are observed. We denote $g_k(x)$ the probability of observing $y = \frac{y_k + y_{k+1}}{2}$ conditional on x and $G_k(x) = \sum\limits_{0 \leq l < k} g_l(x)$ the observed cumulative distribution of y i.e., $Pr(y < y_k \mid x)$.

Within this framework, we consider linear latent models

$$y^* = x\beta + \epsilon, \tag{4}$$

where ϵ is a random variable uncorrelated with x. The distribution of ϵ conditional on x is denoted $F_\epsilon(. \mid x)$. The issue is to characterize B the set of parameters β such that the latent model $(\beta, F_\epsilon(. \mid x))$ generates the distribution $G_k(x)$ ($k = 0, ...K-1$) through model 3. By definition, β is in B if and only if there is an ε uncorrelated with x satisfying,

$$F_\epsilon(y_k - x\beta \mid x) = G_k(x), \text{ for any } k = 0, ...K-1. \tag{5}$$

In other words, the identification set B_1 can be defined as,

$$B_1 = \left\{ \beta \in \mathbb{R}^K \text{ s.t. there is } \varepsilon \text{ uncorrelated with } x \text{ and satisfying Eq. (5)} \right\}$$

For reasons related to consistency arguments in the estimation section, we shall, from now on, use the closure of such a set denoted as $cl(B_1)$. We will also assume that all variables that we consider are in L_2 so that all cross-moments exist.

The following proposition shows that B_1 is not a singleton, but defined by a moment condition similar to moment condition (1).

Proposition 1 *The two following statements are equivalent,*

(i) $\beta \in cl(B_1)$,

(ii) there exists a measurable function $u(x)$ *from* $\mathbb{R}^K$ *to* $\mathbb{R}$ *which takes its values in the interval* $I(x) = [-\Delta(x), \Delta(x)]$, *where*

$\Delta(x) = \frac{1}{2} \sum_{k=0}^{K-1} [(y_{k+1} - y_k)(G_{k+1}(x) - G_k(x))]$

and such that, $E(x^T(x\beta - y)) = E(x^T u(x))$.

Proof: All proofs are in Appendix A.

2.2 Example 2: Categorical Data on Subjective Outcomes

Another interesting set of examples corresponds to categorical data on dependent variables related to individual opinions or attitudes. Public opinion polls typically contain dozens of such outcomes. For example, a poll conducted before presidential elections generally contains a large set of questions measuring binary subjective data, such as ”Which of the two candidates - George W. Bush or Al Gore - do you think would do a better job on the gun control issue?” (Louis Harris, may 2000). Survey on attitudes to public policy issues provides information on similar variables. For example, the International Survey Program conducted in 1992 asks individuals whether they agree with the following statement, ”It is the responsibility of the government to reduce the differences in income between people with high income and those with low income” (see e.g., Corneo and Grüner, 2000). Also survey on job satisfaction or on happiness typically contain categorical data on subjective outcomes, such as ”Taking all things together, how would you say things are these days - would you say you are very happy, fairly happy or not too happy these days?” (see e.g. Di Tella and MacCulloch, 2006).

To analyse such outcomes, researchers typically assume that they are related to a continuous intensity measure $y^* = x\beta + \varepsilon$ and provide estimates of β under specific parametric assumptions on the distribution of ε (ordered probit or logit). Whether or not these models are non parametrically identified is an open question, however. To begin with, consider the very simple case where the categorical outcome under consideration is binary $d \in \{0, 1\}$ (fairly happy vs not too happy) and suppose that it is related to a latent intensity $y^* = x\beta + \varepsilon$ varying between 0 and 1 by the following relationship:

$$d = 1 \text{ iff } y^* \geq y_1$$

where ε is uncorrelated with x whereas y_1 is an unknown threshold in $]0, 1[$.The intercept of the model (say β_0) and the threshold y_1 cannot be jointly identified and we can always chose $y_1 = \frac{1}{2}$ as a normalisation.

Within this framework, y can be defined (exactly in the same spirit as in the previous section) as $y = \frac{1}{4}1(y^* < \frac{1}{2}) + \frac{3}{4}1(y^* \geq \frac{1}{2})$ and we have,

$$E(x^T(x\beta - y)) = E(x^T(y^* - y)) = E(x^T E((y^* - y) \mid x)) = E(x^T u(x))$$

where $u(x)$ satisfies,

$$-\frac{1}{4} < u(x) = -\frac{1}{4} + E(y^* - \frac{1}{2}1(y^* \geq \frac{1}{2})) \mid x) \leq \frac{1}{4}$$

Denoting B_2 the identification set, we have just shown that $\beta \in cl(B_2)$ implies that there is a $u(x)$ taking its value in $\left[\frac{-1}{4}, \frac{1}{4}\right]$ such that $E(x^T(x\beta - y) = E(x^T u(x))$. Using exactly the same argument as for Proposition 1, we can show that the reciprocal holds true too and that the identification set is actually defined by a moment condition similar to condition (1).

Proposition 2 *The two following statements are equivalent,*

(i) $\beta \in cl(B_2)$,

(ii) there is a measurable function $u(x)$ *from* $\mathbb{R}^K$ *to* $\mathbb{R}$ *which takes its values in the interval* $I(x) = \left[-\frac{1}{4}, \frac{1}{4}\right]$ *, such that,* $E(x^T(x\beta - y)) = E(x^T u(x))$.

When the categorical outcome under consideration has $K > 2$ categories, it is not difficult to adapt the above argument. Specifically, let us now assume that d takes its value in $\{0, ..., K\}$ and suppose that it is related to $y^* = x\beta + \varepsilon$ by :

$$d = k \text{ iff } y_k \leq y^* < y_{k+1}$$

where $y_1 = 0 \leq ...y_{k...} < y_{K+1} = 1$ is a set a threshold in $[0, 1]$. If these thresholds are known, we are formally back to Example 1 and the identification set has exactly the same structure as in Propositions 1 or 2. When the thresholds are not known, β is in the identification set if and only if there is a set of thresholds $y_1 = 0 < ...y_k < ... < y_{K+1}$ and a function $u(x) \in [-\Delta(x), \Delta(x)]$ such that $E(x^T(x\beta - y) = E(x^T u(x))$, where y is re-defined as before as $\sum_{k=0}^{K-1}(\frac{y_k + y_{k+1}}{2})1(y^* \in [y_k, y_{k+1}[)$ and where $\Delta(x) = \frac{1}{2}\sum_{k=0}^{K-1}[(y_{k+1} - y_k)(G_{k+1}(x) - G_k(x))]$. In such a case, the identification set is the union of sets defined by moment conditions similar to condition (1).

2.3 Example 3: Binary Models with Discrete or Interval-valued Regressors

A last set of examples correspond to contingent valuation studies where participants are asked whether their willingness-to-pay (y^*) for a good or resource exceeds a bid $-v$ chosen by experimental design (see e.g., McFadden, 1994). The outcome under consideration y equals one if the respondent willingness-to-pay exceed the experimental bid (i.e., $y^* + v > 0$) and the problem is to infer the relationships between y^* and a set of covariates x from available observations on y, x

and v. Related examples correspond to dosage response models where y is one if, for example, a lethal dose y^* exceeds a treatment dose $-v$ chosen by experimental design.

In all these cases, a natural approach is to assume that $y^* = x\beta + \varepsilon$ and to estimate the semiparametric binary model $y = 1(x\beta + v + \varepsilon > 0)$ under the assumption that ϵ is uncorrelated with regressors x and independent of regressor v conditional on x (i.e., $F_\varepsilon(. \mid x, v) = F_\varepsilon(. \mid x)$ if only because of the experimental design). Also, it is often plausible to suppose that the support of y^* is small relative to the support of v (i.e. , Supp$(x\beta + \varepsilon) \subset$ Supp$(-v)$). Assuming that $(x\beta + \varepsilon)$ represents the latent propensity to buy an object and $-v$ the price of this object, it simply amounts assuming that for sufficiently high (low) price no one (everyone) buys the object under consideration.

When v is continuously observed and its support is an interval, we are in the case studied by Lewbel (2000) or Magnac and Maurin (2007a), and β is point identified. In contrast, when v is not observed continuously[3], the set B_4 of observationally equivalent parameters is not a singleton anymore.

To be more specific, let us assume that the data are characterized by (y, v, v^*, x) but that only (y, v^*, x) is observed where v^* is the result of censoring v by interval. The support of v^* is denoted $\{1, ..., K-1\}$ and the support of v conditional on $v^* = k$ is denoted $[v_k, v_{k+1})$.

In such a case, a vector of parameter β is in B_4 if and only if there is (1) a latent distribution function of v conditional on (v^*,x) and (2) a latent random variable ε uncorrelated with x, independent of v conditional on x and satisfying Supp$-(x\beta + \varepsilon) \subset$ Supp(v), such that the latent model $(\beta, F_\varepsilon(. \mid x))$ generates the observed conditional probability of success $Prob(y = 1 \mid x, v^*)$. Specifically, denoting $\bar{y} = \frac{v_{v^*+1} - v_{v^*}}{p_{v^*}(x)} y - v_K$ we have,

Proposition 3 *(Magnac and Maurin, 2007b) Consider β a vector of parameter and $Pr(y = 1 \mid v^*, x, z)$ (denoted $G_{v^*}(x, z)$) a conditional distribution function which is non decreasing in v^*. The two following statements are equivalent,*

(i) $\beta \in cl(B_4)$

(ii) there exists a function $u(x)$ taking its values in $I^*(x) = [\underline{\Delta}^*(x), \overline{\Delta}^*(x)]$ where (by con-

[3] As noted by Lewbel, Linton and McFadden (2006), virtually all existing contingent valuation datasets draw bids from a discrete distribution. Furthermore, it is often the case that variables are censored by intervals.

vention, $G_0(x) = 0$*,* $G_K(x) = 1$*),*

$$\begin{aligned}
\overline{\Delta}^*(x) &= \sum_{k=1,\dots K-1} (G_{k+1}(x) - G_k(x))(v_{k+1} - v_k), \\
\underline{\Delta}^*(x) &= -\sum_{k=1,\dots K-1} (G_k(x) - G_{k-1}(x))(v_{k+1} - v_k),
\end{aligned}$$

and such that,

$$E(x^T(x\beta - \bar{y})) = E(x^T u^*(x))). \tag{6}$$

Note that in this case, the definition interval of $u(x)$ is asymmetric, contrary to the previous examples and contrary to what happens if the distribution of v is discrete (see Magnac and Maurin (2007b) for the extension).

Generally speaking, it is also very easy to see that the moment condition $E(x^T\varepsilon) = 0$ can be easily replaced by the generalized moment condition $E(z^T\varepsilon) = 0$ where z are some instruments. It comes at no cost by replacing x by z in the moment condition (for instance, equation (6))

2.4 The Set-up of Incomplete Linear Models

In this paper, we shall analyze the identification and estimation of parameters satisfying what we call an *incomplete linear model* (denoted ILM) given by *incomplete linear moment conditions*:

$$E(z^T(x\beta - y)) = E(z^T u(z)), \tag{7}$$

where $u(z)$ is any measurable function which takes values in an admissible set $I(z) = [\underline{\Delta}(z), \overline{\Delta}(z)]$ where $\underline{\Delta}(z) < 0 < \overline{\Delta}(z)$. We also assume that there exist two observable variables, $\overline{y}$ and $\underline{y}$ such that:

$$E(\overline{y} - y \mid z) = \overline{\Delta}(z), E(\underline{y} - y \mid z) = \underline{\Delta}(z). \tag{8}$$

It is indeed the case in examples 1, 2 and 3 that were developed before (See Appendix A). Cases where it is less easy construct such variables are certainly more computationally challenging that what we proposed.

We also assume the following regularity conditions:

Assumption R(egularity):

$R.i.$ *(Dependent variables)* $\overline{y}$, $\underline{y}$ and y are scalar random variables.

$R.ii.$ *(Covariates & Instruments)* The support of the distribution, $F_{x,z}$ of (x, z) is $S_{x,z} \subset \mathbb{R}^p \times \mathbb{R}^m$. The dimension of the set $S_{x,z}$ is $r \leq p + m$ where $p + q - r$ are the potential overlaps and functional dependencies.[4]. Furthermore, the conditions of full rank, $rank(E(z^T x)) = p$, and $rank(E(z^T z)) = m$ hold.

$R.iii$. The random vector $(\overline{y}, \underline{y}, y, x, z)$ belongs to the space L_2 of square integrable variables.

Along with equation (7), assumptions $R.i - ii$ defines the linear model where there are p explanatory variables and m instrumental variables (assumption $R.ii$). Assumption $R.ii$, allows for having the standard exogenous case $x = z$ as a particular case. Assumption $R.iii$ implies in particular that all cross-moments and regression parameters are well defined. In particular, it implies that, because of equation (3), we have:

$$\Delta_M = E(\max(\overline{\Delta}(z)^2, \underline{\Delta}(z)^2)) < +\infty,$$

which will be shown in the next section to imply that the set of identified parameters is bounded.

3 The Identified Set of Structural Parameters

This section provides a detailed description of B, the set of observationally equivalent parameters, β, satisfying the incomplete linear model above (IML). We first focus on the case where the number of instruments z is equal to the number of variables x (the exogenous case $z = x$ being the leading example). Second we show how the results can be extended to the case where the number of instruments z is larger that the number of explanatory variables, x.

3.1 No Moment Conditions in Surplus

When the number of instruments is equal to the number of variables, the assumption (*R.ii*) that $E(z^T x)$ is full rank implies that equation (7) has one and only one solution in β for any function $u(z)$. The set of identified parameters, B, is the collection of such parameters when function $u(z)$ varies in the admissible set:

$$B = \{\beta : \beta = (E(z^T x))^{-1} E(z^T (y + u(z))), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\}. \quad (9)$$

We first look at general properties of existence, convexity and boundedness. We continue by characterizing the identified set completely.

[4]With no loss of generality, the p explanatory variables x can partially overlap with the $q \geq p$ instrumental variables z. Variables (x, z) may also be functionally dependent (for instance x, x^2, $\log(x)$,...). A collection $(x_1, ., x_K)$ of real random variables is functionally independent if its support is of dimension K (i.e. there is no set of dimension strictly lower than K which probability measure is equal to 1).

3.1.1 Geometric and Topological Properties of the Identified Set

They are summarized in:

Proposition 4 *The identified set B is non empty, closed, convex and bounded in $\mathbb{R}^p$. It contains the focal value β^* defined as:*

$$\beta^* = E(z^T x)^{-1} E(z^T y)$$

and any $\beta \in B$ satisfies,

$$(\beta - \beta^*)^T W (\beta - \beta^*) \leq \Delta_M = E(\max(\overline{\Delta}(z)^2, \underline{\Delta}(z)^2)),$$

where $W = E(x^T z)(E(z^T z))^{-1} E(z^T x)$.

Proof. See Appendix B. ■

Proposition 4 shows that B lies within an ellipsoïd whose size is bounded by Δ_M in the metric W. The maximum-length index, Δ_M, can be taken as a measure of distance to point identification. Indeed, we can show that:

$$\lim_{\Delta_M \mapsto 0} B = \{\beta^*\},$$

and point identification is restored.

The key result in Proposition 4 is that B is convex because, as such, B can be unambiguously characterized by its *support function*. For any vector $q \in \mathbb{R}^p$, the support function of a set B is defined as:

$$\delta^*(q \mid B) = \sup\{q^T \beta \mid \beta \in B\}.$$

Using that support functions are linear homogenous of degree 1, it is sufficient to define them for any vector q belonging to the unit sphere of $\mathbb{R}^p$ i.e. $\mathbb{S} = \{q \in \mathbb{R}^p; \parallel q \parallel = 1\}$. In our specific case, the support function defined over $\mathbb{S}$ is bounded because B is bounded.

Conversely, convex sets are completely characterized by their support function (for instance, Proposition 13.1 of Rockafellar (1970)).[5] Convex set B is given by:

$$\beta \in B \Leftrightarrow \forall q, \parallel q \parallel = 1, q^T \beta \leq \delta^*(q \mid B).$$

The issue of identification of B is equivalent to the issue of identifying the support function, $\delta^*(q \mid B)$.

[5] See Beresteanu and Molinari (2006) who also use this property in order to apply the theory of random set variables.

3.1.2 The Support Function

This section characterizes the support function of B as a function of simple moments of the data.

Let q any vector of dimension $[p,1]$ and of Euclidean norm $\parallel q \parallel = 1$. Consider Q an orthogonal matrix of dimension $[p,p]$ which last column is vector q. It can be written $Q=(Q_0,q)$ where Q_0 is a matrix of dimension $[p,(p-1)]$.[6] By definition, it satisfies $QQ^T = Q^TQ = I$ and we can always write:

$$x\beta = xQ.Q^T\beta = s\beta_Q$$

where $s = xQ$ and $\beta_Q = Q^T\beta$. The p-th component of β_Q is the scalar, $q^T.\beta$, which is the coefficient of the p-th explanatory variable, $x.q$. By definition, the support function in the direction of q, is the supremum of $q^T.\beta$ when $\beta \in B$.

Most interestingly, $q^T.\beta$ can now be interpreted as the coefficient of the single-dimensional variable $x.q$ in the regression of $y + u(z)$ on s, using characterization (9). The natural tool for identifying a single coefficient in a regression is the Frisch-Waugh theorem. The value of the support function at q is obtained by taking the supremum of the set of all these single-dimensional coefficients when $u(z)$ varies in $[\underline{\Delta}(z), \overline{\Delta}(z)]$, which leads to,

Proposition 5 *Define x_q, the scalar random variable which is the remainder of the IV-projection of $x.q$ onto $x.Q_0$ using instruments z (as formally defined in the proof) and denote $z_q = \frac{x_q}{E(x_q^2)}$. We have,*

$$z_q = z.E(x^Tz)^{-1}.q$$

and the support function of B is equal to:

$$\delta^*(q \mid B) = E(z_q w_q).$$

where $w_q = \mathbf{1}\{z_q > 0\}\overline{y} + \mathbf{1}\{z_q < 0\}\underline{y}$. The function $\delta^*(q \mid B)$ correspond to supremum of $q^T\beta$ at the frontier point β_q of B, with

$$\beta_q = E(z^Tx)^{-1}E(z^Tw_q)$$

Proof. See Appendix B ■

This proposition completely characterizes B. The support function is defined everywhere because of assumption *(R.iii)* that all cross-moments are well defined.

[6] There are several ways to define Q_0 albeit it has no consequence in the following. One way to make it unique is to define Q as the unique rotation in $\mathbb{R}^p$ which maps the last basis vector $(0,...,0,1)$ into q.

Many interesting properties that we derive in the following, depend on the following Lemma. This result is a consequence of the fact that B is a bounded and convex set and thus that its support function $\delta^*(q \mid B)$ is bounded and convex.

Lemma 6 *The support function $\delta^*(q \mid B)$ is differentiable on $\mathbb{S}$ and its derivative is:*

$$\frac{\partial \delta^*(q \mid B)}{\partial q^T} = E(z^T x)^{-1} E(z^T w_q) = \beta_q,$$

This derivative is continuous except at a countable number of points. These points are defined as:

$$D_f = \{q \in \mathbb{S}; \Pr(z_q = 0) > 0\}.$$

Proof. See Appendix C. ■

The shape of the derivative is the result of an envelope theorem since the support function is obtained as the supremum of a linear expression over a convex set.

3.1.3 Implementing the Construction of B

To construct B in practice, we would choose L vectors q_l and we would construct:

$$\overline{B}_L = \cap_{q_l : \|q_l\|=1} \{\beta : q_l^T \beta \leq \frac{E(x_{q_l} w_{q_l})}{E(x_{q_l}^2)}\}.$$

By construction $B \subset \overline{B}_L$ and it is straightforward to show that the Hausdorf distance between these sets

$$d(B, \overline{B}_L) = \sup_{q; \|q\|=1} \left| \delta^*(q \mid B) - \delta^*(q \mid \overline{B}_L) \right|$$

converges to 0 when L tends to infinity provided that q_l are appropriately chosen so that:

$$\sup_{q; \|q\|=1} \min_l \|q - q_l\| \to 0.$$

Alternatively, the support function is attained at the frontier point β_q defined in Proposition 5. The convex hull, $\underline{B}_L$, of L such points, β_{q_l}, is included in convex set B and the Hausdorf distance between these sets $d(B, \underline{B}_L)$ converges to 0 when L tends to infinity provided that q_l are appropriately chosen.

