Set Identified Linear Models

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

First version, August 2006
Version under reconstruction, May 2009
Before quoting, please ask for the fully revised version

Abstract

We analyze the identification and estimation of parameters β satisfying the *incomplete* linear moment restrictions $E(z^\top(x\beta - y)) = E(z^\top u(x))$ where z is a set of instruments and $u(z)$ an unknown bounded scalar function. We first provide empirically relevant examples of such a set-up. Second, we show that these conditions set identify β where the identified set B 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. Third, we provide new results on the asymptotics of analog estimates. When B is a strictly convex set, 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. Results of some Monte Carlo experiments are presented.

[*]Toulouse School of Economics (GREMAQ and IDEI), Toulouse, France, bontemps@cict.fr
[†]Toulouse School of Economics, (GREMAQ and IDEI), Toulouse, France, magnac@cict.fr
[‡]Paris School of Economics, Paris, France. Eric.Maurin@ens.fr

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 in which 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:

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

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 and 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 neither 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

[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, Whitney Newey, Adam Rosen and particularly Francesca Molinari for helpful discussions as well as three anonymous referees for their insighful comments. We also thank the participants at seminars at PUC-Rio, CEMMAP, CREST Malinvaud seminar, Mannheim, Yale, NYU, Paris I, Cornell, MIT, Northwestern, Toulouse and Cambridge as well as in workshops and conferences (ESRC Bristol '07, Montréal Conference on GMM '07, London Cemmap-Northwestern University Conference on "Inference in Partially Identified Models" '08, Marseille Festschrift for Russel Davidson '08) for comments. We thank Mehtap Akguc for her excellent research assistance. The usual disclaimer applies.

identified set using the concept of support functions which is naturally associated with convex sets (Rockafellar, 1970). In each direction of interest, which spans the unit sphere in $\mathbb{R}^p$, we show that the support function of the identified set B is the expectation of an explicit and simple random function. Second, we show that a similar characterization of the identified set also holds true when the incomplete linear moment conditions are written as a function of m instruments z:

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

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 (2008) 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 (2008) 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 (2007) 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 (2005). 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 (Horowitz and Manski, 2006, Beresteanu & Molinari, 2007) 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 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 concludes.

2 The Set-up of Incomplete Linear Models

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

$$E(z^\top (x\beta - y)) = E(z^\top u(z)), \tag{3}$$

where y is a scalar dependent variable, x a vector of p covariates, z a vector of m instruments and $u(z)$ a measurable function which takes values in an admissible set $I(z) = [\underline{\Delta}(z), \overline{\Delta}(z)]$ where $-\infty < \underline{\Delta}(z) < 0 < \overline{\Delta}(z) < +\infty$ and $\overline{\Delta}(z) - \underline{\Delta}(z) \geq \Delta_0 > 0$. These bounds are constructed using two observable variables $\overline{y}$ and $\underline{y}$ with $\overline{y} \geq y \geq \underline{y}$ and such that,

$$E(\overline{y} - y \mid z) = \overline{\Delta}(z), E(\underline{y} - y \mid z) = \underline{\Delta}(z). \tag{4}$$

The next subsections provide examples, the leading one being when y is observed by interval only and $\underline{y}$ and $\overline{y}$ are the lower and upper bounds of the interval that are explicitely reported in the dataset (e.g. Manski and Tamer, 2002).

We 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.[2]. Furthermore, the conditions of full rank, $rank(E(z^\top x)) = p$, and $rank(E(z^\top z)) = m$ 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 (3), 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. As shown in the next section, it implies that the set of identified parameters is bounded.

2.1 Censored Dependent Variables

The first interesting set of examples corresponds to familiar linear regression models 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

[2]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).

that are often reported by interval only in survey or administrative data.[3] 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 $w = (\overline{y}, \underline{y}, x)$ where $\left[\underline{y}, \overline{y}\right]$ represents the interval measurement[4] of a latent variable y^* and x a vector of p covariates. The latent variable y^* and the observed bounds $\underline{y}$ and $\overline{y}$ are assumed to be L^2-integrable.[5] Within this framework, we consider linear latent models :

$$y^* = x\beta + \varepsilon, \tag{5}$$

where ε is a random variable uncorrelated with x, $E(x^\top \varepsilon) = 0$. The issue is to characterize B the set of parameters β such that the latent model defined by equation (5) is consistent with the observed bounds. By definition, β is in B if and only if there is a random variable, ε, uncorrelated with x and such that $x\beta + \varepsilon \in \left[\overline{y}, \underline{y}\right]$.

Assuming that all variables are in L_2 so that all cross-moments exist, the following proposition shows that B is defined by an incomplete linear regression of the center of the interval measurement $y = \frac{\overline{y}+\underline{y}}{2}$ on covariates x.

Proposition 1 *Denote $y = \frac{\overline{y}+\underline{y}}{2}$ the center and $\Delta(x) = E(\frac{\overline{y}-\underline{y}}{2} \mid x)$ half of the average length of the interval measurement $\left[\underline{y}, \overline{y}\right]$. Then β is in B if and only if there exists a measurable function $u(x)$ which takes values in $I(x) = [-\Delta(x); \Delta(x)]$ such that,*

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

Proof. See Appendix A.1. ■

[3]Also, for anonymity reasons,only interval information could be made available to researchers even though the information collected is actually continuous.

[4]When the measurement interval is not closed, it is not the identification set B itself, but the closure of B which is defined by an incomplete linear regression model.

[5]Without this condition, 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).

2.2 Discussion of other Applications

Other interesting examples correspond to contingent valuation studies where participants are asked whether their willingness-to-pay (w^*) for a good or resource exceeds a bid $-v$ chosen by experimental design (see e.g., McFadden, 1994). The outcome under consideration w equals one if the respondent willingness-to-pay exceeds the experimental bid (i.e., $w^* + v > 0$) and the relationship of interest between w^* and a set of covariates x is to be inferred from available observations on w, x and v. Dosage response models are a related example in which w is equal to one when a lethal dose w^* exceeds a treatment dose, $-v$, chosen by experimental design. In all these cases, it is natural to assume that $w^* = x\beta + \varepsilon$ and estimate the semiparametric binary model $w = 1(x\beta + v + \varepsilon > 0)$ under the assumption that ε is uncorrelated with regressors x and is independent of regressor v conditional on x (i.e., $F_\varepsilon(. \mid x, v) = F_\varepsilon(. \mid x)$) if only because of experimental design. Also, it is often plausible to suppose that the support of w^* 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) and β is point identified. In contrast, when the distribution of v is not continuous, the set B of observationally equivalent parameters is a proper set defined by a moment condition similar to condition (1) (see Magnac and Maurin, 2008).

Categorical data on individual opinions or attitudes are another potential field of applications. Surveys on job satisfaction or 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?". It is assumed that these responses are function of a continuous intensity measure $y^* = x\beta + \varepsilon$ where ε has a parametric distribution (ordered probit or logit). Alternatively, if the distribution of ε is unspecified, the identified set of parameters is defined by moment conditions similar to condition (1).

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 (ILM) above. We first focus on the case in which 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 in which 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 $E(z^\top x)$ is full rank implies that equation (3) has one and only one solution in β for any function $u(z)$ varying in the admissible set. The set of identified parameters, B, is the collection of such parameters:

$$B = \{\beta : \beta = (E(z^\top x))^{-1} E(z^\top (y + u(z))), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\}. \tag{7}$$

The identified set B is therefore non empty (set e.g. $u(z) = 0$), convex and closed since the admissible set is convex and closed. This is a key result for this paper because convex set B can be unambiguously characterized by its *support function* defined as (Rockafellar, 1970):[6]

$$\forall q \in \mathbb{R}^p, \delta^*(q \mid B) = \sup\{q^\top \beta \mid \beta \in B\}.$$

and any closed and convex set B satisfies:

$$\beta \in B \Leftrightarrow \forall q \in \mathbb{R}^p,\ q^\top \beta \leq \delta^*(q \mid B).$$

Given that support functions are homogenous of degree 1, it is sufficient to define them over the unit sphere of $\mathbb{R}^p$ i.e. $\mathbb{S} = \{q \in \mathbb{R}^p; \|q\| = 1\}$. Identification of B is equivalent to identifying the support function, $\delta^*(q \mid B)$, over the unit sphere $\mathbb{S}$.

Our key result is that the support function of B can be written as a function of a population moment of two simple random variables. Let β a point which belongs to set B. From (7), there exists some function $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$ such that

$$\beta = \left(E(z^\top x)\right)^{-1} E\left(z^\top \left(y + u(z)\right)\right).$$

[6] Beresteanu and Molinari (2008) also use this property in order to apply the theory of random set variables.

We can multiply this equation by vector q to express:

$$q^\top \beta = \left(q^\top E(z^\top x)\right)^{-1} E\left(z^\top \left(y + u(z)\right)\right) = E\left(z_q \left(y + u(z)\right)\right), \tag{8}$$

where $z_q = q^\top \left(E(z^\top x)\right)^{-1} z^\top$. As the support function in the direction q is the supremum of $q^\top.\beta$ when $\beta \in B$, it is the supremum of (8) over the set of admissible $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. It is a simple single-dimensional optimization problem whose solution is:

Proposition 2 *Let $w_q = \underline{y} + \mathbf{1}\{z_q > 0\}(\overline{y} - \underline{y})$. 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 $\beta_q = E(z^\top x)^{-1} E(z^\top w_q)$ is the frontier point of B such that $\delta^*(q \mid B) = q^\top \beta_q$.

Proof. See Appendix B.1 ■

This proposition sharply characterizes B. The support function is defined everywhere because of assumption *(R.iii)* that all cross-moments are well defined. In particular, the support function is bounded and therefore set B is bounded. Furthermore, as a convex function, $\delta^*(q \mid B)$ is differentiable except at a countable number of points in a set D_f. The following lemma provides geometric properties of set B and an explicit characterisation of D_f .

Lemma 3 *The support function $\delta^*(q \mid B)$ is differentiable on $\mathbb{S}$ except on a set D_f which is composed of directions $q \in \mathbb{S}$ such that $\Pr(z_q = 0)$ is positive. These directions are orthogonal to exposed faces of the identified set. Furthermore, the identified set has kinks if and only if there exist q and $r \neq q$ such that:*

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

Spaces orthogonal to convex combinations of such q and r are tangent to set B at such kinks.

Proof. See Appendix B. ■

Exposed faces of the identification set B are intersections of B and supporting hyperplanes that are not reduced to singletons (see Rockafellar, 1970, pp:162-163) and set D_f is not empty for instance when some variables z have mass points. Defining this set turns out to be important for asymptotic properties derived in the next section. One Monte Carlo experiment in Section 5.3 analyzes the common case in which one explanatory variable is a dummy variable.

3.2 Supernumerary Moment Conditions

We consider now that the dimension, m, of the random vector z is larger than the dimension, p, of covariates x and we denote $x(z)$ the linear projection of x onto instruments z, i.e., $x(z) = zE(z^\top z)^{-1}E(z^\top x)$. Without loss of generality, we assume that the $m-p$ supernumerary instruments $z^s = (z_{p+1}, ..., z_m)$ provide supernumerary moment conditions in the sense that no linear combination of these additional instruments is linearly dependent of $x(z)$. These instruments always exist because of the rank condition ***R.iii***. Formally, if ε^s denotes the vector of residuals of the linear projection of these $m-p$ instruments onto $x(z)$, we assume rank($E(\varepsilon^{sT}\varepsilon^s)) = m-p$. It may very well be the case that other subsets of $m-p$ instruments satisfy this condition, but, as discussed in the appendix, our results do not depend on the choice of a specific subset.

The parameters of interest β satisfy the incomplete linear moment conditions (3):

$$E(z^\top x)\beta = E(z^\top(y + u(z))), \tag{9}$$

and the identified set B is again closed, convex and bounded. The first two properties hold as before because the moment conditions are linear and the admissible set $I(z)$ containing $u(z)$ is closed and convex. To show that *B* is bounded, we can always restrict equation (9) to a subset of p instruments, say $x(z)$, and construct the identified region as in the previous section. The true identified set is included in this identified region.

Yet in contrast to the case in which ($m = p$), the identified 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. To do that, we need new notations. Define first z_F the ortho-normalization of the linear projection $x(z)$ defined above

$$z_F = x(z)E(x(z)^\top x(z))^{-1/2}.$$

Define also z_H the ortho-normalization of the residuals ε_H of the projection of the supernumerary instruments z^s onto z_F. Analytically, $z_H = \varepsilon_H E(\varepsilon_H^\top \varepsilon_H)^{-1/2}$ where,

$$\varepsilon_H = z^s - z_F E(z_F^\top z^s) = z^s - x(z)E(x(z)^\top x(z))^{-1}E(x(z)^\top z^s).$$

After some algebraic manipulations, we have by construction that $E(\varepsilon_H^\top x) = 0_{m-p,p} \implies E(z_H^\top x) = 0_{m-p,p}$.

3.2.1 The Validity of Supernumerary Moment Conditions

Both vectors z_F, of dimension p, and z_H, of dimension $m-p$, are linear combinations of the m instruments z, so that Equation (9) implies,

$$E(z_F^\top x)\beta = E(z_F^\top(y+u(z))) \text{ and } E(z_H^\top x)\beta = E(z_H^\top(y+u(z))).$$

As $E(z_H^\top x) = 0_{m-p,p}$, the second set writes $E(z_H^\top(y+u(z))) = 0$. Not only these two sets of restrictions are necessary, but they can be proven to be sufficient:

Lemma 4 *Parameter β belongs to B if and only if there exists $u(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that:*

$$E(z_F^\top x)\beta = E(z_F^\top(y+u(z))) \tag{10}$$

$$E(z_H^\top(y+u(z))) = 0 \tag{11}$$

Proof. See Appendix B.3. ■

Interestingly enough, the second set of restrictions does not depend on β whereas the first set provides a one-to-one relationship between admissible $u(z)$ and admissible β. It follows that B is non empty if and only if there is $u(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that

$$E(z_H^\top(y+u(z))) = 0. \tag{12}$$

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

$$B_{\text{Sargan}} = \{\gamma : \gamma = E(z_H^\top(y+u(z))), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\} \subset \mathbb{R}^{m-p}. \tag{13}$$

The adapted Sargan condition given by equation (12) means that B_{Sargan} contains the point $\gamma = 0$, that is O_{m-p}, the origin point of $\mathbb{R}^{m-p}$. [7]

Proposition 5 *The two following conditions are equivalent:*

i. B is not empty,

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

[7] As discussed at the end of the proof of Lemma 4, the adapted Sargan condition imposes restrictions on the set of admissible $u(z)$ which do not depend on the choice of supernumerary instruments, z_H.

Proof. Using the previous developments. ■

This extends 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 $E(z_H^\top.y)$. The Sargan or J-test consists in testing $O_{m-p} \in B_{\text{Sargan}} = \{E(z_H^\top.y)\}$ or equivalently that $E(z_H^\top y) = 0$. In section 4, we will construct a general test for the assumption $H_0 : \beta_0 \in B$, when B is the identified region of an incomplete linear moment model. It will provide us with a direct way for testing the Sargan condition given in Proposition 5. Before moving on to this issue, the next subsection provides a characterization of the identified set when there are supernumerary moment conditions.

