Set Identified Linear Models

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

First version, August 2006
This version, December 2007

Abstract

We analyze the identification and estimation of parameters β satisfying the *incomplete* linear moment restrictions $E(z^T(x\beta - y)) = E(z^T u(x))$ where z is a set of instruments and $u(z)$ an unknown bounded scalar function. We first provide several empirically relevant examples of such a set-up. Second, we show that these conditions set identify β where the identified set is bounded and convex. We provide a sharp characterization of the identified set not only when the number of moment conditions is equal to the number of parameters of interest but also in the case in which the number of conditions is strictly larger than the number of parameters. We derive a necessary and sufficient condition of the validity of supernumerary restrictions, which generalizes the familiar Sargan condition. We also construct a test of the null hypothesis, $\beta_0 \in B$, whose level is asymptotically exact and which relies on the minimization of the support function of the set $B - \{\beta_0\}$. Inverting this test makes it possible to construct confidence regions with uniformly exact coverage probabilities. Some Monte Carlo and empirical illustrations are presented.

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

1 Introduction[1]

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 intervals. 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 boundedness 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

[1]This paper was developed for the invited session that one of us gave at ESEM'06 in Vienna. We thank Richard Blundell, Andrew Chesher, Guy Laroque and Adam Rosen for helpful discussions. We thank the participants at seminars at PUC in Rio, IFS London, Paris Malinvaud seminar, Mannheim as well as in workshops and conferences (ESRC Bristol '07, Montréal Conference on GMM '07) for comments. The usual disclaimer applies.

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 the identified set is not empty. We sharply characterize the identified set and 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 Molinari (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 derived. Interestingly enough, our approach reveals that the asymptotic results of Beresteanu and Molinari (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, $H_0{:}B_0 \subset B$). For example, the generalized 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 or is not strictly convex (faces). 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 or inequality restrictions only. 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. Andrews and Guggenberger (2007) studies cases that do not fall under the assumptions of Imbens and Manski (2004) or Stoye (2007).

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, 2008), 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 developed 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 analyze 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 sharply 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 condition. For the sake of simplicity, Section 4 specializes to the case of outcomes measured by intervals. Under general conditions, we derive asymptotic properties of estimates in the case of no moment restrictions in surplus. We develop exact test procedures, construct exact confidence regions by inversion of the tests and derive asymptotic properties of the 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 is 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 a random shock, 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.[2] 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 $w = (y, x)$ where y is the result of censoring a latent variable y^* by intervals, $y^* \in [y_0, y_K)$ being a bounded

[2] 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.

outcome[3], x a vector of L covariates. If $y_0 < y_1 < ... < y_{K-1} < y_K$ denote the bounds of the K intervals, y can be re-defined as the center of the observed intervals,[4] i.e.

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

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 :

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

where ε is a random variable uncorrelated with x, $E(x^T\varepsilon) = 0$. The distribution of ε conditional on x is denoted $F_\varepsilon(. \mid x)$. The issue is to characterize B_1 the set of parameters β such that the latent model $(\beta, F_\varepsilon(. \mid x))$ generates the distribution $G_k(x)$ $(k = 0, ...K-1)$ using equation (3). By definition, β is in B_1 if and only if there is a random variable, ε, uncorrelated with x satisfying,

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

In other words, the identified set B_1 is defined as,

$$B_1 = \left\{\beta \in \mathbb{R}^K \text{ s.t. there exists } F_\varepsilon(. \mid x) \text{ satisfying Eq. (5) and such that } E(x^T\varepsilon) = 0\right\}$$

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

[3] 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 for the mean (see Magnac and Maurin, 2007, for an example).

[4] The choice of the mid-point is just a normalization as proved in Appendix A.2 after the proof of Proposition 1 below. Any choice of y which preserves the ordering of y^* does not affect the developments that we now analyze.

$\Delta(x) = \frac{1}{2}\sum_{k=0}^{K-1}\left[(y_{k+1} - y_k)(G_{k+1}(x) - G_k(x))\right]$
and such that, $E(x^T(x\beta - y)) = E(x^T u(x))$.

Proof: See Appendix A.1.

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). Suppose that this outcome is given as a function of a latent intensity $y^* = x\beta + \varepsilon$, by:

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

where y_1 is an unknown threshold in $(0, 1)$. In addition, we assume that y^* varies between 0 and 1 and that ε is uncorrelated with x.

The intercept of the model (say β_0) and the threshold y_1 cannot be jointly identified and we can always set $y_1 = \frac{1}{2}$ as a normalisation. In the same spirit as in the previous section, a discrete variable, y, can be defined (as $y = \frac{1}{4}1(y^* < \frac{1}{2}) + \frac{3}{4}1(y^* \geq \frac{1}{2})$ so that,

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

Function $u(x)$ 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 identified 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 belongs to the identified set if and only if there is a set of thresholds $y_1 = 0 < \cdots < y_k < \cdots < y_{K+1}$ 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 identified 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 corresponds 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 exceeds 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, $-v$, is the price of this object, it simply amounts to assume that for sufficiently high (respectively low) price no one (respectively 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 (2007), and β is point identified. In contrast, when v is not observed continuously[5], the set B_3 of observationally equivalent parameters is not a singleton anymore.

To be more specific, 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, parameter β belongs to B_3 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, 2008) 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) $\beta \in cl(B_3)$

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

(ii) there exists a function $u(x)$ taking its values in $I^*(x) = [\underline{\Delta}^*(x), \overline{\Delta}^*(x)]$ where (by convention, $G_0(x) = 0$, $G_K(x) = 1$),

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

and such that,

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

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 (Magnac and Maurin, 2008).

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:

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

These variables are easy to construct in examples 1, 2 and 3 that were developed above (See Appendix A.3). Moreover, in the case of Manski and Tamer (2002), where y is observed by interval only, the lower and upper bounds $\underline{y}$ and $\overline{y}$ are part of the dataset and so explicitly given. Where they are not easy to construct, the methods that we will propose are more computationally challenging.

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 $p+m-r$ are the potential overlaps and functional dependencies.[6]. 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 (ILM). 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:

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

[6]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 whose probability measure is equal to 1).

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

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,

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

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

Proof. See Appendix B.1. ■

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:

$$\lim_{\Delta_M \to 0} B = \{\beta^*\}, \tag{10}$$

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:

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

Given 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. $\mathbb{S} = \{q \in \mathbb{R}^p; \|q\| = 1\}$. 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)).[7] Convex set B is sharply characterized by:

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

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

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

3.1.2 The Support Function

We now show that the support function of B can be written as a function of simple moments of the data.

Let q a element of the unit sphere $\mathbb{S} \subset \mathbb{R}^p$ i.e. $\|q\| = 1$. Consider Q an orthogonal matrix of dimension $[p, p]$ whose last column is vector q. It can be written $Q = (Q_0, q)$ where Q_0 is a matrix of dimension $[p, (p-1)]$.[8] 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 (Davidson and McKinnon, 2004, for instance). 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*

$$z_q = z.E(x^Tz)^{-1}.q, \quad w_q = \mathbf{1}\{z_q > 0\}\overline{y} + \mathbf{1}\{z_q < 0\}\underline{y}.$$

The support function of B is equal to:

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

The interior of B is not empty and:

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

is a frontier point of B such that $\delta^*(q \mid B) = q^T\beta_q$.

[8] 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.

Proof. See Appendix B.2 ■

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

Many interesting properties that we derive in the following, depend on the following Lemma. This result is a consequence of the fact that set B is bounded and convex so 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 B. ■

Note that the expression of the derivative results from 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 for which we can derive L frontier points β_{q_l} using proposition 5 and we would construct:

$$\overline{B}_L = \cap_{q_l : \|q_l\|=1} \{\beta : q_l^T \beta \leq q_l^T \beta_{q_l}\}.$$

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 $\{q_l\}_{l=1,.,L}$ are appropriately chosen so that:

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

Alternatively, the convex hull, $\underline{B}_L$, of the L frontier points, β_{q_l}, is included in the 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 $\{q_l\}_{l=1,.,L}$ 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. For reasons that will become clearer later, we rewrite the incomplete linear moment conditions (7) as:

$$E(z^T x)\beta = E(z^T (y + u(z))). \tag{11}$$

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, select p instruments out of the possible m ones and construct the corresponding identified 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 in which ($m = p$) is that set B could be empty. In the next sub-section, we derive a necessary and sufficient condition which generalizes the usual over-identifying condition à la Sargan.

3.2.1 The Validity of Supernumerary Moment Conditions

As some moment conditions are in surplus, we decompose equation (11) into two subsets:

Lemma 9 *We can write $m \times 1$ random vector z as a linear combination of two sets of variables z_F and z_H of respective dimensions, p and $m-p$, such that equation (11) is equivalent to:*

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

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

Additionally:

$$E(z_F^T z_F) = I_p, E(z_H^T z_H) = I_{m-p}, E(z_F^T z_H) = 0_{p,m-p}.$$

Proof. See Appendix B ■

Interestingly enough, the second equation does not depend on β anymore and the first equation identifies a set. 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.:

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

The adapted Sargan condition means simply that B_{Sargan} contains the point $\gamma = 0$, that is O_{m-p}, the origin point of $\mathbb{R}^{m-p}$.

Proposition 10 *The two following conditions are equivalent:*

i. B is not empty,

ii. $B_{Sargan} \ni O_{m-p}$.

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 $O_{m-p} \in B_{\text{Sargan}} = \{E(z_H^T.y)\}$ or equivalently that $E(z_H^T y) = 0$. 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 10.

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:

$$\begin{aligned} E(z^T(x\beta - y)) &= E(z^T u(z)), \\ \text{subject to } u(z) &\in [\underline{\Delta}(z), \overline{\Delta}(z)]. \end{aligned}$$

This program can be rewritten by introducing auxiliary parameters, γ, as:

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

under the same constraint for $u(z)$. Let B_U (U for unconstrained) be the set of (β, γ) satisfying the relaxed program,

$$\begin{aligned} E(z^T(x\beta + z_H\gamma - y)) &= E(z^T u(z)), \\ \text{subject to } u(z) &\in [\underline{\Delta}(z), \overline{\Delta}(z)]. \end{aligned}$$

An interesting feature of this definition of B_U is that the number of explanatory variables is equal to the number of moment conditions, and no more moments are in surplus. The support function of B_U can then be characterized using Proposition 5. Also, we can build on B_U to provide a very simple geometric characterization of B:

Lemma 11 *The identified set B is the intersection of B_U and the hyperplane defined by $\gamma = 0$ (see Figure 4).*

Proof. Straightforward. ■

This intersection is not empty under the condition of Proposition 10, i.e., $B_{\text{Sargan}} \ni O_{m-p}$, that we shall assume from now on.