These two approximations provide a sandwich-type procedure for constructing B:

$$\underline{B}_L \subset B \subset \overline{B}_L.$$

3.1.4 Geometric Properties of the Frontier of the Identified Set

The identified set B is a non empty convex set. Its frontier can thus have two peculiar characteristics, flat faces and kinks whose existence we now investigate:

Proposition 7 *i). The frontier of set B has a flat face, orthogonal to vector q, if:*

$$\Pr(z_q = 0) > 0.$$

The converse is generically true (under a condition on the support of z).

ii). The frontier of set B has kinks if and only if there exist q and $r \neq q$ such that:

$$\Pr(z_q > 0, z_r < 0) = \Pr(z_q < 0, z_r > 0) = 0$$

Proof. See Appendix B ■

One leading example of kinks and faces corresponds to the regression of an outcome observed by intervals on a constant and a dummy explanatory variable. One such example is analyzed in Section 5.3.

3.2 Supernumerary Moment Conditions

Suppose now that z is a random vector of dimension $m > p$, the dimension of covariates x. Rewrite the incomplete moment conditions as:

$$\begin{aligned} E(z^T x)\beta &= E(z^T(y + u(z))) \\ &\iff E(z^T z)^{-1/2} E(z^T x)\beta = E(z^T z)^{-1/2} E(z^T(y + u(z))) \end{aligned} \tag{10}$$

where for reasons that will become clearer later, we normalized the moment conditions by $E(z^T z)^{-1/2}$.

Regarding the general properties, it is straightforward to show that *B*, the set of observationally equivalent parameters β, is still closed, convex and bounded. The first two properties can be shown as in Proposition 4 using the fact that the moment conditions are linear and the admissible set $I(z)$ is closed and convex. To show that *B* is bounded, it suffices to select p instruments out of the possible m ones and construct the corresponding identification region as in the previous section. The true identified set is included in this identification region which is bounded by Proposition 4. These results are summarized by:

Lemma 8 *B is closed, convex and bounded.*

Obviously the additional difficulty with respect to the case where $(m = p)$ is to identify the condition under which B is not empty. The next section derives a necessary and sufficient condition which is the generalization of the usual over-identifying condition à la Sargan.

3.2.1 The Validity of Supernumerary Moment Conditions

A closer look at Equation (10) shows that parameter β lies in the identified set B if and only if the $[m, 1]$ column vector $E(z^T z)^{-1/2}E(z^T(y+u(z))$ belongs to the linear subspace generated by the p column vectors composing the $[m, p]$ matrix $E(z^T z)^{-1/2}E(z^T x)$. Let $F = \{f \in \mathbb{R}^m; f = E(z^T z)^{-1/2}E(z^T x)\beta, \beta \in \mathbb{R}^p\}$ be this subspace and let $F^\perp = H = \{h \in \mathbb{R}^m; h^T f = 0, f \in F\}$ be its orthogonal.Let $G_F = E(z^T z)^{-1/2}E(z^T x)\left[E(x^T z)E(z^T z)^{-1}E(z^T x)\right]^{-1/2}$. G_F is a $m \times p$ matrix such that $P_F = G_F G_F^T$ is the orthogonal projection on F. We can similarly define a $m \times (m-p)$ matrix G_H such that[7] $P_H = G_H G_H^T$ is the orthogonal projection on $H = F^\perp$.

One can easily notice that $G_F^T G_F = I_p$, $G_H^T G_H = I_{p-m}$, $G_H^T G_F = 0$, $G_F^T G_H = 0$ and $I_m = P_F + P_H = G_F G_F^T + G_H G_H^T$ which will help in the following derivations.

We know that we can decompose z into two sets of variables z_H and z_F which are the coefficients (up to some normalization) of the decomposition of the $m \times 1$ random vector z^T onto the subspace F and H. Formally we have the equality:

$$E(z^T z)^{-1/2} z^T = G_F z_F^T + G_H z_H^T, \tag{11}$$

where z_F^T is a $p \times 1$ vector whereas z_H^T is a $(m-p) \times 1$ vector.

We can write z_F^T and z_H^T as simple functions[8] of z:

$$z_F^T = G_F^T E(z^T z)^{-1/2} z^T$$
$$z_H^T = G_H^T E(z^T z)^{-1/2} z^T.$$

[7]Let W be the $m \times (m-p)$ matrix of the $m-p$ vectors which compose a basis of $F^\perp$, $G_H = W(W^T W)^{-1/2}$.

[8]Similarly, we can compute the following expressions:

$$E(z_F^T z_F) = G_F^T E(z^T z)^{-1/2} E(z^T z) E(z^T z)^{-1/2} G_F = I_p,\ E(z_H^T z_H) = G_H^T E(z^T z)^{-1/2} E(z^T z) E(z^T z)^{-1/2} G_H = I_{m-p}$$
$$E(z_F^T x) = G_F^T E(z^T z)^{-1/2} E(z^T x) = \left[E(x^T z)E(z^T z)^{-1}E(z^T x)\right]^{1/2},\ E(z_H^T x) = 0,\ E(z_H^T z_F) = 0.$$

Using these notations and multiplying Equation (10) successively by G_F^T and G_H^T, we obtain that the parameter β lies in B if and only if there is $u(z)$ in $I(z)$ such that

$$G_F^T(E(z^Tz)^{-1/2}E(z^T(y+u(z))) = G_F^T(E(z^Tz)^{-1/2}E(z^Tx)\beta)$$

$$\Leftrightarrow E(z_F^T(y+u(z))) = E(z_F^Tx)\beta) \tag{12}$$

and

$$G_H^T(E(z^Tz)^{-1/2}E(z^T(y+u(z))) = G_H^T(E(z^Tz)^{-1/2}E(z^Tx)\beta$$

$$\Leftrightarrow E(z_H^T(y+u(z))) = 0 \tag{13}$$

Interestingly enough, the second equation does not depend on β anymore and the first equation defines β exactly (this is the identifying equation in the point identified case). It follows that B is non empty if and only if there is $u(z)$ in $I(z)$ such that

$$E(z_H^T(y+u(z))) = 0 \tag{14}$$

where the normalized random vector z_H may be interpreted as the vector of supernumerary instruments.

Denote B_{Sargan} the identified set of parameters of the incomplete regression of y on these supernumerary instruments z_H,i.e.:

$$B_{\text{Sargan}} = \{\gamma : \gamma = (E(z_H^Tz_H))^{-1}E(z_H^T(y+u(z))), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\}. \tag{15}$$

The adapted Sargan condition means simply that B_{Sargan} contains O, the origin point of $\mathbb{R}^m$

Proposition 9 *The two following conditions are equivalent:*

i. B is not empty,

ii. $B_{\mathit{Sargan}} \ni O$, the origin point.

Proof. Using the previous developments. ■

This condition is a simple extension of the usual overidentification restriction. When moment conditions are complete, the set of admissible $u(z)$ is reduced to 0 and the set B_{Sargan} is reduced to the point $E(z_H^T.y)$. The Sargan or J-test then consists in testing $O \in B_{\text{Sargan}} = \{E(z_H^T.y)\}$ or equivalently that $E(z_H^Ty) = 0$. In the next section of the paper, we construct a general test for the assumption $H_0 : \beta_0 \in B$, when B is the identified region of an incomplete moment regression. It will provide us with a direct way for testing the Sargan condition given in Proposition 9.

3.2.2 Geometric and Analytic Characterization of the Identified Set

Assuming that the moment conditions are valid, the next issue is to provide a characterization of B and of its support function. The identified set B is defined by the general moment conditions:

$$E(z^T(x\beta - y)) = E(z^T u(z))$$

under the constraint:

$$u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)].$$

This program can be rewritten as:

$$\begin{aligned} E(z^T(x\beta + z_H\gamma - y)) &= E(z^T u(z)) \\ \gamma &= 0 \end{aligned}$$

under the same constraint for $u(z)$. Within this framework, let B_U (U : unconstrained) be the set of (β, γ) satisfying the relaxed program

$$E(z^T(x\beta + z_H\gamma - y)) = E(z^T u(z))$$

under the constraint:

$$u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)].$$

The interesting feature of B_U is obviously that it corresponds to a just identified set, which means that its support function can be characterized using Proposition 5. Also, we can build on B_U to provide a very simple geometric characterization of B:

Lemma 10 *The identified set B is the intersection of B_U and the hyperplane defined by $\gamma = 0$ (see figure 4).*

It is possible to obtain a more analytic characterization of B using equations (12) and (13). Let P_B be the identified set defined by (12) :

$$P_B = \left\{\beta \in \mathbb{R}^p,\ \exists\, u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)],\ \beta = (E(z_F^T z_F))^{-1} E(z_F^T(y + u(z))),\right\}.$$

B is included in P_B because the function $u(z)$ should also satisfied equation (13). The overidentifying restrictions reduce the size of the identified set due to this constraint. It is also true (see appendix) that P_B is the projection of B_U on the β-space (see figure 4). Assuming that the moment conditions are valid and that B is non empty (i.e., $B_{\text{Sargan}} \ni O$), we can then use general solutions for finding the support function of intersections of convex sets (Rockafellar, 1970). A very interesting characterization is:

Proposition 11 *Let q a vector of $\mathbb{R}^p$ and (q, λ) a vector of $\mathbb{R}^m$. We have:*

$$\delta^*(q \mid B) = \inf_{\lambda} \delta^*((q, \lambda) \mid B_U)). \tag{16}$$

and the infimum is attained at a set of values, $\lambda_m(q)$.

Proof. See Appendix ■

The geometric intuition is quite easy to grasp. The intersection of B_U and $\gamma = 0$ is always included in the projection of B_U onto $\gamma = 0$ using any projection direction. This explains why the support of the identified set can be expressed as a minimum of support functions of the unconstrained set B_U. Furthermore, for any point $\beta_0 \in \partial B$, the frontier of B, there always exists one projection direction such that the projection of B_U onto $\gamma = 0$ into this direction, admits β_0 as a frontier point. This projection direction is simply the tangent space (not necessarily unique) of B_U at β_0.

3.2.3 The Example of a Single Supernumerary Instrument

When there is only one supernumerary instrument, H is a subspace of dimension 1. By taking linear combination, it is possible to redefine z_H so that $z_H = (z_0, 0, ., 0)$.In such a case**,** the overidentifying condition reduces to a one-dimensional condition:

$$E(z_0(y + u(z))) = 0$$

where $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$.

The necessary and sufficient condition, $O \in B_{\text{Sargan}}$, is then equivalent to:

$$E(z_0 y) \in [\underline{U}, \overline{U}]$$

where $\overline{U} = E(|z_0| (\mathbf{1}\{z_0 > 0\}\overline{\Delta}(z) - \mathbf{1}\{z_0 < 0\}\underline{\Delta}(z)))$ and $\underline{U}$= $E(|z_0| (\mathbf{1}\{z_0 > 0\}\underline{\Delta}(z) - \mathbf{1}\{z_0 < 0\}\overline{\Delta}(z)))$. In such a case, testing overidentification boils down to testing that $(E(z_0 y) - \underline{U})(E(z_0 y) - \overline{U})$ is negative.

When O is on the frontier of B_{Sargan} (say $\partial B_{\text{Sargan}}$), we have either $E(z_0 y) = \underline{U}$ or $E(z_0 y) = \overline{U}$. For instance, assume that $E(z_0 y) = \overline{U}$. Assuming that the instrument z_0 is absolutely continuous, there is a unique $u_q(z)$ satisfying the supernumerary condition (Equation 14). It is defined as

$$u_q(z) = \mathbf{1}\{z_0 > 0\}\overline{\Delta}(z) - \mathbf{1}\{z_0 < 0\}\underline{\Delta}(z).$$

In that specific case, exact identification is restored since the parameter of interest is defined uniquely by:

$$E(z^T(x\beta - y)) = E(z^T u_q(z)).$$

The next subsection extents this result to the general case.

3.2.4 Supernumerary Moment Conditions as a Way to Restore Point Identification

We shall consider the implication of the condition that $O \in B_{\text{Sargan}}$ on functions $u(z)$ and consequently on the construction of the identified set B. First, when $O \in ri(B_{\text{Sargan}})$ [9], functions that satisfy condition (14) cannot be unique. We continue to have set identification where B has a non empty (possibly relative) interior.

More interesting cases arise when O belongs to the frontier of B_{Sargan} i.e. $O \in \partial B_{\text{Sargan}}$. Let q_O a vector which is orthogonal to a supporting hyperplane of the convex set B_{Sargan} at O. q_O is not necessarily unique if O is at a kink. Recall also that, by definition (15), B_{Sargan} is constructed as the set of solutions to the incomplete linear moment conditions:

$$E(z_H^T(z_H\gamma - y)) = E(z_H^T u(z)),$$

where $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. We can therefore apply Proposition 7 to B_{Sargan}.

Assume first that O is on a face of B, i.e. q_O is such that $\Pr\{z_{Hq_O} = 0\} > 0$. By using the proof of Proposition 7, the generating function $u_{q_O}^{\text{Sargan}}(z)$ is not unique and the identified set B is not reduced to a singleton.

Second, if B_{Sargan} has no faces, $u_{q_O}^{\text{Sargan}}(z)$ is unique. By Proposition 7 (ii), it is true whether q_O is unique or not. Therefore, the set B is reduced to a singleton generated by such a function:

$$\beta = (E(z_F^T x))^{-1}(E(z_F^T(y + u_{q_O}^{\text{Sargan}}(z)))).$$

We summarize this result in the next proposition:

[9] We use this notation of *relative interior (ri)* because B_{Sargan} is included in the $m - p$ dimensional subspace H.

Proposition 12 *If O is on the frontier ∂B_{Sargan} but not on a face, point identification of β is restored.*

4 Estimation and Inference

This section provides a description of how we estimate the support function of B and how we test hypotheses of interest. We provide asymptotic properties of our estimators and tests. We start with analyzing estimation in the case of no supernumerary moment conditions, continue with inference and finish with the case of supernumerary moment conditions.

4.1 Consistent and Asymptotically Normal Estimation: No Supernumerary Moment Conditions

We will deal only with samples $i = 1, ., n$, where $(\overline{y}_i, \underline{y}_i, y_i, x_i, z_i)$ is independently and identically distributed although proofs could be adapted to non identifal disributions and some dependent cases. In this section, we provide an estimate of the support function of set B using the central result of Proposition 5 :

$$\delta^*(q \mid B) = E(z_q w_q). \tag{17}$$

To apply the analogy principle and construct an estimate, we first use that:

$$z_q = z.E(x^T z)^{-1} q = q^T.E(z^T x)^{-1} z^T,$$

where the second equality comes from the fact that z_q is a random scalar. Let $\hat{\Sigma}_n$ be a cross-moment empirical analogue to $E(x^T z)^{-1}$.[10] Define for any $i = 1, ., n$:

$$\begin{aligned} z_{n,qi} &= z_i.\hat{\Sigma}_n.q \\ w_{n,qi} &= \mathbf{1}\{z_{n,qi} > 0\}\overline{y}_i + \mathbf{1}\{z_{n,qi} < 0\}\underline{y}_i. \end{aligned}$$

to construct the estimate:

$$\hat{\delta}_n^*(q \mid B) = \frac{1}{n}\sum z_{n,qi}.w_{n,qi} = q^T.\hat{\Sigma}_n^T \left(\frac{1}{n}\sum z_i^T.w_{n,qi}\right).$$

the empirical analog of equation (17).

Let $C_D(\mathbb{S})$ be the set of continuously differentiable functions except at a countable number of points, defined on $\mathbb{S}$ associated with the supremum distance (or Hausdorff if these functions

[10]See Appendix C for the exact definition where the usual empirical estimate is trimmed to make it bounded.

are support functions):

$$d(\phi_1, \phi_2) = \sup_{\|q\|=1, q\in\mathbb{R}^p} \mid \phi_1(q) - \phi_2(q) \mid .$$

Then $(C_D(\mathbb{S}), d)$ is a complete and separable metric space which simplifies the measurability issues that we shall simply ignore in the following (van der Vaart and Wellner, 1996).

We first state the results about consistency under very usual conditions (White, 1999, p35).

Assumption C:

$E(\mid x^T z \mid^{1+\delta}) < M < \infty$, $E(\mid z^T \overline{y} \mid^{1+\delta}) < M < \infty$, $E(\mid z^T \underline{y} \mid^{1+\delta}) < M < \infty$, for some $\delta > 0$.

Proposition 13 *Let Assumption C. Then, the estimator of the support function is, uniformly over* $\mathbb{S}$, *strongly consistent*:

$$\hat{\delta}_n^*(q \mid B) \overset{a.s.u.}{\to} \delta^*(q \mid B).$$

Proof. See Appendix C. ■

A sketch of the proof is the following. We first start from the case where Σ, the moment matrix, is known. We then show that the function within the expectation $z_q w_q$ form a parametric class as a function of Σ and q. Under condition C and the boundedness assumption, it thus forms a Glivenko-Cantelli class and the consistency result applies. Second, we replace parameter Σ by a consistent estimate and show consistency using standard results for parametric classes (van der Vaart, 1998).

For inference, we use the uniform version of a central limit theorem. Consider the stochastic process defined on $\mathbb{S}$:

$$\begin{aligned} \tau_n(q) &= \sqrt{n}\left(\frac{1}{n}\sum z_{n,qi}.w_{n,qi} - E(z_q w_q)\right) \\ &= \sqrt{n}\left(\hat{\delta}_n^*(q \mid B) - \delta^*(q \mid B)\right). \end{aligned}$$

and assume the very usual conditions (White, 1999, p118):

Assumption AN: $E(\mid z^T \overline{y} \mid^{2+\eta}) < M < \infty$, $E(\mid z^T \underline{y} \mid^{2+\eta}) < M < \infty$ for some $\eta > 0$.

Proposition 14 *Under Assumptions C and AN, $\tau_n(q)$ tends uniformly in distribution when n tends to ∞ to a Gaussian stochastic process, $\tau(q)$, or Gaussian random system such that:*

$$E(\tau(q)) = 0$$

$$Cov(\tau(q)\tau(r)) = E(x_q\varepsilon_q x_r\varepsilon_r) - E(x_q\varepsilon_q)E(x_r\varepsilon_r)$$

where r is another direction in $\mathbb{R}^p$, $\| r \| = 1$ and where:

$$\varepsilon_q = w_q - x\beta_q.$$

Proof. See Appendix C. ■

A sketch of the proof is the following. As for consistency, we start from the case where Σ, the moment matrix, is known and from the fact that $z_q w_q$ form a parametric class as a function of Σ and q. Under condition AN and the boundedness assumption, it thus forms a Donsker class and the asymptotic result applies. Second, we replace parameter Σ by a consistent estimate and show convergence in distribution using standard results for parametric classes (van der Vaart, 1998). The same argument than in the point identification is used and explains why the distribution of the estimate of Σ does not play any rôle in the asymptotic distribution of the estimates.

This section provides estimates of the support function of interest, as limit of Gaussian stochastic processes defined on the unit sphere. An alternative approach would require to rewrite B as the (Aumann) expectation of a set valued random variable and to use the recent contribution of Beresteanu and Molinari (2006) to construct a sample analog $\widehat{B}_n$ of B, such that the Haussman distance $\sqrt{n}H(\widehat{B}_n,B)$ converges to a Gaussian system. In fact, the two settings are equivalent, even though the formalism and the proofs are very different. By Hörmander's embedding proposition, the set of convex and compact set-valued random variables defined in R^p is actually homeomorphic to the set of Gaussian stochastic processes with continuous sample paths (Beresteanu and Molinari, 2006, Molchanov, 2003). We think, however, that working directly on stochastic processes defined on the unit sphere, as we do, is more adapted to our specific set-up, if only because the support function of interest can be defined by very simple moment conditions[11]. More importantly, we will see that our approach is very easily extended to the more difficult case where there are supernumerary instruments.

[11] In Appendix C, we explain why our approach provides us with an expression of the covariance function $Cov(\tau(q)\tau(r))$ which is simpler than those provided in Beresteanu and Molinari (2006).

4.2 Tests

We propose here a test of the null hypothesis H_0 :

$$\beta_0 \in B$$

whose level is asymptotically equal to a given value α. Let

$$T_n(q;\beta_0) = \hat{\delta}_n^*(q \mid B) - q^T\beta_0$$

and $T_\infty(q;\beta_0) = \delta^*(q \mid B) - q^T\beta_0$. As B is a convex set, an alternative characterization of the null hypothesis is:

$$\forall q, \| q \| = 1, T_\infty(q;\beta_0) \geq 0$$

where $q^T\beta_0$ is the scalar product between q and β_0 (Rockafellar, 1970).

Let us first consider the least favourable case, that is: $\beta_0 \in \partial B$, the frontier of B. In such a case there exists q_0 such that $T_\infty(q;\beta_0) = 0$. Let $\mathcal{Q}_0$ be the set of all q in $\mathbb{S}$ which minimize $T_\infty(q;\beta_0)$:

$$\mathcal{Q}_0 = \{q \in \mathbb{S} \text{ such that } T_\infty(q;\beta_0) = 0\}$$

For any other direction $q \in \mathbb{S}$ which does not belong to $\mathcal{Q}_0$:

$$T_\infty(q;\beta_0) > 0 \tag{18}$$

Let $q_0 \in \mathcal{Q}_0$.We have $\delta^*(q_0 \mid B) = q^T\beta_0$ and we know from proposition 14 that

$$\sqrt{n}T_n(q_0;\beta_0) = \sqrt{n}(\hat{\delta}_n^*(q_0 \mid B) - q_0^T\beta_0)$$

is asymptotically normally distributed with variance $V_{q_0} = q_0^T\Sigma^T V(z^T\varepsilon_{q_0})\Sigma q_0$.