3.2.2 Geometric and Analytic Characterization of the Identified Set

Assuming that the Sargan condition holds true, the identified set B is defined by the incomplete moment conditions:

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

This set of restrictions can be rewritten by introducing auxiliary parameters, γ, as:

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

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

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

An interesting feature of this relaxed program is that the number of explanatory variables (m) is equal to the number of moment conditions, and no more moments are in surplus. Consequently, the support function of B_U can be characterized using Proposition 2. The second interesting feature of B_U is that the identified set B corresponds to the intersection of B_U and the hyperplane defined by $\gamma = 0$. General results for the support function of intersection of convex sets (Rockafellar, 1970) can be used to characterize set B and this yields:

Proposition 6 *Let q a vector of $\mathbb{R}^p$ and (q, λ) a vector of $\mathbb{R}^m$. We have:*

$$\delta^*(q \mid B) = \inf_\lambda \delta^*((q, \lambda) \mid B_U)). \tag{14}$$

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

Proof. Rockafellar (1970) and Appendix B ■

The geometric intuition 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. The vector $\lambda_m(q)$ characterizes this projection direction. It corresponds to a tangent space (not necessarily unique) of B_U at β_f.

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

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

and is equal to the set of unconstrained solutions to equation (10). This projection contains set B since parameters in set B are generated by functions $u(z)$ satisfying also condition (11). Supernumerary restrictions therefore reduce the size of the identified set and generically they strictly do so.[8]

3.2.3 Supernumerary Moment Conditions as a Way to Restore Point Identification

The adapted Sargan condition ($O \in B_{\text{Sargan}}$) imposes restrictions on the size of the set of admissible functions $u(z)$ and, consequently, on the size of the identified set B. This section explores whether B can eventually be reduced to a singleton and point identification be restored.

When the point O_{m-p} belongs to the interior of B_{Sargan}, functions which satisfy the Sargan condition (12) are by construction not unique and set B has necessarily a non empty interior. More interesting cases arise when O_{m-p} belongs to the frontier of B_{Sargan}. Using the proof of Proposition 2, the frontier points of B_{Sargan} are generated by functions $u_q^{\text{Sargan}}(z)$ defined as,

$$u_q^{\text{Sargan}}(z) = \overline{\Delta}(z)\mathbf{1}\{z_q > 0\} + \underline{\Delta}(z)\mathbf{1}\{z_q < 0\} + \mathbf{\Delta}^*(z)\mathbf{1}\{z_q = 0\}$$

where $\mathbf{\Delta}^*(z)$ can be any function taking values in $[\underline{\Delta}(z), \overline{\Delta}(z)]$.

Suppose first that O_{m-p} lies on an exposed face of B_{Sargan} and let q_O the vector of $\mathbb{S}_{m-p}$ orthogonal to this face. Lemma 3 implies that $\Pr\{z_{Hq_O} = 0\} > 0$, the generating function $u_{q_O}^{\text{Sargan}}(z)$ is thus not unique and the identified set is not reduced to a singleton. In contrast, if O_{m-p} is not on an exposed face of B_{Sargan}, $u_{q_O}^{\text{Sargan}}(z)$ is unique and the set B is reduced to a singleton $\{\beta_0\}$ defined as:

$$\beta_0 = (E(z_F^\top x))^{-1}(E(z_F^\top (y + u_{q_O}^{\text{Sargan}}(z)))).$$

[8] Appendix B shows that the resulting set B does not depend on which version of z_H was chosen and thus on which version of unconstrained B_U was selected.

We summarize this result in:

Proposition 7 *If O_{m-p} belongs to the frontier ∂B_{Sargan} albeit not to an exposed face of B_{Sargan}, parameter β is point-identified.*

4 Estimation and Inference

This section provides a description of how we estimate the support function of B and how we test hypotheses of interest. We will deal only with random samples $i = 1, ., n$, where $(\overline{y}_i, \underline{y}_i, y_i, x_i, z_i)$ is observed in the data and independently and identically distributed.[9] We start by analysing the case where there is no supernumerary moment conditions.

4.1 Consistent and Asymptotically Normal Estimation: No Supernumerary Moment Conditions

In this section, we provide an estimate of the support function of the identified set B as characterized in Proposition 2 :

$$\delta^*(q \mid B) = E(z_q w_q). \tag{15}$$

To apply the analogy principle, we first construct $\hat{\Sigma}_n$ a bounded estimate[10] of $E(x^\top z)^{-1}$ and we define for any i:

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

We define the estimate of $\delta^*(q \mid B)$ as:

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

Under usual conditions (White, 1999, p35), the estimate $\hat{\delta}_n^*(q \mid B)$ is uniformly consistent.

Proposition 8 *Assume that there exist $M > 0$ and $\gamma > 0$, such that $E(\||x^\top z\|^{1+\gamma})$, $E(\|z^\top \overline{y}\|^{1+\gamma})$ and $E(\|z^\top \underline{y}\|^{1+\gamma})$ are bounded by M. Then, $\hat{\delta}_n^*(q|B)$ is, uniformly over $\mathbb{S}$, strongly consistent:*

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

[9]Note that it precludes pre-estimation of the bounds as in Magnac and Maurin (2008). We leave this extension for future work.

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

Proof. See Beresteanu and Molinari, 2008 and Additional Appendix E. ■

The proof builds on the fact that the expression $z_q w_q$ within the expectation defining $\delta^*(q \mid B)$ can be written as a random function $f_{(q,\Sigma)}(z_i, \overline{y}_i, \underline{y}_i)$ indexed by parameter $(q, \Sigma) \in \Theta = \mathbb{S} \times \{\|\Sigma\| \leq M\}$. Under the conditions of proposition 8, the parametric class of functions $f_{(q,\Sigma)}$ is Glivenko-Cantelli. If Σ is known, the empirical expectation of $f_{(q,\Sigma)}$ converges almost surely to $\delta^*(q|B)$ uniformly over Θ under the conditions stated above. Using results for parametric classes (van der Vaart, 1998), we can replace Σ by a bounded consistent estimate $\hat{\Sigma}_n \in \{\|\Sigma\| \leq M\}$ and the same result holds true.

We use similar reasonning to derive the asymptotic distribution of the estimate by considering the stochastic process defined on $\mathbb{S}$:

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

whose asymptotic behavior is characterized under usual conditions (White, 1999, p118) in the following.

Proposition 9 *Assume that there exist $M > 0$ and $\eta > 0$ such that $E(\left\|x^\top z\right\|^{2+\gamma})$, $E(\left\|z^\top \overline{y}\right\|^{2+\eta})$ and $E(\left\|z^\top \underline{y}\right\|^{2+\eta})$ are bounded by M. When set B has no exposed faces (see Lemma 3), $\tau_n(q)$ uniformly converges in distribution when n tends to ∞ to a Gaussian stochastic process $\tau_0(q)$ centered at zero. The covariance function of this process for vectors $(q, r) \in \mathbb{S}$ is,*

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

where $\varepsilon_q = w_q - xE(z^\top x)^{-1}E(z^\top w_q)$ are the residuals of the IV regression of w_q on x using instruments z.

When set B has exposed faces, $\tau_n(q)$ uniformly converges in distribution when n tends to ∞ to $\tau_0(q) + \tau_1(q)$, where $\tau_1(q)$ is asymptotically equivalent to :

$$E(\left|\eta^\top W^{1/2}(I_K \otimes q)z_i^\top\right|(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_i\Sigma q = 0\}))/2,$$

where W is the asymptotic variance of $vec(\hat{\Sigma}_n^\top)$ and whereas η is a normally distributed random vector of dimension p^2, independent of z.

Proof. See Appendix C. ■

When there are no exposed faces, Beresteanu and Molinari (2008) already derived the asymptotic distribution of $\tau_n(q)$ using the formalism of set valued random variables. Proposition

9 provides an alternative characterization of the covariance function of the Gaussian process in terms of residuals ε_q that are easy to compute. When there are exposed faces, this Proposition provides the original result that the limit in distribution is the sum of a Gaussian process and a countable point process which takes non zero values at directions q orthogonal to exposed faces of set B.

4.2 Tests

In partially identified models, using set inclusion tests to conduct inference about parameter values leads to conservative tests. In this section, we focus on testing parameter values such as the null hypothesis $H_0 : \beta_0 \in B$, although a similar approach could be used for sets (see Beresteanu and Molinari, 2008). We restrict our analysis to the case in which the identified set B has no exposed faces (see Lemma 3 for deep conditions on the support of z) so that the estimate of the support function is asymptotically Gaussian (see Proposition 9).

Assumption D: *The support function $\delta^*(q \mid B)$ is differentiable everywhere.*

An alternative characterization of H_0 using the support function is:

$$\beta_0 \in B \Longleftrightarrow \forall q \in \mathbb{S}, T_\infty(q;\beta_0) = \delta^*(q|B) - q^\top \beta_0 \geq 0 \Longleftrightarrow \min_{q\in\mathbb{S}} T_\infty(q;\beta_0) \geq 0,$$

as $\mathbb{S}$ is compact. If we knew a minimizer q_0 of $T_\infty(q;\beta_0)$, we could consider the empirical analog of $T_\infty(q_0;\beta_0)$:

$$T_n(q_0;\beta_0) = \hat{\delta}_n^*(q_0|B) - q_0^\top \beta_0,$$

and use that $\sqrt{n}(T_n(q_0;\beta_0) - T_\infty(q_0;\beta_0))$ is asymptotically normally distributed and have variance $V_{q_0} = V(z_{q_0}^\top \varepsilon_{q_0})$. Observe that, when the point β_0 belongs to the frontier of the set, $T_\infty(q_0;\beta_0) = 0$ and $\sqrt{n}T_n(q_0;\beta_0)/V_{q_0}$ qualifies as a test statistics for H_0.

The two issues that we have to deal with are (1) q_0 is not known (2) it needs not be unique if set B has kinks. We thus have to select one admissible q_0 and replace it by an estimate. The next Proposition shows how to address the second issue by perturbing function $T_\infty(q;\beta_0)$ and the first issue by minimizing the empirical analogue of such a function.

Proposition 10 *Under Assumption D and conditions given in Proposition 9, there exist two sequences $v_{0,n} \in \mathbb{S}$ and $a_n \in \mathbb{R}^+$ characterized in the proof, such that any sequence q_n of local* *minimizers of the perturbed program :*

$$\hat{\Psi}_{n,a_n}(q;\beta_0) = T_n(q;\beta_0) - a_n q^\top v_{0,n},$$

converges, when n tends to ∞, to one single minimizer q_0^* of $T_\infty(q;\beta_0)$. Then,

$$\begin{cases} \sqrt{n}T_n(q_n;\beta_0)/\hat{V}_n \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1), & \textit{if } \beta_0 \in \partial B, \\ \sqrt{n}T_n(q_n;\beta_0)/\hat{V}_n \xrightarrow[n\to\infty]{a.s.} +\infty, & \textit{if } \beta_0 \in int(B), \\ \sqrt{n}T_n(q_n;\beta_0)/\hat{V}_n \xrightarrow[n\to\infty]{a.s} -\infty, & \textit{if } \beta_0 \notin B, \end{cases}$$

where $\hat{V}_n = \widehat{V}(z_{n,q_n}\varepsilon_{n,q_n})$ is a consistent estimator of $V_{q_0^*}$.

Proof. See appendix C.2 ■

Two cases in which there is no need to actually perturb the program and a_n can be set to zero are worth noticing: First, when set B has no kinks, for instance when the density of z is positive everywhere, $T_\infty(q;\beta_0)$ is strictly convex and any sequence of local minimizers of $T_n(q_n;\beta_0)$ tends to the unique minimizer of $T_\infty(q;\beta_0)$. Second, we can set a_n to zero when we test a single component of β_0, or a single linear combination of components since there is no kink in the single dimensional case.

Critical regions with asymptotical level α for two interesting null hypotheses can be constructed:

- 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:

$$W_n^1(\alpha) = \{\beta_0 \in \mathbb{R}^p,\ \sqrt{n}T_n(q_n;\beta_0)/\hat{V}_n < \mathcal{N}_\alpha\}$$

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

$$W_n^2(\alpha) = \{\beta_0 \in \mathbb{R}^p,\ |\sqrt{n}T_n(q_n;\beta_0)/\hat{V}_n)| > \mathcal{N}_{1-\frac{\alpha}{2}}\}$$

where $\mathcal{N}_\alpha$ denotes the α-quantile of the standard normal distribution and where q_n is defined by Proposition 10. In addition, the test statistics is asymptotically pivotal so that we could enhance its finite sample properties by bootstrapping it.

We are specifically interested by the first test. The second one is also of practical interest for instance when 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 $100 - 100\alpha$ %. Following Lehmann (1986, Chapter 3), the confidence region CI_α^n is the collection of parameters $\beta \in \mathbb{R}^d$ for which the null hypothesis is not rejected *i.e.* which does not belong to $W_n^1(\alpha)$. The following proposition expresses this statement and Appendix E.2 provides a simple way of constructing the confidence region.

Proposition 11 *Let α be a significance level, and let CI_α^n be the set of points of $\mathbb{R}^p$ such that $\xi_n(\beta) > \mathcal{N}_\alpha$, where*

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

where $\hat{V}_n$ is defined in Proposition 10. Under the conditions of Proposition 10,

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

The limit expressed in the proposition is valid for a fixed data generating process leading to the identification of a proper set B. 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 region is not uniformly asymptotically of nominal size equal to $(1 - \alpha)$. However uniformity is important as we might never know, in practice, how far we are from a just identified case.

For simplicity, assume for the remaining part of the section that set B is strictly convex and smooth i.e. the support function is differentiable and strictly convex. Let us consider the limit case in which set $B = \{\beta_0\}$. If we construct a confidence region for the parameters using the last Proposition the coverage probability will tend to $1 - 2\alpha$ (see Additional Appendix E.3). Indeed, the statistics developed above is discontinuous with respect to the diameter of the identified set at the boundary *i.e.* when the diameter is equal to zero.[11]

The construction of Imbens and Manski (2004) is uniformly valid in a context of single dimensional sets and more recently, Stoye (2009) clarified the conditions under which this result can be obtained. We can adapt Lemma 4 in Imbens and Manski (2004) to our set-up where the length of the interval is replaced by the diameter of the set and construct a uniform confidence

[11] The diameter of the set B is the maximum of $\delta(q|B) + \delta(-q|B)$ on the compact unit sphere.

region (see Additional Appendix E.4).[12]. In appendix B.1, we indeed showed that the diameter is strictly positive.

4.4 Consistent and Asymptotically Normal Estimation: The Supernumerary Case

We use the characterization given in equation (14) in Proposition 6 and . 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 $\delta^*((q, \lambda)|B_U)$ as derived in Section 4.1 and such that, by Proposition 8:

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

and, by Proposition 9, under condition D:

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

uniformly converges to a Gaussian process when n tends to infinity.[13]

For any q, define:

$$\hat{\lambda}_n(q) \in \arg\min_{\lambda} [\hat{\delta}_n^*((q, \lambda)|B_U) + a_n \lambda^T \lambda]$$

where a_n is a sequence converging to zero with n, defined in the proof below. The estimate $\hat{\lambda}_n(q)$ is a solution to a perturbed objective function as in Section 4.2. Define the estimate of the support function of the identified set as:

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

The same kind of proof as in Sections 4.1 and 4.2 then applies. Estimating $\hat{\lambda}_n(q)$ does not affect the consistency and asymptotic normality of the support function estimates.

[12] 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.

[13] The difficulty with dealing with the additional point processs is that the limit is not asymptotically equicontinuous. We leave this subject for future research and assume that set B has no exposed faces (condition D).

Proposition 12 *Under the conditions stated in Proposition 9 and condition D we have:*

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

and:

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

converges to a Gaussian process when n tends to infinity. The Gaussian process has expectation equal to zero and its covariance operator for two directions $(q, r) \in \mathbb{S}$ is given by:

$$E(z_{(q,\hat{\lambda}_n(q))}\varepsilon_{(q,\hat{\lambda}_n(q))}\varepsilon_{(r,\hat{\lambda}_n(r))}z_{(r,\hat{\lambda}_n(r))})$$

Proof. See appendix C. ■

5 Monte-Carlo Experiments

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 and the identified set is of dimension 2 for simplicity. In the first two experiments, the frontier of the identified set has no kinks and no exposed 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 is neither smooth nor strictly convex in the third experiment.