We can use general solutions for finding the support function of the intersection of convex sets. 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 the hyperplane $\gamma = 0$ using any projection direction. As the identified set is an intersection, its support function can be expressed as the minimum of support functions of the unconstrained set B_U. The analytic characterization of the identified set is indeed given by:

Proposition 12 *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. Rockafellar (1970) and Appendix B ■

The meaning of $\lambda_m(q)$ is the following. For any point $\beta_f \in \partial B$, 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_f as a frontier point. This projection direction, described by $\lambda_m(q)$, is simply the tangent space (not necessarily unique) of B_U at β_f.

Note also that, the specific orthogonal projection of B_U onto $\gamma = 0$ is:

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

the set of unconstrained solutions to equation (12). It contains set B because the function $u(z)$ should also satisfied equation (13). Supernumerary restrictions reduce the size of the identified set.

3.2.3 The Example of a Single Supernumerary Instrument

When there is only one supernumerary instrument, z_H is a real random variable and:

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

where $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. The necessary and sufficient condition, $O \in B_{\text{Sargan}}$, is equivalent to:

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

$$\begin{aligned} \overline{U} &= E(|z_H|\,(\mathbf{1}\{z_H > 0\}\overline{\Delta}(z) - \mathbf{1}\{z_H < 0\}\underline{\Delta}(z))) \\ \underline{U} &= E(|z_H|\,(\mathbf{1}\{z_H > 0\}\underline{\Delta}(z) - \mathbf{1}\{z_H < 0\}\overline{\Delta}(z))). \end{aligned}$$

Testing overidentification boils down to testing that $(E(z_H y) - \underline{U})(E(z_H y) - \overline{U})$ is negative.

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

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

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

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

The next subsection extends 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 $O_{m-p} \in int(B_{\text{Sargan}})$, functions that satisfy the Sargan 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_{m-p} belongs to the frontier of B_{Sargan} i.e. $O_{m-p} \in \partial B_{\text{Sargan}}$. 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} to characterize the geometric properties of its frontier.

Consider first that O_{m-p} belongs to a face of B_{Sargan} and let q_O the vector of $\mathbb{S}_{m-p}$, orthogonal to this face. By Proposition 7 $\Pr\{z_{Hq_O} = 0\} > 0$ and the generating function $u_{q_O}^{\text{Sargan}}(z)$ is not unique. The identified set B is generically not reduced to a singleton. In contrast, if O_{m-p} is not on a face of B_{Sargan}, $u_{q_O}^{\text{Sargan}}(z)$ is unique when q_O is the vector, orthogonal to the supporting hyperplane of B_{Sargan} at O_{m-p}. By Proposition 7 (ii), it is true whether q_O is unique or not (i.e. B_{Sargan} has a kink at O_{m-p}). In consequence, 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:

Proposition 13 *If O belongs to the frontier ∂B_{Sargan} albeit not to a face of B_{Sargan}, point identification of β is restored.*

4 Estimation and Inference

This section provides a description of how we estimate the support function of B, how we test hypotheses of interest and how we construct confidence intervals. 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 identical distributions 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 :

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

where:

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

The second equality comes by transposition and using that z_q is a random scalar.

To apply the analogy principle and construct an estimate, we define $\hat{\Sigma}_n$ as a cross-moment empirical analogue to $E(x^T z)^{-1}$.[9] Define also 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:

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

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

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 are support functions):

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

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:

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

Proposition 14 *Let Assumption C. Then, the estimator of the support function is, uniformly over* $\mathbb{S}$, *strongly consistent*:

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

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 any function, $z_q w_q$, appearing within the expectation, belongs to a parametric class as a function of Σ and q. Under condition C and the boundedness assumption, this class is a Glivenko-Cantelli class and the uniform consistency result applies. Second, we replace parameter Σ by a consistent estimate and show consistency using 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 $\mathbb{S}$:

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

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

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

Proposition 15 *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 as for point identification is used and this explains why the distribution of the estimate of Σ does not play any rôle in the asymptotic distribution of the estimates (see Appendix C.1.3).

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 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[10]. More importantly, we will see that our approach is easily extended to the more difficult case where there are supernumerary instruments.

[10] In Appendix C.1.4, we explain why our approach provides us with an expression of the covariance function $Cov(\tau(q)\tau(r))$ which is a simplification of what is provided by Beresteanu and Molinari (2006).

4.2 Tests

We propose here a test of the null hypothesis $H_0 : \beta_0 \in B$, which is not conservative. Its level is asymptotically equal to a given value α. It is worth noting that the test proposed here is asymptotically pivotal. We could therefore increase the finite sample properties of the test by bootstrapping it. We first rewrite the null hypothesis in terms of support functions and consider its empirical analogue. We then derive its asymptotic distribution under the null and different alternatives.

As B is a convex set, an alternative characterization of the null hypothesis is (see above and Rockafellar, 1970):

$$\forall q, \|q\| = 1, T_\infty(q; \beta_0) = \delta^*(q|B) - q^T \beta_0 \geq 0.$$

The least favorable case under H_0, is given by $\beta_0 \in \partial B$, the frontier of B. In such a case there exists q_0 such that $T_\infty(q_0; \beta_0) = 0$, the minimum value of $T_\infty(q; \beta_0)$ over $\mathbb{S}$. Let $\mathcal{Q}_0$ the set of all such minimizers:

$$\mathcal{Q}_0 = \arg\min_{q \in \mathbb{S}} T_\infty(q_0; \beta_0).$$

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:

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

For any $q_0 \in \mathcal{Q}_0$, proposition 15 tells us that,

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

is asymptotically normally distributed with variance $V_{q_0} = q_0^T \Sigma^T V(z^T \varepsilon_{q_0}) \Sigma q_0$. Yet, as q_0 in $\mathcal{Q}_0$ are unknown, we have to replace it by an estimate defined by analogy as:

$$q_n \in \arg\min_{q \in \mathbb{S}} T_n(q; \beta_0).$$

The test that we propose, is based on the statistic $T_n(q_n; \beta_0)$ and uses the following result:

Proposition 16 *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 $\hat{V}_n = q_n^T \hat{\Sigma}_n^T \hat{V}(z^T \varepsilon_{q_n}) \hat{\Sigma}_n q_n$.

Proof. See appendix C.2 ■

The sketch of the proof is the following. First, $\sqrt{n}T_n(q_n;\beta_0)$ is the sum of two terms that can be handled separately:

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

By proposition 15, 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 because q_n is a minimizer of it and its asymptotic distribution is non standard. As q_0 and q_n are minimizers, q_n is a superconsistent estimate of q_0 and replacing q_0 by q_n has no influence on the asymptotic distribution to the first order. Moreover, 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 at a kink of B or not). What matters is not the uniqueness of the solution to the minimization problems but the uniqueness of the value of the criterion function.

Let us now examine alternative cases in which $\beta_0 \in Int(B)$ or $\beta_0 \notin B$. When $\beta_0 \in Int(B)$, $\forall q \in \mathbb{S}$, $T_\infty(q;\beta_0) > 0$, and when $\beta_0 \notin B$, 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 $\beta_0 \notin B$:

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

In consequence, when β_0 is not on the frontier of B, $T_n(q_n;\beta_0)$ is either strictly positive or strictly negative, for n large enough. Combining this result with Proposition 15 yields:

Proposition 17 *Let $\beta_0 \in \mathbb{R}^d$ and*

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

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 $\hat{V}_n = q_n^T\hat{\Sigma}_n^T\hat{V}(z^T\varepsilon_{q_n})\hat{\Sigma}_n q_n$ is a consistent estimator of V_{q_0}.

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

$$\xi_n(\beta_0) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1),$$

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

$$\xi_n(\beta_0) \xrightarrow[n\to\infty]{a.s.} +\infty$$

and if β_0 *does not belong to* B,

$$\xi_n(\beta_0) \xrightarrow[n\to\infty]{a.s} -\infty.$$

Proof. See Appendix C.3. ■

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:

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

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

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

- 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:

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

We are specifically 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 $O \in \partial B_{\text{Sargan}}$).

4.3 Confidence Regions

By inverting the first test developed previously with a level of significance equal to α, we can construct confidence regions of nominal size asymptotically equal to $100 - 100\alpha$ %. Lehmann (1986, Chapter 3) defines the confidence region CI_α^n as the collection of parameters $\beta \in \mathbb{R}^d$ for which the null hypothesis is rejected *i.e.* which does not belong to $W_n^1(\alpha)$. The following proposition expresses this statement and Appendix C.4 provides a simple way of constructing the confidence region.

Proposition 18 *Let* α *be a significance level, and let* CI_α^n *be the set of points of* $\mathbb{R}^d$ *such that* $\xi_n(\beta) > \mathcal{N}_\alpha$*:*

$$CI_\alpha^n = \left\{\beta \in \mathbb{R}^d \ s.t \ \xi_n(\beta) > \mathcal{N}_\alpha\right\}.$$

Then,

$$\lim_{n\to+\infty} \inf_{\beta\in B} Pr\left(\beta \in CI_{\alpha}^{n}\right) = 1-\alpha.$$

The limit expressed in the proposition is valid for a fixed data generating process leading to the identification of a set B which is not a singleton. It is not uniformly valid for all data generating processes even if they satisfy the condition, under which we work, that the corresponding identified set B has a non-empty interior. As a consequence, the confidence interval is not uniformly asymptotically of nominal size equal to $(1-\alpha)$. When the set is arbitrarily close to a singleton in a sense made precise below, the nominal size tends to $1-2\alpha$. In a context of interval estimation, Imbens and Manski (2004) propose another construction which is uniformly valid. More recently, Stoye (2007) clarified the conditions under which this result can be obtained. What we show now is that it can be adapted to our set-up.[11].

In our framework, the analogue of the length of the interval Δ in Imbens and Manski (2004) and Stoye (2007) is the maximal *thickness* of set B in any direction, that is, the maximum of $\delta(q|B)+\delta(-q|B)$ on the unit sphere.[12] Assumption R*ii* ensures that this parameter is strictly positive and that the estimated set is almost surely non empty (see also Appendix B.2).

The empirical counterpart $\hat{\Delta}_n$ of the maximal thickness of set B is:

$$\hat{\Delta}_n = \max_{q\in\mathbb{S}} \left(\hat{\delta}_n^*(q|B) + \hat{\delta}_n^*(-q|B)\right). \tag{20}$$

Because it is a maximum, $\hat{\Delta}_n$ is a superefficient estimator of Δ, *i.e.*

$$\sqrt{n}\left(\hat{\Delta}_n - \Delta\right) \xrightarrow[n\to\infty]{P} 0,$$

using a proof analogue to the one developed in Proposition 16.

The next proposition provides an extension of Lemma 4 of Imbens and Manski (2004) in the multivariate case for constructing a uniform confidence region:

Proposition 19 *Let*

$$\hat{\sigma}_n = \sqrt{\hat{V}_{q_n}} = \sqrt{q_n^T \hat{\Sigma}_n \hat{V}(z^T \varepsilon_{q_n}) \hat{\Sigma}_n q_n},$$

where q_n is the argument of the maximum of equation (20).