If we knew at least one direction q_0 in $\mathcal{Q}_0$then this result would provide us with a very standard procedure for testing H_0. The problem is that we do not know q_0 and have to rely on a proxy. Let q_n be a direction which minimizes $T_n(q;\beta_0)$. Under the null, β_{q_n} should not be too far from β_0 when β_0 is on the frontier of B for n large enough. The test proposed is based on the comparison between $T_n(q_n;\beta_0)$ and $T_\infty(q_0;\beta_0) = 0$:

Proposition 15 *Let*

$$q_n \in \{\arg\min_q T_n(q;\beta_0) = (\hat{\delta}_n^*(q \mid B) - q^T\beta_0)\},$$

if $\beta_0 \in \partial B$,

$$\sqrt{n}T_n(q_n;\beta_0) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, V_{q_0}).$$

Moreover, V_{q_0} *is consistently estimated by* $\hat{V}_n = q_n^T \hat{\Sigma}_n^T \hat{V}(z^T \epsilon_{q_n}) \hat{\Sigma}_n q_n$.

The proof is provided in the appendix C. The previous result raises few comments. It is straightforward to show that $\sqrt{n}T_n(q_n;\beta_0)$ is the sum of two terms that one can handle separately:

$$\sqrt{n}T_n(q_n;\beta_0) = \sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right) + \sqrt{n}T_n(q_0;\beta_0). \tag{19}$$

Proposition 14 proves that the second term is asymptotically normally distributed with variance V_{q_0}. The first term is related to the uncertainty associated to the estimation of the direction q_0 by q_n. This term is always negative and its asymptotic distribution is non standard. However, the proposition shows that it is indeed negligible with respect to the second one. It has no influence on the asymptotic distribution (at a first order)[12] and one can estimate the variance by the empirical estimator taken at the estimated parameter. Note that the proposition is valid regardless of whether the directions q which minimize $T_n(.;\beta_0)$ or $T_\infty(.;\beta_0)$ are unique or not (i.e., regardless of whether β_0.is a kink or not). What matters is not the unicity of the arg min but the unicity of the value function.

Let us now examine the case where $\beta_0 \in Int(B)$.In such a case, $T_\infty(q;\beta_0) > 0, \forall q$ and for β_0 outside B there exists at least one direction q such that $T_\infty(q;\beta_0) < 0$. Thus, for $\beta_0 \in Int(B)$ (using the compacity of the unit sphere)

$$\inf_{q\in\mathbb{S}} T_\infty(q;\beta_0) > 0,$$

and, for β_0 outside B:

$$\inf_{q\in\mathbb{S}} T_\infty(q;\beta_0) < 0.$$

[12] This result is similar to the standard result of the LR test. In a likelihood framework, we know that the LR test is chi-squared distributed. If we call L_n the log-likelihood of the sample we know that $2L_n(\frac{\hat{\theta}_n}{\theta_0}) = 2(L_n(\hat{\theta}_n) - L_n(\theta_0))$ tends asymptotically to a chi-squared distribution:

$$2(L_n(\hat{\theta}_n) - L_n(\theta_0)) = Z + o(1)$$

where $Z \sim \chi^2$. However : $L_n(\theta) = \sum_i f_i(\theta) = n(\frac{1}{n}\sum_i f_i(\theta))$. The last expansion could be rewritten as:

$$\sqrt{n}\left(\frac{1}{n}\sum_i f_i(\hat{\theta}_n) - \sum_i f_i(\theta_0)\right) = \frac{1}{2\sqrt{n}}Z + o(\frac{1}{\sqrt{n}})$$

Our result is similar except that we do not have a M-estimator. First, w_q still depends on n in the expression of $\hat{\delta}_n^*$ and second, the arg min is not necessary unique when the set B has a kink.

Hence, when β_0 is not on the frontier of B, it is clear that $T_n(q_n;\beta_0)$ will remain either strictly positive or strictly negative, for n large enough. The combination of this result with the results given in Proposition 15 yields the following basic proposition:

Proposition 16 *Let $\beta_0 \in \mathbb{R}^d$ and*

$$\xi_n = \sqrt{n}\frac{T_n(q_n;\beta_0)}{\sqrt{\hat{V}_n}}$$

where $T_n(.;\beta_0)$ is the empirical estimator of $T_\infty(.;\beta_0)$, q_n minimizes $T_n(.;\beta_0)$ on the unit sphere and $\hat{V}_n = q_n^T\hat{\Sigma}_n^T\hat{V}(z^T\epsilon_{q_n})\hat{\Sigma}_n q_n$ is a consistent estimator of V_{q_0}.

Then, if $\beta_0 \in \partial B$,

$$\xi_n \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1),$$

if $\beta_0 \in int(B)$,

$$\xi_n \underset{n\to\infty}{\longrightarrow} +\infty$$

and if β_0 does not belong to B,

$$\xi_n \underset{n\to\infty}{\longrightarrow} -\infty.$$

The proof is straightforward. The first assertion is exactly proposition (15) where V_{q_0} is replaced by a consistent estimator $\hat{V}_n$. The last two assertions come for the positiveness (resp. the negativeness) of the second term in the equation (19) and the fact that the first one is asymptotically negligible (even if β_0 is not on the frontier, see appendix C).

We can construct a critical region with asymptotical level α for three different tests ($\mathcal{N}_\alpha$ denotes the α-quantile of the standard normal distribution):

- Test 1: $H_0 : \beta_0 \in B$ against $H_a :\ \beta_0 \notin B$. The critical region $W_n^1(\alpha)$ is defined by:

$$W_n^1(\alpha) = \{\beta_0 \in \mathbb{R}^d,\ \xi_n < \mathcal{N}_\alpha\}$$

- Test 2: $H_0 : \beta_0 \notin B$ against $H_a :\ \beta_0 \in B$. The critical region $W_n^2(\alpha)$ is:

$$W_n^2(\alpha) = \{\beta_0 \in \mathbb{R}^d,\ \xi_n > \mathcal{N}_{1-\alpha}\}$$

- Test 3: $H_0 : \beta_0 \in \partial B$ against $H_a :\ \beta_0 \notin \partial B$. The critical region $W_n^3(\alpha)$ is:

$$W_n^3(\alpha) = \{\beta_0 \in \mathbb{R}^d,\ |\xi_n| > \mathcal{N}_{1-\frac{\alpha}{2}}\}$$

We are particularly interested by the first test. The third one is also of practical interest for testing whether supernumerary instruments help in recovering point identification (i.e., for testing $O \in \partial B_{\text{Sargan}}$).

4.2.1 Consistent and Asymptotically Normal Estimation: Some Supernumerary Conditions

We use the characterization given by Proposition 11 and equation (16). If q is a vector of $\mathbb{R}^p$ and (q, λ) a vector of $\mathbb{R}^m$, we have:

$$\delta^*(q \mid B) = \inf_{\lambda} \delta^*((q, \lambda) \mid B_U)).$$

and the infimum is attained at a set of values, $\lambda_m(q)$.

Let $\hat{\delta}_n^*((q, \lambda) \mid B_U)$ the estimate of $\delta_n^*((q, \lambda) \mid B_U)$ as derived in Section 4.1 and such that:

$$\hat{\delta}_n^*((q, \lambda) \mid B_U) \overset{a.s.u.}{\rightarrow} \delta_n^*((q, \lambda) \mid B_U),$$

and:

$$\tau_n^U((q, \lambda)) = \sqrt{n}(\hat{\delta}_n^*((q, \lambda) \mid B_U) - \delta_n^*((q, \lambda) \mid B_U))$$

converges to a Gaussian process when n tends to infinity.

For any q, define:

$$\hat{\lambda}_n \in \arg\min_{\lambda} \hat{\delta}_n^*((q, \lambda) \mid B_U)$$

as was developed very similarly in Section 4.2. Define:

$$\hat{\delta}_n^*(q \mid B) = \hat{\delta}_n^*((q, \hat{\lambda}_n) \mid B_U) = \min_{\lambda} \hat{\delta}_n^*((q, \lambda) \mid B_U).$$

The same kind of proof than in Sections 4.1 and 4.2 then applies. As the estimation of $\hat{\lambda}_n$ is superconsistent, it does not affect consistency and asymptotic normality of the support function estimates.

Proposition 17 *Under the respective conditions C and AN we have:*

$$\hat{\delta}_n^*(q \mid B) \overset{a.s.u.}{\rightarrow} \delta_n^*(q \mid B),$$

and:

$$\tau_n(q) = \sqrt{n}(\hat{\delta}_n^*(q \mid B) - \delta_n^*(q \mid B))$$

converges to a Gaussian process when n tends to infinity.

5 Monte-Carlo Simulations of Testing Procedures

In this section, we provide examples where we can explicitly construct set B in order to assess the performance of our inference and test procedures. We consider three cases where the dimension of the parameters to be estimated is equal to 2. The simulations are performed in the set-up of Example 1 (Section 2) where the dependent variable is bounded and censored by intervals. In the first two series of simulations, set B is smooth and strictly convex in the sense that its frontier has no kinks and no faces. The two series differ from the number of instruments used for estimating the set. The first series consider the same number of instruments than the dimension of the parameters whereas the second one considers the case of one supernumerary instrument. In the third series of simulations, set B has kinks and faces.

5.1 Smooth and Strictly Convex Sets

5.1.1 The Model

The simulated model is:

$$y = 0.x_1 + 0.x_2 + \varepsilon$$

The true value of β is therefore $(0,0)^T$.

In the simulations, $x^T = (x_1, x_2)^T$ is a standard normal vector where as ε is uniformly distributed on $[0,1]$ and is independent from x. We assume that, instead of observing the exact value y, we only observe it by interval (K intervals $I_k = [k/K; (k+1)/K]$, $k = 0...K-1$). $\Delta = \frac{1}{2K}$ is half the length of every interval.

Using the same notations as previously, the support function is (see Appendix D.1 for the expressions of the intermediate variables):

$$\delta^*(q \mid B) = \frac{2\Delta}{\sqrt{2\pi}}$$

B is a circle whose radius is equal to $\frac{2\Delta}{\sqrt{2\pi}}$ (see Figure D.3). When K tends to infinity, Δ tends to 0 and the identified set shrinks to a singleton.

When $q = (\cos\theta, \sin\theta)^T$,

$$\beta_q = \Sigma^T E(z^T w_q) = \frac{2\Delta}{\sqrt{2\pi}} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

Figure D.3 displays set B for $K = 2$. We draw 1000 simulations of the model for 4 sample sizes : n =100, 500, 1000 and 2500. In these simulations, the number K of interval is equal to

2. In practice, the results are robust to the choice of the number of intervals. For each draw, we estimate the set B and implement the three tests defined in Section 4.2 for various values of β_0.

5.1.2 Set Estimation

Figure 5 displays set B as well as the average and quartiles of the set estimates for the four sample sizes. For a given direction, q, we construct the mean of the estimated support function (dashed line) as well as its first and third quartiles in direction q. Including the true value, four curves are therefore displayed. We could notice that even for small sample size the set is very well estimated. Of course the interquartile range decreases when the sample size increases.

As Proposition 14 shows, $\tau_n(q)$ is asymptotically normally distributed. Table 7 displays a Normality test (see Bontemps and Meddahi (2005)) on $\tau_n(q_0)$ when $q_0 = (1, 0)$, for the four sample sizes. These tests are based on the third and fourth Hermite polynomials.

Except for the sample size equal to 100, the normal approximation is not rejected at the 5% level.

5.1.3 Testing Procedures

We also study the performance of the three test procedures developed in Section 4.2.

Test 1 examines whether β_0 is inside set B or not, Test 2 whether it is outside or not, Test 3 whether β_0 is on the frontier of B or not. For each test, we report the percentage of rejections at a 5% level test.

Observe that the set B is isotropic at the true value β. We therefore present the results for points being on the same radial. β^a is a point on a radial line whose distance from the origin of the circle is equal to a times the value of the radius of the set B. With these notations, β^0 is the center of the circle, β^a is inside B for $a \leq 1$ and outside in the other case. β^1 is on the frontier.

Table 1 displays the percentage of rejections. We emphasize in bold the Figures corresponding to the point on the frontier of B. The size of the three tests is very accurate and very close to 5% even for $n = 100$. Of course, the power increases with the sample size. Nevertheless, observe that, even for a sample size equal to 500, the power properties are very good.

5.2 Smooth set with one supernumerary instrument

The simulated model is the same than the example before. However we assume that the second explanatory variable x_2 is expressed as:

$$x_2 = \lambda e_2 + \sqrt{1-\lambda^2} e_3$$

where (e_2, e_3) are i.i.d. standard normal variables. Moreover let $w = \nu e_3 + \sqrt{1-\nu^2} e_4$ be a supernumerary instrument (e_4 being i.i.d. standard normal). λ and ν measure the strength of the correlation between e_2 and w with respect to x_2. In this simulation series, we set $\lambda = \nu = 0.6$. When $q = (\cos\theta, \sin\theta)^T$, the support function (see Appendic D.2 for further details) can be expressed as:

$$\delta^*(q \mid B) = \frac{2\Delta}{\sqrt{2\pi}} \sqrt{\cos^2\theta + \frac{\sin^2\theta}{\lambda^2 + \nu^2(1-\lambda^2)}}$$

Figure D.3 displays the set B. Observe that when $\nu = 1$, the set B is the same than in the previous example due to the fact that the knowledge of e_2 and w gives all the information on x_2. Moreover when ν and λ are strictly positive but stricly lower than 1, there is some information loss due to the use of e_2 and w instead of x_2. The set B is therefore deformed along the second axis.

Like before, we draw 1000 simulations of the model for 4 sample sizes : n =100, 500, 1000 and 2500. For each draw, we estimate the set B and implement the three tests defined in Section 4.2 for various values of β_0.

Figure 6 displays set B as well as the average and quartiles of the estimated sets for the four sample sizes. Tables **??** displays the percentage of rejections for the three tests for different points along a radius. The results are insensitive to the choice of the radius. We only present the results along the x_1 axis.

As before, the estimation and test procedures work well and there is no significative distorsion while using supernumerary instruments in the estimation and test procedures.

5.3 Unsmoothed and Flat B

In this example, x has mass points (so that set B has kinks) and the support of x is not an interval (so that set B has faces). The simulated model is:

$$y = \frac{1}{2} + \frac{x}{4} + \varepsilon$$

The true value of β is therefore $(\frac{1}{2}, \frac{1}{4})^T$.

In the simulations, x is equal to $-\frac{1}{2}$ with probability $\frac{1}{2}$ and to $\frac{1}{2}$ with probability $\frac{1}{2}$. ε is uniformly distributed on $[-\frac{1}{4}, \frac{1}{4}]$ and is independent from x. We assume that, instead of observing the exact value y, we only observe it by interval (2 intervals: $I_1 = [0, \frac{1}{2}]$ and $I_2 = [\frac{1}{2}, 1]$).

After standard computations, set B is the convex envelop of the four points $(\frac{3}{4}, \frac{1}{4})$, $(\frac{1}{2}, \frac{3}{4})$, $(\frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, -\frac{1}{4})$ (see Figure 3 and Appendix D.3 for the expressions of the intermediate variables).

As in the previous example, we simulate 1000 draws for 4 sample sizes: 100, 500, 1000 and 2500.

Figure 7 displays set B as well as the average and quartiles of the estimated sets for the four sample sizes. Tables 2 and 3 display the percentage of rejections for the three tests for different points (shown in bold in Figure 3). As in the previous example, the estimation and test procedures work well even for small sample sizes.

6 Empirical Illustrations

This section provides a description of empirical illustrations.

6.1 Income band data

In the French Labor Force Survey, interviewees are offered the choice of responding to questions of monthly income either by reporting their exact values or by bracketing them between two bounds. We thus have two samples, one of *exact responses* for individuals reporting the exact amount of their earnings, the other of *interval responses* for individuals reporting earnings by bands. Out of the 25000 observations of males between 25 and 55 in the issue of 1999, we found that around 2000 chose to follow the second route. One question of interest is the selection of such samples. Let β_e the parameter in the sub-population corresponding to the first sample and B_i the identified set in the interval response sample. The null hypothesis that we have just described is $\beta_e \in B_i$ or more exactly $0 \in B_i - \{\beta_e\}$.

The equation of interest is expressed in logs:

$$\log R_i = X_i\beta + \varepsilon_i$$

where R_i is monthly earnings and X_i includes education, age and age squared.

There are ten bounds and because the dependent variable is in logs, these bounds are both open on the left and on the right. In order to conform with the setting of interval valued dependent variables, as developed in Example 1 and detailed in Appendix A, we chose to fix the lower and upper bound to the arbitrary values of 200FF (30€) and 70000FF (10500€) by looking at the distribution of incomes that are declared in exact amounts.

In Table 4, we report estimated coefficients in three cases. The first column reports OLS coefficients which are obtained using the sample of exact responses. The second column reports OLS coefficients using the second sample of interval responses by considering (incorrectly) that the dependent variable is equal to the mid-value of the reported band. Finally, the third column reports the estimated intervals for each coefficient and the confidence intervals corresponding to the parameters of interest (not of the intervals of interest, see Imbens and Manski, 2004). We also plotted the estimated set concerning the two coefficients of age and education in Figure 8.

Several results emerge from this table. As expected in an income equation, the confidence intervals are very small in the exact response sample. It is also the case when using the interval response sample and the length of the confidence intervals relative to those in column 1, is approximately in a ratio of $\sqrt{n_e/n_i}$ as expected. The results of the procedure described in this text is thus striking. The length of the confidence interval increases by a factor of 10 in case of the coefficient of age and a bit less for the coefficient of education. As returns to education are much more precisely estimated, they are still significantly positive in a large range though of 4.5% and 12%. Any significance of the age coefficient is utterly lost.

We did not perform formally the test that the two samples are the same with respect to the income projection on explanatory variables because the result is obvious. It is impossible to reject such an hypothesis considering the large errors due to interval reporting.

In order to illustrate the use of supernumerary moment conditions and instruments, we also estimated the following earnings equation:

$$\log R_i = X_i\beta + \gamma \log H_i + \varepsilon_i$$

where H_i is the number of hours worked. We assume that hours of work are endogenous and we use as instruments the income declared by their spouse (if any).

Results are reported in Table 5. The same picture than previously emerges. It is in particular impossible to reject the validity of the overidentifying restrictions in the interval data case even if in the complete sample they are rejected very strongly indeed.

This is why it is interesting to use other supernumerary moment conditions in order to shrink the size of the estimated intervals.

6.2 Artificial Data: Mean Independence

There is a particularly simple way of adding instruments to a model which is to use an assumption of mean independence instead of the assumption of non correlation.

We consider one explanatory variable x and one random term ε which are truncated to lie into $[0, 1]$ and which are uniform, normally or mixed normally distributed. We have:

$$y^* = \beta_0 + \beta_1 x + \varepsilon$$

The number of categories for y is equal to 10 and we assume that they are equally spaced i.e.

$$y = 1/20, 3/20, ..., 19/20$$

The number of observations varies between 100, 1000 and 10000. We use the fact that $E(\varepsilon \mid x) = 0$ to write that for any function $h(x)$, $E(h(x)\varepsilon) = 0$. Those are the supernumerary moment conditions.

The functions $h(x)$ that we consider are either polynomials of increasing degree or sinusoids of equally spaced frequencies.

Results are reported in Table 6. We see that the gain in terms of the length of estimated intervals or confidence intervals, is strong using two additional restrictions at most in the case of polynomials and a bit more in the case of sinusoids.

7 Extension: Two-sample models

The approach developped in this paper can be easily extended to other setting where the identified set is non empty and convex. For example, let us assume that a set of L explanatory variables (say v) are not observed in the database that contains information on y. Specifically, the data are given by $w = (y, x, v)$ but realizations of $w = (y, x)$ only are observed in one sample and realizations of $w = (v, x)$ only are observed in the other sample. It is one problem known as ecological inference (Manski, 2004 and Moffitt & Ridder, 2003). Economists are very often confronted to the case where there is no single database that contains all relevant variables. Administrative databases typically contain a very limited set of variables (i.e., only those that are relevant for the administration). Also it is often the case that survey data do not contain information on the exact date of birth or the exact income of respondents, for confidentiality reasons. Within this framework, we consider the familiar linear model,

$$(y = x\beta + v\alpha + \epsilon)$$

where ϵ is uncorrelated with the set of $(K + L)$ regressors (x, v). The issue is to characterize B_3, the set of observationally equivalent β. The existence of two samples is a necessary condition for obtaining bounds. Without information about v, little can be said on B_3.