5.1 Smooth and Strictly Convex Sets

Consider the model:

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

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

The support function of the identified set B is and equal to (see Appendix D.1):

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

where $\Delta = \frac{1}{2K}$. In other words, the identified set B is a circle whose radius is $\frac{2\Delta}{\sqrt{2\pi}}$ (see Table E.4).

We draw 1000 simulations in four different sample size experimentss : n =100, 500, 1000 and 2500. We report results when the number of intervals, K, is equal to 2 as our results are robust when K is increased. The three quartiles as well as the mean of the distribution of the estimated support function at one angle are displayed in Table E.4 although all angles give the same results. Even for small sample size, the identified set is well estimated and unsurprisingly, the interquartile interval decreases when the sample size increases.

Regarding the performance of test procedures, 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, a definition that is valid for any ray since set B is a disk around the true value $\beta^0 = 0$. Point 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, we computed the rejection frequencies at a 5% level for the two tests developed in Section 4.2: Whether β^r belongs to B against the alternative that it does not (Test 1); Whether it belongs to the frontier of B against the alternative that it does not (Test 2). Results are reported in Table 2 in the Additional Appendix. These results show that the size of the three tests is very accurate and remains very close to 5% even for $n = 100$ and that the power of these tests is very good even in small samples.

5.2 Smooth set with one supernumerary instrument

The simulated model is as before except that the second explanatory variable x_2 is now generated 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. Variables x_1, e_2 and w are used for estimating set B instead of x_1 and x_2 and we have therefore one supernumerary instrument. Note that parameter π (respectively ν) measures the strength of the correlation between x_2 and e_2 (respectively x_2 and w).

Setting $q = (\cos\theta, \sin\theta)^\top$, the support function can be expressed as (see Appendix D.2):

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

When $\nu = 1$, set B is the same as in the previous example because x_2 is a deterministic function of e_2 and w. Moreover when π and ν are positive and strictly lower than 1, there is some information loss due to the use of e_2 and w instead of x_2 and set B is stretched along the second axis (see Figure, Table E.4).

As before, we draw 1000 simulations in four sample size experiments : n =100, 500, 1000 and 2500. Table E.4 displays descriptive statistics (Mean, quartiles) related to the distribution of the estimated support function at one angle. Table 4 displays the percentage of rejections for the tests for different points along the x-axis. The line which corresponds to the frontier point ($r = 1$) is reported in bold. As before, there is no significative distortion when 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 exposed faces and its support is discrete so that the identified set has kinks. The simulated model is:

$$y^* = \frac{1}{2} + \frac{x}{8} + \varepsilon$$

where x is equal to -1 with probability $\frac{1}{2}$ and to 1 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 $(\frac{1}{2}, \frac{1}{8})^\top$. As before, we only observe y^* by intervals ($I_1 = [0, \frac{1}{2}]$ and $I_2 = [\frac{1}{2}, 1]$). The identified set B_2 can be shown to be the convex envelop of the four points $(\frac{3}{4}, \frac{1}{8})$, $(\frac{1}{2}, \frac{3}{8})$, $(\frac{1}{4}, \frac{1}{8})$ and $(\frac{1}{2}, -\frac{1}{8})$ (see Appendix D.3). As in the previous example, we simulate 1000 draws for 4 sample sizes: 100, 500, 1000 and 2500 and the same conclusions concerning the estimation of the set remain valid here (see Table E.4 in the Additional Appendix)

One attractive feature of this toy example is that, despite the presence of exposed faces, the additional term $\tau_1(q)$ in the asymptotic distribution of the support function (see Proposition 9) vanishes (see Appendix D.3) and we can apply the test procedures developed in the Gaussian case. We focus on the points belonging to the half-line starting from the central point $\beta^* = (1/2, 1/8)$ and parallel to the x-axis. Like before, we index the points by r the fraction of the distance to the frontier along this axis and $\beta^1 = (3/4, 1/8)$ the frontier point is now a kink of set B.

Table 5 displays the rejection rate for the test of the frontier for different values of r (from 0.01 to 2) at a 5%-level test (Table 7 in the Additional Appendix reports results for the test for the interior). In the first panel of columns (labeled $a_n = 0$), we display results ignoring that there is a kink whereas by Proposition 10 we should be using perturbed programs ($a_n > 0$). Unsurprisingly we overreject at the frontier point . In the second panel of columns, we display the rejection rates using the perturbed program defined in Proposition 10 with $a_n = \frac{0.5}{n^{1/3}}$. Rejection rates, though smaller than in the previous case, are still too large and the reason is the estimate of the variance at q_0^*. Sample sizes properties can indeed be improved while estimating the variance with i.i.d. bootstrap techniques. Rejection rates that are not reported here are reduced by a factor of around 40% and are thus much more in line with the nominal size.

6 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.

There remains many pending questions. Adapting our test procedure to the case in which the set has exposed faces is high on the agenda. Various other extensions were also out of the scope of this paper. First, some examples that we developed require more work in terms of estimation and asymptotic theory even if our set-up provides a building block to study the asymptotic properties of these estimates. For instance, 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). Second, other examples about categorical data or two-sample combination need also some adaptation of the identification analysis.

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., 1994, ”Empirical Process Methods in Econometrics”, eds R.Engle and D.McFadden, *Handbook of Econometrics*, IV:2247-2294, North Holland: Amsterdam.

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, 2007, ”Asymptotic Properties for a Class of Partially Identified Models”, forthcoming in *Econometrica*.

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., 2005, ”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.

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.

Horowitz, J.L. and C.F. Manski, 2006, ”Identification and Estimation of Statistical Functionals using Incomplete Data”, *Journal of Econometrics*, 132:445-459.

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.

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.

McFadden, D., 1974,

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.

Ridder G. E. and R., Moffitt, 2007, ”The econometrics of data combination”, in J. J.Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, North-Holland, Amsterdam.

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.

Stoye, J., 2008,

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 there is a latent random variable ε uncorrelated with x such that the latent variable $y^* \equiv x\beta + \varepsilon$ lies within the observed bounds, i.e., $x\beta + \varepsilon \in \left[\underline{y}; \overline{y}\right]$. Denoting $y = (\overline{y} + \underline{y})/2$ and using that ε is uncorrelated with x, we have,

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

We also have :

$$-\frac{(\overline{y} - \underline{y})}{2} \leq y^* - y \leq \frac{(\overline{y} - \underline{y})}{2}$$

which yields bounds on $u(x) \equiv E(y^* - y|x)$,

$$-E(\frac{(\overline{y} - \underline{y})}{2} \mid x) \leq u(x) \leq E(\frac{(\overline{y} - \underline{y})}{2} \mid x)$$

Setting $\Delta(x) = E(\frac{\overline{y} - \underline{y}}{2} \mid x)$, there thus exists a measurable $u(x) \in [-\Delta(x), \Delta(x)]$ such that $E(x^\top(x\beta - y)) = E(x^\top u(x))$.

(Sufficiency) Conversely, let us assume that there exists $u(x)$ in $[-\Delta(x), \Delta(x)]$ such that $E(x^\top(x\beta - y)) = E(x^\top u(x))$. We are going to construct a random variable ε which is uncorrelated with x and which is such that $y^* \equiv x\beta + \varepsilon$ lies within the observed bounds.

First, consider λ a random variable whose support is $[0, 1]$, which is independent of $\underline{y}$ and $\overline{y}$ and whose conditional mean given x is:

$$E(\lambda|x) = \frac{1}{2}\frac{u(x)}{\Delta(x)} + \frac{1}{2}.$$

Second, define ε as :

$$\varepsilon = -x\beta + (1 - \lambda)\underline{y} + \lambda\overline{y}$$

By construction, $y^* \equiv x\beta + \varepsilon$ is consistent with the observed censoring mechanism i.e. $y^* \in \left[\underline{y}; \overline{y}\right]$. Let us prove that ε is also uncorrelated with x. Consider, for almost any x,

$$\begin{aligned} E(y|x) - E(x\beta + \varepsilon|x) &= E(\frac{(\overline{y} + \underline{y})}{2} \mid x) - E((1 - \lambda)\underline{y} + \lambda\overline{y} \mid x) \\ &= E((1 - 2\lambda)\frac{(\overline{y} - \underline{y})}{2} \mid x) = E((1 - 2\lambda) \mid x)E(\frac{(\overline{y} - \underline{y})}{2} \mid x) \\ &= E(-\frac{u(x)}{\Delta(x)}\Delta(x) \mid x) = -u(x). \end{aligned}$$

where we used that λ is independent of $\underline{y}$ and $\overline{y}$. Therefore, we have $E(\varepsilon|x) = E(y - x\beta|x) + u(x)$, which implies:

$$E(x^\top\varepsilon) = E(x^\top(y - x\beta)) + E(x^\top u(x)) = -E(x^\top u(x)) + E(x^\top u(x)) = 0.$$

using the moment condition (ii) involving y, β and $u(x)$.

B Proofs in Section 3

B.1 Proof of Proposition 2

The support function in direction $q \in \mathbb{S}$ is obtained as the supremum of the expression

$$q^\top \beta = E\left(z_q\left(y + u(z)\right)\right), \tag{B.1}$$

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:

$$u_q(z) = \underline{\Delta}(z) + (\overline{\Delta}(z) - \underline{\Delta}(z))\mathbf{1}\{z_q > 0\} + \mathbf{\Delta}^*(z)\mathbf{1}\{z_q = 0\} \tag{B.2}$$

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 (4), $E(\overline{y} - y|z) = \overline{\Delta}(z)$, $E(\underline{y} - y|z) = \underline{\Delta}(z)$, so that the support function or the supremum of (B.1) is equal to:

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

where:

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

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 $\Pr(z_q = 0) > 0$,an since $\mathbf{\Delta}^*(z)$ varies in $[\underline{\Delta}(z), \overline{\Delta}(z)]$,the functions $u_q(z)$ defined by equation (B.2) generate all points $\beta = \left(E(z^\top x)\right)^{-1} E(z^\top(y + u_q(z))$ which belong to the tangent space to B that is orthogonal to q (an exposed face in the vocabulary used in the next Proposition).

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

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

and, by definition:

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

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

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

or equivalently that:

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

Start from consequences of definitions:

$$z_q = q^\top E(z^\top x)^{-1} z^\top = -z_{-q}, w_q - w_{-q} = (\bar{y} - \underline{y})(\mathbf{1}\{z_q > 0\} - \mathbf{1}\{z_q < 0\}),$$

so that:

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

because $E(\bar{y} - \underline{y}|z) = \overline{\Delta}(z) - \underline{\Delta}(z) \geq \Delta_0 > 0$ and $|z_q| > 0$ with positive probability because of the full rank assumption in $R.iii$.

This quantity $\delta^*(q|B) + \delta^*(-q|B)$ is the diameter of B in direction q, and by using the same argument:

$$\max_{q \in \mathbb{S}} (\delta^*(q|B) + \delta^*(-q|B)) > 0$$

since $\mathbb{S}$ is compact.

B.2 Proof of Lemma 3

We use the expression derived in Proposition 2:

$$\delta^*(q|B) = E(z_q w_q) = E(z_q \underline{y}) + E(z_q \mathbf{1}\{z_q > 0\}(\overline{y} - \underline{y})). \tag{B.3}$$

First of all, the support function of a convex set is convex and therefore is differentiable except at a countable number of directions q denoted D_f. We characterize this set and identify it with the set of directions orthogonal to exposed faces and finally turn to characterizing kink points of set B.

Characterization of D_f The first term on the RHS of equation (B.3) is linear in q since (see the previous proof) :

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

and thus is continuously differentiable on $\mathbb{S}$. As $E(\overline{y} - \underline{y} \mid z) \geq \Delta_0 > 0$, the second term can be written as:

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

where $z^* = z(E(x^\top z))^{-1}(\overline{y} - \underline{y})$. The set of points D_f is the set of points where $\psi(q)$ is not differentiable.

Fix $q \in \mathbb{S}$. For any $t \in \mathbb{S}$:

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

so that:

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

Points of non differentiability are obtained when the expression in RHS is NOT $o(\|t - q\|)$. It is the sum of three terms :

$$\begin{aligned}
A_1 &= E(z^*.t.\mathbf{1}\{z^*.t > 0, z^* q < 0\}), \\
A_2 &= -E(z^*.t.\mathbf{1}\{z^*.q > 0, z^* t \leq 0\}) \\
A_3 &= E(z^*.t.\mathbf{1}\{z^*.q = 0, z^* t > 0\})
\end{aligned}$$

Regarding A_1 and A_2, when $z^*.t > 0$ and $z^* q < 0$, we have,

$$0 < z^*.t = z^*.(t-q) + z^* q < z^*.(t-q),$$

whereas when $z^*.q > 0$ and $z^* t \leq 0$, we have,

$$z^*.(t-q) < z^* t \leq 0.$$

Hence, we get,

$$\begin{aligned} 0 &\leq |A_1| \leq E(\|z^*\|)\, \|t-q\| \Pr(z^*.t>0, z^*q<0), \\ 0 &\leq |A_2| \leq E(\|z^*\|)\, \|t-q\| \Pr(z^*.q>0, z^*t\leq 0). \end{aligned}$$

As $\Pr(z^*.t>0, z^*q<0) = \Pr(z^*.(t-q) > -z^*q > 0)$ we have $\lim_{t\to q}\Pr(z^*.t>0, z^*q<0)=0$. Similarly, $\lim_{t\to q}\Pr(z^*.q>0, z^*q\leq 0)=0$, so that these inequalities imply:

$$A_1 = o(\|t-q\|) \text{ and } A_2 = o(\|t-q\|),$$

since R.iii implies that $E(\|z^*\|)$ is bounded.

Regarding the last term A_3, note that in the case in which $\Pr(z^*.q=0)=0$, we have $A_3=0$ and thus $\psi(q)$ is differentiable at q. Consider now the case in which $\Pr(z^*.s=0)>0$. We have

$$\begin{aligned} A_3 &= E(z^*.t.\mathbf{1}\{z^*.q=0, z^*t>0\}) = E(z^*\mathbf{1}\{z^*.q=0, z^*(t-q)>0\}).(t-q) \\ &= E(z^*\mathbf{1}\{z^*.q=0, z^*h>0\}).h \end{aligned}$$

where $h=t-q$ can be any vector provided that q and t belong to the unit sphere. It follows that ψ has different derivatives in different directions h, derivatives which depend on the term,

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

It turns out that this term is equal to zero for any h if and only if $E(z^*1\{z^*.h>0\}|z^*.q=0)$ tends to zero for any $\|h\|\to 0$. It happens if and only if the support of z^* conditional on $(z^*.q=0)$ is $\{0\}$ which is impossible given the rank condition stated in $R.iii$.

Concluding the points of non differentiability of the support function are directions q such that $\Pr(z^*.q=0)=\Pr(z_q=0)>0$. There can be no more than a countable number of such points.

Exposed faces Using lemma 3 we obtain for any q which does not belong to D_f :

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

As $\delta^*(q|B)=q^\top\beta_q$, and $\beta_q \in \arg\max_{\beta\in B}(q^\top\beta)$, this result is a disguise of the envelope theorem.

Assume now that B has an exposed face B_f. By definition, B_f is the intersection of B with one of its supporting hyperplane H_f which is not reduced to a singleton. If q_f denotes the vector orthogonal to H_f, we have for any β_f in B_f :

$$\delta^*(q_f|B) = q_f^\top\beta_f,$$

which means (see equation (B.2)) that there exists $\mathbf{\Delta}_f^*(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that:

$$\begin{aligned} \beta_f &= \beta_{q_f} + \left(E(z^\top x)\right)^{-1} E(z^\top \mathbf{\Delta}_f^*(z)\mathbf{1}\{z_{q_f}=0\}) \\ &= \beta_{q_f} + \left(E(z^\top x)\right)^{-1} E(z^\top \mathbf{\Delta}_f^*(z)|z_{q_f}=0)\Pr(z_q=0) \end{aligned}$$

For the set of all β_f not to be reduced to the singleton $\left\{\beta_{q_f}\right\}$, we clearly need $\Pr(z_q=0)>0$ and the support of z conditional on $(z_{q_f}=0)$ not to be reduced to $\{0\}$.