[11] Stoye also extends the construction of the confidence region to the case in which the estimated size is not a superefficient estimator of the true one although it remains asymptotically normal. In a more general case, Andrews and Guggenberger (2007) focus on the construction of confidence regions using subsampling techniques when the assumption of asymptotic normality is no longer valid.

[12] It is indeed a maximum because of the compacity of the unit sphere.

A confidence region $\tilde{CI}_\alpha^n$ of asymptotic level equal to $1-\alpha$ is defined by the collection of the points such that $\xi(\beta) \geq \tilde{\mathcal{N}}_{1-\alpha}$ where $\tilde{\mathcal{N}}_{1-\alpha}$ satisfies the equation

$$\Phi\left(\tilde{\mathcal{N}}_{1-\alpha} + \sqrt{n}\frac{\hat{\Delta}_n}{\hat{\sigma}_n}\right) - \Phi(-\tilde{\mathcal{N}}_{1-\alpha}) = \alpha.$$

$$\lim_{n\to+\infty} \inf_{\beta\in B,\Delta\geq 0} Pr\left(\beta \in \tilde{CI}_\alpha^n\right) = 1-\alpha.$$

4.4 Consistent and Asymptotically Normal Estimation: Some Supernumerary Conditions

We use the characterization given by Proposition 12 and equation (16). If q is a vector of $\mathbb{R}^p$ and (q,λ) a vector of $\mathbb{R}^m$, we have:

$$\delta^*(q|B) = \inf_\lambda \delta^*((q,\lambda)|B_U)).$$

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

Let $\hat{\delta}_n^*((q,\lambda)|B_U)$ the estimate of $\delta_n^*((q,\lambda)|B_U)$ as derived in Section 4.1 and such that, by Proposition 14:

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

and, by Proposition 15:

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

converges to a Gaussian process when n tends to infinity.

For any q, define:

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

as it was developed quite similarly in Section 4.2. Define:

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

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 20 *Under the respective conditions C and AN we have:*

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

and:

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

converges to a Gaussian process when n tends to infinity.

5 Monte-Carlo Simulations of Testing Procedures

In this section, we develop three simple experiments to assess the performance of our inference and test procedures. In these experiments, the dependent variable is bounded and censored by intervals, as in the first example of Section 2 and we focus on the case where the identified set is of dimension 2 for simplicity. In the first two experiments, the identified set is smooth and strictly convex since the frontier of the identified set has no kinks and no faces. In the first experiment, the number of instruments is the same as the number of parameters while we use one supernumerary instrument in the second experiment. We explore the case of an identified set that has kinks and faces in the third experiment.

5.1 Smooth and Strictly Convex Sets

Consider the model:

$$y^* = 0.x_1 + 0.x_2 + \varepsilon,$$

where $x^T = (x_1, x_2)^T$ is a standard normal vector while ε is independent of x and uniformly distributed on $[0, 1]$. As a consequence, the true value of β is $(0, 0)^T$. We assume that y^* is observed by intervals, as in Example 1 of Section 2, where there are K such intervals ($I_k = [k/K; (k + 1)/K]$, $k = 0...K - 1$).

In Appendix D.1, we show that the support function of the identified set B boils down to a constant :

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

where $\Delta = \frac{1}{2K}$ is equal to the half length of the intervals. In other words, the identified set B is a circle whose radius is $\frac{2\Delta}{\sqrt{2\pi}}$ (see Figure D.3). When the number of intervals, K, tends to infinity, Δ tends to 0 and the identified set shrinks to a singleton.

5.1.1 Set Estimation

We draw 1000 simulations of the model for four different sample sizes : n =100, 500, 1000 and 2500. For simplicity, the number of intervals, K, is hold constant and equal to 2. We investigated, in practice, that the results are robust for larger values of K. Figure 5 displays the true support function $\delta^*(q \mid B)$ as well as the mean (dashed line) and quartiles of the distribution of $\hat{\delta}_n^*(q \mid B)$, when q varies over the unit sphere. Even for small sample size, the identified set is well estimated and unsurprisingly, the interquartile interval decreases when the sample size increases.

By Proposition 15, $\sqrt{n}\left(\hat{\delta}_n^*(q \mid B) - \delta^*(q \mid B)\right) = \tau_n(q)$ is asymptotically normally distributed. Table 8 reports a normality test (see Bontemps and Meddahi, 2005) of $\tau_n(q_0)$ when $q_0 = (1, 0)$, for the four sample sizes. The normal approximation is not rejected at the 5% level, except in the smallest sample, n =100.

5.1.2 Testing Procedures

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

Let $\beta^0 = 0$ be the center of B and let β^r a point on a ray such that the distance between 0 and β^r is equal to r times the value of the radius of B. As B is a circle around the true value $\beta^0 = 0$, this procedure and definitions are valid for any ray. Moreover, β^r belongs to B if and only if $r \leq 1$ and β^1 belongs to the frontier. For r varying stepwise from 0 to 3, Table 1 reports the rejection frequencies at a 5% level of 3 different tests; Whether β^r belongs to B against the alternative that it does not (Test 1); whether it does not belong to B (Test 2); whether it belongs to the frontier of B against the alternative that it does not (Test 3). The results that correspond to the frontier point ($r = 1$) are reported in bold.

The main conclusion of these experiments is first that the size of the three tests is very accurate and remain very close to 5% even for $n = 100$. Unsurprisingly, the power of the tests increases with the sample size, yet, it remains very good even for small sample sizes.

5.2 Smooth set with one supernumerary instrument

The simulated model is the same as before although we now assume that the second explanatory variable x_2 is expressed as:

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

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 another variable where e_4 is i.i.d. standard normal. Parameter π (respectively ν) measures the strength of the correlation between x_2 and e_2 (respectively w).

We assume in this part that we use x_1, e_2 and w for estimating the set B instead of x_1 and x_2. We have therefore one supernumerary instrument.

When $q = (\cos\theta, \sin\theta)^T$, the support function (see Appendix D.2 for further details) can be expressed as:

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

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 π and ν are strictly positive but strictly lower than 1, there is some information loss due to the use of e_2 and w instead of x_2. Set B is therefore stretched along the second axis. Figure D.3 displays set B when $\pi = \nu = 0.6$.

As before, we draw 1000 simulations of the model for four sample sizes : $n =$100, 500, 1000 and 2500. For each draw, we estimate B and implement the three tests as defined in Section 4.2.

Figure 6 displays set B as well as the mean and quartiles of the estimated sets for the four sample sizes. Tables 2 displays the percentage of rejections for the three tests for different points along a radius.

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

5.3 A set with kinks and faces

In this experiment, the explanatory variable has mass points (so that the identified set has faces) and its support is discrete (so that the identified set has kinks). The simulated model is:

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

where x is equal to $-\frac{1}{2}$ with probability $\frac{1}{2}$ and to $\frac{1}{2}$ with probability $\frac{1}{2}$ and where ε is independent of x and is uniformly distributed on $[-\frac{1}{4}, \frac{1}{4}]$. The true value of β is $(\frac{1}{2}, \frac{1}{4})^T$. As before, we assume that, instead of observing the exact value, y^*, we only observe it by intervals (2 intervals:

$I_1 = [0, \frac{1}{2}]$ and $I_2 = [\frac{1}{2}, 1]$). In this setting, the identified set B_2 can be shown to be 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).

As in the previous example, we simulate 1000 draws for 4 sample sizes: 100, 500, 1000 and 2500. Figure 7 displays set B_2 as well as the average and quartiles of the estimated sets for the four sample sizes. Tables 3 and 4 display the percentage of rejections for the three tests at different points (shown in bold in Figure 3) as in the previous example. The estimation and test procedures seems to work as well, even for small sample sizes.

6 Empirical Illustrations

This section provides an illustration of our estimation and test procedures using real survey data. In the French Labor Force Survey (LFS), respondents can either report their exact labor income or indicate the interval within which their income lies. We focus on males aged 25 to 55 ($N \simeq 25000$) and we distinguish two samples, one of *exact responses* ($N \simeq 23000$) and one of *interval responses* ($N \simeq 2000$). We consider the familiar earning model:

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

where R_i represents monthly labor income whereas education, age and age squared are the explanatory variables. One question of interest is whether the two populations are similar, at least with respect to parameter β. Let β_e the parameter in the 'exact response' population and B_i the identified set in the 'interval response' sample. The null hypothesis is $\beta_e \in B_i$ or more exactly $0 \in B_i - \{\beta_e\}$.

There are ten income intervals in the survey and, because the dependent variable is set in logs, the extreme intervals are open on the left and on the right. In order to conform with the setting developed in Example 1 and Appendix A, we chose to fix the lower and upper bound of the income range to the arbitrary values of 200FF (30€) and 70000FF (10500€). These lower and upper bound corresponds to those actually observed in the 'exact response' population.

6.1 Results

Table 5 reports the 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 sample of interval responses and imputing (incorrectly) the mid-value of the reported band for the dependent variable. The third column reports the estimated intervals for each coefficient according to our estimation procedures. The confidence intervals refer to the parameters of interest, not to the intervals of interest (see Imbens and Manski, 2004). We also plotted the estimated set of dimension 2 that corresponds to the coefficients of age and education (see 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 slightly less for the coefficient of education. As returns to education are much more precisely estimated, they are still significantly positive in a large range however, from 4.5% to 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.

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 number of children less than 18 years old in the household and a dummy variable for the presence of a spouse as well as education and income reported by the spouse (if any).

Results are given in Table 6. The same picture as previously emerges. In particular, it is impossible to reject the validity of the supernumerary restrictions in the interval data case, even if in the complete sample they are rejected very strongly indeed.

6.2 Mean Independence as a Source of Identification

A very simple way of adding instruments to a model is to use an assumption of mean independence rather than one of absence of correlation. Specifically, we consider one explanatory variable x and one random term ε which are both truncated to lie into $[0, 1]$ and which are both truncated normally distributed. We assume that,

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

is censored by intervals and that the number of equally spaced intervals is equal to 10 i.e.

$$y^* \in [(i-1)/10, i/10).$$

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 7. 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 developed in this paper can be easily extended to other settings 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 context, we consider the familiar linear model,

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

where ε is uncorrelated with the set of regressors (x, v) of respective dimensions K and L. The issue is to characterize B_4, 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_4.

As we are analyzing linear projections, the only useful information about v comes from linear predictions. 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 covariance matrix of the disturbances 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 identified set B_4 if and only if there is an α and a η such that,

$$E(x^T(y - x\beta)) = E(x^T(x\rho\alpha) \text{ and } E(\eta^T y) = \Omega^{1/2}\alpha. \tag{21}$$

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 identified set B_4 if and only if there is an α^* in the compact ball $\|\alpha^*\| \leq 1$ such that,

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

where B_4 is clearly non empty ($\beta^* \in B_4$) and convex. It is sharply characterized by its support function,

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

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 21 *The identified set B_4 is non empty, closed, convex and bounded. It is characterised by the following support function,*

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

The function $\delta^*(q|B_4)$ 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 develop in this paper a class of models defined by incomplete linear moment conditions and we provide examples of how this set up can be applied to economic data. In the most prominent one, the dependent variable in a linear model is censored by intervals. We present simple ways that lead to a sharp characterization of the identified sets. We generalize previous results about estimating such sets and we construct asymptotic tests for null hypotheses concerning the true value of the parameter of interest. These procedures are easy to implement and we can invert them and derive confidence regions for the parameter of interest. We also generalize the simple setting of linear prediction using explanatory variables to the case in which supernumerary moment conditions are available. Specifically, we provide an extension to the usual Sargan test that can be performed using the asymptotic tests that we develop. Asymptotic properties of these generalized estimates are derived.