As we are analyzing linear projections, the only useful information about v comes from linear projections. Denote ρ, the matrix of dimension $[K, L]$ of the regression coefficients of v on x that one can derive from using the second (infinite) sample. Denote Ω the variance-covariance of the residuals of these regressions of v on x. We can thus write:

$$v = x\rho + \eta\Omega^{1/2}$$

where η is a random vector of row dimension L such that $E(\eta^T x) = 0, E(\eta^T\eta) = I_L$. This random variable describes what is ignored on v when using the first sample.

Using these notations, the original model can be rewritten as:

$$y = x\beta + x\rho\alpha + \eta\Omega^{1/2}\alpha + \varepsilon$$

and a parameter β belongs to the identification set B_3 if and only if there is an α and a η such that,

$$E(x^T(y - x\beta)) = E(x^T(x\rho\alpha) \text{ and } E(\eta^T y) = \Omega^{1/2}\alpha. \tag{20}$$

Denoting $y_x = y - xE(x^T x)E(x^T y)$ the residual of the regression of y on x, we have $E(\eta^T y) = E(\eta^T y_x)$ and $\alpha^* = \frac{\Omega^{1/2}.\alpha}{(Vy_x)^{1/2}}$satisfies necessarily $\|\alpha^*\| \leq 1$. Hence a parameter β belongs to the identification set B_3 if and only if there is an α^* in the compact ball $\|\alpha^*\| \leq 1$ such that,

$$\beta = \beta^* + (Vy_x)^{1/2}\,\rho\Omega^{-1/2}\alpha^*, \tag{21}$$

where β^*= B_3 is clearly non empty ($\beta^* \in B_3$) and convex. It is completly characterized by its support function,

$$\delta^*(q \mid B_3) = q^T\beta^* + (Vy_x)^{1/2} \sup_{\|\alpha^*\|\leq 1} (q^T\rho\Omega^{-1/2}\alpha^*)$$

It is not difficult to show that the RHS supremum is attained when $\alpha^* = \frac{\Omega^{-1/2}\rho^T q}{\sqrt{q^T\rho\Omega\rho^T q}}$.

Proposition 18 *The identified set B_3 is non empty, closed, convex and bounded. It is characterised by the following support function,*

$$\delta^*(q \mid B_3) = q^T\beta^* + (Vy_x)^{1/2}\sqrt{q^T\rho\Omega\rho^T q}.$$

The function $\delta^*(q \mid B_3)$ is continuously differentiable on the unit sphere and it is possible to develop an estimation procedure following the same lines as in the previous sections.

8 Conclusion

We developed in this paper simple ways to identify and estimate sets in the linear model using linear restrictions. We generalize previous results when estimating linear predictions. We provide asymptotic tests for null hypotheses concerning a single point which are easy to implement. We also generalize this simple seeting of linear prediction to the egenral case with supernumerary meoment conditions. Examples 1 and 4 can be treated using this framework. Examples 2 and 3 needs more care for instance because the identified set is the union of the sets which we showed are identified in the framework.

Various questions are open and seem worth pursuing. First, the question of the optimality of inference in the supernumerary restriction case is worth posing. Second, the gain of the direct approach that we used with respect to the approach followed by Chernozhukov et al. (2006) is an interesting question. In particular, it might be the case that our results help selecting the best criterion in the latter framework.

REFERENCES

Andrews, D.W.K., S., Berry and P. Jia, 2004, ”Confidence Regions for Parameters in Discrete Games with Multiple Equilibria with an Application to Discount Chain Store Location”, working paper.

Beresteanu, A. and F., Molinari, 2006, ”Asymptotic Properties for a Class of Partially Identified Models”, working paper.

Blundell, R.W., A. Gosling, H. Ichimura and C. Meghir, 2007, ”Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds”, *Econometrica,* 75:323-363.

Bollinger, C.R., 1996, ”Bounding mean regressions when a binary regressor is mismeasured”, *Journal of Econometrics*, 73:387-399.

Bontemps, C., and N., Meddahi, 2005, ”Testing Normality: a GMM Approach”, *Journal of Econometrics*, 124:149-186

Chernozhukov, V., H. Hong, E. Tamer, 2007, ”Inference on Parameter Sets in Econometric Models”, *Econometrica,* 75:1243-1284.

Chesher, A., 2003, ”Non Parametric Identification under Discrete Variation”, *Econometrica,* 73:1525-1550.

Ciliberto F. and E. Tamer, 2004, ”Market structure and multiple equilibria in airline markets”, WP.

Corneo G. and H.P. Grüner, 2000, ”Social Limits to Redistribution”, *American Economic Review*, 90:1491-1507.

Davidson, R., and J.G., McKinnon, 2004, *Econometric Theory and Methods,* Oxford UP: Oxford.

Di Tella, R. and R. MacCulloch, 2006, ”Some Uses of Happiness Data in Economics”, *Journal of Economic Perspectives*, 20:25-46.

Erickson, T., 1993, ”Restricting Regression Slopes in the Errors in Variables Model by Bounding the Error Correlation”, *Econometrica,* 61:959-69.

Fréchet, **M.,** 1951, ”Sur les tableaux de corrélation dont les marges sont données”, *Annales de l'Université de Lyon*, III° Série Sci. A, 53-77.

Frisch, R., 1934, *Statistical Confluence Analysis by Means of Complete Regression Systems,* Oslo, Norway: University Institute of Economics.

Galichon A., and M., Henry, 2006, ”Inference in Incomplete Models”, Unpublished Manuscript.

Gini, C., 1921, ”Sull'interpoliazone di una retta quando i valori della variabile indipendente sono affetti da errori accidentali”, *Metroeconomica,* 1:63-82.

Haile P. and E.Tamer, 2003, ”Inference with an Incomplete Model of English Auctions”, *Journal of Political Economy*, 2003, 111:1-51.

Hoeffding, W., 1940, ”Masstabinvariante Korrelationstheorie”, *Shriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin,* 5(3): 179-233.

Honoré, B. and A. Lleras-Muney, 2004, ”Bounds in Competing Risks Models and the War on Cancer”, NBER WP 10963, Cambridge.

Horowitz, J.L. and C.F. Manski, 1995, ”Identification and Robustness with Contaminated and Corrupted Data”, *Econometrica*, 63:281-302.

Imbens, G., and C.F., Manski, 2004, “Confidence Intervals for Partially Identified Parameters”, *Econometrica,* 72:1845-1859.

Klepper S., and E.E., Leamer, 1984, ”Consistent Sets of Estimates for Regressions with Errors in all Variables”, *Econometrica,* 52:163-183.

Leamer, E.E., 1987, ”Errors in Variables in Linear Systems”, *Econometrica,* 55(4): 893-909.

Lewbel, A., 2000, “Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables”, *Journal of Econometrics*, 97:145-77.

Lewbel, A., O., Linton and D., McFadden, 2006, ”Estimating Features of a Distribution From Binomial Data”, unpublished manuscript.

McFadden, D., 1994, **XXX**

Magnac, T. and E. Maurin, 2007a, ”Identification and Information in Monotone Binary Models”, forthcoming *Journal of Econometrics*.

Magnac, T. and E. Maurin, 2007b, ”Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data”, WP IDEI.

Manski, C.F., 1989, ”Anatomy of the Selection Problem”, *Journal of Human Resources,* 24:343-60.

Manski, C.F., 2003, *Partial Identification of Probability Distributions*, Springer-Verlag: Berlin.

Manski, C.F., 2006, ”Partial Identification in Econometrics”, *New Palgrave Dictionary of Economics*, 2nd edition, London:McMillan.

Manski, C.F., and E. Tamer, 2002, “Inference on Regressions with Interval Data on a Regressor or Outcome”, *Econometrica,* 70:519-546.

Moffitt, R. and G. E. Ridder, (2006), ”The econometrics of data combination”, in J. J.Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, North-Holland, Amsterdam.

Molchanov, **I.S.,** 2005, *Theory of Random Sets,* Springer Verlag: London.

Pakes, A., J. Porter, K. Ho and J., Ishii, 2005, ”Moment Inequalities and their Applications”, working paper.

Piketty, T., (2005), ”Top Income Share in the Long Run: An Overview”, *Journal of the European Economic Association*, 3:1-11.

Rockafellar, R.T., 1970, *Convex Analysis,* Princeton University Press: Princeton.

Shaikh, A.M., 2005, ”Inference for partially Identified Econometric Models”, working paper.

van der Vaart, A.W., 1998, *Asymptotic Statistics,* Cambridge University Press: Cambridge.

Vazquez-Alvarez, R., B., Melenberg and A., van Soest, 2001, ”Nonparametric bounds in the presence of item nonresponse, unfolding brackets, and anchoring”, working paper.

White, **H.,** 1999, *Asymptotic Theory for Econometricians*, Academic Press: San Diego.

Appendices

Appendices

A Proofs in Section 2

A.1 Proof of Proposition 1

(Necessity) Consider β in $\mathbb{R}^K$ and assume that $\beta \in B$ so that there is a latent random variable ε which is uncorrelated with x such that $(\beta, F_\varepsilon(. \mid x))$ generates $\{G_k(x)\}_{k=1,.,K}$ through model (4). By definition, the distribution of ε satisfies,

$$\forall k; G_k(x) = \Pr(y^* \in [y_0, y_k)) = \int_{y_0 - x\beta}^{y_k - x\beta} f_\varepsilon(\varepsilon \mid z) d\varepsilon = F_\epsilon(y_k - x\beta \mid x). \tag{A.1}$$

Also, if $\varepsilon \equiv y^* - x\beta$ is uncorrelated with x, we have necessarily,

$$E(x^T(x\beta - y)) = E(x^T(y^* - y)) = E(x^T E(y^* - y \mid x))$$

By construction, $u(x) = E(y^* - y \mid x)$ is a measurable function which can be rewritten,

$$u(x) = \sum_{k=0}^{K-1} E(y^* - y \mid x, y^* \in [y_k, y_{k+1})). \Pr(y^* \in [y_k, y_{k+1})).$$

Using Equation (3), we easily obtain bounds:

$$y_k - \frac{y_k + y_{k+1}}{2} \leq E(y^* - y \mid x, y^* \in [y_k, y_{k+1})) < y_{k+1} - \frac{y_k + y_{k+1}}{2}$$

which yields bounds on $u(x)$,

$$\sum_{k=0}^{K-1} (\frac{y_k - y_{k+1}}{2})(G_{k+1}(x) - G_k(x)) \leq u(x) < \sum_{k=0}^{K-1} (G_{k+1}(x) - G_k(x))(\frac{y_{k+1} - y_k}{2})$$

By considering the limits of any converging sequence $\beta_n \in B$, we obtain any point of the closure of B by replacing the strict inequality for $u(x)$ in,to a loose inequality. Hence, if $\beta \in cl(B)$, there exists a $u(x) \in I(x)$ such that $E(x^T(x\beta - y)) = E(x^T u(x))$, meaning that (i) implies (ii).

(Sufficiency) Conversely, let us prove that statement (ii) implies statement (i). We first assume that there exists $u(x)$ in $[-\Delta(x), \Delta(x)) \subset I(x)$ such that statement (ii) holds true and we construct a distribution function $F_\varepsilon(. \mid x)$ such that ϵ is uncorrelated with x and such that the image of $(\beta, F_\varepsilon(. \mid x))$ through model (4) is $\{G_k(x)\}_{k=0,.,K-1}$.

First, let us consider λ a random variable whose support is $[0, 1)$, whose conditional density given x is:

$$E(\lambda \mid x) = (u(x) + \Delta(x))/(2\Delta(x)).$$

Second, let κ a discrete random variable whose support is $\{0,.,K-1\}$ and whose conditional distribution given x is :

$$Pr(\kappa = k \mid x) = G_{k+1}(x) - G_k(x).$$

For any $k \in \{0,.,K-1\}$, consider K random variables, say $\epsilon(\lambda, k)$ which are constructed from λ by:

$$\epsilon(\lambda, k) = -x\beta + (1-\lambda)y_k + \lambda y_{k+1}$$

Given that $\lambda \in [0,1)$, the support of $\epsilon(\lambda, k)$ is $[y_k - x\beta, y_{k+1} - x\beta)$. Finally, consider the random variable:

$$\varepsilon = \epsilon(\lambda, \kappa) \tag{A.2}$$

which support is $[y_0 - x\beta, y_K - x\beta)$. Because of Equation (A.1), the image of $(\beta, F_\varepsilon(. \mid x))$ through model (4) is $\{G_k(x)\}_{k=0,\ldots,K-1}$. The last condition to prove is that ε is uncorrelated with x. Consider, for almost any x,

$$\begin{aligned}
E(y^* \mid x) - E(x\beta + \varepsilon \mid x) = \sum_{k=0}^{K-1} [\, (\frac{y_{k+1} + y_k}{2}) \qquad & (G_{k+1}(x) - G_k(x)) \\
- \int_{y_k - x\beta}^{y_{k+1} - x\beta} & E(x\beta + \varepsilon \quad \mid \quad x, \kappa = k)) f(\varepsilon \mid x, \kappa = k) d\varepsilon . \Pr(\kappa = k \mid x)] \\
= \sum_{k=0}^{K-1} (\frac{y_{k+1} + y_k}{2} & - E((1-\lambda)y_k + \lambda y_{k+1}) \mid x)(G_{k+1}(x) - G_k(x)) \\
= \sum_{k=0}^{K-1} & E(1/2 - \lambda \mid x).(y_{k+1} - y_k).(G_{k+1}(x) - G_k(x)) \\
& = (-u(x)/(2\Delta(x)))(2\Delta(x)) = -u(x).
\end{aligned}$$

Therefore, we have $E(y - x\beta \mid x) = -u(x) + E(\varepsilon \mid x)$, which implies:

$$E(x^T y) = E(x^T x)\beta + E(x^T \varepsilon) - E(x^T u(x)).$$

Given (ii), this equation implies that $E(x^T \varepsilon) = 0$.

To finish the proof, it suffices to consider sequences $u_n(x)$ converging to $u(x) \in I(x)$ (in the L^2 norm for instance, see below). Any sequence generates a parameter $\beta_n \in B$ satisfying the moment condition and which converges to $\beta \in cl(B)$ because the mapping from $u_n()$ to β_n is continuous.

A.2 The mid-interval normalization in the interval outcome case

We start from:

$$E(x^T(y^* - x\beta)) = 0,$$

and the definition:

$$y = \sum_{k=0}^{K-1} v_k.\mathbf{1}\{y^* \in [y_k, y_{k+1})\}.$$

where, to study the general case, we consider an arbitrary set of values $\{v_0, ., v_{K-1}\}$. Thus:

$$y^* - y = \sum_{k=0}^{K-1} (y^* - v_k).\mathbf{1}\{y^* \in [y_k, y_{k+1})\} \in [\underline{\Gamma}, \overline{\Gamma}]$$

where

$$\begin{aligned}
\overline{\Gamma} &= \sum_{k=0}^{K-1} (y_{k+1} - v_k).\mathbf{1}\{y^* \in [y_k, y_{k+1})\}, \\
\underline{\Gamma} &= \sum_{k=0}^{K-1} (y_k - v_k).\mathbf{1}\{y^* \in [y_k, y_{k+1})\}.
\end{aligned}$$

Then:

$$E(x^T(x\beta - y)) = E(x^T(y^* - y)) = E(x^T E((y^* - y) \mid x))$$

where:

$$E((y^* - y) \mid x) = u(x) \in [E(\underline{\Gamma} \mid x), E(\overline{\Gamma} \mid x)]$$

and where $E(\underline{\Gamma} \mid x) = \underline{\Delta}(x)$ for instance.

Thus:

$$\begin{aligned}
\overline{\Delta}(x) &= \sum_{k=0}^{K-1} (y_{k+1} - v_k).(G_{k+1}(x) - G_k(x)), \\
\underline{\Delta}(x) &= \sum_{k=0}^{K-1} (y_k - v_k).(G_{k+1}(x) - G_k(x)),
\end{aligned}$$

where:

$$G_k(x) = \Pr(y < k \mid x).$$

Remark that the length of this interval does not depend on $\{v_k\}_k$ since it is equal to:

$$\sum_{k=0}^{K-1} (y_{k+1} - y_k).(G_{k+1}(z) - G_k(z)) = 2\Delta(z) \text{ (say)},$$

so that the choice of the sequence $\{v_k\}_k$ is arbitrary. The most convenient choice is $v_k = \frac{y_k + y_{k+1}}{2}$ since it implies that $-\underline{\Delta}(z) = \bar{\Delta}(z) = \Delta(z)$.

A.3 The existence of two variables $\bar{y}$ and $\underline{y}$.

Examples 1 and 2: Using the two previous subsections, it is immediate that

$$\bar{y} = \sum_{k=0}^{K-1} y_{k+1}\mathbf{1}\{y^* \in [y_k, y_{k+1})\}, \underline{y} = \sum_{k=0}^{K-1} y_k\mathbf{1}\{y^* \in [y_k, y_{k+1})\},$$

verify the conditions. For instance,

$$E(\bar{y}-y \mid x)=\sum_{k=0}^{K-1}(y_{k+1}-\frac{y_k+y_{k+1}}{2})E(\mathbf{1}\{y^*\in[y_k,y_{k+1})\}\mid x)=\Delta(x).$$

Example 3: The supremum of the function $(Vy_x)^{1/2}x\rho\Omega^{-1/2}\alpha^*$ when $\|\alpha^*\|\leq 1$ is obtained when:

$$\alpha^*=(x\rho\Omega^{-1/2})^T/\left\|x\rho\Omega^{-1/2}\right\|,$$

so that:

$$\Delta(x)=(Vy_x)^{1/2}\left\|x\rho\Omega^{-1/2}\right\|.$$

We would have to construct $\bar{y}=y+\Delta(x)$ and this is less easy.

Example 4: See Magnac and Maurin (2007b). It is very similar to Examples 1 and 2 since it consists in setting y^* to its larger or lower possible value in the interval.

B Proofs in Section 3

B.1 Proof of Proposition 4

First, B contains β^* because $u(z)=0$ belongs to the admissible set $I(z)$. Second, B is closed and convex because $I(z)$ is closed and convex and equation (7) is linear. Furthermore, as (7) can be written as:

$$E(z^Tx)(\beta-\beta^*)=E(z^Tu(z))$$

and using the definition of W, we have:

$$(\beta-\beta^*)^TW(\beta-\beta^*)=E(u^T(z)z)E(z^Tz)^{-1}E(z^Tu(z)).$$

Using the generalized Cauchy-Schwartz inequality,

$$E(u^T(z)z)E(z^Tz)^{-1}E(z^Tu(z))\leq E(u(z)^2).$$

By the definition of the admissible set,

$$E(u(z)^2)\leq E(\max(\overline{\Delta}(z)^2,\underline{\Delta}(z)^2))=\Delta_M$$

which is bounded by Assumption *R.iii* since $E(\bar{y}-y\mid z)=\bar{\Delta}(z)$.

B.2 Proof of Proposition 5

For the sake of clarity, we start with the simple case where $z=x$. Then we study the general case and finish by the proof of the alternative characterization of the support function.

Simple case: $z = x$. Let q any vector of dimension p of Euclidean norm $\| q \| = 1$. Consider Q an orthogonal matrix of dimension p which is such that $Q = (Q_0, q)$ where Q_0 a matrix of dimension $[p, (p-1)]$. We have:

$$x\beta = xQQ^T\beta = (xQ_0, xq) \begin{pmatrix} Q_0^T\beta \\ q^T\beta \end{pmatrix}$$

The p-th component of the parameter of interest, $q^T.\beta$, is associated to the p-th explanatory variable, $x.q$. The support function is the supremum of $q^T.\beta$ when $\beta \in B$. As it is the coefficient of one explanatory variable only, $x.q$, the natural tool is the Frisch-Waugh theorem (Davidson and McKinnon, 2004, for instance). It states that in the regression where x_0 is a regressor whereas all other regressors are x_{-0}, the coefficient of x_0 can be obtained in the simple linear regression on the residual of the projection of x_0 onto x_{-0} defined as:

$$x_0 - x_{-0}\left(E(x_{-0}^T x_{-0})\right)^{-1} E(x_{-0}^T x_0) \tag{B.3}$$

In our case, denote x_q the residual of the projection of $x_0 = x.q$ onto $x_{-0} = x.Q_0$ (the other regressors). Replace in equation (B.3) to obtain the real random variable:

$$\begin{aligned} x_q &= x.q - x.Q_0\left(Q_0^T E(x^T x) Q_0\right)^{-1} Q_0^T E(x^T x) q \\ &= x(q - Q_0\left(Q_0^T E(x^T x) Q_0\right)^{-1} Q_0^T E(x^T x) q) \\ &= x.a(q) \end{aligned} \tag{B.4}$$

Note that the definition is independent of Q_0 since we can replace Q_0 by any matrix $Q_0 M$ where M is full rank $(p-1)$.