Conversely, suppose that there exists a direction q such that $\Pr(z_q = 0) > 0$ and such that the support of z conditional on $(z_q = 0)$ is not reduced to $\{0\}$. Denote $\beta_q = \left(E(z^\top x)\right)^{-1} E(z^\top w_q)$ and H_q the supporting hyperplane at β_q orthogonal to q. Consider the set B_f of all β_f such that there exists $\mathbf{\Delta}_f^*(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that:

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

B_f is clearly included in $B\cap H_q$.Also, as the conditional support of z is not reduced to $\{0\}$ and $\Pr(z_q = 0)$ is non zero, the second term in the RHS is itself non zero for at least some $\mathbf{\Delta}_f^*(z)$, which implies that B_f is not reduced to the singleton $\{\beta_q\}$ and that B has an exposed face.

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

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

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

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

using Proposition 2. Write the decomposition:

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

to get:

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

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

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

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

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

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

Conversely, it is straightforward to see that if:

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

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

B.3 Proof of Lemma 4

We have already proven that conditions (10) and (11) are necessary, we want to prove that they are sufficient. Specifically, we suppose that conditions (10) and (11) hold true and we want to prove that,

$$E(z^\top(x\beta - (y + u(z))) = 0$$

To begin with, condition (10) implies :

$$E(x(z)^\top(x\beta - (y + u(z))) = 0, \tag{B.4}$$

whereas condition (11) implies (using $E(\varepsilon_H^\top x = 0)$,

$$E(\varepsilon_H^\top(x\beta - (y + u(z))) = 0.$$

Using $\varepsilon_H = z^s - z_F E(z_F^\top z^s)$ and condition (10), Equation (B.4) implies,

$$E(z^{sT}(x\beta - (y + u(z))) = 0. \tag{B.5}$$

Hence, using that $x(z)$ is (by construction) a linear combination of the $m-p$ instruments z^s and of the first p instruments $z^0 = (z_1, ., z_p)$, Equations (B.4) and (B.5) imply necessarily,

$$E(z^{0T}(x\beta - (y + u(z))) = 0.$$

Combining the last two equations yields,

$$E(z^\top(x\beta - (y + u(z))) = 0.$$

which finishes the proof.

We can now show that the choice of z_s is without loss of generality. Suppose that z_H associated with a given subset of supernumerary instruments z^s satisfies condition (11). Then, B is non empty because condition (11) is sufficient. Yet, if B is non empty and since condition (11) is necessary, condition (11) is necessarily satisfied by any other subset of $(m-p)$ instruments (say z_H^*) constructed from an alternative z^{*s} satisfying the same condition as z^s. Overall, because condition (11) is both necessary and sufficient for the condition that B is not empty, when it is satisfied by a given subset of supernumery instruments, it is necessarily satisfied by any alternative subsets.

There is another interesting way to see why restrictions involved by condition (11) are invariant to the choice of the specific subset of supernumerary instruments. Note that as z_H is (1) a linear combination of z (2) uncorrelated with $x(z)$ (3) of variance equal to the identity matrix, it can be written as $zE(z^\top z)^{-1/2}Q_H$ where the $m-p$ columns of matrix Q_H are an orthonormal basis of the kernel of the orthogonal projection onto $x(z)$. Changing one specific subset of supernumerary instruments z_H into an alternative subset z_H^* boils down to moving from one orthonormal basis Q_H to an alternative basis Q_H^* (i.e., to $Q_H^* = Q_H Q$, where Q is an orthogonal matrix). In other words, for any z_H^* satisfying the same conditions as z_H, there exists necessarily an orthogonal matrix Q (with $Q = Q_H^\top Q_H^*$) such that $z_H^* = z_H Q$. This basic linear relationship between all possible subsets of supernumerary instruments implies that when linear moment condition (11) is satisfied by a given subset it is necessarily satisfied by any alternative one.

B.4 Proof of Proposition 6

We assume that the Sargan condition (as given by Proposition 5) is satisfied so that the intersection of the set B_U and the hyperplane, $\gamma = 0$, is not empty. Both sets $\{\gamma = 0\}$ and B_U are convex. The support function of B_U is $\delta^*(x_1^*|B_U)$ where $x_1^* = (q_1, \lambda_1)$. The support function of $\{\gamma = 0\}$ is as follows if $x_2^* = (q_2, \lambda_2)$:

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

Corollary 16.4.1 page 146 Rockafellar (1970) states that the intersection of two convex sets such that their relative interiors[14] have one point in common, we have:

$$\delta^*(x^*|B_U \cap \{\gamma = 0\}) = \inf_{(x_1^*,x_2^*):x_1^*+x_2^*=x^*} (\delta^*(x_1^*|B_U) + \delta^*(x_2^*|\{\gamma = 0\})) \tag{B.6}$$

and the infimum is attained. Remark also that when the hyperplane $\gamma = 0$ is not tangent to set B_U and their intersection is not empty, their relative interiors of have all the points of the relative interior of the intersection in common. Thus:

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

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

$$\delta^*(q|B) = \inf_\lambda \delta^*((q,\lambda)|B_U)). \tag{B.7}$$

Furthermore, the infimum in λ is attained.

On the other hand, when the hyperplane $\{\gamma = 0\}$ is tangent to B_U, the relative interiors have no points in common. Corollary 16.4.1 page 146 Rockafellar (1970) states that we should replace Equation (B.6) 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.7) applies also to this case and the infimum is attained.

B.5 The Construction of B_U

Let $s = (q, \lambda)$ be the direction used for estimating B_U, λ being the components relative to the variables z_H. By definition of B_U, we have that:

$$\begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \left[E(z^\top x) : E(z^\top z_H)\right]^{-1} E\left[z^\top(y + u(z))\right].$$

The support function of B_U is as in proposition 2

[14] 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:

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

$$\delta^*(s|B_U) = E(z_s w_s),$$

where $z_s = s^\top \Omega^\top z^\top, \quad w_s = \underline{y} + (\overline{y} - \underline{y})\mathbf{1}\{z_s > 0\}$ and

$$\Omega = \left[E(z^\top x) : E(z^\top z_H)\right]^{-1},$$

is well defined because of the rank conditions $R.ii$ and Appendix B.3.

The invariance of this construction to the specific choice of z_H follows the same argument as before. Write:

$$z_H \gamma = z_H Q Q^\top \gamma, \lambda^\top \gamma = \lambda^\top Q Q^\top \gamma$$

for any arbitrary orthogonal matrix Q. The solution is thus invariant to the choice of Q provided that (z_H, γ, λ) is changed into $(z_H Q, Q^\top \gamma, Q^\top \lambda)$. Minimizing with respect to λ or $Q^T \lambda$ is equivalent.

C Proofs in Section 4

We denote M a generic majorizing constant.

C.1 Proof of Proposition 9

We use that:

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

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

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

where:

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

where $\hat{\Sigma}_n$ is an estimate of Σ.

Define $\|\Sigma\| = Tr(\Sigma)$ and choose M arbitrarily such that $M > Tr(\Sigma)$. We now show that we can construct an estimate of Σ satisfying $\left\|\hat{\Sigma}_n\right\| \leq M$. Define $\hat{\Sigma}_n^u$ the sample analog of Σ:

$$\hat{\Sigma}_n^u = \left(\frac{1}{n} \sum x_i^\top.z_i\right)^{-1}, \tag{C.8}$$

and define $\hat{\Sigma}_n$, the estimate of Σ, as:

$$\begin{cases} \hat{\Sigma}_n = \hat{\Sigma}_n^u & \text{if } \left\|\hat{\Sigma}_n^u\right\| \leq M, \\ \hat{\Sigma}_n = \hat{\Sigma}_n^u\left(\frac{M}{\left\|\hat{\Sigma}_n^u\right\|}\right) & \text{if not.} \end{cases} \tag{C.9}$$

The element $(q, \hat{\Sigma}_n)$ always belongs to the bounded set $\Theta = \mathbb{S} \times \{\|\Sigma\| \leq M\}$. Under the conditions $R.iii$, $\hat{\Sigma}_n$ is almost surely consistent to Σ:

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

and that, under the conditions of Proposition 9, $\hat{\Sigma}^u_n$ and $\hat{\Sigma}_n$ are asymptotically equivalent:

$$\sqrt{n}\left(\hat{\Sigma}_n - \hat{\Sigma}^u_n\right) \underset{n\to\infty}{\overset{P}{\to}} 0, \tag{C.10}$$

and asymptotically normal:

$$\sqrt{n}\left(vec(\hat{\Sigma}^T_n - \Sigma^T)\right) \Longrightarrow N(0, W). \tag{C.11}$$

We proceed in two steps. As the first step is simple, we give the proof of and consistency and asymptotic normality at the same time.

C.1.1 Consistency and Asymptotic Normality: Σ is known

Suppose that Σ is known and denote:

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

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

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

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.

As the proof of Lemma 3 shows, the function is convex in Σq and therefore Lipschitzian. Adapting the proof, we have:

$$\begin{aligned}
\left|f_{\theta_1}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_2}(z_i, \underline{y}_i, \overline{y}_i)\right| &\leq \max(\left\|z_i^\top \underline{y}_i\right\|, \left\|z_i^\top \overline{y}_i\right\|) \left\|q_1^\top\Sigma_1^\top - q_2^\top\Sigma_2^\top\right\|, \\
&\leq M\max(\left\|z_i^\top \underline{y}_i\right\|, \left\|z_i^\top \overline{y}_i\right\|) \left\|\theta_1 - \theta_2\right\|.
\end{aligned} \tag{C.12}$$

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

Under conditions $R.iii$, we have:

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

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

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

Also, under the conditions of Proposition 9, we have:

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

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, 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 of the proof of Proposition 9 consists in replacing Σ by the almost sure limit $\hat{\Sigma}_n$ defined above. Consistency is proved in the Additional Appendix. We will rely heavily on Section 19.4 of van der Vaart (1998) where relevant properties are proposed.

C.1.2 Proof of Asymptotic Normality

We analyze the asymptotic behavior of $\tau_n(q)$ defined as

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

Note that $\tau_n(q) \equiv A_n(q) + B_n(q)$ where,

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

To begin with, denote $\theta = (q, \Sigma)$ the true value and $\hat{\theta}_n = (q, \hat{\Sigma}_n)$ the estimate. Let us prove that if $\hat{\theta}_n \underset{n\to\infty}{\overset{P}{\to}} \theta$ uniformly in q:

$$E(z_{n,qi}w_{n,qi} - z_{qi}w_{qi})^2 = E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_\theta(z_i, \underline{y}_i, \overline{y}_i))^2 \underset{n\to\infty}{\overset{P}{\to}} 0. \tag{C.13}$$

Using equation (C.12), we have:

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

so that:

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

Under the conditions of Proposition 9, $E\left(\max(\left\|z_i^\top \underline{y}_i\right\|, \|z_i^\top \overline{y}_i\|)\right)^2 < \infty$ and is independent of q. As $\left\|\hat{\theta}_n - \theta\right\|^2$ tends in distribution to 0 uniformly in $q \in \mathbb{S}$ (equation (C.10)), it tends also in probability to 0, uniformly in $q \in \mathbb{S}$, which finishes the proof. Hence, we can apply Lemma 19.24 of van der Vaart (1998), so that $A_n(q)$ has the same distribution as:

$$C_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} z_{qi}w_{qi} - E(z_{qi}w_{qi})\right). \tag{C.14}$$

uniformly in $q \in S$. Therefore the problem boils down to compute the limit of these processes given in the following lemma:

Lemma 13 *We have, uniformly in $q \in \mathbb{S}$:*

(i) $B_n(q) - \sqrt{n}E(\left|q^\top(\Sigma_n^\top - \Sigma^T)z_i^\top\right|(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_i\Sigma q = 0\}))/2 - \sqrt{n}q^\top(\hat{\Sigma}_n^\top(\Sigma^T)^{-1} - I)\beta_q^* \underset{n\to\infty}{\overset{P}{\to}} 0,$

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

where $\beta_q^* = \Sigma^T E(z_i^\top w_{qi}^*)$, $\varepsilon_{qi}^* = w_{qi} - x_i\beta_q^*$, *and* $w_{qi}^* = w_{qi} + \frac{1}{2}(\bar{y}_i - \underline{y}_i)\mathbf{1}\{z_{qi} = 0\}$.

Proof. For convenience sake, we first rewrite w_{qi}^*:

$$w_{qi}^* = \underline{y}_i + (\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{qi} > 0\} + \mathbf{1}\{z_{qi} \geq 0\})/2.$$

and note that $E(z_{qi}w_{qi}) = E(z_{qi}w_{qi}^*)$.

We first prove (i). Write:

$$\begin{aligned} B_n(q) &= \sqrt{n}(E(z_{n,qi}w_{n,qi}) - E(z_{qi}w_{qi}^*)) = \sqrt{n}E(z_{n,qi}(w_{n,qi} - w_{qi}^*)) + E((z_{n,qi} - z_{qi})w_{qi}^*) \\ &\equiv B_n^1(q) + B_n^2(q). \end{aligned}$$

By definition of $z_{n,qi} = q^\top\hat{\Sigma}_n^\top z_i^\top$ and $z_{qi} = q^\top\Sigma^T z_i^\top$, the second term on the RHS is equal to:

$$\begin{aligned} B_n^2(q) &= \sqrt{n}(q^\top(\hat{\Sigma}_n - \Sigma)^\top E(z_i^\top w_{qi}^*) = \sqrt{n}q^\top(\hat{\Sigma}_n - \Sigma)^\top(\Sigma^\top)^{-1}\beta_q^*, \\ &= \sqrt{n}q^\top(\hat{\Sigma}_n(\Sigma^\top)^{-1} - I)^\top\beta_q^* \end{aligned}$$

using the definition of β_q^*. The first term on the RHS is equal by replacement of $w_{n,qi}$ and w_{qi}^* to:

$$B_n^1(q) = \sqrt{n}E(z_{n,qi}\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{n,qi} > 0\} - \frac{1}{2}(\mathbf{1}\{z_{qi} > 0\} + \mathbf{1}\{z_{qi} \geq 0\})\right]).$$

Because $z_{n,qi}$ is a root n consistent estimator of z_{qi}, uniformly for any z_i in a compact set, this expression converges to 0 in probability when $z_{qi} \neq 0$ since the last term, equal in this case to $\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\})$, is identically 0 out of a (root n) decreasing neighboorhood of the true value. We thus have:

$$\sqrt{n}E(z_{n,qi}(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\})\mathbf{1}\{z_{qi} \neq 0\}) = o_P(1),$$

and we are left with the term when $z_{qi} = 0$:

$$\begin{aligned} B_n^1(q) &= \sqrt{n}E(z_{n,qi}\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{n,qi} > 0, z_{qi} = 0\} - \frac{1}{2}\mathbf{1}\{z_{qi} = 0\})\right]) + o_P(1) \\ &= \sqrt{n}E(z_{n,qi}\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{n,qi} > 0, z_{qi} = 0\}\right]/2 \\ &\quad -\sqrt{n}E(z_{n,qi}\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{n,qi} \leq 0, z_{qi} = 0\})\right]/2 + o_P(1) \\ &= \sqrt{n}E(|z_{n,qi}|\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{qi} = 0\})\right])/2 + o_P(1) \\ &= \sqrt{n}E(|z_{n,qi} - z_{qi}|\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{qi} = 0\})\right])/2 + o_P(1) \\ &= \sqrt{n}E(\left|q^\top(\Sigma_n^\top - \Sigma^T)z_i^\top\right|\left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{qi} = 0\})\right])/2 + o_P(1). \end{aligned}$$

Adding $B_n^2(q)$ and $B_n^1(q)$ finishes the proof of (i).