Various extensions are possible but out of the scope of this paper. First, some examples that we developed require more work in terms of estimation and asymptotic theory. For categorical data on subjective outcomes, the identified set is the union of elementary identified sets which are the focus of this paper. In this case, our set-up provides a building block to study the asymptotic properties of estimates constructed by analogy. For binary data with discrete or interval-valued regressors, the asymptotic properties of estimation would be the result of marrying the results of this paper with those of Lewbel (2000). Finally, the two sample-model requires a more specific approach although it remains in the framework that we developed.

Econometric assumptions can be questioned and extended. For simplicity, we focus on the case in which instruments and errors are not correlated. In structural settings, we would rather impose a stronger condition of mean independence between instruments and errors or even stronger of independence between instruments and errors. As is well known, mean independence (respectively independence) generates an infinite number of moment conditions given by the absence of correlation between any function of instruments and errors (respectively any function of errors). We presumably could use our framework by using only a finite number of moment conditions although the extension to the general case is worth pursuing. It also begs the question of the optimality of inference in the supernumerary restriction case and how it differs from the usual point-identified case.

Along a different vein, our setting remains global and semi-parametric. For non parametric estimation, it would be interesting to adapt our set-up to local approaches such as local linear regression. Other questions are open and seem worth pursuing. The gain of the direct approach that we used with respect to the approach followed by Chernozhukov et al. (2007) using a criterion is an interesting question. It is easy to write a criterion function using support functions (see Magnac and Maurin, 2008). It might be the case that our results help select the best criterion in the latter framework but this is left for future work.

Finally and more ambitiously, the deep foundation of our approach is a convexity argument. It indeed allows to replace the problem of identify a set in a very general space of sets by a problem which is finite dimensional since it requires to identify and estimate a function using finitely many parameters, the vectors of the unit sphere of $\mathbb{R}^p$. This approach can presumably be extended to any set identified problem when the set is convex. The problem of identifying the frontier of this set might be highly non linear although the real issue is to construct the support function, or the limits of the projection of the identified set in any direction q. Estimation and inference would likely follow from our arguments under adapted conditions.

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.

Andrews, D.W.K. and P., Guggenberger, 2007, ”Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities”, Unpublished manuscript, Cowles Foundation, Yale University.

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, ”Contingent Valuation and Social Choice”, *American Journal of Agricultural Economics*, 76:689-708.

Magnac, T. and E. Maurin, 2007, ”Identification and Information in Monotone Binary Models”, *Journal of Econometrics,* 139:76-104.

Magnac, T. and E. Maurin, 2008, ”Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data”, forthcoming *Review of Economic Studies*.

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.

Romano J.P., and A.M. Shaikh, 2005, ”Inference for Identifiable Parameters in Partially Identified Econometric Models”, working paper.

Stoye, J., 2007, ”More on Confidence Intervals for Partially Identified Parameters”, unpublished manuscript.

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

van der Vaart, A.W., and J.A., Wellner, 1995, *Weak Convergence and Empirical Processes*, Springer Verlag: Berlin.

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

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 $(\beta, F_\varepsilon(.|x))$ generates $\{G_k(x)\}_{k=1,.,K}$ through model (4). By definition, the distribution of ε satisfies,

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

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

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

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

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

Using Equation (3), we easily obtain bounds:

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

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 upper strict inequality for $u(x)$ by a weak inequality. This equation implies that $u(x)$ is uniformly bounded so that $E(x^T u(x))$ exists under condition $R.iii$. 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 $F_\varepsilon(.|x)$ such that ε is uncorrelated with x and such that the image of $(\beta, F_\varepsilon(.|x))$ 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:

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

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

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

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

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

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

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

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

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

$$\begin{aligned} &= \sum_{k=0}^{K-1} (\frac{y_{k+1}+y_k}{2} - E((1-\lambda)y_k + \lambda y_{k+1})|x)(G_{k+1}(x) - G_k(x)) \\ &= \sum_{k=0}^{K-1} E(1/2 - \lambda|x).(y_{k+1} - y_k).(G_{k+1}(x) - G_k(x)) \\ &= (-u(x)/(2\Delta(x)))(2\Delta(x)) = -u(x). \end{aligned}$$

Therefore, we have $E(y - x\beta|x) = -u(x) + E(\varepsilon|x)$, 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

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

Then:

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

where:

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

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

Thus:

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

where:

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

Remark that the length of the interval, $\overline{\Delta}(x)-\underline{\Delta}(x)$, does not depend on $\{v_k\}_k$ since it is equal to:

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

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 $-\underline{\Delta}(x) = \bar{\Delta}(x) = \Delta(x)$.

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,

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

Example 3: See Magnac and Maurin (2008). 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:

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

and using the definition of W, we have:

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

Using the generalized Cauchy-Schwartz inequality,

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

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.*

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 $\|q\| = 1$. By definition, the support function is the supremum of $q^T.\beta$ when $\beta \in B$. Furthermore, equation (9) implies that for any $\beta \in B$, there exists a function $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$ such that the regression of $y + u(z)$ on x yields parameter β.

We shall first write $q^T.\beta$ when $\beta \in B$ as the result of such a regression. Second, we will analyze the supremum of $q^T.\beta$ over $\beta \in B$.

First, fix $u(z)$ and $\beta \in B$. 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, $x.q$. Using equation (9), parameter, $q^T.\beta$, is the coefficient of one explanatory variable only, $x.q$, in the regression of $y + u(z)$ on x.

The natural tool to derive such a scalar coefficient is the Frisch-Waugh theorem. 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:

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

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 $(p-1)$. Applying the Frisch-Waugh theorem, we get:

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

where $z_q = \frac{x_q}{E({x_q}^2)}$. We will prove later that the alternative definition of z_q provided in the proof is valid.

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

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

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 $\delta^*(q|B)$ 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:

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

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:

$$\begin{aligned} x_0 &= z(E(z^Tz))^{-1}E(z^Tx_0) - \\ &\quad z(E(z^Tz))^{-1}E(z^Tx_{-0})\left[E(x_{-0}^Tz)(E(z^Tz))^{-1}E(z^Tx_{-0})\right]^{-1}E(x_{-0}^Tz)(E(z^Tz))^{-1}E(z^Tx_0). \end{aligned}$$

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

$$\begin{aligned} x_q &= z(E(z^Tz))^{-1}E(z^Tx).a(q) \qquad \text{(B.8)} \\ &= zD.a(q) \end{aligned}$$

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

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

$$\begin{aligned} a(q) &= q - Q_0\left(Q_0^TE(x^Tz)E(z^Tz)^{-1}E(z^Tx)Q_0\right)^{-1}Q_0^TE(x^Tz)E(z^Tz)^{-1}E(z^Tx)q \\ &= q - Q_0\left(Q_0^TRQ_0\right)^{-1}Q_0^TRq, \qquad \text{(B.9)} \end{aligned}$$

setting $R = E(x^Tz)E(z^Tz)^{-1}E(z^Tx)$. Note that this is also what we find in equation (B.4) when $z = x$. By Assumption R.ii and the equal dimensions 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 $z = x$.

We now return to the proof that z_q can be written $zE(x^Tz)^{-1}q$. The following Lemma leads to the proof.

Lemma 22 *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:*

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

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 $Q_0^TRQ_0$ 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 $a(q)^T$:

$$a(q)^TR.a(q) = \lambda a(q)^Tq = \lambda$$

because $q^T.a(q) = 1$ and λ is positive since R is positive definite and $a(q) \neq 0$. ■

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

$$E(x_q^2) = a(q)^T D^T E(z^T z) Da(q) = a(q)^T Ra(q).$$

We also have:

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

using the previous Lemma and because,

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

Finally, the interior of B is not empty, if we can prove that, for any $q \in \mathbb{S}$,

$$\sup_{\beta \in B} q^T \beta > \inf_{\beta \in B} q^T \beta$$

or equivalently that:

$$\delta^*(q|B) > -\delta^*(-q|B).$$

Start from consequences of definitions:

$$z_q = q^T \Sigma^T z = -z_{-q}, w_q - w_{-q} = (\bar{y} - \underline{y})(\mathbf{1}\{z_q > 0\} - \mathbf{1}\{z_q > 0\}),$$

so that:

$$\delta^*(q|B) + \delta^*(-q|B) = E(|z_q| .(\bar{y} - \underline{y})) > 0,$$

because $E(\bar{y} - \underline{y}|z) = \overline{\Delta}(z) - \underline{\Delta}(z) > 0$ and $|z_q| > 0$ with positive probability because of the full rank assumption in $R.iii$.

B.3 Proof of Lemma 6

We use the expression derived in Proposition 5:

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

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

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

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

so that:

$$\psi(t) - \psi(s) - E(z^*\mathbf{1}\{z^*.s > 0\})(t-s) = E((z^*.t.(\mathbf{1}\{z^*.t > 0\} - \mathbf{1}\{z^*.s > 0\}),$$

This expression is the sum of two terms, A_1 and A_2 :

$$A_1 = E(z^*.t.\mathbf{1}\{z^*.t > 0, z^*s \leq 0\}), A_2 = -E(z^*.t.\mathbf{1}\{z^*.s > 0, z^*t \leq 0\})$$

Consider first A_1 and $z^*.t > 0$ and $z^*.s \leq 0$:

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

so that:

$$0 < A_1 < E(z^*(t-s)|z^*.t > 0, z^*.s \leq 0). \Pr(z^*.t > 0, z^*.s \leq 0).$$

As:

$$\lim_{t \to s}\{z^* \text{ such that } z^*.t > 0 \text{ and } z^*s \leq 0) = \varnothing.$$

we have,

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

Furthermore, assumption R.iii implies that $E(|z^*|)$ is bounded (White, 1994, p32-33), so that we get:

$$A_1 = o(\|s-t\|).$$

The other case is similar so that $A_2 = o(\|s-t\|)$. We thus obtain:

$$\psi(t) - \psi(s) - E(z^*\mathbf{1}\{z^*.s > 0\})(t-s) = o(\|s-t\|).$$

In consequence, 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 can be no 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:

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

by Proposition 5. It is continuous except at points in D_f. As $\delta^*(q|B) = q^T\beta_q$, 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:

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

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:

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

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:

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

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:

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

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

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

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 The construction of z_F and z_H

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 the subspace generated by $E(z^T z)^{-1/2} E(z^T x)$ and let $F^\perp = H = \{h \in \mathbb{R}^m; h^T f = 0, f \in F\}$ its orthogonal.