Characterization (9) implies that for any $\beta \in B$, there exists a function $u(z)$ such that the regression of $y + u(z)$ on x yields parameter β. Applying the Frisch-Waugh theorem implies that:

$$q^T\beta = \left(E((x_q)^2)\right)^{-1} E(x_q(y + u(z)) = E(z_q(y + u(z)). \tag{B.5}$$

where $z_q = \frac{x_q}{E(x_q{}^2)}$. The support function in the direction q is obtained by looking for the supremum of this expression when $u(z)$ varies in $[\underline{\Delta}(z), \overline{\Delta}(z)]$. The supremum of the scalar $E(z_q u(z))$ is obtained by setting $u(z)$ to its maximum (resp. minimum) value when z_q is positive (resp. negative) because $0 \in]\underline{\Delta}(z), \overline{\Delta}(z)[$ and by setting $u(z)$ to any value when z_q is equal to 0. It yields a set of "supremum" functions:

$$u_q(z) = \overline{\Delta}(z)\mathbf{1}\{z_q > 0\} + \underline{\Delta}(z)\mathbf{1}\{z_q < 0\} + \mathbf{\Delta}^*(z)\mathbf{1}\{z_q = 0\} \tag{B.6}$$

where $\mathbf{\Delta}^*(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. Note that $u_q(z)$ is unique (a.e. P_z) if $\Pr(z_q = 0) = 0$. From now on, the uniqueness of $u_q(z)$ should always be understood as "almost everywhere P_z".

Recall that by equation (8), $E(\overline{y} - y \mid z) = \overline{\Delta}(z)$, $E(\underline{y} - y \mid z) = \underline{\Delta}(z)$, so that the support function or the supremum of (B.5) is equal to:

$$\delta^*(q \mid B) = E(z_q w_q),$$

where:

$$w_q = \underline{y}\mathbf{1}\{z_q > 0\} + \overline{y}\mathbf{1}\{z_q < 0\}.$$

Note that the term $\mathbf{\Delta}^*(z)$ in $u_q(z)$ disappears because it is multiplied within the second expectation by z_q which is equal to 0 at these values. It implies, as expected, that $\delta^*(q \mid B)$ is unique even though $u_q(z)$ is not.

Furthermore, when $\mathbf{\Delta}^*(z)$ varies in $[\underline{\Delta}(z), \overline{\Delta}(z)]$,the functions $u_q(z)$ defined by equation (B.6) generate all points $\beta \in B$ which belong to the tangent space to B that is orthogonal to q (a face of B see below):

$$\left(E(z^T x)\right)^{-1} E(z^T(y + u_q(z)). \tag{B.7}$$

If we select the particular value of $u_q(z)$ such that $\mathbf{\Delta}^*(z) = 0$, we get the particular value of β:

$$\beta_q = \left(E(z^T x)\right)^{-1} E(z^T w_q),$$

and, by definition:

$$\delta^*(q \mid B) = q^T \beta_q.$$

Further developments along these lines will be undertaken in Section B.4.

IV case This case is a simple adaptation of the previous one. It requires to replace (B.3) by the generalized transformation using an IV projection:

$$\begin{aligned} x_0 &= z(E(z^T z))^{-1} E(z^T x_0) - \\ &\quad z(E(z^T z))^{-1} E(z^T x_{-0}) \left[E(x_{-0}^T z)(E(z^T z))^{-1} E(z^T x_{-0})\right]^{-1} E(x_{-0}^T z)(E(z^T z))^{-1} E(z^T x_0). \end{aligned}$$

In that case, equation (B.4) is transformed into:

$$\begin{aligned} x_q &= z(E(z^T z))^{-1} E(z^T x).a(q) \\ &= zD.a(q) \end{aligned} \tag{B.8}$$

where $D = (E(z^T z))^{-1} E(z^T x)$. Note that when $z = x$, $D = I$ so that expression (B.4) is a particular case.

The definition of $a(q)$ is now:

$$\begin{aligned} a(q) &= q - Q_0 \left(Q_0^T E(x^T z) E(z^T z)^{-1} E(z^T x) Q_0\right)^{-1} Q_0^T E(x^T z) E(z^T z)^{-1} E(z^T x) q \\ &= q - Q_0 \left(Q_0^T R Q_0\right)^{-1} Q_0^T R q, \end{aligned} \tag{B.9}$$

setting $R = E(x^T z) E(z^T z)^{-1} E(z^T x)$. Note that this is also what we find in equation (B.4) when $z = x$. Anyway, by Assumption R.ii and the equal dimension of z and x, we have $rank(R) = rank(D) = p$. Furthermore, R is a positive definite matrix.

Given this definition of x_q, we can adapt without change the proof used in the previous subsection treating the case $z = x$.

The last property to be proved is that z_q can be written $zE(x^T z)^{-1} q$. The following Lemma leads to the proof.

Lemma 19 *The vectorial function $a(q)$ from $\mathbb{S}$ to $\mathbb{R}^p$ as defined in (B.9), is bounded and continuously differentiable. Furthermore, $a(q) \neq 0$, and the range of $\frac{a(q)}{\|a(q)\|}$ is $\mathbb{S}$. Function $a(q)$ is invertible from $Range(\mathbb{S})$ to $\mathbb{S}$ and:*

$$q = \frac{R.a(q)}{a(q)^T R.a(q)}.$$

Proof. The definition of Q_0 implies that it is a continuously differentiable mapping of q. As seen above, R is full rank and the inverse of $Q_0^T R Q_0$ exists. By the continuous differentiability of the inverse, the first part of the Lemma is proved. Second, notice that because $Q = (Q_0, q)$ is orthogonal, $q^T.a(q) = 1$, which implies that $a(q) \neq 0$ for any $q \in \mathbb{S}$ and that $a(q)$ is bounded. Third, let $s \in \mathbb{S}$ and look for $q \in \mathbb{S}$ such that $s = \frac{a(q)}{\|a(q)\|}$. Note that $Q_0^T.R.a(q) = 0$ so that $R.a(q)$ lies in the null space of Q_0 which is composed by all vectors colinear to q which means that:

$$R.a(q) = \lambda.q.$$

The scalar λ can be obtained by premultiplying by $a(q)^T$:

$$a(q)^T R.a(q) = \lambda a(q)^T q = \lambda$$

because $q^T.a(q) = 1$. ■

As $x_q = zDa(q)$ by Equation (B.8), we have:

$$\begin{aligned} E(x_q^2) &= a(q)^T D^T E(z^T z) D a(q) \\ &= a(q)^T R a(q), \end{aligned}$$

which is positive since R is positive definite and $a(q) \neq 0$.

We also have:

$$\begin{aligned} z_q &= \frac{x_q}{E(x_q^2)} = \frac{zDR^{-1}Ra(q)}{a(q)^T Ra(q)} \\ &= zDR^{-1}q = zE(x^T z)^{-1}q. \end{aligned}$$

using the previous Lemma and because,

$$DR^{-1} = (E(z^T z))^{-1} E(z^T x)(E(x^T z)E(z^T z)^{-1}E(z^T x))^{-1} = E(x^T z)^{-1}.$$

B.3 Proof of Lemma 6

We use the expression derived in Proposition 5:

$$\begin{aligned} \delta^*(q \mid B) = E(z_q w_q) &= E(z_q(\mathbf{1}\{z_q > 0\}\overline{y} + \mathbf{1}\{z_q < 0\}\underline{y})) \\ &= E(z_q \underline{y}) + E(z_q \mathbf{1}\{z_q > 0\}(\overline{y} - \underline{y})) \end{aligned}$$

The first term on the RHS is linear in q since:

$$z_q = z(E(x^T z))^{-1} q.$$

and thus is continuously differentiable on $\mathbb{S}$.

As $(\overline{y} - \underline{y}) > 0$, the second term can be written as:

$$\psi(q) = E(z^*.q.\mathbf{1}\{z^*.q > 0\})$$

where $z^* = z(E(x^T z))^{-1}(\overline{y} - \underline{y})$. Lemma 6 shall be proven if we prove:

Lemma 20 *$\psi(q)$ is continuously differentiable in q on the unit sphere $\mathbb{S}$ except at a countable number of points.*

Proof. Consider for any $s, t \in \mathbb{S}$:

$$\begin{aligned} \psi(s) - \psi(t) &= E((z^*.(s-t).\mathbf{1}\{z^*.s > 0\}) \\ &+ E((z^*.t.(\mathbf{1}\{z^*.s > 0\} - \mathbf{1}\{z^*.t > 0\}). \end{aligned}$$

Assumption R.iii implying that $E(|z^*|)$ is bounded (White, 1994, p32-33), we get by the triangular inequality :

$$\begin{aligned} |\psi(s) - \psi(t)| &\leq E(|z^*|\,\mathbf{1}\{x^*.s > 0\}).\,\|s-t\| \\ &\quad + E(|z^* t|\,|\mathbf{1}\{z^*.s > 0\} - \mathbf{1}\{z^*.t > 0\}|). \end{aligned}$$

Note that

$|\mathbf{1}\{z^*.s > 0\} - \mathbf{1}\{z^*.t > 0\}| = \mathbf{1}(\{z^*.s > 0\}$ and $\{z^*.t \leq 0\}) + \mathbf{1}(\{z^*.s \leq 0\}$ and $\{z^*.t > 0\})$

so that the second term can be re-written,

$E(|z^* t| \mid z^*.t > 0, z^*.s \leq 0)Pr(z^*.t > 0, z^*.s \leq 0) + E(|z^* t| \mid z^*.t \leq 0, z^*.s > 0)Pr(z^*.t \leq 0, z^*.s > 0)$

Consider first the case where $z^*.t > 0$ and $z^*.s \leq 0$. Then:

$$0 < z^*.t = z^*.(t-s) + z^* s \leq z^*.(t-s)$$

so that:

$$E(|z^* t| \mid z^*.t > 0, z^*.s \leq 0) < E(|z^*|).\,\|s-t\|\,.$$

The other case is similar so that $\psi(s)$ is Lipschitzian on $\mathbb{S}$:

$$|\psi(s) - \psi(t)| \leq E(|z^*|)\,\|s-t\|$$

and thus uniformly continuous.

Using similar arguments, it is differentiable because:

$$\lim_{t \to s} \Pr(z^*.t > 0, z^*.s \leq 0) = 0.$$

as:

$$\lim_{t \to s} \{z^* \text{ such that } ((z^*.s)(z^*.t) < 0) = \varnothing.$$

The differential at s is :

$$E((z^*.\mathbf{1}\{z^*.s > 0\})$$

and is continuous except at points such that $\Pr(z^*.s = 0) > 0$. There cannot be more than a countable number of such points. Furthermore:

$$\mathbf{1}\{z^*.q = 0\} = \mathbf{1}\{z_q = 0\},$$

which justifies the definition of D_f. ■

Adding the linear term (that we dropped before the previous Lemma) yields:

$$\frac{\partial \delta^*(q \mid B)}{\partial q^T} = E(z^T x)^{-1} E(z^T w_q) = \beta_q,$$

by Proposition 5. It is continuous except at points in D_f. As $\delta^*(q \mid B) = q^T \beta_q$, and $\beta_q \in \arg\max_{\beta \in B}(q^T \beta)$, this result is a disguise of the envelope theorem.

B.4 Proof of Proposition 7

Proof of i) If B has a flat face (say B_f), define q as the vector orthogonal to B_f. We then have:

$$\forall \beta_f \in B_f, \delta^*(q \mid B) = q^T \beta_f.$$

Using equation (B.7), we have that there exist $u_f(z)$, defined by equation (B.6) and such that:

$$\beta_f = \left(E(z^T x)\right)^{-1} E(z^T(y + u_f(z))$$

As β_f is not unique, $u_f(z)$ is not unique. The only possibility is $\Pr(x_q = 0) = \Pr(z_q = 0) > 0$.

Conversely, suppose that $\Pr(z_q = 0) > 0$ and use equations (B.6) and (B.7) to write:

$$\begin{aligned} \beta_f &= \beta_q + \left(E(z^T x)\right)^{-1} E(z^T \mathbf{\Delta}^*(z) \mathbf{1}\{z_q = 0\}) \\ &= \beta_q + \left(E(z^T x)\right)^{-1} E(z^T \mathbf{\Delta}^*(z) \mid z_q = 0) \Pr(z_q = 0). \end{aligned}$$

As $\mathbf{\Delta}^*(z)$ is an arbitrary function in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ and as z_q is a linear function of z, generically (under some assumption on the support of z) the second term in the RHS is non zero for at least some $\beta_f \neq \beta_q$. As β_f and β_q belong to convex B, the segment $[\beta_f, \beta_q]$ belongs to B so that B has a flat face.

Proof of ii) A kink at $\beta_k \in \partial B$ is obtained when there exist vectors q and r ($r \neq q$) whose orthogonal hyperplanes are supporting hyperplanes of B at β_k. There exist $u_q(z)$ and $u_r(z)$ and thus $\mathbf{\Delta}_q^*(z)$ and $\mathbf{\Delta}_r^*(z)$ such that:

$$\beta_k = \beta_q + \left(E(z^T x)\right)^{-1} E(z^T \mathbf{\Delta}_q^*(z)\mathbf{1}\{z_q = 0\}) = \beta_r + \left(E(z^T x)\right)^{-1} E(z^T \mathbf{\Delta}_r^*(z)\mathbf{1}\{z_r = 0\}).$$

As B is convex, any hyperplane orthogonal to a (interior) convex combination of q and r is a supporting hyperplane of B at that point. Therefore, any q', r' on the arc $]q, r[$ on $\mathbb{S}$ are such that $\Pr(z_{q'} = 0) = \Pr(z_{r'} = 0) = 0$ (if not there will be a face orthogonal to these vectors) and are such that:

$$\begin{aligned}
\beta_{q'} &= \beta_{r'}, \\
&\Longrightarrow E(z^T w_{q'}) = E(z^T w_{r'}) \\
&\Longrightarrow E(z^T(\underline{y} + (\bar{y} - \underline{y})\mathbf{1}\{z_{q'} > 0\})) = E(z^T(\underline{y} + (\bar{y} - \underline{y})\mathbf{1}\{z_{r'} > 0\}))
\end{aligned}$$

using Proposition 5. Write the decomposition:

$$\mathbf{1}\{z_{q'} > 0\} = \mathbf{1}\{z_{r'} > 0\} + \mathbf{1}\{z_{q'} > 0, z_{r'} < 0\} - \mathbf{1}\{z_{q'} < 0, z_{r'} > 0\},$$

to get:

$$E(z^T(\bar{y} - \underline{y})\left(\mathbf{1}\{z_{q'} > 0, z_{r'} < 0\} - \mathbf{1}\{z_{q'} < 0, z_{r'} > 0\}\right)) = 0.$$

Premultiply by $q'^T E(x^T z)$ to get:

$$E(z_{q'}(\bar{y} - \underline{y})\left(\mathbf{1}\{z_{q'} > 0, z_{r'} < 0\} - \mathbf{1}\{z_{q'} < 0, z_{r'} > 0\}\right)) = 0.$$

This term is necessarily non negative because $\bar{y} - \underline{y} > 0$. It is equal to zero if and only if:

$$\Pr\{z_{q'} > 0, z_{r'} < 0\} = \Pr\{z_{q'} < 0, z_{r'} > 0\} = 0.$$

As it is true for any q', r' on the arc $]q, r[$ on $\mathbb{S}$, it is also true for q and r.

Conversely, it is straightforward to see that if:

$$\Pr\{z_q > 0, z_r < 0\} = \Pr\{z_q < 0, z_r > 0\} = 0$$

then almost everywhere, $w_q = w_r$ and thus $\beta_q = \beta_r$. B has a kink at this point.

B.5 Construction of the identified set in the supernumerary case

Let (q, λ) be the direction used for estimating B_U, λ being the components relative to the z_H space. From previous result, we know that:

$$\begin{aligned}
\beta_{q,\lambda} &= \left[E(z^T x) : E(z^T z_H)\right]^{-1} E(z^T w_{q,\lambda}) \\
w_{q,\lambda} &= 1_{\{z_{q,\lambda} > 0\}}\bar{y} + 1_{\{z_{q,\lambda} < 0\}}\underline{y} \\
z_{q,\lambda} &= (q^T, \lambda^T)\left[E(z^T x) : E(z^T z_H)\right]^{-1} z^T = q^T E(z_F^T x)^{-1} z_F^T + \lambda^T z_H^T
\end{aligned}$$

The expression of $z_{q,\lambda}$ comes from the expression of $\left[E(z^T x) : E(z^T z_H)\right]^{-1}$ which can be proved to be:

$$\left[E(z^T x) : E(z^T z_H)\right]^{-1} = \left[\begin{array}{c} \left[E(x^T z)E(z^T z)^{-1}E(z^T x)\right]^{-1/2} F^T E(z^T z)^{-1/2} \\ H^T E(z^T z)^{-1/2} \end{array} \right]$$

Consequently we are able to derive a closed-form expression for the support function:

$$\delta_{q,\lambda} = q^T E(z_F^T x)^{-1} E(z_F^T w_{q,\lambda}) + \lambda^T E(z_H^T w_{q,\lambda}),$$

as well as for the point $\beta_{q,\lambda}$ which achieved the value of the support function in the given direction:

$$\beta_{q,\lambda} = \left[\begin{array}{c} E(z_F^T x)^{-1} E(z_F^T w_{q,\lambda}) \\ E(z_H^T w_{q,\lambda}) \end{array} \right].$$

The identified set is the collection of the points whose last components are equal to zero:

$$B = \left\{\beta_{q,\lambda}, E(z_H^T w_{q,\lambda}) = 0\right\}$$

The orthogonal projection of B_U on the space $\{\gamma = 0\}$ is the collection of points β in this space such that the support function is the same than the support function $\delta_{q,0}$ of B_U:

$$\left\{E(z_F^T x)^{-1} E(z_F^T w_{q,0}), q \in \mathbb{S}_p\right\}.$$

This does not depend on z_H since $z_{q,0} = q^T E(z_F^T x)^{-1} z_F^T$. It is the same set than P_B which is defined by the instruments z_F with no additional restrictions on $u(z)$.

B.6 Proof of Proposition 11

We first prove the result when the hyperplane $\gamma = 0$ is not tangent to B_U.

Corollary 16.4.1 page 146 Rockafellar (1970) states: Let C_1 and C_2 be non empty convex sets in $\mathbb{R}^n$ and let $ri(C_i)$ have one point in common.[13] Then, first:

$$\delta^*(x^* \mid C_1 \cap C_2) = \inf_{(x_1^*, x_2^*): x_1^* + x_2^* = x^*} (\delta^*(x_1^* \mid C_1) + \delta^*(x_2^* \mid C_2)) \tag{B.10}$$

where (x_1^*, x_2^*, x^*) are vectors of $\mathbb{R}^n$. Second, the infimum is attained.

Set $C_1 = B_U$ where:

$$B_U = \{(\beta, \gamma); E(z^T(x\beta + z_H\gamma - y) = E(z^T u(z)), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\}$$

[13] Let the smallest affine set containing C, be $aff(C)$. Let $B(x, \varepsilon)$ be the ball centered at x and of diameter $\varepsilon/2$. The relative interior of a set C is defined as:

$$ri(C) = \{x \in aff(C); \exists \varepsilon > 0, B(x, \varepsilon) \cap aff(C) \subset C\}$$

where we do not impose the restriction that $\gamma = 0$. C_1 is a convex set with a non empty interior.

Set $C_2 = \{\gamma = 0\}$ which is a convex hyperplane. Its support function is as follows if $x_2^* = (q_2, \lambda_2)$:

$$\begin{aligned}\delta^*(x_2^* \mid \{\gamma = 0\}) &= \sup_{(\beta,\gamma)\in C_2} \beta^T q_2 + \gamma^T \lambda_2 = \sup_{\beta\in\mathbb{R}^{m-p}} \beta^T q_2 \\ &= \begin{cases} 0 \text{ if } q_2 = 0 \\ +\infty \text{ if } q_2 \neq 0 \end{cases}\end{aligned}$$

Remark that B is the intersection of C_1 and C_2. Remark, further on, that except in the case of point identification, the relative interiors of C_1 and C_2 have all the points of $ri(C_1 \cap C_2)$ in common. The previous result can then be applied to $x^* = (q, \lambda)$ so that:

$$\begin{aligned}\delta^*((q,\lambda) \mid B) &= \inf_{(x_1^*,x_2^*):x_1^*+x_2^*=x^*} \delta^*(x_1^* \mid B_U) + \delta^*(x_2^* \mid \{\gamma = 0\}) \\ &= \inf_{(\lambda_1,\lambda_2):\lambda_1+\lambda_2=\lambda} \delta^*((q,\lambda_1) \mid B_U)).\end{aligned}$$

As expected, the RHS is independent of λ and we can write:

$$\delta^*(q \mid B) = \inf_{\lambda} \delta^*((q,\lambda) \mid B_U)). \tag{B.11}$$

Furthermore, the infimum is attained.