To prove (ii), use $z_q = q^\top \Sigma^\top z_i^\top$ to write :

$$C_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n z_{qi}w_{qi} - E(q^\top \Sigma^\top z_i^\top w_{qi}^*)\right).$$

Using $w_{qi} = x_i\beta_q^* + \varepsilon_{qi}^*$, we have:

$$C_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n z_{qi}\varepsilon_{qi}^*\right) + \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n q^\top \Sigma^\top z_i^\top x_i\beta_q^* - E(q^\top \Sigma^\top z_i^\top w_{qi}^*)\right)$$

Using $E(z_{qi}^\top w_{qi}^*) = E(z_{qi}^\top w_{qi}) = E(z_{qi}^\top x_i \beta_q^*)$, the second term on the right hand side is equal to:

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

The third line uses that $\sqrt{n}(\hat{\Sigma}_n^u - \hat{\Sigma}_n) \underset{n\to\infty}{\overset{P}{\to}} 0$ by equation (C.10) and uniform bounds on q, Σ and β_q^*. Moreover, as $\hat{\Sigma}_n$ is bounded and its inverse exists, $\Sigma^\top(\hat{\Sigma}_n^\top)^{-1} \underset{n\to\infty}{\overset{a.s}{\to}} I$, and we have, uniformly in q:

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

■

Summing the different terms in the Lemma implies that $\tau_n(q)$ is asymptotically equivalent to :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n z_{qi}\varepsilon_{qi}^*\right) + \sqrt{n}E(\left|q^\top(\Sigma_n^\top - \Sigma^\top)z_i^\top\right| \left[(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_i\Sigma q = 0\})\right])/2$$

If there are no exposed faces (i.e., $\Pr(z_i\Sigma q = 0) = 0$), the second term is identically equal to zero and $\tau_n(q)$ converges in distribution, uniformly in q, to a Gaussian process centered at zero and of covariance function:

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

Suppose that there exist exposed faces ($\Pr(z_i\Sigma q = 0) > 0$). Write:

$$\Sigma q = (I_K \otimes q^T)vec(\Sigma^T)$$

so that, using the asymptotic normality of the estimate of $vec(\Sigma^T)$ in equation (C.11) we have:

$$\sqrt{n}q^\top(\Sigma_n^\top - \Sigma^\top)z_i^\top = \sqrt{n}(vec(\Sigma_n^\top)^\top - vec(\Sigma^\top)^\top)(I_K \otimes q)z_i^\top = \sqrt{n}\eta^\top W^{1/2}(I_K \otimes q)z_i^\top + o_P(1),$$

where η is a unit normal variate of dimension p^2 independent of z_i.

C.2 Proof of Proposition 10

Let $\beta_0 \in \partial B$ and let $\mathcal{Q}(\beta_0)$ the set of all $q_0 \in \mathbb{S}$ which minimize $T_\infty(q; \beta_0)$, i.e., the set of all $q_0 \in \mathbb{S}$ satisfying $\delta^*(q_0 \mid B) = q_0^\top \beta_0$. $\mathcal{Q}(\beta_0)$ is a non-empty, compact subset of $\mathbb{S}$. We first consider the case where $\mathcal{Q}(\beta_0)$ is a singleton. In the second part, the proof is expanded to the case in which $\mathcal{Q}(\beta_0)$ may contain more than one element of $\mathbb{S}$.

***(i)* $\mathcal{Q}(\beta_0)$ is a singleton $\mathcal{Q}(\beta_0) = \{q_0\}$.** In the case where $\delta^*(q \mid B)$ is differentiable (condition D), the empirical stochastic process $v_n(.)$, defined for $q \in \mathbb{S}$ as,

$$v_n(q) = \sqrt{n}\,(T_n(q; \beta_0) - T_\infty(q; \beta_0)) = \sqrt{n}(\hat{\delta}_n^*(q|B) - \delta^*(q|B)),$$

converges to a Gaussian process (Proposition 9) whose sample paths are uniformly continuous on the unit sphere $\mathbb{S}$ endowed with the usual Euclidean norm. Hence $v_n(.)$ is stochastically equicontinuous (for instance, p.2251 of Andrews, 1994).

Let $q_n \in \mathbb{S}$ be any sequence of directions defined as near minimizers of the empirical counterpart $T_n(q; \beta_0)$ defined as,

$$T_n(q_n; \beta_0) \leq \min_q T_n(q; \beta_0) + o_P(1).$$

Standard arguments employed for Z-estimators (e.g. van der Vaart, 1998) when the objective function has a unique minimum, imply that:

$$\text{plim}_{n\to\infty} q_n = q_0.$$

Because (i) $v_n(.)$ is stochastically equicontinuous (ii) $q_n \in \mathbb{S}$ (iii) $\text{plim}_{n\to\infty} q_n = q_0$, Andrews (1994, equation (3.36), p.2265) shows that:

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

The proof finishes by using the asymptotic distribution of $\sqrt{n} T_n(q_0; \beta_0)$ as stated in the text.

(ii) $\mathcal{Q}(\beta_0)$ is not a singleton We are first going to select and characterize a unique q_0^* from $\mathcal{Q}(\beta_0)$. Specifically, if β^* denotes any point in the interior of B, for instance the "center" of B (obtained by setting $u(z) = \frac{\underline{\Delta}(z)+\overline{\Delta}(z)}{2}$), we have,

$$\beta^* = E(\Sigma^\top z^\top \frac{\overline{y} + \underline{y}}{2}),$$

and we are going to select the direction q_0^* in $\mathcal{Q}(\beta_0)$ which is the closest to the one which connects β^* to β_0.

The selection of a single $q_0^* \in \mathcal{Q}(\beta_0)$ For any $q_0 \in \mathcal{Q}(\beta_0)$, we have by definition:

$$\delta^*(q_0 \mid B) - q_0^\top \beta_0 = 0 \Longrightarrow q_0^\top(\beta_0 - \beta^*) = \delta^*(q_0 \mid B) - q_0^\top \beta^*.$$

As β^* lies in the interior of set B, we have for all $q \in \mathbb{S}$,

$$\delta^*(q \mid B) - q^\top \beta^* > 0.$$

In particular, for any q_0 in $\mathcal{Q}(\beta_0)$, $\delta^*(q_0 \mid B) - q_0^\top \beta^* > 0$. Consequently, we get:

$$\forall q_0 \in \mathcal{Q}(\beta_0),\ q_0^\top(\beta_0 - \beta^*) > 0. \tag{C.15}$$

Consider now the smallest convex cone which includes $\mathcal{Q}(\beta_0)$:

$$\mathcal{C}(\beta_0) = \{\lambda q_0; q_0 \in \mathcal{Q}(\beta_0), \lambda \geq 0\}.$$

and consider the projection of $\beta_0 - \beta^*$ on $\mathcal{C}(\beta_0)$. Note that $\beta_0 - \beta^*$ is different from zero since $\beta_0 \in \partial B$ whereas β^* lies in the interior of B. This projection is unique and defined by $\lambda^* q_0^*$ where (λ^*, q_0^*) is the argument of the minimum:

$$\min_{(\lambda \geq 0, q \in \mathcal{Q}(\beta_0))} (\beta_0 - \beta^* - \lambda q)^\top(\beta_0 - \beta^* - \lambda q) \propto \min_{(\lambda \geq 0, q \in \mathcal{Q}(\beta_0))} \{-2\lambda q^\top(\beta_0 - \beta^*) + \lambda^2\}$$

which yields $\lambda^* = q_0^{*\top}(\beta_0 - \beta^*)$ (which is positive, see equation C.15) whereas q_0^* is the argument of the maximum:

$$\max_{q \in Q(\beta_0)} q^\top(\beta_0 - \beta^*).$$

Vector q_0^* is unique because it is a (normalized) projection.

Denote $v_0 = \frac{(\beta_0 - \beta^*)}{\|\beta_0 - \beta^*\|}$ the normalized vector connecting β^* to β_0 and which points ouwards of set B. When $v_0 \in \mathcal{Q}(\beta_0)$, we have $q_0^* = v_0$ whereas in other cases q_0^* belongs to the frontier of $\mathcal{Q}(\beta_0)$.

The construction of the perturbed program Note that by construction, for any $a > 0$, the point $\beta_0 + av_0$ does not belong to B, since it satisfies

$$q_0^{*T}(\beta_0 + av_0) = \delta^*(q_0^* \mid B) + a\frac{q_0^{*T}(\beta_0 - \beta^*)}{\|\beta_0 - \beta^*\|} > \delta^*(q_0^* \mid B).$$

Given this fact, we can define a sequence of perturbed programs such that q_0^* corresponds to the limit of the sequence of minima. Specifically, for any $a > 0$, let :

$$\Psi_a(q) = \delta^*(q \mid B) - q^\top \beta_0 - a q^\top v_0.$$

Denote q_a one element of the argument of its minimum:

$$q_a \in \arg\min_q \Psi_a(q).$$

Since B is assumed to have no exposed faces (Condition D), $\delta^*(. \mid B)$ and $\Psi_a(.)$ are differentiable and $\nabla_q \Psi_a(.)$ satisfies,

$$\|\nabla_q \Psi_a(q)\| = \|\beta_q - \beta_0 - av_0\| > 0,$$

since $\beta_0 + av_0 \notin B$. Therefore $\Psi_a(q)$ is a strictly convex function on a compact set and the minimum q_a is unique.

Convergence of q_a to q_0^* when $a \to 0$.

Lemma 14 *The limit of the sequence $\{q_a\}_{a>0}$ exists when $a \to 0$ and is equal to q_0^*.*

Proof. To begin with, it is useful to note that $-a$ provides a lower bound of $\Psi_a(q)$,

$$\Psi_a(q) = \delta^*(q \mid B) - q^\top \beta_0 - a q^\top v_0 \geq -a,$$

because $\beta_0 \in B$ and q and v_0 belong to $\mathbb{S}$.

We are going to consider in turn two cases:

- Assume first that $v_0 \in \mathcal{Q}(\beta_0)$. In such a case, $q_0^* = v_0$ and $\Psi_a(q_0^*) = -a$. Hence, given that q_a is unique and that $-a$ is a lower bound for $\Psi_a(q)$, we have necessarily $q_a = q_0^*$ for any $a > 0$.

- Assume now that $v_0 \notin \mathcal{Q}(\beta_0)$. by definition of q_a as a minimum,

  $$\Psi_a(q_a) = \delta^*(q_a \mid B) - q_a^\top \beta_0 - a q_a^\top v_0 \leq \Psi_a(q_0^*) - a q_0^{*T} v_0,$$

  since $\delta^*(q_0^* \mid B) = q_0^{*T} \beta_0$. It implies that:

  $$\delta^*(q_a \mid B) - q_a^\top \beta_0 \leq a(q_a - q_0^*)^\top v_0 \leq 2a, \tag{C.16}$$

  since $\|q_a - q_0^*\| \leq 2$. Since $\beta_0 \in B$, the left-hand side, $\delta^*(q_a \mid B) - q_a^\top \beta_0$, is non-negative. Consequently, we have,

  $$\lim_{a \to 0} (\delta^*(q_a \mid B) - q_a^\top \beta_0) = 0.$$

  Also, the distance between set $\mathcal{Q}(\beta_0)$ and q_a tends to zero by continuity of function $\delta^*(q \mid B) - q^\top \beta_0$.

  Consider now q_m any accumulation point of the sequence q_a i.e., any point satisfying, $\forall \eta > 0$, $\exists a > 0$ such that $\|q_a - q_m\| < \eta$. Because $\mathcal{Q}(\beta_0)$ is compact, $q_m \in \mathcal{Q}(\beta_0)$. We are going to show that $q_m = q_0^*$. By definition of q_a and q_0^*, we have

  $$\frac{\Psi_a(q_a)}{a} \leq \frac{\Psi_a(q_0^*)}{a} = -q_0^{*T} \frac{(\beta_0 - \beta^*)}{\|\beta_0 - \beta^*\|} \leq -q_m^\top \frac{(\beta_0 - \beta^*)}{\|\beta_0 - \beta^*\|}.$$

  where the first inequality holds true because q_a minimizes Ψ_a on the unit sphere whereas the second inequality holds true because $q_m \in \mathcal{Q}(\beta_0)$ and q_0^* maximizes $q^\top(\beta_0 - \beta^*)$ on $\mathcal{Q}(\beta_0)$. Furthermore, since $\delta^*(q \mid B) \geq q^\top \beta_0$ for any q on the unit sphere, we have,

  $$\frac{\Psi_a(q_a)}{a} = \frac{\delta^*(q_a \mid B) - q_a^\top \beta_0}{a} - q_a^\top v_0 \geq -q_a^\top v_0.$$

  Combining this inequality with the two previous ones, we have,

  $$-q_a^\top v_0 \leq -q_0^{*T} v_0 \leq -q_m^\top v_0$$

  By taking limits and using that q_a tends to q_m when a tends to zero, we obtain that $q_m = q_0^*$ and therefore:

  $$\|q_a - q_0^*\| = O(a). \tag{C.17}$$

  Furthermore, as:

  $$0 \leq \frac{\Psi_a(q_0^*) - \Psi_a(q_a)}{a} = \frac{\delta^*(q_a \mid B) - q_a^\top \beta_0}{a} \leq (q_a - q_0^*)^\top v_0$$

  we have when $a \to 0$:

  $$\frac{\Psi_a(q_0^*) - \Psi_a(q_a)}{a} = O(a). \tag{C.18}$$

■

The minimum of the perturbed program is well separated

Lemma 15

$$\forall \varepsilon > 0, \exists \eta > 0, \exists a_0 > 0, \text{ such that } \inf_{\|q-q_0^*\|>\varepsilon, a<a_0} \frac{\Psi_a(q) - \Psi_a(q_0^*)}{a} > \eta. \tag{C.19}$$

Proof. By definition since $\beta_0 \in \partial B_0$, for any q :

$$\frac{\Psi_a(q) - \Psi_a(q_0^*)}{a} = \frac{\delta^*(q \mid B) - q^\top \beta_0}{a} - (q^\top - q_0^{*T})v_0 \geq -(q^\top - q_0^{*T})v_0.$$

(i) Assume first that $v_0 \in Q(\beta_0)$. In such a case, we have $v_0 = q_0^*$ and, for any q in the unit sphere,

$$\|q\|^2 = 1 = \|q - v_0 + v_0\|^2 = 1 - 2(q^\top - q_0^{*T})v_0 + \|q - v_0\|^2,$$

which implies that $-(q^\top - q_0^{*T})v_0 = \frac{\|q-v_0\|^2}{2}$. Hence, for any $\varepsilon > 0$, we have,

$$\inf_{\|q-q_0^*\|>\varepsilon} \frac{\Psi_a(q) - \Psi_a(q_0^*)}{a} \geq \inf_{\|q-q_0^*\|>\varepsilon} (-(q^\top - q_0^{*T})v_0) \geq \frac{\varepsilon^2}{2}.$$

We only have to set $\eta < \frac{\varepsilon^2}{2}$ for condition (C.19) to be satisfied for all values of a.