We can write z as a linear combination of two sets of variables z_F and z_H of respective dimensions, p and $m-p$, which are the normalized orthogonal projections of the $m \times 1$ random vector z onto the subspace F and H. Let:

$$z_F^T = \left[E(x^T z)E(z^T z)^{-1}E(z^T x)\right]^{-1/2} E(x^T z)E(z^T z)^{-1} z^T = M_F^T z^T$$

so that $E(z_F^T z_F) = I_p$. Let also $z_H = z.M_H$ where M_H is a matrix of dimension $[m, m-p]$ such that:

$$E(z_F^T z_H) = 0, E(z_H^T z_H) = I_{m-p}.$$

The first condition implies that:

$$E(x^T z)M_H = 0$$

so that the columns of M_H are the $m-p$ eigenvectors associated to the $m-p$ zero eigenvalues of $E(z^T x)E(x^T z)$.

The relationship between z and (z_F, z_H) is one-to-one since:

$$I_m = E(z_F, z_H)^T (z_F, z_H) = (M_F, M_H)^T E(z^T z)(M_F, M_H)$$

and $E(z^T z)$ is full rank. Thus, (M_F, M_H) is full rank and:

$$z = (z_F, z_H)(M_F, M_H)^{-1}.$$

Multiplying equation (11) successively by M_F^T and M_H^T, we obtain that the parameter β lies in B if and only if there is $u(z)$ in $I(z)$ such that

$$M_F^T E(z^T x)\beta = M_F^T E(z^T(y + u(z))) \Leftrightarrow E(z_F^T x)\beta = E(z_F^T(y + u(z)))$$

$$M_H^T E(z^T x)\beta = M_H^T E(z^T(y + u(z))) \Leftrightarrow 0 = E(z_H^T(y + u(z)))$$

since $M_H^T E(z^T x) = 0$.

B.6 Proof of Proposition 12

We assume that the Sargan condition, given by Proposition 10 is satisfied so that the intersection of the set B_U and the hyperplane, $\gamma = 0$, is not empty. 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:

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

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

Set $C_1 = B_U$ where:

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

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 $x_2^* = (q_2, \lambda_2)$:

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

Remark that B is the intersection of C_1 and C_2. Remark also that when the hyperplane $\gamma = 0$ is not tangent to set B_U, the relative interiors of C_1 and C_2 have all the points of $ri(C_1 \cap C_2)$ in common. Rockafellar's corollary can then be applied to set B using $x^* = (q, \lambda)$:

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

As expected, the RHS is independent of λ_2 and λ, and we can write:

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

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.

Specifically, let (q, λ) be the direction used for estimating B_U, λ being the components relative to the z_H space. From Proposition 5, we know that:

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

and:

$$\delta_{q,\lambda} = E(z_{q,\lambda} w_{q,\lambda}).$$

For any q, minimize this expression wrt λ to get the support function of B.

C Proofs in Section 4

C.1 Proof of Propositions 14 and 15

We use that:

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

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

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

where:

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

where $\hat{\Sigma}_n$ is an estimate of Σ_0 that we now define.

Suppose that 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 chosen as $\|\Sigma_0\| \ll M$.

The estimate $\hat{\Sigma}_n$ belongs to $\{\|\Sigma\| \leq M\}$ 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 14, $\hat{\Sigma}_n$ 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 15 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 of propositions 14 and 15.

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:

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

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

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

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

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 15, 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.

We denote generically M as any majorizing constant.

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 14, 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 $\|\Sigma_n\| \leq M$:

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

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

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

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:

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

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 $q \in \mathbb{S}$ (equation (C.14)) 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 24 *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 $\beta_q = \Sigma_0^T E(z_i^T w_{qi})$.

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 $\sqrt{n}(\hat{\Sigma}_n^u - \hat{\Sigma}_n) \underset{n\to\infty}{\overset{P}{\longrightarrow}} 0$ by equation (C.14) and uniform bounds on q, Σ_0 and β_q. Moreover, as $\Sigma_0^T(\hat{\Sigma}_n^T)^{-1} \underset{n\to\infty}{\overset{a.s}{\longrightarrow}} I$, 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 vis à vis 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:

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

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 16

Let $q_0 \in \mathcal{Q}_0$ any vector that minimizes $T_\infty(q;\beta_0)$. q_0 is not necessary unique. Let q_n be any direction which minimizes its empirical counterpart $T_n(q;\beta_0)$.

First, for all $q \in \mathbb{S}$, our definitions give:

$$\sqrt{n}\left(T_n(q;\beta_0) - T_\infty(q;\beta_0)\right) = \sqrt{n}(\hat{\delta}^*_n(q|B) - \delta^*(q|B) = \upsilon_q + \omega_q, \tag{C.20}$$

where υ_q is the Gaussian process given by Proposition 15 and ω_q is a process which tends uniformly in q to zero:

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

Second, let us analyze $\sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right)$ for any (q_n, q_0). It is bounded from above by zero since q_n minimizes $T_n(.;\beta_0)$. ¿From below, decompose this expression in three terms:

$$\begin{aligned} 0 \geq \sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right) &= \sqrt{n}\left(T_n(q_n;\beta_0) - T_\infty(q_n;\beta_0)\right) \\ &+ \sqrt{n}\left(T_\infty(q_n;\beta_0) - T_\infty(q_0;\beta_0)\right) + \sqrt{n}\left(T_\infty(q_0;\beta_0) - T_n(q_0;\beta_0)\right), \\ &= \upsilon_{q_n} + \omega_{q_n} + \sqrt{n}\left(T_\infty(q_n;\beta_0) - T_\infty(q_0;\beta_0)\right) - \upsilon_{q_0} - \omega_{q_0}, \end{aligned}$$

using equation (C.20). Thus:

$$0 \geq \sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right) \geq \upsilon_{q_n} - \upsilon_{q_0} + \omega_{q_n} - \omega_{q_0}$$

since q_0 minimizes $T_\infty(.;\beta_0)$. As $\lim_{n\to\infty} \Pr(\sup_{n>N} \sup_{q_0,q_n\in\mathbb{S}} |\omega_{q_n} - \omega_{q_0}| \geq \varepsilon) = 0$, the difference $\upsilon_{q_n} - \upsilon_{q_0}$ is asymptotically bounded from above by 0. As $\upsilon_{q_n} - \upsilon_{q_0}$ is normally distributed, its variance is asymptotically converging to zero and the two random variables υ_{q_n} and υ_{q_0} are asymptotically perfectly correlated with the same variance. We thus have:

$$\upsilon_{q_n} - \upsilon_{q_0} \underset{n\to\infty}{\overset{P}{\to}} 0 \Rightarrow \sqrt{n}\left(T_n(q_n;\beta_0) - T_n(q_0;\beta_0)\right) \underset{n\to\infty}{\overset{P}{\to}} 0.$$

The previous result means two things. 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 $q_n^T \Sigma^T V(z^T \varepsilon_{q_n}) \Sigma q_n$ consistently estimates $V_{q_0} = q_0^T \Sigma^T V(z^T \varepsilon_{q_0}) \Sigma q$. We can of course replace Σ and $V(z^T \varepsilon_{q_n})$ by their empirical counterpart to provide a feasible consistent estimator.

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 set of minimizers is not necessary unique when set B has a kink.

C.3 Proof of Proposition 17

The first assertion comes from proposition (16) 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). Let us prove the last assertion, when $\beta_0 \notin B$, the proof of the other assertion being similar.

Write:

$$\begin{aligned}\xi_n(\beta_0) &= \sqrt{n}\frac{T_n(q_n;\beta_0)}{\sqrt{\hat{V}_n}} = \sqrt{n}\frac{(T_n(q_n;\beta_0) - T_n(q_0;\beta_0))}{\sqrt{\hat{V}_n}} \\ &+\sqrt{n}\frac{(T_n(q_0;\beta_0) - T_\infty(q_0;\beta_0))}{\sqrt{\hat{V}_n}} + \sqrt{n}\frac{T_\infty(q_0;\beta_0))}{\sqrt{\hat{V}_n}},\end{aligned}$$

where $q_0 \in \arg\min_q T_\infty(q_0;\beta_0)$. When $\beta_0 \notin B$, we have $T_\infty(q_0;\beta_0) < 0$.

As in the proof of Proposition (16), the first term is asymptotically negligible, the second term is asymptotically equivalent to a random variable distributed $N(0,1)$. The last term is negative and diverges to $-\infty$ because of the $\sqrt{n}$ factor.

C.4 Construction of the Confidence Region in Proposition 18

We here provide a simple way to construct $CI^n_\alpha = \{\beta; \xi_n(\beta) > \mathcal{N}_\alpha\}$ when $\alpha < 1/2$. Recall first the definitions of $\xi_n(\beta)$ and $T_n(q;\beta)$ in Section 4.2:

$$\xi_n(\beta) = \frac{\sqrt{n}}{\sqrt{\hat{V}_{q_n}}}(T_n(q_n;\beta)) \text{ where } T_n(q_n;\beta) = \min_{q\in\mathbb{S}}(\hat{\delta}^*_n(q|B) - q^T\beta),$$

where q_n is one argument of the minimum. Therefore, the confidence region is also given by $CI^n_\alpha = \{\beta; \min_{q\in\mathbb{S}}(T_n(q;\beta)) > \frac{\sqrt{\hat{V}_{q_n}}}{\sqrt{n}}\mathcal{N}_\alpha\}$

Second, the estimated set $\hat{B}_n$ is included in CI^n_α as $\mathcal{N}_\alpha < 0$ for any $\alpha < 1/2$ and as for all β belonging to the the estimated set, $\hat{B}_n$:

$$\min_{q\in\mathbb{S}}(\hat{\delta}^*_n(q|B) - q^T\beta) \geq 0,$$

Consider any point $\beta_f \in \partial\hat{B}_n \subset CI^n_\alpha$, the frontier of the estimated set $\hat{B}_n$. There exists at least one, and possibly a set (which is the intersection of a cone and $\mathbb{S}$) denoted $\mathcal{C}(\beta_f)$, of vectors $q_f \in \mathbb{S}$ such that:

$$\begin{aligned}T_n(q_f;\beta_f) = \hat{\delta}^*_n(q_f|B) - q_f^T\beta_f = 0, \\ \forall q \in \mathbb{S}, T_n(q;\beta_f) \geq T_n(q_f;\beta_f) = 0\end{aligned}$$

Choose such a q_f and consider the points $\beta_f(\lambda)$, where $\lambda \geq 0$, on the half-line defined by β_f and direction q_f:

$$\beta_f(\lambda) = \beta_f + \lambda q_f.$$

We have:

$$\begin{aligned}T_n(q;\beta_f(\lambda)) &= T_n(q;\beta_f) + q^T(\beta_f - \beta_f(\lambda)) \\ &= T_n(q;\beta_f) - \lambda q^T q_f\end{aligned}$$

where $-\lambda q^T q_f \geq -\lambda q_f^T q_f = -\lambda$ and $T_n(q;\beta_f) \geq T_n(q_f;\beta_f) = 0$ for any q, as seen above. As a consequence,

$$T_n(q;\beta_f(\lambda)) \geq -\lambda = T_n(q_f;\beta_f(\lambda)).$$

where vector q_f which minimizes $T_n(q;\beta_f)$ minimizes also $T_n(q;\beta_f(\lambda))$.