In the case where the hyperplane $\gamma = 0$ is tangent to B_U, the relative interiors have no point in common. Corollary 16.4.1 page 146 Rockafellar (1970) states that we should replace Equation (B.10) by its closure and the infimum is not necessarily attained. In our case though, B_U is a compact and closed set and in consequence, Equation (B.11) applies also to this case and the infimum is attained.

C Proofs in Section 4

C.1 Proof of Propositions 13 and 14

We use that:

$$\delta^*(q \mid B) = E(z_q w_q) = q^T E(z^T x)^{-1} E(z^T w_q) = q^T \Sigma_0^T E(z^T w_q).$$

where $\Sigma_0 = E(x^T z)^{-1}$. The estimator that we consider is:

$$\hat{\delta}_n^*(q \mid B) = \frac{1}{n} \sum z_{n,qi} . w_{n,qi},$$

where $\hat{\Sigma}_n$ is an estimate of Σ_0 defined below:

$$z_{n,qi} = q^T . \hat{\Sigma}_n^T z_i^T$$

$$w_{n,qi} = \underline{y}_i + \mathbf{1}\{z_{n,qi} > 0\}(\overline{y}_i - \underline{y}_i).$$

We suppose that the parameter $\Theta = (q, \Sigma) \in \Theta = \mathbb{S} \times \{\|\Sigma\| \leq M\}$ where $\|\Sigma\|$ is (for instance) equal to the sum of the eigenvalues of Σ and where M is an arbitrary large constant. By the full rank assumption (R.iii), the true value Σ_0 is such that $(q, \Sigma_0) \in int(\Theta)$, so that the constant is chosen as $\|\Sigma_0\| \ll M$.

The estimate $\hat{\Sigma}_n$ belongs to Θ by trimming if it is necessary. First, let:

$$\hat{\Sigma}_n^u = \left(\frac{1}{n} \sum x_i^T . z_i\right)^{-1}. \tag{C.12}$$

and define the estimate of Σ_0 as:

$$\begin{cases} \hat{\Sigma}_n = \hat{\Sigma}_n^u & \text{if } \hat{\Sigma}_n^u \in \Theta, \\ \hat{\Sigma}_n = \hat{\Sigma}_n^u (\frac{M}{\|\hat{\Sigma}_n^u\|}) & \text{if not.} \end{cases} \tag{C.13}$$

It is then straightforward to show that under the conditions of Proposition 13, $\hat{\Sigma}_n$ is almost surely consistent to Σ_0:

$$\lim_{n \to \infty} \Pr(\sup_{n > N} \left\| \hat{\Sigma}_n - \Sigma_0 \right\| \geq \varepsilon) = 0.$$

and under the conditions of Proposition 14 that $\hat{\Sigma}_n^u$ and $\hat{\Sigma}_n$ are asymptotically equivalent:

$$\sqrt{n} \left(\hat{\Sigma}_n - \hat{\Sigma}_n^u \right) \underset{n \to \infty}{\overset{P}{\to}} 0, \tag{C.14}$$

and asymptotically normal:

$$\sqrt{n} \left(vec(\hat{\Sigma}_n - \Sigma_0) \right) \Longrightarrow N(0, W).$$

We proceed in two steps. As the first step is simple, we proceed in parallel for the two proofs.

C.1.1 Consistency and Asymptotic Normality: Σ is known

Suppose that Σ is known and denote:

$$\begin{aligned} z_{qi} &= z_i . \Sigma . q \\ w_{qi} &= \underline{y}_i + \mathbf{1}\{z_{qi} > 0\}(\overline{y}_i - \underline{y}_i). \end{aligned}$$

Consider function f_θ indexed by $\theta \in \Theta$ from the the support of $(z_i, \underline{y}_i, \overline{y}_i)$ to $\mathbb{R}$ such that:

$$f_\theta(z_i, \underline{y}_i, \overline{y}_i) = z_{qi} w_{qi} = q^T \Sigma^T z_i^T (\underline{y}_i + \mathbf{1}\{z_{qi} > 0\}(\overline{y}_i - \underline{y}_i)).$$

Note that $\mathcal{F} = \{f_\theta; \theta \in \Theta\}$ is a parametric class and is indexed by a parameter θ lying in a bounded set Θ. By a slight modification of the proof of Lemma 6 (take derivatives with respect to $q^T \Sigma^T$ instead of q^T), this function is differentiable with respect to $q^T \Sigma^T$ everywhere except at the points where $z_{qi} = 0$ where the left and right derivatives exist but differ. At other points, the derivative is equal to either $z_i^T \underline{y}_i$ or $z_i^T \overline{y}_i$ whether z_{qi} is non negative or not. We deduce that, for any $\theta_1, \theta_2 \in \Theta$, we have:

$$\begin{aligned}
\left| f_{\theta_1}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_2}(z_i, \underline{y}_i, \overline{y}_i) \right| &\leq \max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \left\| q_1^T \Sigma_1^T - q_2^T \Sigma_2^T \right\| \\
&= \max(z_i^T \underline{y}_i, z_i^T \overline{y}_i). \left\| (q_1 - q_2)^T \Sigma_1^T - q_2^T (\Sigma_1 - \Sigma_2)^T \right\| \\
&= \max(z_i^T \underline{y}_i, z_i^T \overline{y}_i).M \left\| \theta_1 - \theta_2 \right\| \qquad \text{(C.15)}
\end{aligned}$$

where the last equality (and the constant $M < \infty$) is derived from the bounds on Θ.

First, under the conditions of Proposition 13, we have that:

$$E \left| \max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \right| < \infty$$

so that $\mathcal{F} = \{f_\theta; \theta \in \Theta\}$ is a Glivenko-Cantelli class (for instance, van der Vaart, 1998, page 271). By the definition of such a class, it shows that, uniformly over Θ:

$$\frac{1}{n} \sum_{i=1}^n f_\theta(z_i, \underline{y}_i, \overline{y}_i) = \frac{1}{n} \sum_{i=1}^n z_{qi} w_{qi} \underset{n \to \infty}{\overset{a.s}{\to}} E(z_{qi} w_{qi}).$$

Second, under the conditions of Proposition 14, we have that:

$$E \left(\max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \right)^2 < \infty$$

so that $\mathcal{F} = \{f_\theta; \theta \in \Theta\}$ is a Donsker class (for instance, van der Vaart, 1998, page 271). By the definition of such a class, it shows that the empirical process:

$$\sqrt{n} \tau_n(q) = \sqrt{n} (\frac{1}{n} \sum_{i=1}^n z_{qi} w_{qi} - E(z_{qi} w_{qi})),$$

converges in distribution to a Gaussian process with zero mean and covariance function:

$$E(z_{qi} w_{qi} z_{ri} w_{ri}) - E(z_{qi} w_{qi}) E(z_{ri} w_{ri}).$$

The second step consists in replacing Σ_0 by the almost sure limit $\hat{\Sigma}_n$ defined above. It is more involved and we thus separate the two proofs.

C.1.2 Proof of Consistency

To prove consistency and asymptotic normality, we rely heavily on Section 19.4 of van der Vaart (1998) where useful properties to show convergence results are proposed.

First, the estimate of the support function is:

$$\frac{1}{n} \sum_{i=1}^n z_{n,qi} w_{n,qi} = \frac{1}{n} \sum_{i=1}^n f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i)$$

where $\hat{\theta}_n = (q, \hat{\Sigma}_n)$ by definitions of $z_{n,qi}$ and $w_{n,qi}$.

First, under the conditions of Proposition 13, the class $\mathcal{F} = \{f_\theta; \theta \in \Theta\}$ is a Glivenko-Cantelli class. By construction of the estimate $\hat{\Sigma}_n$ (see above), $\hat{\theta}_n$ belongs to Θ. It is thus immediate that, for every sequence of functions $f_{\hat{\theta}_n} \in \mathcal{F}$, and uniformly in $q \in \mathbb{S}$, we have:

$$\left| \frac{1}{n} \sum_{i=1}^{n} f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i)) \right| \underset{n\to\infty}{\overset{a.s}{\to}} 0. \tag{C.16}$$

Second, as matrix Σ_0 is estimated by its almost surely consistent empirical analogue $\hat{\Sigma}_n$:

$$\lim_{n\to\infty} \Pr(\sup_{n>N} \left\| \hat{\Sigma}_n - \Sigma_0 \right\| \geq \varepsilon) = 0,$$

we have:

$$\lim_{n\to\infty} \Pr(\sup_{n>N} \sup_{q\in\mathbb{S}} \left\| \hat{\theta}_n - \theta_0 \right\| \geq \varepsilon) = 0.$$

Use equation (C.15):

$$\left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_0}(z_i, \underline{y}_i, \overline{y}_i) \right| = |z_{n,qi}w_{n,qi} - z_{qi}w_{qi}| \leq \left| \max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \right| .M \left\| \hat{\theta}_n - \theta_0 \right\|.$$

to conclude that, uniformly over $q \in \mathbb{S}$, we have:

$$\left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_0}(z_i, \underline{y}_i, \overline{y}_i) \right| \underset{n\to\infty}{\overset{a.s}{\to}} 0. \tag{C.17}$$

To finish the proof, notice that the sequence $f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i)$ is uniformly bounded for $q \in \mathbb{S}$, because, by majorization and triangular inequality, we have:

$$f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) = |z_{n,qi}w_{n,qi}| \leq \left\| q^T \Sigma_n^T \right\| (\left\| z_i^T \overline{y}_i \right\| + \left\| z_i^T \underline{y}_i \right\|) = \|\Sigma_n\| (\left\| z_i^T \overline{y}_i \right\| + \left\| z_i^T \underline{y}_i \right\|)$$

since $\|q\| = 1$. Therefore, as $\|\Sigma_n\| \leq M$:

$$\sup_{q\in\mathbb{S}} \left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) \right| \leq M_1(\left\| z_i^T \overline{y}_i \right\| + \left\| z_i^T \underline{y}_i \right\|)$$

As $z_i, \overline{y}, \underline{y}_i$ are in L^2 (Assumption R.iii), it implies that:

$$E \sup_{q\in\mathbb{S}} \left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) \right| \leq M_2 < +\infty.$$

Thus, equation (C.17) implies that, by the dominated convergence theorem, uniformly over q,

$$E \left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_0}(z_i, \underline{y}_i, \overline{y}_i) \right| \underset{n\to\infty}{\overset{a.s}{\to}} 0.$$

¿From the latter equation, equation (C.16) and the triangular inequality, we thus conclude that, uniformly for $q \in \mathbb{S}$:

$$\frac{1}{n} \sum_{i=1}^{n} z_{n,qi}w_{n,qi} \underset{n\to\infty}{\overset{a.s}{\to}} E(z_{qi}w_{qi}).$$

C.1.3 Proof of Asymptotic Normality

Let us first prove that, uniformly in q:

$$E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_0}(z_i, \underline{y}_i, \overline{y}_i))^2 \underset{n\to\infty}{\overset{P}{\to}} 0. \tag{C.18}$$

where $\theta_0 = (q, \Sigma_0)$. Use equation (C.15):

$$\left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_0}(z_i, \underline{y}_i, \overline{y}_i) \right| \leq \left| \max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \right| M \left\| \hat{\theta}_n - \theta_0 \right\|.$$

so that:

$$E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_0}(z_i, \underline{y}_i, \overline{y}_i))^2 \leq E\left(\max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \right)^2 M^2 \left\| \hat{\theta}_n - \theta_0 \right\|^2$$

Under the conditions stated before equation (C.15), $E\left(\max(z_i^T \underline{y}_i, z_i^T \overline{y}_i) \right)^2 < \infty$ and is independent of q. As $\left\| \hat{\theta}_n - \theta_0 \right\|^2$ tends in distribution to 0 uniformly in $q \in \mathbb{S}$ then it tends in probability to 0, uniformly in $q \in \mathbb{S}$.

We can then apply Lemma 19.24 of van der Vaart (1998), so that:

$$\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} z_{n,qi} w_{n,qi} - E(z_{n,qi} w_{n,qi}) \right).$$

has the same distribution than:

$$\tau_n(q) = \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} z_{qi} w_{qi} - E(z_{qi} w_{qi}) \right).$$

uniformly in $q \in S$. Thus:

$$A_n(q) = \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} z_{n,qi} w_{n,qi} - E(z_{qi} w_{qi}) \right)$$

is an empirical process asymptotically equivalent to:

$$\tau_n(q) + \sqrt{n}(E(z_{n,qi} w_{n,qi}) - E(z_{qi} w_{qi})).$$

To compute the limit of this process, we use the following:

Lemma 21 *We have, uniformly in $q \in \mathbb{S}$:*

i. $\sqrt{n}(E(z_{n,qi} w_{n,qi}) - E(z_{qi} w_{qi})) - \sqrt{n} q^T (\hat{\Sigma}_n^T (\Sigma_0^T)^{-1} - I)\beta_q \underset{n\to\infty}{\overset{P}{\to}} 0,$

ii. $\tau_n(q) - \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^n q^T \Sigma_0^T z_i^T \varepsilon_{qi} \right) - \sqrt{n} q^T (I - \hat{\Sigma}_n^T (\Sigma_0^T)^{-1})\beta_q \underset{n\to\infty}{\overset{P}{\to}} 0,$

where if $\beta_q = \Sigma_0^T E(z_i^T w_{qi})$.

Proof. We first prove (i). $\tau_n^e(q) = \sqrt{n}(E(z_{n,qi}w_{n,qi}) - E(z_{qi}w_{qi}))$ can be written as an empirical process, $\sqrt{n}(g(\Sigma_n q) - g(\Sigma_0 q))$ where g is a function from $\mathbb{R}^K$ to $\mathbb{R}$:

$$g(\Sigma q) = E(z_{qi}w_{qi}).$$

By Lemma 6, this function is differentiable:

$$\frac{\partial g}{\partial(\Sigma q)^T}(\Sigma q) = E(z_i^T w_{qi}),$$

and the differential is uniformly bounded over $q \in \mathbb{S}$. Thus:

$$\tau_n^e(q) = \sqrt{n}(q^T(\hat{\Sigma}_n - \Sigma_0)^T E(z_i^T w_{qi}) + o_P(1) = \sqrt{n}q^T(\hat{\Sigma}_n - \Sigma_0)^T({\Sigma_0}^T)^{-1}\beta_q + o_P(1).$$

which proves (i).

To prove (ii), write:

$$\tau_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^T\Sigma_0^T z_i^T w_{qi} - E(q^T\Sigma_0^T z_i^T w_{qi})\right)$$

and define $\varepsilon_{qi} = w_{qi} - x_i\beta_q$. Note that $E(z_i^T\varepsilon_{qi}) = 0$ by definition of β_q. Thus:

$$\tau_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^T\Sigma_0^T z_i^T \varepsilon_{qi}\right) + \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^T\Sigma_0^T z_i^T x_i\beta_q - E(q^T\Sigma_0^T z_i^T w_{qi})\right)$$

The second term on the right hand side is equal to:

$$\begin{aligned}
\sqrt{n}q^T\Sigma_0^T\left(\frac{1}{n}\sum_{i=1}^n z_i^T x_i\right)\beta_q - \sqrt{n}q^T\Sigma_0^T E(z_i^T x_i)\beta_q &= \sqrt{n}q^T(\Sigma_0^T(\hat{\Sigma}_n^{uT})^{-1} - I)\beta_q \\
&= \sqrt{n}q^T(\Sigma_0^T(\hat{\Sigma}_n^T)^{-1} - I)\beta_q + o_p(1) \\
&= \sqrt{n}q^T\Sigma_0^T(\hat{\Sigma}_n^T)^{-1}(I - \hat{\Sigma}_n^T(\Sigma_0^T)^{-1})\beta_q + o_p(1)
\end{aligned}$$

The first line uses definition (C.12) and the definition of Σ_0, the second line uses that $\sqrt{n}(\hat{\Sigma}_n^u - \hat{\Sigma}_n) \underset{n\to\infty}{\overset{P}{\to}} 0$ by equation (C.14) and uniform bounds on q, Σ_0 and β_q. Moreover, as $\Sigma_0^T(\hat{\Sigma}_n^T)^{-1} \underset{n\to\infty}{\overset{a.s}{\to}} I$, we have that, uniformly in q:

$$\tau_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^T\Sigma_0^T z_i^T\varepsilon_{qi}\right) + \sqrt{n}q^T(I - \hat{\Sigma}_n\Sigma_0^{-1})^T\beta_q + o_p(1).$$

■

Wrapping up,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n z_{n,qi}w_{n,qi} - E(z_{qi}w_{qi})\right)$$

is distributed as:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^T\Sigma_0^T z_i^T\varepsilon_{qi}\right)$$

It converges in distribution, uniformly in q, to a Gaussian process centered at zero and of covariance function:

$$E(z_{qi}\varepsilon_{qi}\varepsilon_{ri}z_{ri}).$$

C.1.4 Covariance Matrix

The intuition of the simplification of the expression of the covariance matrix difference in Beresteanu and Molinari (2006) can be understood using the standard OLS example where the same simplification occurs. Let:

$$\hat{\beta}_n = \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' y_i \right)$$

the OLS estimate of:

$$\beta_0 = \left(E(x_i' x_i) \right)^{-1} E(x_i' y_i).$$

Set:

$$\hat{\Sigma}_n = \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1}, \Sigma = \left(E(x_i' x_i) \right)^{-1}$$

so that:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta) &= \sqrt{n} \left[\hat{\Sigma}_n. \left(\frac{1}{n} \sum_{i=1}^{n} x_i' y_i \right) - \beta_0 \right] \\ &= \sqrt{n} \hat{\Sigma}_n . \Sigma^{-1} \left[\Sigma \left(\frac{1}{n} \sum_{i=1}^{n} x_i' y_i \right) - \Sigma . \hat{\Sigma}_n^{-1} . \beta_0 \right] \\ &= \sqrt{n} \hat{\Sigma}_n . \Sigma^{-1} \left[\Sigma \left(\frac{1}{n} \sum_{i=1}^{n} x_i' y_i - E(x_i' y_i) \right) - (\Sigma . \hat{\Sigma}_n^{-1} - I) . \beta_0 \right] . \end{aligned} \tag{C.19}$$

As $\hat{\Sigma}_n \overset{a.s.}{\to} \Sigma$, this expression has the same distribution than:

$$\sqrt{n} \left[\Sigma \left(\frac{1}{n} \sum_{i=1}^{n} x_i' y_i - E(x_i' y_i) \right) - (\Sigma . \hat{\Sigma}_n^{-1} - I) . \beta_0 \right] .$$

Beresteanu and Molinari (2006) can then derive the asymptotic distribution of $\sqrt{n}(\hat{\beta}_n - \beta)$ from the distribution of $\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' y_i - E(x_i' y_i) \right)$ and $\sqrt{n}(\Sigma . \hat{\Sigma}_n^{-1} - I) . \beta_0$ as they do (Proof of their Theorem 3). Yet this is hardly necessary since by replacing y_i by $x_i \beta_0 + \varepsilon_i$ in the first line of the previous computation we directly get:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta_0) &= \sqrt{n} \left[\hat{\Sigma}_n. \left(\frac{1}{n} \sum_{i=1}^{n} x_i' (x_i \beta_0 + \varepsilon_i) \right) - \beta_0 \right] \\ &= \sqrt{n} \left[\hat{\Sigma}_n. \left(\frac{1}{n} \sum_{i=1}^{n} x_i' \varepsilon_i \right) \right] \end{aligned}$$

that has the same distribution that:

$$\sqrt{n} \left[\Sigma. \left(\frac{1}{n} \sum_{i=1}^{n} x_i' \varepsilon_i \right) \right]$$

where we only need the distribution of $\left(\frac{1}{n} \sum_{i=1}^{n} x_i' \varepsilon_i \right)$. It is quite straightforward to show that the first method gives exactly the same distribution as the second one. The same remark than OLS applies to linear set identification as shown by the slightly longer proof above.

C.2 Proof of Proposition 15:

Let $q_0 \in \mathcal{Q}$, which minimizes $T_\infty(q;\beta_0)$. q_0 is not necessary unique. Let q_n be a direction which minimizes its empirical counterpart $T_n(q;\beta_0)$.

¿From Proposition 14, we know that $\sqrt{n}(\hat{\delta}_n^*(q) - \delta^*(q)$ tends in distribution uniformly in q to some Gaussian Process υ_q.