(ii) Second, assume that $v_0 \notin Q(\beta_0)$. We have:

$$\frac{\Psi_a(q) - \Psi_a(q_0^*)}{a} = \frac{\Psi_a(q) - \Psi_a(q_a)}{a} + \frac{\Psi_a(q_a) - \Psi_a(q_0^*)}{a}. \tag{C.20}$$

Fix $\varepsilon > 0$. As by equation (C.17), $\|q_a - q_0^*\| = O(a)$ when $a \to 0$, there exists a_1 such that $\forall a < a_1$, $\|q_a - q_0^*\| < \varepsilon/2$. Therefore if $a < a_1$,

$$\|q - q_0^*\| > \varepsilon \Longrightarrow \|q - q_a\| > \|q - q_0^*\| - \|q_a - q_0^*\| = \frac{\varepsilon}{2}.$$

It implies,

$$\inf_{\|q-q_0^*\|>\varepsilon, a<a_1} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} > \inf_{\|q-q_a\|>\varepsilon/2, a<a_1} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} \equiv 2\eta,$$

where η is positive because Ψ_a is a strictly convex function and q_a is a well separated minimum.

Using equation (C.18), $\frac{\Psi_a(q_0^*) - \Psi_a(q_a)}{a} = O(a)$, we can always choose $a_0 \leq a_1$, such that, for any $a < a_0$,

$$\frac{\Psi_a(q_a) - \Psi_a(q_0^*)}{a} > -\eta \Longrightarrow \inf_{\|q_a-q_0^*\|<\varepsilon/2, a<a_0} \frac{\Psi_a(q_a) - \Psi_a(q_0^*)}{a} > -\eta$$

since $a_0 \leq a_1$, for any $a < a_1$, we have $\|q_a - q_0^*\| < \varepsilon/2$. Furthermore,

$$\inf_{\|q-q_a\|>\varepsilon/2, a<a_0} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} > \inf_{\|q-q_a\|>\varepsilon/2, a<a_1} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} = 2\eta,$$

so that using equation (C.20),

$$\begin{aligned} \inf_{\|q-q_0^*\|>\varepsilon, a<a_0} \frac{\Psi_a(q) - \Psi_a(q_0^*)}{a} &> \inf_{\|q-q_a\|>\varepsilon/2, a<a_0} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} \\ &\quad + \inf_{\|q_a-q_0^*\|<\varepsilon/2, a<a_0} \frac{\Psi_a(q_a) - \Psi_a(q_0^*)}{a} \;>\; 2\eta - \eta = \eta. \end{aligned}$$

■

Estimation and convergence to the unique q_0^* Finally, we construct the estimate of q_a. Fix $a > 0$. Define the perturbed estimated program as:

$$\Psi_{n,a}(q;\beta_0) = \hat{\delta}_n^*(q \mid B) - q^\top \beta_0 - aq^\top v_{0,n}$$

where $v_{0,n} = \frac{\beta_0 - \hat{\beta}_n^*}{\|\beta_0 - \hat{\beta}_n^*\|}$ and $\hat{\beta}_n^* = \frac{1}{n}\sum_{i=1}^n \hat{\Sigma}_n^\top z_i \frac{\bar{y}_i + \underline{y}_i}{2}$ and where some innocuous normalization for $v_{0,n}$ is adopted when $\beta_0 - \hat{\beta}_n^* = 0$.

Define and estimate $q_{n,a}$:

$$\begin{aligned}\Psi_{n,a}(q_{n,a}) &\leq \Psi_{n,a}(q_a) + O_P(n^{-1/2}) \\ &\leq \Psi_a(q_a) + O_P(n^{-1/2}).\end{aligned}$$

Therefore:

$$\begin{aligned}\Psi_a(q_{n,a}) - \Psi_a(q_a) &\leq \Psi_a(q_{n,a}) - \Psi_{n,a}(q_{a,n}) + O_P(n^{-1/2}) \\ &\leq \sup_q |\Psi_a(q) - \Psi_{n,a}(q)| + O_P(n^{-1/2}).\end{aligned}$$

We thus have:

$$\frac{\Psi_a(q_{n,a}) - \Psi_a(q_a)}{a} \leq \frac{\sup_q |\Psi_a(q) - \Psi_{n,a}(q)| + O_P(n^{-1/2})}{a}.$$

Let $a_n = O(n^{-\alpha})$ a sequance such that $\alpha < 1/2$. Because of equicontinuity and $n^{1/2}$ convergence of $\hat{\delta}_n^*(q \mid B)$ to $\delta^*(q \mid B)$ and of $v_{0,n}$ to v_0, we have that:

$$n^\alpha . \sup_q |\Psi_{a_n}(q) - \Psi_{a_n,n}(q)| \underset{n\to\infty}{\overset{P}{\to}} 0.$$

Then:

$$\frac{\Psi_{a_n}(q_{n,a_n}) - \Psi_{a_n}(q_{a_n})}{a_n} \leq o_P(1).$$

Using equation (C.18):

$$\frac{\Psi_a(q_0^*) - \Psi_a(q_a)}{a} = O(a) \Longrightarrow \frac{\Psi_{a_n}(q_{n,a_n}) - \Psi_{a_n}(q_0^*)}{a_n} \leq o_P(1).$$

We thus have:

$$\forall \eta, \lim_{n\to\infty} \Pr\left(\frac{\Psi_{a_n}(q_{n,a_n}) - \Psi_{a_n}(q_0^*)}{a_n} > \eta\right) = 0$$

so that given condition (C.19) and as $a_n \to 0$, we obtain that:

$$\forall \varepsilon, \lim_{n\to\infty} \Pr(d(q_{n,a_n}, q_0^*) > \varepsilon) = 0 \Longrightarrow q_{n,a_n} \underset{n\to\infty}{\overset{P}{\to}} q_0^*.$$

Then the same argument than in part *(i)* applies and:

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

C.3 Proof of Proposition 12

The proof consists in three steps:

1. Under the assumption that the unconstrained set B_U has no faces, the estimate of the unconstrained support function is a consistent and asymptotically Gaussian random process.

2. The minimization of the estimate $\hat{\delta}_n^*((q,\lambda)|B_U)$ with respect to λ holding q constant for any q can be analyzed as in Proposition 16.

   (a) If $\lambda_0(q)$, the minimizer of the true support function, is unique, then any near minimizer in λ of $\hat{\delta}_n^*((q,\lambda)|B_U)$ is a $\sqrt{n}-$consistent and asymptotically normal estimate of $\delta^*(q|B)$.

   (b) If $\lambda_0(q)$ is not unique, we define a perturbed criterion to minimize to construct an estimate of one particular element $\lambda_0^*(q)$. Then $\hat{\delta}_n^*((q,\lambda_n(q))|B_U)$ is a $\sqrt{n}-$consistent and asymptotically normal estimate of $\delta^*(q|B)$.

   In both cases, we can do that for any finite list of q and show that the vector of those estimates are jointly asymptotically normal.

3. The process $v_n(q) = \sqrt{n}(\hat{\delta}_n^*((q,\lambda_n(q))|B_U) - \delta^*(q|B))$ is stochastically equicontinuous.

Using Andrews (1994, p2251), it then proves that $v_n(q)$ is a consistent and asymptotically Gaussian random process.

Step 1: According to what was developed above, the empirical stochastic process $v_n^U(.)$, defined for $s = (q,\lambda) \in \mathbb{S}_m$, the unit sphere in $\mathbb{R}^m$, as,

$$v_n^U(s) = \sqrt{n}\left(\hat{\delta}_n^*(s \mid B_U) - \delta^*(s \mid B_U)\right),$$

converges to a Gaussian process whose sample paths are uniformly continuous on the unit sphere $\mathbb{S}$, using the usual Euclidean norm. Hence $v_n^U(.)$ is stochastically equicontinuous (for instance, p.2251 of Andrews, 1994).

Step 2: Fix $q \in \mathbb{S}_p$ the unit sphere in $\mathbb{R}^p$ and let $\mathcal{S}(q)$ the set of all $s(q) = (q, \lambda_0(q))$ that minimize $\delta^*(s \mid B_U)$ with respect to λ i.e.:

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

$\mathcal{S}(q)$ is a non-empty subset of $\mathbb{R}_m$ by the above. Note also that to obtain the standard evaluation on the unit sphere some renormalization is necessary and this is done using the homotheticity property of support functions:

$$\delta^*(s \mid B_U) = \sqrt{s^\top s}.\delta^*(\frac{s}{\sqrt{s^\top s}} \mid B_U).$$

where $\frac{s}{\sqrt{s^\top s}} \in \mathbb{S}_m$. In the following, we will directly deal with the support function $\delta^*(s \mid B_U)$ extended to $\mathbb{R}^m$ in this way.

We first consider the case where $\mathcal{S}(q)$ is a singleton. The second part of the proof will extend the result to the case where $\mathcal{S}(q)$ potentially contains more than one element of $\mathbb{R}^m$, the issue being to select one specific element of $\mathcal{S}(q)$ and to construct a consistent estimate of it.

a) Suppose that $\mathcal{S}(q)$ is a singleton, $\mathcal{S}(q) = \{s_0 = (q, \lambda_0)\}$ where $q \in \mathbb{S}_p$.

Let $s_n = (q, \lambda_n)$ be any sequence of directions defined as near minimizers of the empirical counterpart $\hat{\delta}_n^*(s_n \mid B_U)$ defined as,

$$\hat{\delta}_n^*(s_n \mid B_U) \leq \min_{\lambda} \hat{\delta}_n^*(s = (q, \lambda) \mid B_U) + o_P(1).$$

Define the estimate of $\delta^*(q \mid B)$ as the value at the near minimizer:

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

First, standard arguments employed for Z-estimators (see van der Vaart, 1998, for instance) imply that:

$$\text{plim}_{n\to\infty} \lambda_n = \lambda_0.$$

Second, because i) $v_n^U(.)$ is stochastically equicontinuous ii) $s_n \in \mathbb{S}$ iii) $\text{plim}_{n\to\infty} s_n = s_0$, Andrews (1994, equation (3.36), p:2265) shows that:

$$\sqrt{n}\left(\hat{\delta}_n^*(s_n \mid B_U) - \hat{\delta}_n^*(s_0 \mid B_U)\right) \underset{n\to\infty}{\overset{P}{\rightarrow}} 0.$$

The proof finishes by using the asymptotic distribution of $\hat{\delta}_n^*(s_0 \mid B_U)$:

$$\sqrt{n}\left(\hat{\delta}_n^*(s_0 \mid B_U) - \delta^*(s_0 \mid B_U)\right) \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, V_{s_0}),$$

which implies that:

$$\sqrt{n}\left(\hat{\delta}_n^*(q \mid B) - \delta^*(q \mid B)\right) \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, V_{s_0}),$$

where V_{s_0} is consistently estimated by V_{s_n}.

Remark that the same result applies to a finite vector $(\hat{\delta}_n^*(q_1 \mid B), \hat{\delta}_n^*(q_2 \mid B), ., \hat{\delta}_n^*(q_J \mid B))$ using the same arguments.

ii) Suppose now that $\mathcal{S}(q)$ is not a singleton. There are various minimizers of $\delta^*(s \mid B_U)$ in λ. We are first going to select and characterize a unique (q, λ_0^*) from $\mathcal{S}(q)$. Consider the smallest convex cone which includes $\mathcal{S}(q)$:

$$\mathcal{CS}(q) = \{c.s_0; s_0 \in \mathcal{S}(q), c \geq 0\}.$$

and consider the projection of $(q, 0)$ on $\mathcal{CS}(q)$. This projection is unique and defined by $c^* s_0^*$ where (c^*, s_0^*) is the argument of the minimum:

$$\min_{(c\geq 0,(q,\lambda)\in\mathcal{S}(q))} \|((1-c)q, -c\lambda\|^2 = \min_{(c\geq 0, s\in\mathcal{S}(q))} \{(1-c)^2 + c^2\lambda^\top\lambda\}$$

since $q^\top q = 1$. It yields $c^* = \frac{1}{1+\lambda^\top\lambda} > 0$ whereas λ_0^* is the argument of:

$$\min_{(q,\lambda)\in\mathcal{S}(q))} \lambda^\top\lambda.$$

Vector λ_0^* is unique because it is a (normalized) projection. Given this fact, we can define a sequence of perturbed programs such that s_0^* corresponds to the limit of the sequence of minima. Specifically, for any $a > 0$, let :

$$\Psi_a(s) = \delta^*(s \mid B_U) + a\lambda^\top \lambda.$$

Because $\delta^*(s \mid B_U)$ is convex in λ and $\lambda^\top\lambda$ is strictly convex in λ, $\Psi_a(s)$ is a strictly convex function.

It is useful to restrict the set of s over which we minimize by choosing for instance the set:

$$S_B = \{s; \forall a \leq 1, \Psi_a(s) \leq \delta^*(s_0^* \mid B_U) + \lambda_0^{*T}\lambda_0^*\},$$

which is non empty ($s_0^* \in S_B$) and compact. Note therefore that $\mathcal{S}(q) \cap S_B$ is not empty.

The minimum $s_a = (q, \lambda_a)$ of $\Psi_a(s)$ is unique because we minimize a convex function on a compact set. Furthermore, we now show that λ_a tends to λ_0^* when $a \to 0$.

Lemma 16 *The limit of the sequence $\{\lambda_a\}_{a>0}$ exists when $a \to 0$ and is equal to λ_0^*.*

Proof. To begin with, it is useful to note that $\delta^*(s_0^* \mid B_U)$ provides a lower bound of $\Psi_a(s)$,

$$\Psi_a(s) \geq \delta^*(s_0^* \mid B_U).$$

Given this fact, we are going to consider in turn two cases:

- Assume first that $(q, 0) \in \mathcal{S}(q)$. In such a case, $\lambda_0^* = 0$ and $\Psi_a(s_0^*) = \delta^*(s_0^* \mid B_U)$. Hence, given that λ_a is unique and that $\delta^*(s_0^* \mid B_U)$ is a lower bound for $\Psi_a(s)$, we have necessarily $s_a = s_0^*$ for any $a > 0$ and hence $\lambda_a = \lambda_0^*$.

- Assume now that $(q, 0) \notin \mathcal{S}(q)$. By definition of λ_a as a minimum,

$$\begin{aligned} \Psi_a(s_a) &= \delta^*(s_a \mid B_U) + a\lambda_a^\top \lambda_a \\ &\leq \Psi_a(s_0^*) = \delta^*(s_0^* \mid B_U) + a\lambda_0^{*T}\lambda_0, \end{aligned}$$

  It implies that:

$$\delta^*(s_a \mid B_U) - \delta^*(s_0^* \mid B_U) \leq a(\lambda_0^{*T}\lambda_0^* - \lambda_a^\top\lambda_a) \leq a\lambda_0^{*T}\lambda_0^*, \tag{C.21}$$

  so that the distance between s_a and the set $\mathcal{S}(q)$ tends to zero when a tends to zero by the continuity of the function $\delta^*(s \mid B_U)$.

  Consider now λ_m any accumulation point of the sequence λ_a i.e., any point satisfying, $\forall \eta > 0, \exists a > 0$ such that $\|\lambda_a - \lambda_m\| < \eta$. Because $\mathcal{S}(q) \cap S_B$ is compact, $s_m \in \mathcal{S}(q) \cap S_B$. We are going to show that $s_m = s_0^*$. By definition of λ_a and λ_0^*, we have

$$\frac{\Psi_a(s_a) - \delta^*(s_0^* \mid B_U)}{a} \leq \frac{\Psi_a(q_0^*) - \delta^*(s_0^* \mid B_U)}{a} = \lambda_0^{*T}\lambda_0^* \leq \lambda_m^{*T}\lambda_m^*.$$

where the first inequality holds true because s_a minimizes Ψ_a whereas the second inequality holds true because $s_m \in \mathcal{S}(q) \cap S_B$ is compact and λ_0^* minimizes $\lambda_0^{*T}\lambda_0^*$ on $\mathcal{S}(q) \cap S_B$. Furthermore, because $s_0^* \in \mathcal{S}(q) \cap S_B$

$$\frac{\Psi_a(s_a) - \delta^*(s_a \mid B_U)}{a} \leq \frac{\Psi_a(s_a) - \delta^*(s_0^* \mid B_U)}{a}.$$

Combining the two equations,

$$\lambda_a^\top \lambda_a \leq \lambda_0^{*T}\lambda_0^* \leq \lambda_m^{*T}\lambda_m^*.$$

By taking limits and using that λ_a tends to λ_m when a tends to zero, we obtain that $\lambda_m = \lambda_0^*$. We thus have shown that:

$$\|\lambda_a - \lambda_0^*\| = O(a).$$

Furthermore, we check:

$$0 \leq \frac{\Psi_a(s_0^*) - \Psi_a(s_a)}{a} \leq \lambda_0^{*T}\lambda_0^* - \lambda_a^\top \lambda_a$$

so that, since $\lambda_a \to \lambda_0^*$ when $a \to 0$:

$$\Psi_a(s_0^*) - \Psi_a(s_a) = o(a). \tag{C.22}$$

■

The next step is to construct an estimate of λ_a. Before moving on to this step, we are going to prove a lemma that will be useful for showing that the estimate of λ_a actually converges to λ_0^*.