We can therefore characterize the points of the half-line which belongs to CI_α^n. Given that λ is positive,

$$\beta_f(\lambda) \in CI_\alpha^n \text{ if and only if } \lambda \leq -\frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha,$$

so that segment $(\beta_f, \beta_f + \frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_{1-\alpha} q_f]$ is included in CI_α^n. We thus proved that:

$$\hat{B}_n \cup \{\cup_{\beta_f \in \partial B_n} \cup_{q_f \in \mathcal{C}(\beta_f)} (\beta_f, \beta_f + \frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_{1-\alpha} q_f)\} \subset CI_\alpha^n, \tag{C.21}$$

where $\mathcal{C}(\beta_f)$ is the cone defined above.

Conversely, let us prove that CI_α^n is included in the set on the LHS. Let β_c a point in CI_α^n. If β_c belongs to $\hat{B}_n$, the inclusion is proved. Assume that β_c is outside the estimated set and let β_f the point on the frontier of $\hat{B}_n$ which is the projection of β_c on $\hat{B}_n$. The projection is unique because set $\hat{B}$ is convex.

Write $\beta_c - \beta_f = \lambda q_f$ for some direction $q_f \in \mathbb{S}$ and some $\lambda > 0$. We have that:

$$q_f^T(\beta_c - \beta_f) \leq q_f^T(\beta_c - \beta),$$

for any $\beta \in \hat{B}_n$ because β_f is the projection of β_c on set $\hat{B}_n$ along the direction q_f. We thus have $q_f^T \beta_f \geq q_f^T \beta$ which proves that $\hat{\delta}_n^*(q_f|B) = q_f^T \beta_f$. The pair (β_f, q_f) satisfies the condition of the previous paragraphs.

As β_c is a point of CI_α^n, λ is necessary less or equal than the value $-\frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha$. Thus it belongs to the LHS of equation (C.21). As a consequence, equation (C.21) is an equality.

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 computing $\delta^*(q|B)$. As $E(x^T x) = I_2$, we have:

$$\begin{cases} z_q = x.q = \cos\theta x_1 + \sin\theta x_2, \\ w_q = y - \Delta + 2\Delta \mathbf{1}\{z_q > 0\}. \end{cases}$$

Using

$$\begin{matrix} x_1 \\ x_2 \\ z_q \end{matrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \cos\theta \\ 0 & 1 & \sin\theta \\ \cos\theta & \sin\theta & 1 \end{bmatrix} \right),$$

we obtain:

$$Ex_1\mathbf{1}_{z_q>0} = \frac{1}{\sqrt{2\pi}}\cos\theta \text{ and } Ex_2\mathbf{1}_{z_q>0} = \frac{1}{\sqrt{2\pi}}\sin\theta.$$

As y^* and x_1, x_2 are independent and as $Ex_1 = Ex_2 = 0$, we get:

$$\delta^*(q|B) = E(z_q w_q) = \frac{2\Delta}{\sqrt{2\pi}}.$$

The corresponding points on the frontier are:

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

The estimates are obtained using the empirical analogs as developed in Section 4.

D.2 Example of Section 5.2

The simulated model is:

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

$x_2 = \pi e_2 + \sqrt{1-\pi^2}e_3$, $w = \nu e_3 + \sqrt{1-\nu^2}e_4$ where (e_2, e_3, e_4) is a standard normal vector. For simplicity, define $\mu = \nu\sqrt{1-\pi^2}$ and $a^2 = \pi^2 + \mu^2 = \pi^2 + \nu^2(1-\pi^2)$.

To conform with the general notations, let $x = (x_1, x_2)$ and $z = (x_1, e_2, w)$. As there exists one supernumerary restriction, we first evaluate z_F and z_H as defined in Appendix B. As $E(z^T z) = I_3$, we have:

$$E(x^T z) = \begin{pmatrix}1 & 0 & 0\\ 0 & \pi & \mu\end{pmatrix}, \quad \left[E(x^T z)E(z^T z)^{-1}E(z^T x)\right]^{-1/2} = \begin{pmatrix}1 & 0\\ 0 & a^{-1}\end{pmatrix},$$

and:

$$z_F^T = \left[E(x^T z)E(z^T z)^{-1}E(z^T x)\right]^{-1/2} E(x^T z)E(z^T z)^{-1}z^T = \begin{pmatrix}x_1\\ \frac{\pi e_2 + \mu w}{a}\end{pmatrix},$$

which is standardized bivariate normally distributed. Moreover to compute z_H:

$$E(z_F^T z) = \begin{pmatrix}1 & 0 & 0\\ 0 & \frac{\pi}{a} & \frac{\mu}{a}\end{pmatrix}$$

has an eigenvector associated to the eigenvalue 0, equal to $\begin{pmatrix}0 & \frac{\mu}{a} & -\frac{\pi}{a}\end{pmatrix}^T$ which is normalized. In consequence, $z_H = \frac{\mu e_2 - \pi w}{a}$.

To construct B_U, we use (z_F, z_H) and we write:

$$\Sigma^T = \left[E\begin{pmatrix}x_1\\ a^{-1}(\pi e_2 + \mu w)\\ z_H\end{pmatrix}\begin{pmatrix}x_1 & x_2 & z_H\end{pmatrix}\right]^{-1} = \begin{pmatrix}1 & 0 & 0\\ 0 & a^{-1} & 0\\ 0 & 0 & 1\end{pmatrix}.$$

$$\begin{aligned} z_{q,\lambda} &= \begin{pmatrix}q^T & \lambda\end{pmatrix}\begin{pmatrix}x_1\\ a^{-1}(\pi e_2 + \mu w)\\ z_H\end{pmatrix}\\ &= x_1 q_1 + (a^{-1}(\pi e_2 + \mu w))q_2 + z_H\lambda. \end{aligned}$$

where $q_1^2 + q_2^2 = 1$ (If we use the polar coordinates, $q_1 = \cos\theta$ and $q_2 = \sin\theta$). Furthermore, as in the previous example,

$$w_q = y - \Delta + 2\Delta \mathbf{1}\{z_{q,\lambda} > 0\}.$$

The correlation of $z_{q,\lambda}$ with the various unit-normal variables are:

$$E(z_{q,\lambda}x_1) = q_1, E(z_{q,\lambda}(a^{-1}(\pi e_2 + \mu w))) = a^{-1}q_2, E(z_{q,\lambda}z_H) = \lambda,$$

so that, using the normality assumption, for instance,

$$Ex_1\mathbf{1}_{z_q>0} = \frac{1}{\sqrt{2\pi}}q_1.$$

In consequence, a closed-form expression for $\delta^*(q, \lambda|B_U)$:

$$\delta^*(q, \lambda|B_U) = \frac{2\Delta}{\sqrt{2\pi}}\left(q_1^2 + \frac{q_2^2}{a^2} + \lambda^2\right).$$

It is obviously minimized when $\lambda = 0$. B_U is an ellipsoid orthogonal to the hyperplane $\gamma = 0$. Its projection on the hyperplane is also an ellipse:

$$\delta^*(q|B) = \frac{2\Delta}{\sqrt{2\pi}}\sqrt{\cos^2\theta + \frac{\sin^2\theta}{a^2}}.$$

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 $\delta^*(q|B)$. Since

$$\Sigma = E(z^Tz)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix},$$

we can similarly derive the variables of interest:

$$\begin{cases} z_q = z\Sigma q = \cos\theta + 4x_1\sin\theta, \\ w_q = \underline{y} + \frac{1}{2}\mathbf{1}\{z_q > 0\} \\ \underline{y} = \frac{1}{2}\mathbf{1}\{y \geq 0.5\}. \end{cases}$$

Since $E(\underline{y}) = \frac{1}{4}$ and $E(x_1\underline{y}) = \frac{1}{16}$, we can derive some expression for β_q:

$$\begin{aligned} \beta_q &= \Sigma E(z^Tw_q) \\ &= \Sigma E(z^T\underline{y}) + \frac{1}{2}\Sigma E(z^T\mathbf{1}\{z_q > 0\} \\ &= \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}E(\mathbf{1}\{z_q > 0\}) \\ 2E(x_1\mathbf{1}\{z_q > 0\}) \end{bmatrix}. \end{aligned}$$

Let $\theta_0 \in [0; \pi/2]$ such that $\tan \theta_0 = \frac{1}{2}$. For θ being between $-\theta_0$ and θ_0 z_q is always positive whatever the value of x_1:

$$E\mathbf{1}\{z_q > 0\} = 1$$
$$Ex_1\mathbf{1}\{z_q > 0\} = 0$$

and $\beta_q = \left[\frac{3}{4}; \frac{1}{4}\right]^T$.

For θ being between θ_0 and $-\theta_0 + \pi$, z_q is negative when $x_1 = -\frac{1}{2}$, otherwise positive:

$$E\mathbf{1}\{z_q > 0\} = \frac{1}{2}$$
$$Ex_1\mathbf{1}\{z_q > 0\} = \frac{1}{4},$$

and $\beta_q = \left[\frac{1}{2}; \frac{3}{4}\right]^T$.

We obtain similarly $\beta_q = \left[\frac{1}{4}; \frac{1}{4}\right]^T$ when θ is between $\theta_0 + \pi$ and $\theta_0 + \pi$ and $\beta_q = \left[\frac{1}{2}; -\frac{1}{4}\right]^T$ for θ being between $\theta_0 - \pi$ and $-\theta_0$.