For all $q \in \mathbb{S}$, we can write the following equality:

$$\begin{aligned}\sqrt{n}\left(T_n(q;\beta_0) - T_\infty(q;\beta_0)\right) &= \sqrt{n}(\hat{\delta}_n^*(q) - \delta^*(q) \\ &= \upsilon_q + \omega_q\end{aligned}$$

where υ_q is the Gaussian process of Proposition 14 and ω_q is a process which tends uniformly in q to zero:

$$\lim_{n\to\infty} \Pr(\sup_{n>N}\sup_{q\in\mathbb{S}} |\omega_q| \geq \varepsilon) = 0.$$

We can now bound $\sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right)$ easily. The upper bound is zero because q_n minimizes $T_n(.;\beta_0)$. For a lower bound we can express the former expression in three terms:

$$\begin{aligned}\sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right) &= \sqrt{n}\left(T_n(q_n;\beta_0) - T_\infty(q_n;\beta_0)\right) + \sqrt{n}\left(T_\infty(q_n;\beta_0) - T_\infty(q_0;\beta_0)\right) + \sqrt{n}\,(T_\infty \\ &= \upsilon_{q_n} - \upsilon_{q_0} + \omega_{q_n} - \omega_{q_0} + \sqrt{n}\left(T_\infty(q_n;\beta_0) - T_\infty(q_0;\beta_0)\right) \\ &\geq \upsilon_{q_n} - \upsilon_{q_0} + \omega_{q_n} - \omega_{q_0},\end{aligned}$$

the latter inequality coming from the definition of q_0. It means that $\upsilon_{q_n} - \upsilon_{q_0}$ is asymptotically upward bounded by 0 as $\lim_{n\to\infty}\Pr(\sup_{n>N}|\omega_{q_n} - \omega_{q_0}| \geq \varepsilon) = 0$. This is possible only if the two random variables υ_{q_n} and υ_{q_0} are perfectly correlated with the same variance.

The previous result means two things. The first one is that $\sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right)$ tends to zero asymptotically. We can base our testing procedure on $T_n(q_n;\beta_0)$ (which can be computed in the sample) in replacement of $T_n(q_0;\beta_0)$ which we cannot compute. The second important conclusion is that $q_n^T \Sigma^T V(z^T \epsilon_{q_n})\Sigma q_n$ consistently estimates $V_{q_0} = q_0^T \Sigma^T V(z^T \epsilon_{q_0})\Sigma q$. We can of course replace Σ and $V(z^T\epsilon_{q_n})$ by their empirical counterpart to provide a feasible consistent estimator. Q.E.D.

D Computations of Section 5

D.1 Example of Section 5.1

The simulated model is:

$$y = 0.x_1 + 0.x_2 + \varepsilon$$

We use $z = x$ as instruments for estimating $\delta^*(q|B)$.

$$\begin{aligned}
E(x^T x) &= I_2 \\
D &= E(z^T z)^{-1} E(z^T x) = I_2 \\
R &= E(x^T z) E(z^T z)^{-1} E(z^T x) = I_2 \\
a(q) &= \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \\
x_q &= zDa(q) = \cos\theta x_1 + \sin\theta x_2 \\
w_q &= y - \Delta + 2\Delta \mathbf{1}\{x_q > 0\}
\end{aligned}$$

As

$$\begin{matrix} x_1 \\ x_q \end{matrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \cos\theta \\ \cos\theta & 1 \end{bmatrix} \right),$$

we can compute:

$$Ex_1 \mathbf{1}_{x_q>0} = \frac{1}{\sqrt{2\pi}} \cos\theta$$

and also the following by the same way:

$$Ex_2 \mathbf{1}_{x_q>0} = \frac{1}{\sqrt{2\pi}} \sin\theta.$$

$$\beta_q = E(x^T w_q) = \frac{2\Delta}{\sqrt{2\pi}} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

At the end, we have a closed-form expression for $\delta^*(q|B)$:

$$\delta^*(q|B) = q^T \beta_q = \frac{2\Delta}{\sqrt{2\pi}}.$$

D.2 Example of Section 5.2

The simulated model is:

$$y = 0.x_1 + 0.x_2 + \varepsilon$$

$x_2 = \lambda e_2 + \sqrt{1-\lambda^2} e_3$, $w = \nu e_3 + \sqrt{1-\nu^2} e_4$ where (e_2, e_3, e_4) is a standard normal vector.

Let $a_{\nu,\lambda} = \lambda^2 + (1-\lambda^2)\nu^2$,

$$
\begin{aligned}
E(z^T z) &= I_3 \\
D = E(z^T z)^{-1} E(z^T x) &= \begin{bmatrix} 1 & 0 \\ 0 & \lambda \\ 0 & \sqrt{1-\lambda^2}\nu \end{bmatrix} \\
R = E(x^T z) E(z^T z)^{-1} E(z^T x) &= \begin{bmatrix} 1 & 0 \\ 0 & a_{\nu,\lambda} \end{bmatrix} \\
a(q) &= \frac{1}{\sin^2\theta + a_{\nu,\lambda}\cos^2\theta} \begin{bmatrix} a_{\nu,\lambda}\cos\theta \\ \sin\theta \end{bmatrix} \\
x_q = zDa(q) &= \frac{a_{\nu,\lambda}\cos\theta x_1 + \sin\theta(\lambda e_2 + \sqrt{1-\lambda^2}\nu w)}{\sin^2\theta + a_{\nu,\lambda}\cos^2\theta} \\
w_q &= y - \Delta + 2\Delta \mathbf{1}\{x_q > 0\}
\end{aligned}
$$

As

$$
\begin{matrix} x_1 \\ x_q \end{matrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \frac{a_{\nu,\lambda}\cos\theta}{\sin^2\theta + a_{\nu,\lambda}\cos^2\theta} \\ \frac{a_{\nu,\lambda}\cos\theta}{\sin^2\theta + a_{\nu,\lambda}\cos^2\theta} & \frac{a_{\nu,\lambda}}{\sin^2\theta + a_{\nu,\lambda}\cos^2\theta} \end{bmatrix} \right),
$$

we can compute:

$$
Ex_1 \mathbf{1}_{x_q>0} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{a_{\nu,\lambda}}{\sin^2\theta + a_{\nu,\lambda}\cos^2\theta}} \cos\theta
$$

and also the following by the same way:

$$
Ee_2 \mathbf{1}_{x_q>0} = \frac{1}{\sqrt{2\pi}} \frac{\lambda}{\sqrt{a_{\nu,\lambda}(\sin^2\theta + a_{\nu,\lambda}\cos^2\theta)}} \sin\theta.
$$

$$
Ew \mathbf{1}_{x_q>0} = \frac{1}{\sqrt{2\pi}} \frac{\nu\sqrt{1-\lambda^2}}{\sqrt{a_{\nu,\lambda}(\sin^2\theta + a_{\nu,\lambda}\cos^2\theta)}} \sin\theta.
$$

$$
\beta_q = \frac{2\Delta}{\sqrt{2\pi a_{\nu,\lambda}(\sin^2\theta + a_{\nu,\lambda}\cos^2\theta)}} \begin{bmatrix} a_{\nu,\lambda}\cos\theta \\ \sin\theta \end{bmatrix}
$$

At the end, we have a closed-form expression for $\delta^*(q|B)$:

$$
\delta^*(q|B) = \frac{2\Delta}{\sqrt{2\pi}} \sqrt{\cos^2\theta + \frac{\sin^2\theta}{a_{\nu,\lambda}}}
$$

D.3 Example of section 5.3

The simulated model is:

$$
y = \frac{1}{2} + \frac{x_1}{4} + \varepsilon
$$

We use $z = x \equiv (1, x_1)^T$ as instruments for estimating $\delta^*(q|B)$.

$$\Sigma = E(z^T z)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \\ & \end{bmatrix}$$
$$D = E(z^T z)^{-1} E(z^T x) = I_2$$
$$a(q) = \frac{1}{\frac{1}{4}\cos^2\theta + \sin^2\theta} \begin{bmatrix} \frac{1}{4}\cos\theta \\ \sin\theta \end{bmatrix}$$
$$x_q = zDa(q) = \frac{\frac{1}{4}\cos\theta + x_1 \sin\theta}{\frac{1}{4}\cos^2\theta + \sin^2\theta}$$
$$w_q = \underline{y} + \frac{1}{2}\mathbf{1}\{x_q > 0\}$$
$$\underline{y} = \frac{3}{8} \text{ if } x_1 = \frac{1}{2}, = \frac{1}{8} \text{ if } x_1 = -\frac{1}{2}$$

Let $\theta_0 \in [0; \pi/2]$ such that $\tan\theta_0 = \frac{1}{2}$. Between $-\theta_0$ and θ_0 x_q is always positive whatever the value of x:

$$E\mathbf{1}_{x_q>0} = 1$$
$$Ex_1\mathbf{1}_{x_q>0} = 0$$

and $\beta_q = (\frac{3}{4}; \frac{1}{4})^T$.

We can therefore derive similar expressions for β_q, $q = (\cos\theta, \sin\theta)^T$.

For θ being between θ_0 and $-\theta_0 + \pi$, $\beta_q = (\frac{1}{2}; \frac{3}{4})^T$, for θ being between $\theta_0 + \pi$ and $\theta_0 + \pi$, $\beta_q = (\frac{1}{4}; \frac{1}{4})^T$ and for θ being between $\theta_0 - \pi$ and $-\theta_0$, $\beta_q = (\frac{1}{2}; -\frac{1}{4})^T$.

FIGURES AND TABLES

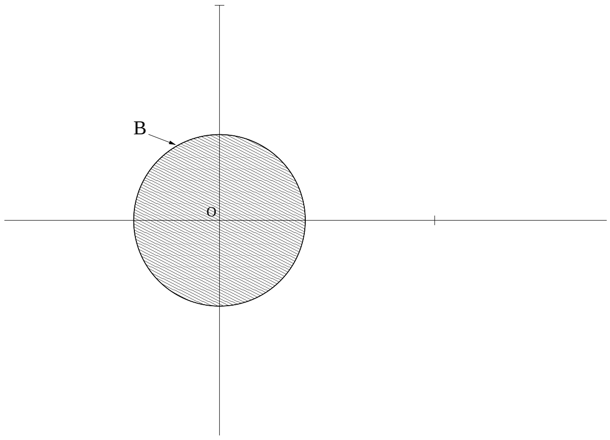


Figure 1: Set B, $y = 0.x_1 + 0.x_2 + \varepsilon$, $(x_1, x_2)^T \sim N(0, I_2)$

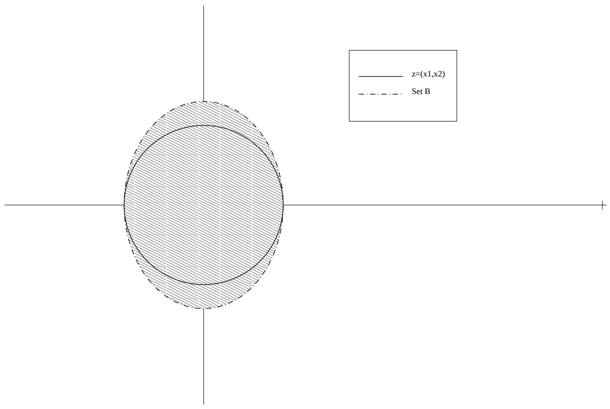


Figure 2: Set B, $y = 0.x_1 + 0.x_2 + \varepsilon$, $z = (x_1, e_2, w)$

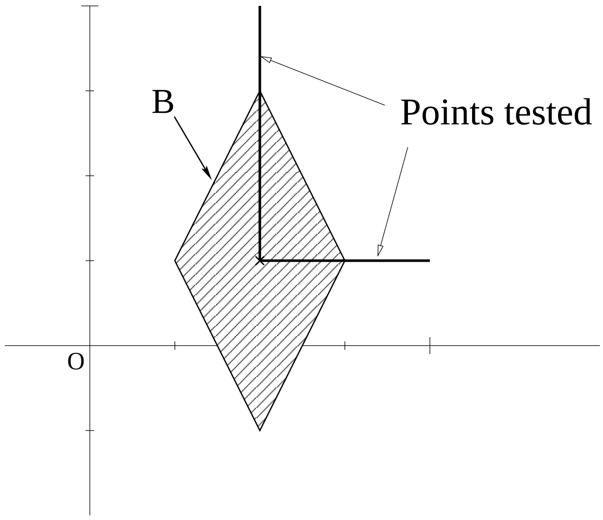


Figure 3: Set B, $y = \frac{1}{2} + \frac{x}{4} + \varepsilon$, $x \in \{-\frac{1}{2}, \frac{1}{2}\}$

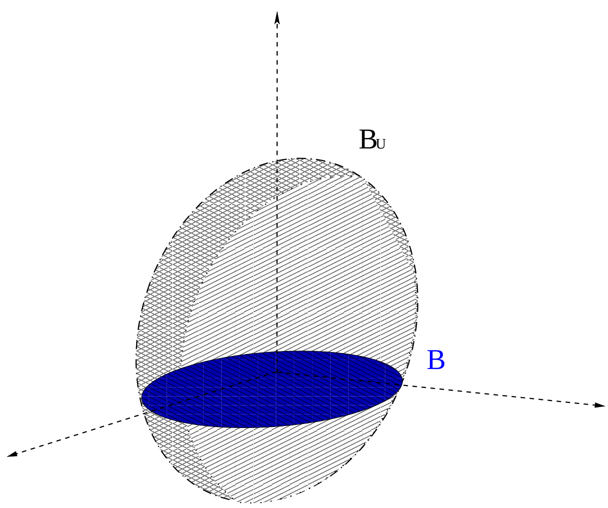
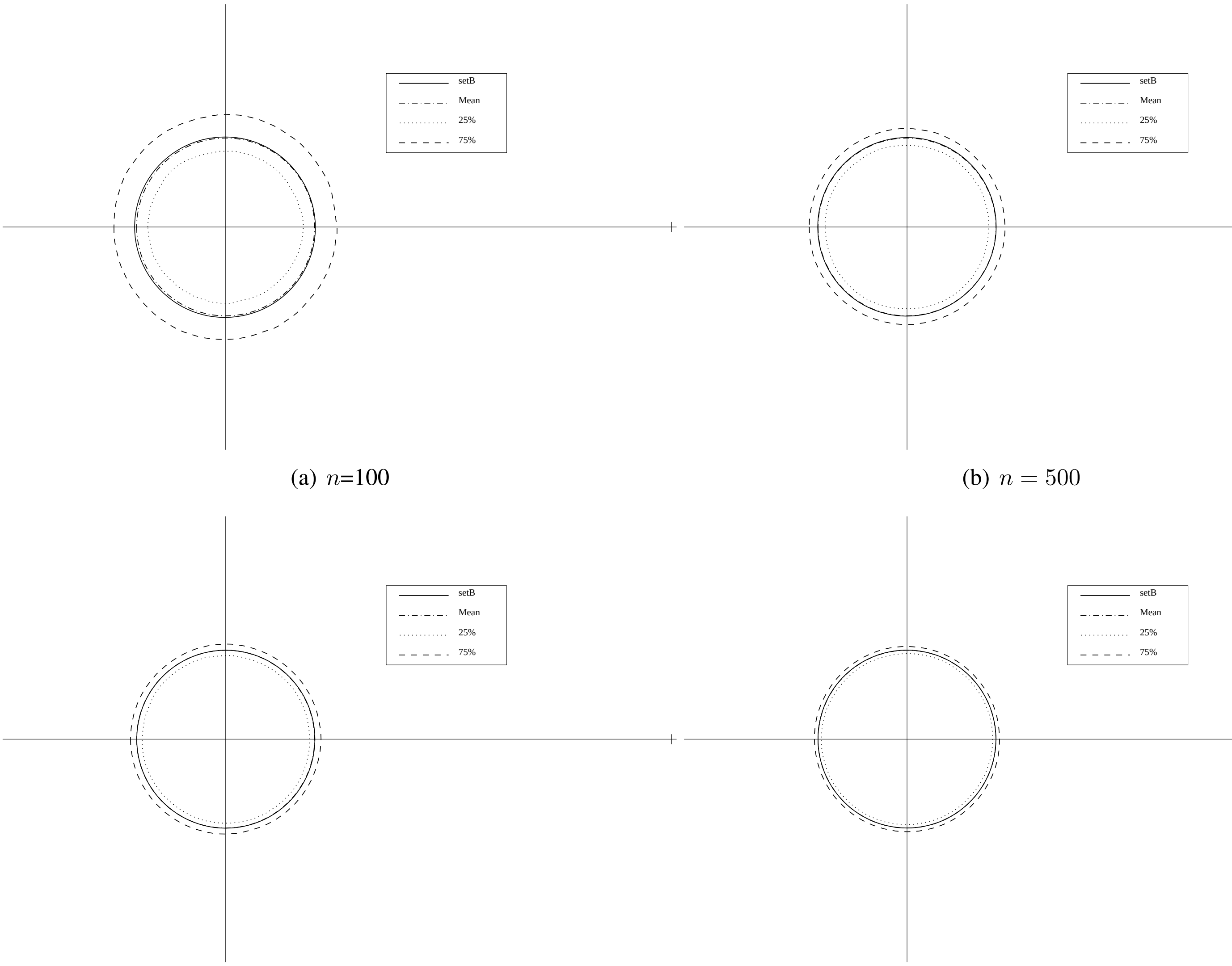


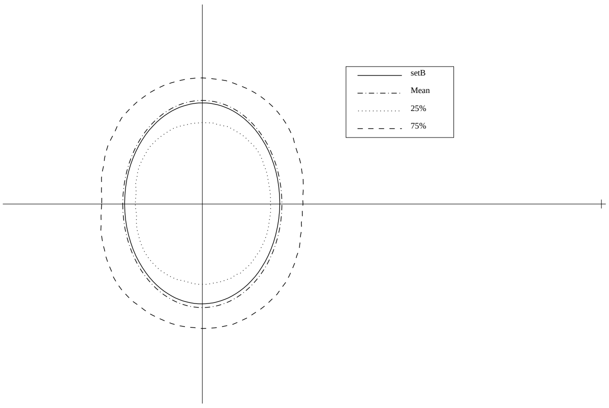
Figure 4: Geometric Characterization of the Identified Set



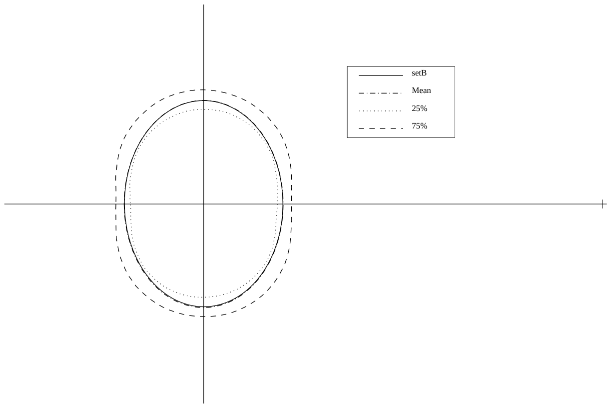
(a) n=100 (b) $n = 500$

(c) $n = 1000$ (d) $n = 2500$

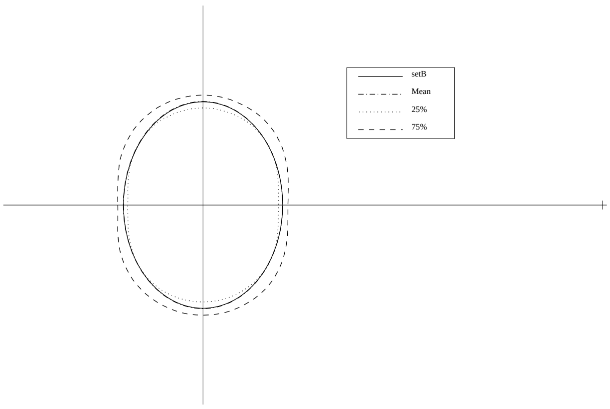
Figure 5: Estimation of B for various sample sizes n.