Lemma 17

$$\forall \varepsilon > 0, \exists \eta > 0, \exists a_0 > 0, \textit{ such that } \inf_{\|s - s_0^*\| > \varepsilon, a < a_0} \frac{\Psi_a(s) - \Psi_a(s_0^*)}{a} > \eta. \tag{C.23}$$

Proof. By definition, for any $s = (q, \lambda)$ where $q \in \mathbb{S}_p$:

$$\frac{\Psi_a(s) - \Psi_a(s_0^*)}{a} = \frac{\delta^*(s \mid B_U) - \delta^*(s_0^* \mid B_U)}{a} + \lambda^\top \lambda - \lambda_0^{*T}\lambda_0^* \geq \lambda^\top \lambda - \lambda_0^{*T}\lambda_0^*.$$

Assume first that $(q, 0) \in \mathcal{S}(q)$. In such a case, $\lambda_0^* = 0$ and $\frac{\Psi_a(s) - \Psi_a(s_0^*)}{a} \geq \lambda^\top \lambda$. As $\|s - s_0^*\| = (\lambda^\top \lambda)^{1/2}$, we have:

$$\inf_{\|q - q_0^*\| > \varepsilon} \frac{\Psi_a(q) - \Psi_a(q_0^*)}{a} \geq \varepsilon^2.$$

We only have to set $\eta < \varepsilon^2$ for condition (C.19) to be satisfied for all values of a.

Second, assume that $(q, 0) \notin \mathcal{S}(q)$. We have:

$$\inf_{\|q - q_0^*\| > \varepsilon} \frac{\Psi_a(s) - \Psi_a(s_0^*)}{a} = \inf_{\|q - q_0^*\| > \varepsilon} \left(\frac{\Psi_a(s) - \Psi_a(s_a)}{a} + \frac{\Psi_a(s_a) - \Psi_a(s_0^*)}{a}\right). \tag{C.24}$$

As λ_a tends to λ_0^* when $a \to 0$, for any $\varepsilon > 0$ there exists a_1 such that $\forall a < a_1$, $d(s_a, s_0^*) < \varepsilon/2$.

Conditional on $a < a_1$, $\|s - s_0^*\| > \varepsilon$ implies that $\|s - s_a\| > \|s - s_0^*\| - d(s_a, s_0^*) > \frac{\varepsilon}{2}$. It yields,

$$\inf_{\left\|s-s_0^*\right\|>\varepsilon, a<a_1} \frac{\Psi_a(s) - \Psi_a(s_a)}{a} > \inf_{\|s-s_a\|>\varepsilon/2, a<a_1} \frac{\Psi_a(s) - \Psi_a(s_a)}{a}.$$

Overall, if we denote $\eta = \frac{1}{2}\inf_{\|s-s_a\|>\varepsilon/2, a<a_1} \frac{\Psi_a(s)-\Psi_a(s_a)}{a}$, η is positive because s_a is a well separated minimum and η satisfies $\inf_{\left\|s-s_0^*\right\|>\varepsilon, a<a_1} \frac{\Psi_a(s)-\Psi_a(s_a)}{a} > 2\eta$.

Using the fact that $\Psi_a(s_0^*) - \Psi_a(s_a) = o(a)$, we can also choose $a_0 \leq a_1$, such that, for any $a < a_0$,

$$\frac{\Psi_a(s_a) - \Psi_a(s_0^*)}{a} > -\eta.$$

Given that $a_0 \leq a_1$, for any $a < a_0$, we have $d(s_a, s_0^*) < \varepsilon/2$ and

$$\inf_{\|s-s_a\|>\varepsilon/2, a<a_0} \frac{\Psi_a(s) - \Psi_a(s_a)}{a} > \inf_{\|s-s_a\|>\varepsilon/2, a<a_1} \frac{\Psi_a(s) - \Psi_a(s_a)}{a},$$

so that,

$$\begin{aligned}
\inf_{\left\|s-s_0^*\right\|>\varepsilon, a<a_0} \frac{\Psi_a(s) - \Psi_a(s_0^*)}{a} &> \inf_{\|s-s_a\|>\varepsilon/2, a<a_0} \frac{\Psi_a(s) - \Psi_a(s_a)}{a} + \inf_{\left\|s_a-s_0^*\right\|<\varepsilon/2, a<a_0} \frac{\Psi_a(s_a) - \Psi_a(s_0^*)}{a} \\
&> \inf_{\|s-s_a\|>\varepsilon/2, a<a_1} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} - \eta > 2\eta - \eta = \eta.
\end{aligned}$$

■

Finally, we construct the estimate of λ_a. Fix $a > 0$. Define the perturbed estimated program as:

$$\Psi_{n,a}(s) = \hat{\delta}_n^*(s \mid B_U) + a\lambda^\top\lambda.$$

and restrict the set over which we take the supremum as $s \in S_B$

Define:

$$\begin{aligned}
\Psi_{n,a}(s_{a,n}) &\leq \Psi_{n,a}(s_a) + O_P(n^{-1/2}) \\
&\leq \Psi_a(s_a) + O_P(n^{-1/2}).
\end{aligned}$$

Therefore:

$$\begin{aligned}
\Psi_a(s_{a,n}) - \Psi_a(s_a) &\leq \Psi_a(s_{a,n}) - \Psi_{n,a}(s_{a,n}) + O_P(n^{-1/2}), \\
&\leq \sup_{s\in S_B} |\Psi_a(s) - \Psi_{n,a}(s)| + O_P(n^{-1/2}), \\
&\leq \sup_{s\in S_B} \left|\delta^*(s \mid B_U) - \hat{\delta}_n^*(s \mid B_U)\right| + O_P(n^{-1/2}),
\end{aligned}$$

We thus have:

$$\frac{\Psi_a(s_{a,n}) - \Psi_a(s_a)}{a} \leq \frac{\sup_{s\in S_B} \left|\delta^*(s \mid B_U) - \hat{\delta}_n^*(s \mid B_U)\right| + O_P(n^{-1/2})}{a}.$$

Let $a_n = O_P(n^{-\alpha})$ where $\alpha < 1/2$. Because of equicontinuity and $n^{1/2}$ convergence of $\hat{\delta}_n^*(s \mid B_U)$ to $\delta^*(s \mid B_U)$, we have that:

$$n^{\alpha}. \sup_{s \in S_B} \left| \delta^*(s \mid B_U) - \hat{\delta}_n^*(s \mid B_U) \right| \underset{n \to \infty}{\overset{P}{\to}} 0.$$

Then:

$$\frac{\Psi_{a_n}(s_{a_n,n}) - \Psi_{a_n}(s_{a_n})}{a_n} \leq o_P(1).$$

As by equation (C.22):

$$\frac{\Psi_{a_n}(s_0^*) - \Psi_{a_n}(s_{a_n})}{a_n} = o(1)$$

we get:

$$\frac{\Psi_{a_n}(s_{a_n,n}) - \Psi_{a_n}(s_0^*)}{a_n} \leq o_P(1).$$

We thus have:

$$\forall \eta, \lim_{n \to \infty} \Pr\left(\frac{\Psi_{a_n}(s_{a_n,n}) - \Psi_{a_n}(s_0^*)}{a_n} > \eta\right) = 0$$

so that given condition (C.23) and as $a_n \to 0$, we obtain that:

$$\forall \varepsilon, \lim_{n \to \infty} \Pr(d(s_{a_n,n}, s_0^*) > \varepsilon) = 0,$$

and therefore

$$s_{a_n,n} \underset{n \to \infty}{\overset{P}{\to}} s_0^*.$$

Then the same argument than in part *(i)* applies and:

$$\sqrt{n} \left(\hat{\delta}_n^*(s_n \mid B_U) - \hat{\delta}_n^*(s_0^* \mid B_U) \right) \underset{n \to \infty}{\overset{P}{\to}} 0.$$

We can then use the asymptotic distribution of $\hat{\delta}_n^*(s_0^* \mid B_U)$ in place of $\hat{\delta}_n^*(s_n \mid B_U)$.

By the same development, it applies to a finite vector of such estimates defined at values $q_1, q_2, .., q_J$.

Step 3: We now turn to equicontinuity. As the process $v_n^U(s)$ is equicontinuous, we know that for any $\varepsilon > 0$ and $\eta > 0$ there exists δ such that:

$$\lim_{n \to \infty} \Pr\left[\sup_{s_1, s_2 \in \mathbb{S}, \|s_1 - s_2\| < \delta} \left| v_n^U(s_1) - v_n^U(s_2) \right| > \eta \right] < \varepsilon.$$

Let s_{1n} and s_{2n} be defined as:

$$\hat{\delta}_n^*(s_{1n} \mid B_U) = \hat{\delta}_n^*(q_1, \lambda_n(q_1) \mid B_U), \hat{\delta}_n^*(s_{2n} \mid B_U) = \hat{\delta}_n^*(q_2, \lambda_n(q_2) \mid B_U)$$

where for $j = 1, 2$, $\lambda_n(q_j)$ are minimizers of $\hat{\delta}_n^*(q_j, \lambda_n(q_j) \mid B_U)$ defined as:

$$\hat{\delta}_n^*(s_{jn} \mid B_U) = \min_{\lambda} \hat{\delta}_n^*(s_j \mid B_U) = \min_{\lambda} \hat{\delta}_n^*((q_j, \lambda) \mid B_U),$$

if they are unique or by teh argument used in Step 2 b) if they are not. Consider the difference:

$$\begin{aligned}
\hat{\delta}_n^*(s_{1n} \mid B_U) - \hat{\delta}_n^*(s_{2n} \mid B_U) &= \min_\lambda \hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \min_\lambda \hat{\delta}_n^*((q_2,\lambda) \mid B_U) \\
&= \min_\lambda (\hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_2,\lambda) \mid B_U) + \hat{\delta}_n^*((q_2,\lambda) \mid B_U)) - \min_\lambda \hat{\delta}_n^*((q_2,\lambda) \mid B_U) \\
&\geq \inf_\lambda (\hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_2,\lambda) \mid B_U)).
\end{aligned}$$

Or alternatively:

$$\begin{aligned}
\hat{\delta}_n^*(s_{1n} \mid B_U) - \hat{\delta}_n^*(s_{2n} \mid B_U) &= \min_\lambda \hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \min_\lambda \hat{\delta}_n^*((q_2,\lambda) \mid B_U) \\
&= \min_\lambda \hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \min_\lambda (\hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_1,\lambda) \mid B_U) + \hat{\delta}_n^*((q_2,\lambda) \mid B_U)) \\
&\leq -\inf_\lambda (-\hat{\delta}_n^*((q_1,\lambda) \mid B_U) + \hat{\delta}_n^*((q_2,\lambda) \mid B_U)) \\
&= \sup_\lambda (\hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_2,\lambda) \mid B_U)).
\end{aligned}$$

In consequence:

$$\left| \hat{\delta}_n^*(s_{1n} \mid B_U) - \hat{\delta}_n^*(s_{2n} \mid B_U) \right| \leq \sup_\lambda \left| \hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_2,\lambda) \right|.$$

By definition:

$$v_n(q_1) - v_n(q_2) = v_n^U(s_{1n}) - v_n^U(s_{2n}) = \sqrt{n}(\hat{\delta}_n^*(s_{1n} \mid B_U) - \hat{\delta}_n^*(s_{2n} \mid B_U))$$

so that:

$$\begin{aligned}
& \lim_{n\to\infty} \Pr\left[\sup_{\|q_1-q_2\|<\delta} |v_n(q_1) - v_n(q_2)| > \eta \right] \\
= & \lim_{n\to\infty} \Pr\left[\sup_{\|q_1-q_2\|<\delta} \left| \sqrt{n}(\hat{\delta}_n^*(s_{1n} \mid B_U) - \hat{\delta}_n^*(s_{2n} \mid B_U)) \right| > \eta \right] \\
< & \lim_{n\to\infty} \Pr\left[\sup_{\|q_1-q_2\|<\delta} \left| \sqrt{n} \sup_\lambda \hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_2,\lambda) \mid B_U) \right| > \eta \right] \\
= & \lim_{n\to\infty} \Pr\left[\sup_{\|s_1-s_2\|<\delta} \left| \sqrt{n} \left(\hat{\delta}_n^*((q_1,\lambda) \mid B_U) - \hat{\delta}_n^*((q_2,\lambda) \right) \mid B_U) \right| > \eta \right] \\
= & \lim_{n\to\infty} \Pr\left[\sup_{\|s_1-s_2\|<\delta} \left| \sqrt{n} \left(\hat{\delta}_n^*(s_1 \mid B_U) - \hat{\delta}_n^*(s_2 \mid B_U) \right) \right| > \eta \right] < \varepsilon.
\end{aligned}$$

that proves that the process $v_n(q)$ is equicontinuous.

The proof when the minimizers are replaced by near-minimizers can be adapted in a straighforward way.

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 compute $\delta^*(q|B)$ using $z = x$ as instruments. As $\Sigma^{-1} = E(x^\top x) = I_2$, 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

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

we obtain:

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

and therefore:

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

The frontier points are:

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

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 unit normal vector. It is convenient to define $\mu = \nu\sqrt{1-\pi^2}$ and $a^2 = \pi^2 + \mu^2 = \pi^2 + \nu^2(1-\pi^2)$.

To conform with 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 $E(z^\top z) = I_3$, we have:

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

and:

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

which is standard unit bivariate normally distributed. Moreover as:

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

the normalized vector $\begin{pmatrix} 0 & \frac{\mu}{a} & -\frac{\pi}{a} \end{pmatrix}^\top$ belongs to the kernel of this operator and in consequence, $z_H = \frac{\mu e_2 - \pi w}{a}$.

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

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

Let $q = (q_1, q_2)$ such that $q_1^2 + q_2^2 = 1$ and define:

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

and 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 variables of interest 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 for instance,

$$E x_1 \mathbf{1}_{z_q > 0} = \frac{1}{\sqrt{2\pi}} q_1,$$

using the normality assumptions. In consequence, a closed-form expression for $\delta^*(q, \lambda | B_U)$ is:

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

This function is minimized when $\lambda = 0$ and B_U is an ellipsoid orthogonal to the hyperplane $\gamma = 0$. Its projection on the hyperplane is also an ellipse and the identified set is an ellipse:

$$\delta^*(q|B) = \frac{2\Delta}{\sqrt{2\pi}} \sqrt{q_1^2 + \frac{q_2^2}{a^2}}.$$

D.3 Example of section 5.3

The simulated model is:

$$y^* = \frac{1}{2} + \frac{x}{8} + \varepsilon$$

and variable $z \equiv (1, x_1)^\top$ are the instruments. As $\Sigma = E(z^\top z)^{-1} = I_2$,we can derive the variables of interest:

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

$E(\underline{y}) = \frac{1}{4}$ and $E(x\underline{y}) = \frac{1}{8}$ so we can derive the frontire points β_q:

$$\begin{aligned} \beta_q = \Sigma E(z^\top w_q) &= E(z^\top \underline{y}) + \frac{1}{2} E(z^\top \mathbf{1}\{z_q > 0\} \\ &= \begin{bmatrix} \frac{1}{4} \\ \frac{1}{8} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}E(\mathbf{1}\{z_q > 0\}) \\ \frac{1}{2}E(x_1 \mathbf{1}\{z_q > 0\}) \end{bmatrix}. \end{aligned}$$

Let $\theta_0 = \pi/4$. For θ being between $-\theta_0$ and θ_0 z_q is always positive whatever the value of x:

$$\begin{aligned} E\mathbf{1}\{z_q > 0\} &= 1 \\ Ex\mathbf{1}\{z_q > 0\} &= 0 \end{aligned}$$

and $\beta_q = \left[\frac{3}{4}; \frac{1}{8}\right]^\top$.