FIGURES AND TABLES

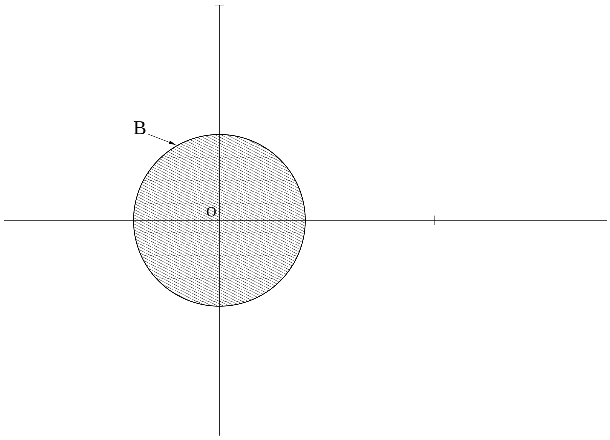


Figure 1: Set B, $y = 0.x_1 + 0.x_2 + \varepsilon$, $(x_1, x_2)^T \sim N(0, I_2)$

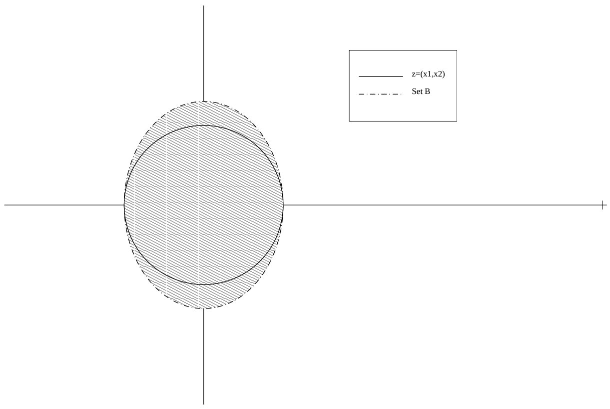


Figure 2: Set B, $y = 0.x_1 + 0.x_2 + \varepsilon$, $z = (x_1, e_2, w)$

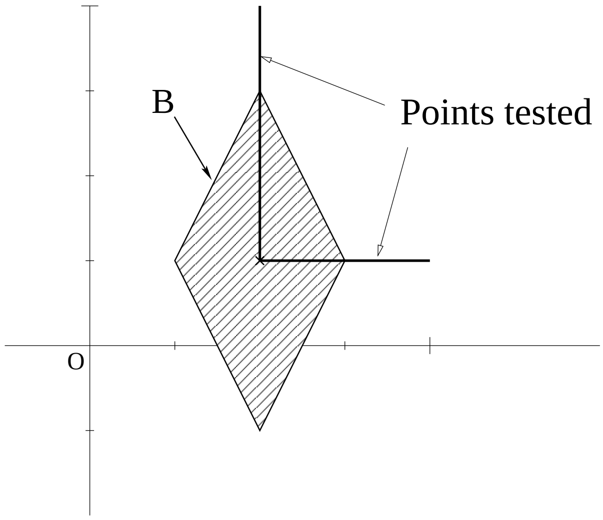


Figure 3: Set B, $y = \frac{1}{2} + \frac{x}{4} + \varepsilon$, $x \in \{-\frac{1}{2}, \frac{1}{2}\}$

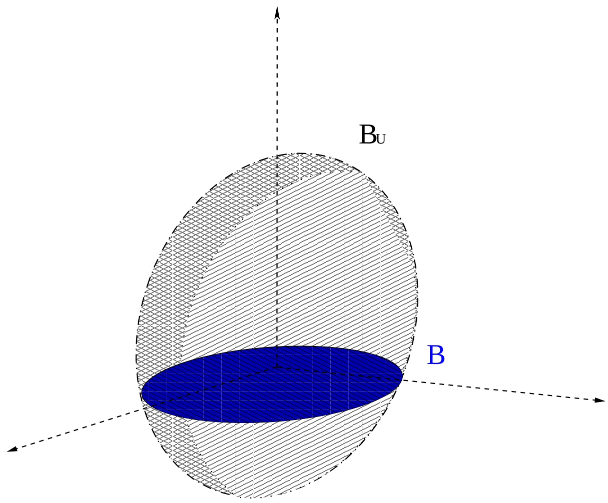
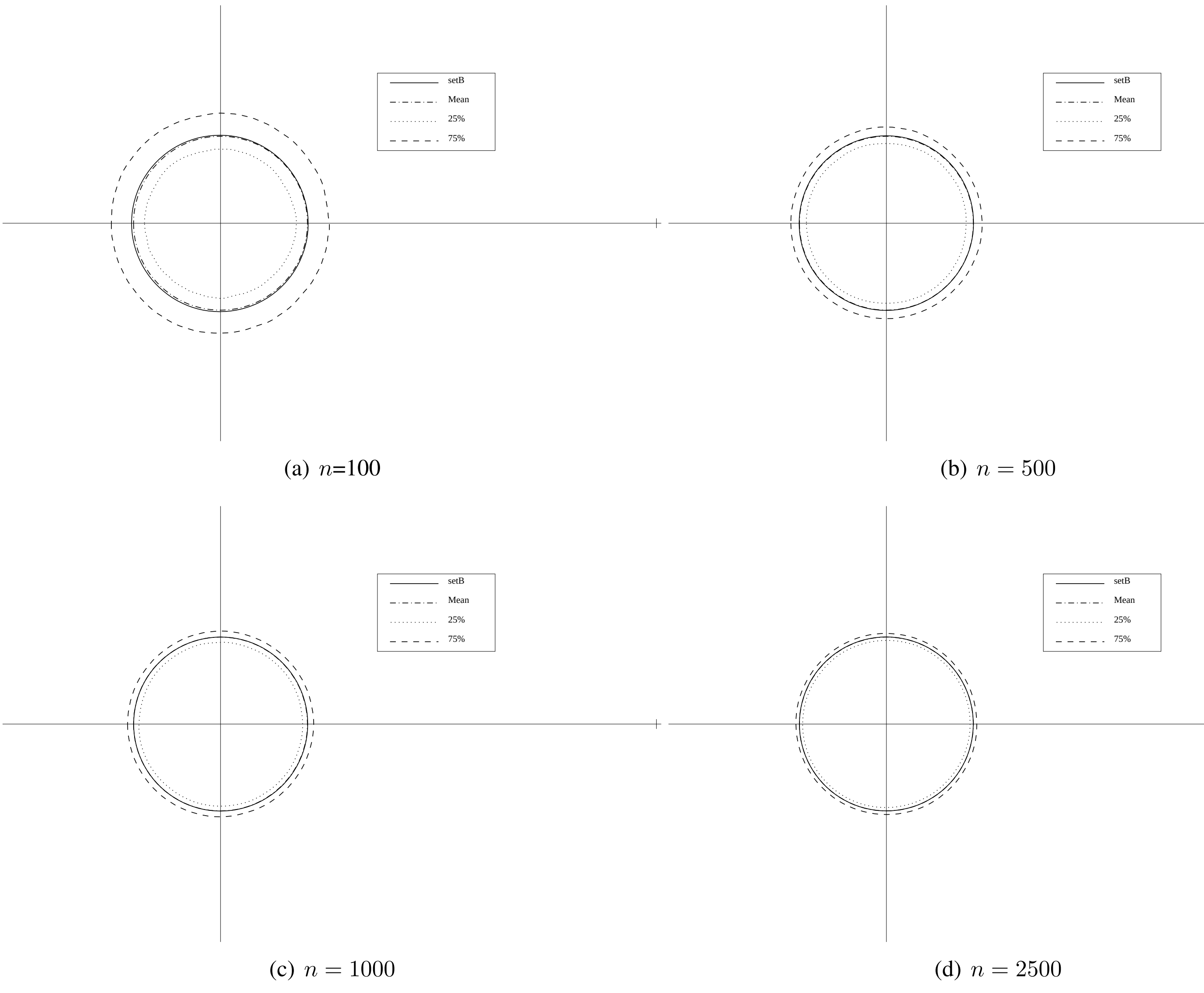


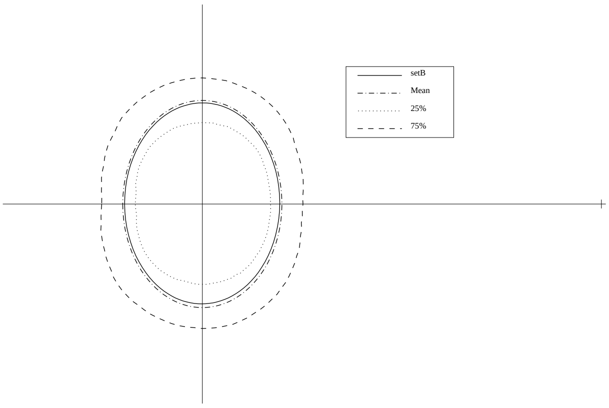
Figure 4: Geometric Characterization of the Identified Set



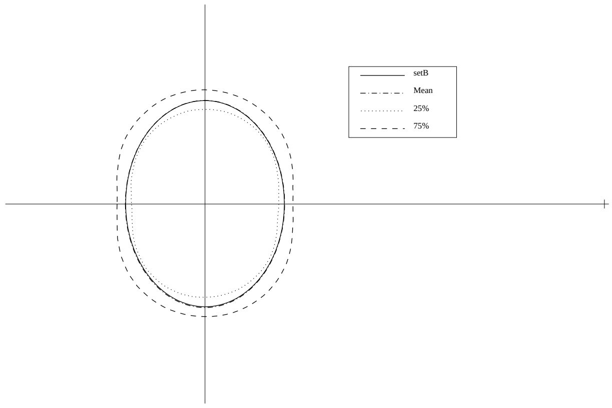
(a) n=100 (b) $n = 500$

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

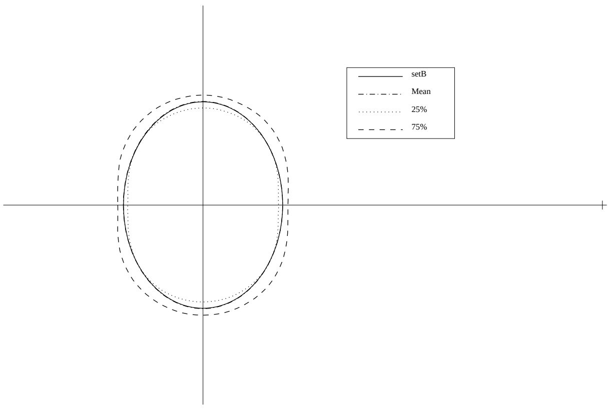
Figure 5: Estimation of B for various sample sizes n.