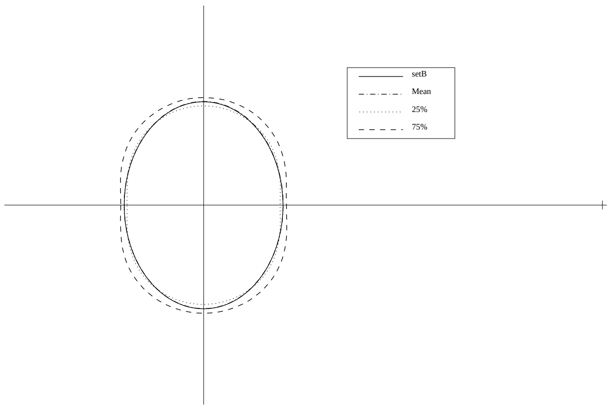
(a) n=100



(b) $n = 500$

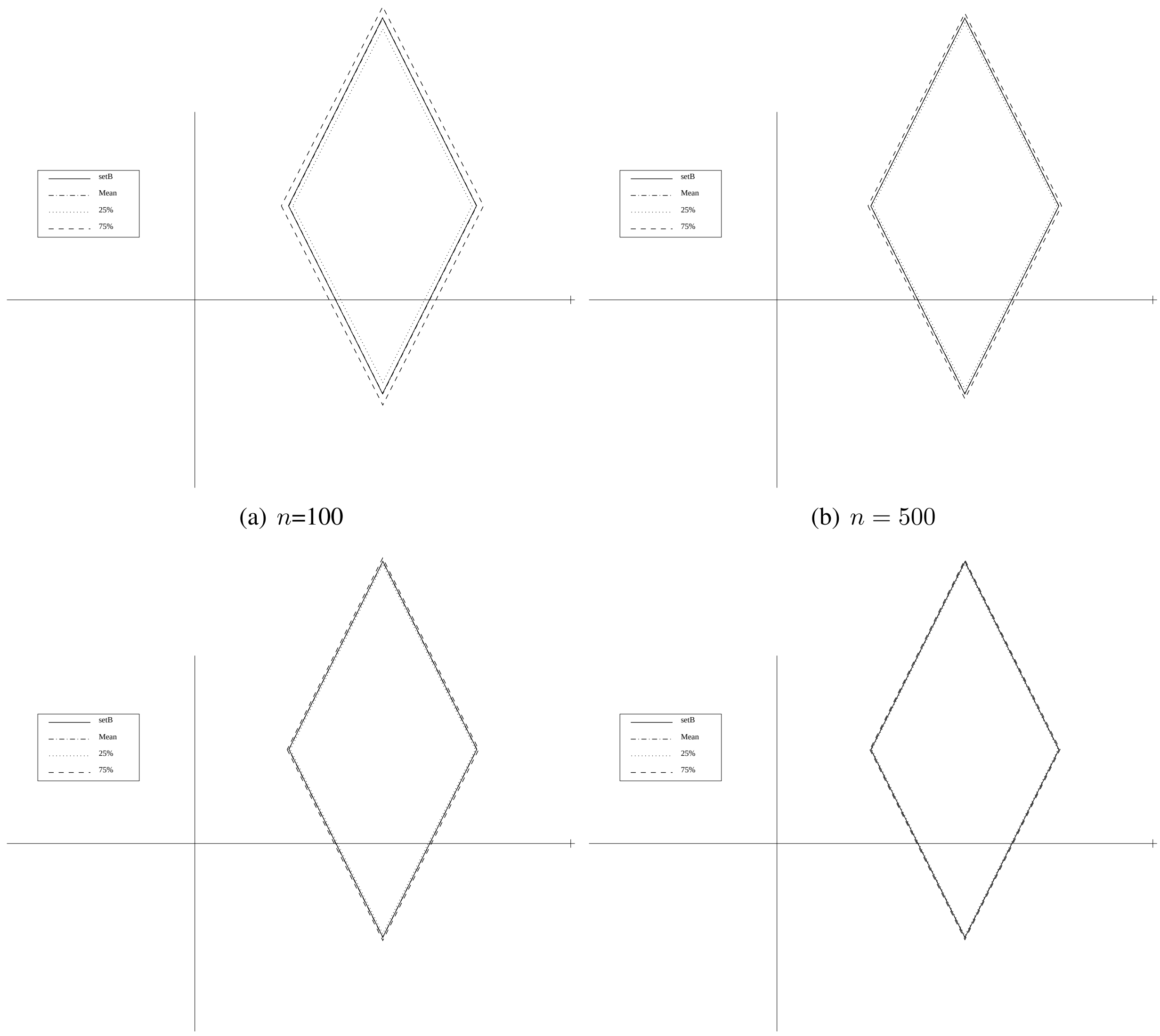


(c) $n = 1000$



(d) $n = 2500$

Figure 6: Estimation of B for various sample sizes n.



(a) n=100

(b) $n = 500$

(c) $n = 1000$

(d) $n = 2500$

Figure 7: Estimation of B for various sample sizes n.

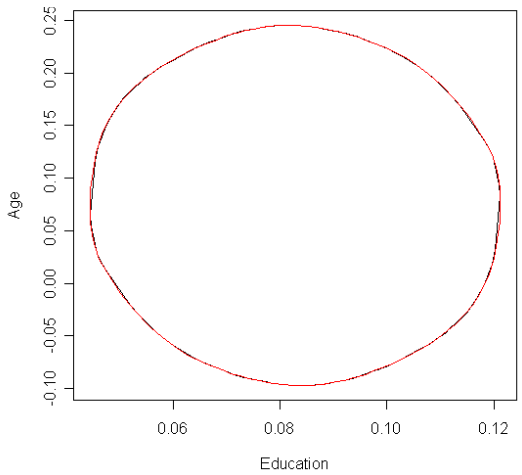


Figure 8: Estimated set for Age and Education

Table 1: Percentage of rejections for the three tests

	Test 1 ($H_0 : \beta^a \in B$)				Test 2 ($H_0 : \beta^a \notin B$)				Test 3 ($H_0 : \beta^a \in \partial B$)			
a	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.01	0%	0%	0%	0%	83.7%	100%	100%	100%	70.9%	100%	100%	100%
0.05	0%	0%	0%	0%	83.4%	100%	100%	100%	69.9%	100%	100%	100%
0.1	0%	0%	0%	0%	82.8%	100%	100%	100%	67.7%	100%	100%	100%
0.2	0%	0%	0%	0%	76%	100%	100%	100%	60.1%	100%	100%	100%
0.3	0%	0%	0%	0%	64.5%	100%	100%	100%	51.6%	99.9%	100%	100%
0.4	0%	0%	0%	0%	53.9%	99.9%	100%	100%	40.5%	99.6%	100%	100%
0.5	0%	0%	0%	0%	42.1%	98.5%	100%	100%	29.4%	97.3%	99.9%	100%
0.6	0.5%	0%	0%	0%	29.4%	92.7%	99.5%	100%	19.6%	85.4%	99%	100%
0.65	0.7%	0%	0%	0%	24.1%	83.9%	98.3%	100%	16.2%	73.3%	97.1%	100%
0.7	1%	0%	0%	0%	19.4%	71.6%	94.7%	99.9%	12.7%	61.1%	89.8%	99.9%
0.75	1.3%	0.1%	0%	0%	15.3%	57.9%	85.8%	99.5%	9.7%	45.8%	76.2%	99%
0.8	1.6%	0.1%	0%	0%	12%	43.6%	69.8%	95.5%	7.9%	31.5%	58.2%	92.3%
0.85	2.6%	0.3%	0.2%	0%	9.5%	29.3%	48.3%	82%	6.5%	19.7%	36.5%	73.2%
0.9	3.2%	0.7%	0.5%	0.1%	6.7%	17.3%	28.1%	51.3%	5.7%	10.4%	19.7%	39.9%
0.95	5.1%	2%	1.5%	0.6%	5.1%	8.7%	13.6%	20.4%	5.3%	5.1%	8.5%	13.6%
1	**6.9%**	**5%**	**5.2%**	**5.5%**	**3.5%**	**3.7%**	**5.1%**	**4%**	**5.6%**	**4.1%**	**5.2%**	**5%**
1.05	10.1%	10.7%	14%	22.9%	2.7%	1.4%	1.8%	0.5%	6.5%	6.4%	9.4%	15.3%
1.1	14%	21.5%	29.9%	54.1%	1.8%	0.8%	0.3%	0%	8.4%	12.3%	20.8%	43.2%
1.15	17.7%	33.9%	50.7%	82.8%	1%	0.2%	0.1%	0%	11.2%	24%	37.1%	74.4%
1.2	21.5%	47.1%	70.7%	97.1%	0.6%	0.1%	0%	0%	14.9%	35.9%	58.7%	93.3%
1.25	25%	62.3%	85.6%	99.6%	0.5%	0%	0%	0%	19.1%	50.4%	78.1%	99.1%
1.3	30.6%	75.2%	94.7%	100%	0.4%	0%	0%	0%	22.3%	64.7%	89.9%	100%
1.35	36.4%	86.4%	98.1%	100%	0.2%	0%	0%	0%	26.2%	77.4%	96.3%	100%
1.4	43.9%	93.4%	99.6%	100%	0.2%	0%	0%	0%	31.7%	87.6%	98.8%	100%
1.45	49.8%	97.6%	99.9%	100%	0.1%	0%	0%	0%	37.4%	94%	99.7%	100%
1.5	57.8%	98.8%	100%	100%	0%	0%	0%	0%	45.1%	97.9%	99.9%	100%
2	96.3%	100%	100%	100%	0%	0%	0%	0%	93.8%	100%	100%	100%
2.25	99.3%	100%	100%	100%	0%	0%	0%	0%	98.6%	100%	100%	100%
2.5	99.9%	100%	100%	100%	0%	0%	0%	0%	99.7%	100%	100%	100%
2.75	100%	100%	100%	100%	0%	0%	0%	0%	99.9%	100%	100%	100%
3	100%	100%	100%	100%	0%	0%	0%	0%	100%	100%	100%	100%

The point tested is $\beta^a = \frac{a}{\sqrt{2\Pi}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. β^1 is on the frontier of B.

Table 2: Percentage of rejections for the three tests, $\beta^a = (a, \frac{1}{4})^T$

		Test 1 ($H_0 : \beta^a \in B$)				Test 2 ($H_0 : \beta^a \notin B$)				Test 3 ($H_0 : \beta^a \in \partial B$)			
β^a		$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.5	0.25	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.52	0.25	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.54	0.25	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.56	0.25	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%
0.58	0.25	0.0%	0.0%	0.0%	0.0%	99.9%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%
0.6	0.25	0.0%	0.0%	0.0%	0.0%	99.9%	100.0%	100.0%	100.0%	99.5%	100.0%	100.0%	100.0%
0.62	0.25	0.0%	0.0%	0.0%	0.0%	99.1%	100.0%	100.0%	100.0%	98.0%	100.0%	100.0%	100.0%
0.64	0.25	0.0%	0.0%	0.0%	0.0%	95.6%	100.0%	100.0%	100.0%	90.1%	100.0%	100.0%	100.0%
0.66	0.25	0.0%	0.0%	0.0%	0.0%	83.9%	100.0%	100.0%	100.0%	72.7%	100.0%	100.0%	100.0%
0.68	0.25	0.0%	0.0%	0.0%	0.0%	54.9%	100.0%	100.0%	100.0%	39.3%	99.9%	100.0%	100.0%
0.7	0.25	0.3%	0.0%	0.0%	0.0%	26.3%	96.7%	100.0%	100.0%	16.1%	92.2%	99.9%	100.0%
0.72	0.25	1.4%	0.0%	0.0%	0.0%	7.0%	51.6%	86.8%	99.9%	4.5%	36.1%	77.2%	99.5%
0.74	0.25	5.1%	1.7%	0.6%	0.1%	1.2%	3.4%	6.3%	23.3%	3.6%	2.1%	3.5%	12.9%
0.75	**0.25**	**11.8%**	**10.9%**	**11.5%**	**11.1%**	**0.5%**	**0.4%**	**0.3%**	**0.2%**	**7.8%**	**6.1%**	**5.8%**	**5.1%**
0.76	0.25	18.7%	33.9%	51.8%	80.9%	0.1%	0.0%	0.0%	0.0%	11.5%	23.1%	37.4%	68.8%
0.78	0.25	47.3%	94.2%	99.5%	100.0%	0.0%	0.0%	0.0%	0.0%	33.8%	88.4%	99.1%	100.0%
0.8	0.25	78.3%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	65.8%	100.0%	100.0%	100.0%
0.82	0.25	95.3%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	90.2%	100.0%	100.0%	100.0%
0.84	0.25	99.4%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	98.6%	100.0%	100.0%	100.0%
0.86	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	99.9%	100.0%	100.0%	100.0%
0.88	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
0.9	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
0.92	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
0.94	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
0.96	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
0.98	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
1	0.25	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%

The point tested is $\beta^a = (a, \frac{1}{4})^T$ for various values of a. $\beta^1 = (\frac{3}{4}, \frac{1}{4})^T$ is on the frontier of B.

Table 3: Percentage of rejections for the three tests, $\beta^a = (\frac{1}{2}, a)^T$

		Test 1 ($H_0 : \beta^a \in B$)				Test 2 ($H_0 : \beta^a \notin B$)				Test 3 ($H_0 : \beta^a \in \partial B$)			
β^a		$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.5	0.25	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.5	0.27	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.5	0.29	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.5	0.31	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.5	0.33	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.5	0.35	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
0.5	0.37	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%
0.5	0.39	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%
0.5	0.41	0.0%	0.0%	0.0%	0.0%	99.9%	100.0%	100.0%	100.0%	99.8%	100.0%	100.0%	100.0%
0.5	0.43	0.0%	0.0%	0.0%	0.0%	99.9%	100.0%	100.0%	100.0%	99.8%	100.0%	100.0%	100.0%
0.5	0.45	0.0%	0.0%	0.0%	0.0%	99.8%	100.0%	100.0%	100.0%	99.0%	100.0%	100.0%	100.0%
0.5	0.47	0.0%	0.0%	0.0%	0.0%	99.0%	100.0%	100.0%	100.0%	97.6%	100.0%	100.0%	100.0%
0.5	0.49	0.0%	0.0%	0.0%	0.0%	97.9%	100.0%	100.0%	100.0%	95.5%	100.0%	100.0%	100.0%
0.5	0.51	0.0%	0.0%	0.0%	0.0%	95.7%	100.0%	100.0%	100.0%	91.6%	100.0%	100.0%	100.0%
0.5	0.53	0.0%	0.0%	0.0%	0.0%	92.1%	100.0%	100.0%	100.0%	82.1%	100.0%	100.0%	100.0%
0.5	0.55	0.0%	0.0%	0.0%	0.0%	82.9%	100.0%	100.0%	100.0%	74.9%	100.0%	100.0%	100.0%
0.5	0.57	0.0%	0.0%	0.0%	0.0%	74.8%	100.0%	100.0%	100.0%	64.2%	100.0%	100.0%	100.0%
0.5	0.59	0.0%	0.0%	0.0%	0.0%	65.4%	100.0%	100.0%	100.0%	50.3%	100.0%	100.0%	100.0%
0.5	0.61	0.0%	0.0%	0.0%	0.0%	50.2%	100.0%	100.0%	100.0%	38.0%	99.9%	100.0%	100.0%
0.5	0.63	0.0%	0.0%	0.0%	0.0%	37.1%	100.0%	99.3%	100.0%	26.6%	97.6%	100.0%	100.0%
0.5	0.65	0.0%	0.0%	0.0%	0.0%	26.4%	100.0%	94.8%	99.8%	16.8%	89.1%	99.7%	100.0%
0.5	0.67	0.1%	0.0%	0.0%	0.0%	15.9%	100.0%	77.9%	98.3%	9.2%	65.1%	96.3%	100.0%
0.5	0.69	0.3%	0.0%	0.0%	0.0%	8.5%	99.8%	47.3%	85.3%	4.1%	32.8%	77.0%	99.2%
0.5	0.71	1.1%	0.0%	0.0%	0.0%	3.8%	88.6%	18.5%	46.1%	1.7%	9.9%	32.4%	79.9%
0.5	0.73	3.0%	0.2%	0.6%	0.2%	1.2%	22.6%	3.0%	8.6%	1.2%	1.5%	3.9%	13.6%
0.5	**0.75**	**7.3%**	**7.3%**	**7.7%**	**9.0%**	**0.1%**	**0.3%**	**0.1%**	**0.3%**	**3.0%**	**3.1%**	**4.2%**	**3.9%**
0.5	0.77	14.4%	33.3%	46.6%	81.5%	0.1%	0.0%	0.0%	0.0%	7.4%	20.4%	33.1%	70.5%
0.5	0.79	28.2%	73.4%	91.8%	100.0%	0.1%	0.0%	0.0%	0.0%	16.1%	61.8%	85.5%	99.7%
0.5	0.81	43.8%	95.2%	99.8%	100.0%	0.0%	0.0%	0.0%	0.0%	31.8%	91.3%	99.3%	100.0%
0.5	0.83	62.9%	99.4%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	48.6%	99.0%	100.0%	100.0%
0.5	0.85	80.8%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	66.4%	100.0%	100.0%	100.0%
0.5	0.95	100.0%	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%

The point tested is $\beta^a = (\frac{1}{2}, a)^T$ for various values of a. $\beta^1 = (\frac{1}{2}, \frac{3}{4})^T$ is on the frontier of B.

Table 4: Income Regression: A Comparison between Exact and Partial Information

LogIncome	Exact amount: OLS	Midbands: OLS	Bands: Set OLS
Education	0.0678	0.0828	$[0.0445, 0.121]$
	$(0.0663, 0.0693)$	$(0.0775, 0.0881)$	$(0.0423, 0.123)$
Age	0.0513	0.0741	$[-0.097, 0.245]$
	$(0.0455, 0.0571)$	$(0.0517, 0.0965)$	$(-0.106, 0.254)$
$\text{Age}^2 : 10^{-3}$	-0.408	-0.652	$[-2.74, 1.44]$
	$(-0.479, -0.336)$	$(-0.925, -0.378)$	$(-2.86, 1.56)$
Intercept	6.62	5.95	$[2.57, 9.32]$
	$(6.50, 6.73)$	$(5.49, 6.41)$	$(2.39, 9.51)$
R^2	30.2%	36.4%	–
Observations	22917	2065	2065

Notes:

Table 5: Income Regression: A Comparison between Exact and Partial Information

LogIncome	Exact amount: 2SLS	Midbands: 2SLS	Band information: S2SLS
Education	0.0652	0.0758	$[0.0379, 0.114]$
	$(0.0635, 0.0668)$	$(0.0653, 0.0864)$	$(0.0355, 0.116)$
Age	0.0403	0.0575	$[-0.107, 0.222]$
	$(0.0344, 0.0462)$	$(0.0144, 0.1007)$	$(-0.116, 0.232)$
$\text{Age}^2 : 10^{-3}$	-0.275	-0.493	$[-2.51, 1.53]$
	$(-0.349, -0.202)$	$(-1.017, 0.031)$	$(-2.63, 1.64)$
LogHours	0.990	3.62	$[-0.50, 7.75]$
	$(0.752, 1.23)$	$(1.73, 5.52)$	$(-2.86, 8.18)$
Intercept	3.24	-6.93	$[-28.61, 14.75]$
	$(2.42, 4.07)$	$(-13.74, -.12)$	$(-30.14, 16.29)$
Sargan	514.92 (3 d.f.)	3.77 (3 d.f.)	P-value = 25.1%
Observations	22486	2015	2015

Notes:

Table 6: Supernumerary Instruments: Simple Regression

Polynoms			
Nobs	100	1000	10000
$m-p=0$	$[0.181, 0.529]$	$[0.109, 0.468]$	$[0.117, 0.489]$
	$(0.153, 0.560)$	$(0.107, 0.471)$	$(0.117, 0.490)$
$m-p=1$	$[0.189, 0.521]$	$[0.165, 0.412]$	$[0.168, 0.438]$
	$(0.161, 0.552)$	$(0.162, 0.415)$	$(0.167, 0.439)$
2	$[0.165, 0.412]$	$[0.166, 0.411]$	$[0.168, 0.439]$
	$(0.162, 0.415)$	$(0.163, 0.414)$	$(0.167, 0.439)$
3	$[0.271, 0.439]$	$[0.166, 0.411]$	$[0.168, 0.439]$
	$(0.242, 0.468)$	$(0.163, 0.414)$	$(0.167, 0.439)$

Sinusoids			
Nobs	100	1000	10000
$m-p=0$	$[0.171, 0.539]$	$[0.109, 0.468]$	$[0.117, 0.489]$
	$(0.143, 0.569)$	$(0.106, 0.471)$	$(0.117, 0.490)$
$m-p=1$	$[0.217, 0.493]$	$[0.169, 0.408]$	$[0.181, 0.426]$
	$(0.188, 0.523)$	$(0.166, 0.411)$	$(0.181, 0.426)$
2	$[0.220, 0.490]$	$[0.170, 0.407]$	$[0.188, 0.418]$
	$(0.191, 0.520)$	$(0.167, 0.410)$	$(0.188, 0.419)$
3	$[0.326, 0.384]$	$[0.178, 0.399]$	$[0.188, 0.418]$
	$(0.296, 0.413)$	$(0.176, 0.402)$	$(0.188, 0.419)$

Notes:

Table 7: Normality Test of $\tau_n((1,0)^T)$

n	$\chi^2(2)$ statistics	p-value
100	7.866	(0.020)
500	0.275	(0.871)
1000	0.310	(0.856)
2500	1.367	(0.505)