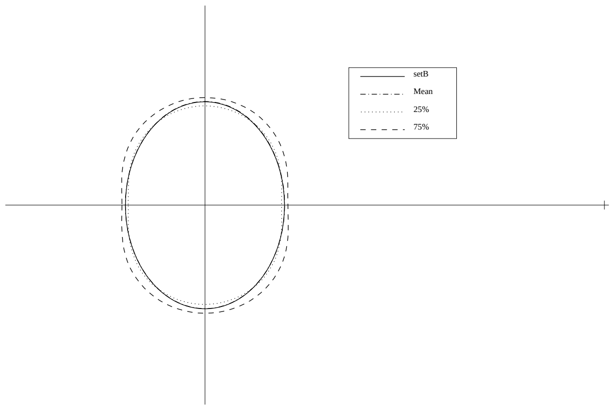
(a) n=100



(b) $n = 500$

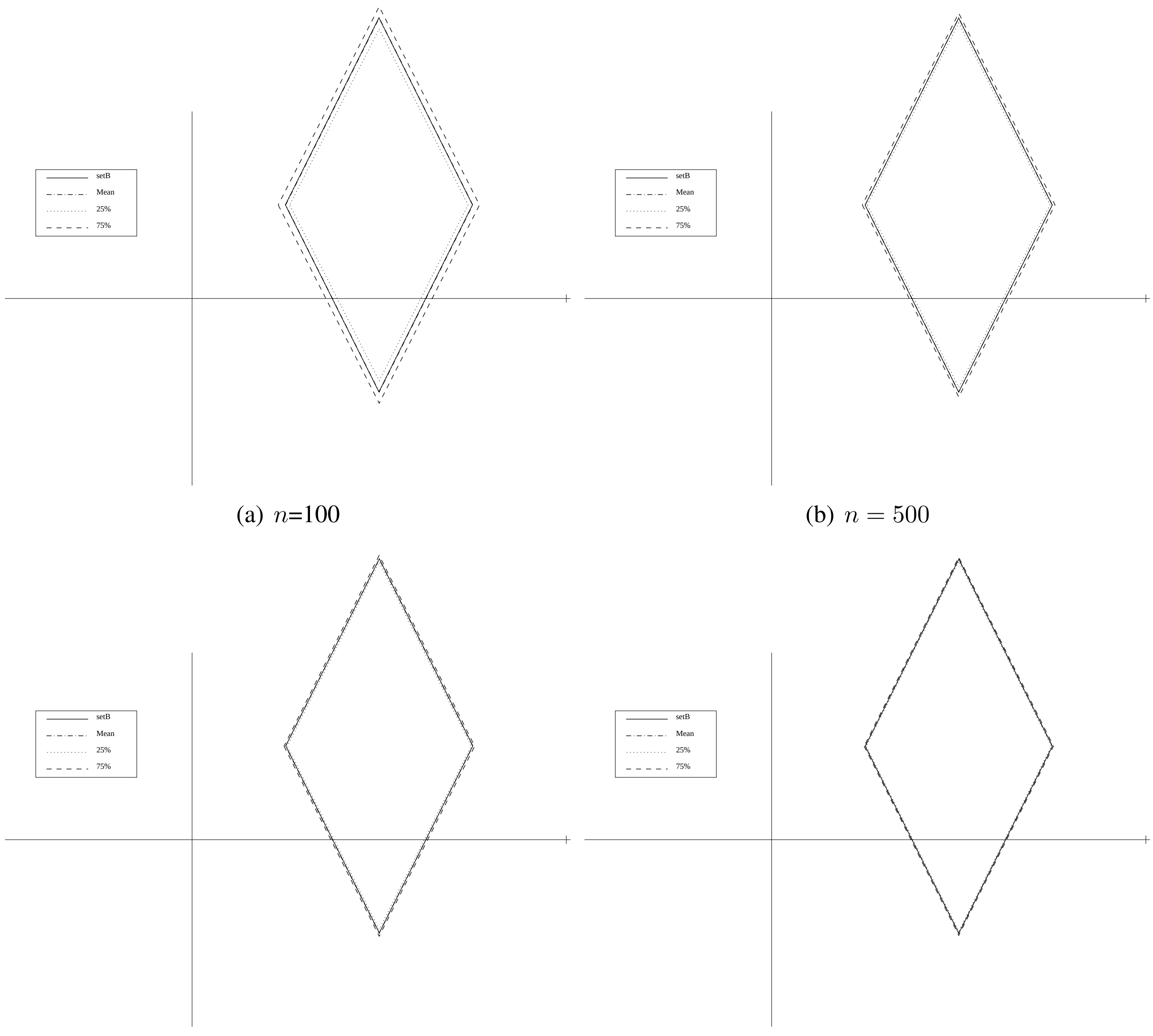


(c) $n = 1000$



(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.

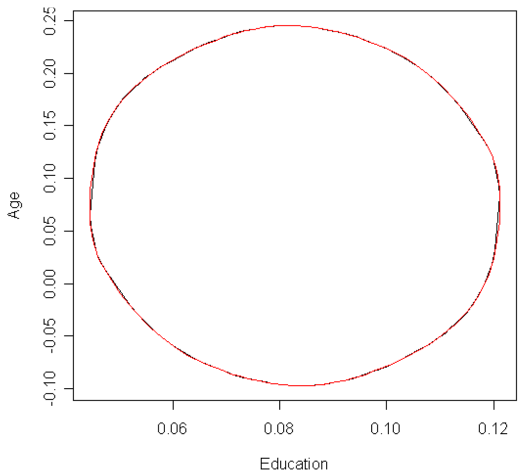


Figure 8: Estimated set for Age and Education

Table 1: Percentage of rejections for the three tests - first example

	Test 1 ($H_0 : \beta^r \in B$)				Test 2 ($H_0 : \beta^r \notin B$)				Test 3 ($H_0 : \beta^r \in \partial B$)			
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
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 $\beta^r = \frac{r}{\sqrt{2\Pi}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. β^1 is on the frontier of B.

Table 2: Percentage of rejections for the three tests - second example

	Test 1 ($H_0 : \beta^r \in B$)				Test 2 ($H_0 : \beta^r \notin B$)				Test 3 ($H_0 : \beta^r \in \partial B$)			
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0	0%	0%	0%	0%	78.5%	100%	100%	100%	62.1%	100%	100%	100%
0.05	0%	0%	0%	0%	77.5%	100%	100%	100%	62%	100%	100%	100%
0.1	0%	0%	0%	0%	76.5%	100%	100%	100%	59.7%	100%	100%	100%
0.2	0%	0%	0%	0%	68.9%	100%	100%	100%	54%	100%	100%	100%
0.3	0%	0%	0%	0%	61.3%	100%	100%	100%	45.6%	100%	100%	100%
0.4	0%	0%	0%	0%	49.5%	99.7%	100%	100%	34.4%	99.5%	100%	100%
0.5	0.1%	0%	0%	0%	36.9%	98.4%	100%	100%	23.2%	96.4%	99.9%	100%
0.6	0.4%	0%	0%	0%	25%	91.2%	99.6%	100%	15.7%	83.5%	99.3%	100%
0.7	1%	0%	0%	0%	15.9%	69.7%	92.5%	99.9%	9.7%	59.1%	87.8%	99.8%
0.8	2.8%	0%	0%	0%	9.5%	39.6%	65.4%	95.2%	6.4%	28%	52.1%	90.7%
0.85	3.6%	0.3%	0.1%	0%	7.3%	26.2%	43.6%	78.9%	5.7%	15.6%	32.9%	70%
0.9	4.6%	0.9%	0.5%	0.1%	5.4%	14.5%	26.5%	48%	5.4%	8.9%	15.4%	33.7%
0.92	5.2%	1.5%	0.7%	0.2%	4.6%	11.8%	18.5%	32.9%	5.3%	6.3%	9.5%	23%
0.94	5.4%	2.1%	1%	0.8%	4.2%	8.9%	11.8%	22.1%	5.6%	5%	6.1%	14.6%
0.96	5.6%	2.8%	2%	1.3%	3.5%	6.1%	7.2%	14.1%	5.5%	4.7%	4.6%	7.8%
0.98	6.8%	3.5%	3.2%	3.4%	3.1%	4.5%	5%	6.8%	5.9%	4.4%	4.4%	4.8%
0.99	7.1%	4.4%	4.4%	4.1%	2.9%	3.4%	4%	4.8%	5.8%	4.4%	4.4%	5%
1	**7.9%**	**5.4%**	**5.9%**	**5.5%**	**2.9%**	**3.2%**	**3.3%**	**3.1%**	**6.1%**	**4.8%**	**3.9%**	**5.2%**
1.01	8.3%	6.3%	7.2%	8.5%	2.4%	3.1%	3%	2.3%	6.3%	4.8%	4.7%	5.6%
1.02	8.5%	7.3%	8.4%	11.5%	2%	2.7%	2.4%	1.7%	6.4%	5%	5.8%	6.7%
1.04	9.7%	9.7%	12.1%	18.7%	1.9%	2.1%	1.6%	0.8%	6.6%	6%	8%	12.4%
1.06	10.2%	12.9%	16.6%	28.5%	1.6%	1.6%	1.1%	0.1%	7.3%	7.6%	10.1%	19.5%
1.08	11.3%	17.4%	22.4%	40.4%	1.2%	0.9%	0.5%	0%	7.9%	9.9%	14.3%	28.9%
1.1	12.3%	20.3%	29.3%	55.8%	1%	0.5%	0.2%	0%	8.5%	12.7%	20.1%	41.4%
1.2	21.6%	47.5%	70.6%	97.3%	0.6%	0%	0%	0%	13.8%	35.2%	58.6%	94.3%
1.3	33.6%	75.3%	95.9%	100%	0.3%	0%	0%	0%	22.9%	64.9%	92.2%	100%
1.4	46.1%	93.3%	99.5%	100%	0%	0%	0%	0%	34.7%	87.5%	98.8%	100%
1.5	60.9%	98.3%	100%	100%	0%	0%	0%	0%	47%	97.2%	100%	100%
1.6	69.6%	99.9%	100%	100%	0%	0%	0%	0%	60.9%	99.6%	100%	100%
1.8	88.5%	100%	100%	100%	0%	0%	0%	0%	81.5%	100%	100%	100%
2.05	97.9%	100%	100%	100%	0%	0%	0%	0%	94.8%	100%	100%	100%
2.3	99.8%	100%	100%	100%	0%	0%	0%	0%	99.3%	100%	100%	100%
2.55	100%	100%	100%	100%	0%	0%	0%	0%	100%	100%	100%	100%
2.8	100%	100%	100%	100%	0%	0%	0%	0%	100%	100%	100%	100%

The point tested β^r is located on the first bisector. r is the fraction of the distance from the origin w.r.t. to the distance origin-frontier point on this bisector. $r = 1$ is therefore associated to the frontier point (results in bold), $r = 0$ to the origin.

Table 3: Percentage of rejections for the three tests, example 3, $\beta^r = (r, \frac{1}{4})^T$

β^r		Test 1 ($H_0 : \beta^r \in B$)				Test 2 ($H_0 : \beta^r \notin B$)				Test 3 ($H_0 : \beta^r \in \partial B$)			
		$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
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 $\beta^r = (r, \frac{1}{4})^T$ for various values of r. $\beta^1 = (\frac{3}{4}, \frac{1}{4})^T$ is on the frontier of B.

Table 4: Percentage of rejections for the three tests, example 3, $\beta^r = (\frac{1}{2}, r)^T$

β^r		Test 1 ($H_0 : \beta^r \in B$) $n = 100$	$n = 500$	$n = 1000$	$n = 2500$	Test 2 ($H_0 : \beta^r \notin B$) $n = 100$	$n = 500$	$n = 1000$	$n = 2500$	Test 3 ($H_0 : \beta^r \in \partial B$) $n = 100$	$n = 500$	$n = 1000$	$n = 2500$
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 $\beta^r = (\frac{1}{2}, r)^T$ for various values of a. $\beta^1 = (\frac{1}{2}, \frac{3}{4})^T$ is on the frontier of B.

Table 5: 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 6: 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 7: 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 8: 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)