For θ being between θ_0 and $-\theta_0 + \pi$, z_q is negative when $x = -1$, otherwise positive:

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

and $\beta_q = \left[\frac{1}{2}; \frac{3}{8}\right]^\top$.

We obtain similarly $\beta_q = \left[\frac{1}{4}; \frac{1}{8}\right]^\top$ when θ is between $\theta_0 + \pi$ and $\theta_0 + \pi$ and $\beta_q = \left[\frac{1}{2}; -\frac{1}{8}\right]^\top$ for θ being between $\theta_0 - \pi$ and $-\theta_0$.

The term $\tau_1(q)$ defined in proposition 9 is equal to zero when $P(z_q = 0) = 0$, *i.e.* when $\theta \neq \frac{(2k+1)\Pi}{4}$. When $\theta = \Pi/4$, $z_q = 0$ when $x = -1$ which occurs with probability 1/2. However the term $q^T(\hat{\Sigma}_n - \Sigma)z^T$ is equal to $\frac{1}{\sqrt{2}}(1+x)\frac{1}{n}\sum_{i=1}^n x_i$ which is equal to zero when $x = -1$. $\tau_1(q)$ the additional term in the asymptotic distribution is therefore equal to zero. The proof is similar for other values of θ.

E Additional Appendix

E.1 Proof of Proposition 8

We denote M a generic majorizing constant. 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 $\hat{\theta}_n = (q, \hat{\Sigma}_n)$. First, under the conditions of Proposition 8, 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:

$$\left|\frac{1}{n}\sum_{i=1}^n f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i))\right| \underset{n\to\infty}{\overset{a.s}{\to}} 0. \tag{E.25}$$

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

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

we have:

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

Use equation (C.12):

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

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

$$\left|f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_\theta(z_i, \underline{y}_i, \overline{y}_i)\right| \underset{n\to\infty}{\overset{a.s}{\to}} 0. \tag{E.26}$$

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:

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

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

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

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 (E.26) implies that, by the dominated convergence theorem, uniformly over q,

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

From the latter equation, equation (E.25) and the triangular inequality, we thus conclude that, uniformly for $q \in \mathbb{S}$:

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

E.2 Construction of the Confidence Region in Proposition 11

Like before, for the simplicity of the exposition, we focus on the case where B is strictly convex. We here provide a simple way to construct $CI_\alpha^n = \{\beta; \xi_n(\beta) > \mathcal{N}_\alpha\}$ when $\alpha < 1/2$. Recall first the definitions of $\xi_n(\beta)$ and $T_n(q;\beta)$ in Section 4.2:

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

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

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

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

Consider any point $\beta_f \in \partial\hat{B}_n \subset CI_\alpha^n$, 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:

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

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:

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

where $-\lambda q^\top q_f \geq -\lambda q_f^\top q_f = -\lambda$ 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 $(\beta_f, \beta_f - \frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha q_f]$ is included in CI_α^n. We thus proved that:

$$\hat{B}_n \cup \{\cup_{\beta_f \in \partial B_n} \cup_{q_f \in \mathcal{C}(\beta_f)} (\beta_f, \beta_f - \frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha q_f)\} \subset CI_\alpha^n, \tag{E.27}$$

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:

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

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 $q_f^\top \beta_f \geq q_f^\top \beta$ which proves that $\hat{\delta}_n^*(q_f|B) = q_f^\top \beta_f$. 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 (E.27). As a consequence, equation (E.27) is an equality.

E.3 Behaviour of $\xi_n(\beta)$ when the set is a singleton

When $B = \{\beta_0\}$, it means that w_q is constant, equal to y_e (either $\overline{y}$ or $\underline{y}$). Consequently, $\beta_0 = E(z^\top x)^{-1}E(z^\top y_e)$. Let β_n be the point where the previous expectations are replaced by their empirical counterpart: $\hat{\delta}_n^* = q^\top \beta_n$. A CLT can therefore be applied to β_n:

$$\sqrt{n}(\beta_n - \beta_0) \underset{n\to+\infty}{\longrightarrow} N(0, V),$$

where V is some p.d matrix.

If we test a point $\beta \neq \beta_0$, $\xi_n(\beta)$ tends to $-\infty$ (q_0 is in this case $\frac{\beta-\beta_0}{\|\beta-\beta_0\|}$).

When $\beta = \beta_0$,

$$\begin{aligned} T_n(q;\beta_0) &= \left(\hat{\delta}_n(q) - q^\top \beta_0\right) \\ &= q^\top (\beta_n - \beta_0) \end{aligned}$$

In this case $q_n = -\frac{\beta_n-\beta_0}{\|\beta_n-\beta_0\|}$ and $T_n(q_n;\beta_0) = -\|\beta_n - \beta_0\|$.

And, after standardization:

$$\xi_n(\beta_0) = -\|u\|,$$

where u tends asymptotically toward a standard normal distribution. If we use the usual critical values to construct the confidence region, *i.e.* $\mathcal{N}_\alpha$, the probability that $\xi_n(\beta_0)$ is greater than this value is not $1-\alpha$ but $1-2\alpha$.

E.4 Uniform confidence regions

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

$$\hat{\Delta}_n = \max_{q \in \mathbb{S}} \left(\hat{\delta}^*_n(q|B) + \hat{\delta}^*_n(-q|B)\right). \tag{E.28}$$

Using a proof analogue to the one developed in Proposition 10:

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

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 18 *Let*

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

where q_n is the argument of the maximum of equation (E.28).

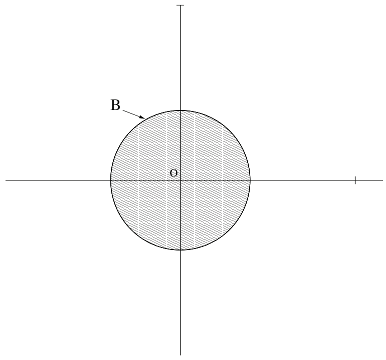
A confidence region $\tilde{CI}^n_\alpha$ of asymptotic level equal to $1-\alpha$ is defined by the collection of the points such that $\xi(\beta) \geq \tilde{\mathcal{N}}_\alpha$ where $\tilde{\mathcal{N}}_\alpha$ satisfies the equation

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

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

FIGURES and TABLES

Table 1: Results related to the Monte Carlo simulations - example 1



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

Support function $\delta(q)$ for $q = (0, 1)^T$
True unknown value 0.199

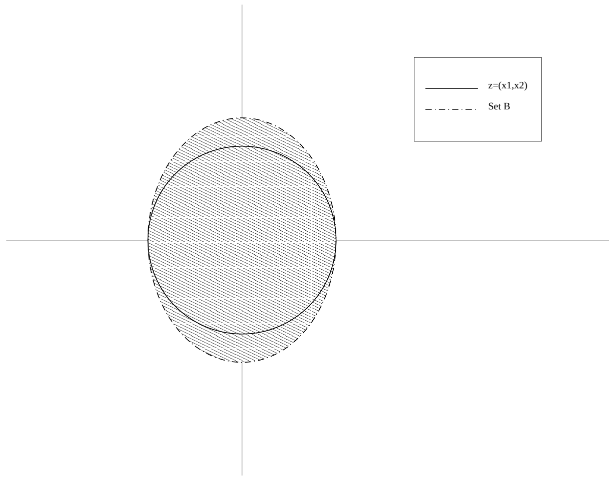
n	Mean	Q1	Q2	Q3
100	0.198	0.178	0.197	0.216
500	0.199	0.190	0.199	0.208
1000	0.199	0.193	0.199	0.206
2500	0.199	0.196	0.199	0.203

Table 2: Percentage of rejections for the two tests - first example

	Test 1 ($H_0 : \beta^r \in B$)				Test 2 ($H_0 : \beta^r \in \partial B$)			
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.01	0%	0%	0%	0%	70.9%	100%	100%	100%
0.05	0%	0%	0%	0%	69.9%	100%	100%	100%
0.1	0%	0%	0%	0%	67.7%	100%	100%	100%
0.2	0%	0%	0%	0%	60.1%	100%	100%	100%
0.3	0%	0%	0%	0%	51.6%	99.9%	100%	100%
0.4	0%	0%	0%	0%	40.5%	99.6%	100%	100%
0.5	0%	0%	0%	0%	29.4%	97.3%	99.9%	100%
0.6	0.5%	0%	0%	0%	19.6%	85.4%	99%	100%
0.65	0.7%	0%	0%	0%	16.2%	73.3%	97.1%	100%
0.7	1%	0%	0%	0%	12.7%	61.1%	89.8%	99.9%
0.75	1.3%	0.1%	0%	0%	9.7%	45.8%	76.2%	99%
0.8	1.6%	0.1%	0%	0%	7.9%	31.5%	58.2%	92.3%
0.85	2.6%	0.3%	0.2%	0%	6.5%	19.7%	36.5%	73.2%
0.9	3.2%	0.7%	0.5%	0.1%	5.7%	10.4%	19.7%	39.9%
0.95	5.1%	2%	1.5%	0.6%	5.3%	5.1%	8.5%	13.6%
1	**6.9%**	**5%**	**5.2%**	**5.5%**	**5.6%**	**4.1%**	**5.2%**	**5%**
1.05	10.1%	10.7%	14%	22.9%	6.5%	6.4%	9.4%	15.3%
1.1	14%	21.5%	29.9%	54.1%	8.4%	12.3%	20.8%	43.2%
1.15	17.7%	33.9%	50.7%	82.8%	11.2%	24%	37.1%	74.4%
1.2	21.5%	47.1%	70.7%	97.1%	14.9%	35.9%	58.7%	93.3%
1.25	25%	62.3%	85.6%	99.6%	19.1%	50.4%	78.1%	99.1%
1.3	30.6%	75.2%	94.7%	100%	22.3%	64.7%	89.9%	100%
1.35	36.4%	86.4%	98.1%	100%	26.2%	77.4%	96.3%	100%
1.4	43.9%	93.4%	99.6%	100%	31.7%	87.6%	98.8%	100%
1.45	49.8%	97.6%	99.9%	100%	37.4%	94%	99.7%	100%
1.5	57.8%	98.8%	100%	100%	45.1%	97.9%	99.9%	100%
2	96.3%	100%	100%	100%	93.8%	100%	100%	100%
2.25	99.3%	100%	100%	100%	98.6%	100%	100%	100%
2.5	99.9%	100%	100%	100%	99.7%	100%	100%	100%
2.75	100%	100%	100%	100%	99.9%	100%	100%	100%
3	100%	100%	100%	100%	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 3: Results related to the Monte Carlo simulations - example 2 with supernumerary instruments



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

Support function $\delta(q)$ for $q = (0, 1)^T$
True unknown value: 0.243

n	Mean	Q1	Q2	Q3
100	0.244	0.216	0.242	0.268
500	0.244	0.232	0.244	0.256
1000	0.243	0.234	0.243	0.252
2500	0.243	0.238	0.243	0.248

Table 4: Percentage of rejections for the two tests - second example

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

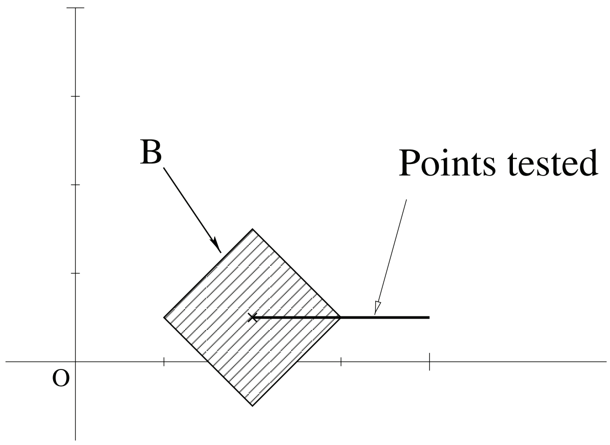
The point tested β^r is located on the x-axis. r is the fraction of the distance from the origin w.r.t. to the distance origin-frontier point on this axis. $r = 1$ is the frontier point (results in bold), $r = 0$ to the origin.

Table 5: Percentage of rejections for the test $H_0 : \beta^r \in \partial B$. Non-smooth set.

	Test with a_n=0				Test with $a_n = \frac{0.5}{n^1/3}$			
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.010	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
0.050	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
0.100	100.0 %	100.0 %	100.0 %	100.0 %	99.8 %	100.0 %	100.0 %	100.0 %
0.200	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
0.300	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
0.400	100.0 %	100.0 %	100.0 %	100.0 %	99.7 %	100.0 %	100.0 %	100.0 %
0.500	100.0 %	100.0 %	100.0 %	100.0 %	98.8 %	100.0 %	100.0 %	100.0 %
0.600	100.0 %	100.0 %	100.0 %	100.0 %	93.3 %	100.0 %	100.0 %	100.0 %
0.700	100.0 %	100.0 %	100.0 %	100.0 %	80.1 %	100.0 %	100.0 %	100.0 %
0.800	100.0 %	97.4 %	100.0 %	100.0 %	49.6 %	98.8 %	100.0 %	100.0 %
0.850	98.6 %	80.7 %	98.7 %	100.0 %	31.5 %	93.2 %	99.6 %	100.0 %
0.900	73.4 %	43.6 %	76.3 %	99.6 %	17.2 %	74.1 %	93.6 %	100.0 %
0.910	63.4 %	35.7 %	65.4 %	98.1 %	14.9 %	64.7 %	89.6 %	99.9 %
0.920	52.6 %	28.8 %	53.5 %	95.1 %	13.1 %	55.2 %	82.9 %	99.8 %
0.930	40.9 %	20.8 %	40.8 %	88.8 %	12.3 %	47.8 %	73.5 %	98.8 %
0.940	28.2 %	15.3 %	29.9 %	75.0 %	11.6 %	38.0 %	62.4 %	95.7 %
0.950	19.0 %	11.2 %	19.6 %	54.0 %	11.1 %	30.7 %	49.5 %	88.0 %
0.960	13.4 %	9.1 %	14.0 %	33.8 %	11.1 %	24.1 %	37.5 %	71.1 %
0.970	9.5 %	7.7 %	9.7 %	19.2 %	10.8 %	17.8 %	24.3 %	48.1 %
0.980	10.1 %	9.1 %	9.4 %	9.0 %	10.0 %	13.7 %	15.2 %	27.8 %
0.990	12.8 %	12.3 %	11.7 %	6.9 %	10.9 %	12.4 %	12.5 %	13.6 %
1	**18.4 %**	**16.1 %**	**16.1 %**	**14.2 %**	**11.7 %**	**12.0 %**	**13.0 %**	**10.6 %**
1.010	28.1 %	21.4 %	24.4 %	28.1 %	12.7 %	13.7 %	18.0 %	18.8 %
1.020	38.6 %	29.0 %	34.7 %	50.3 %	14.6 %	17.6 %	26.2 %	38.4 %
1.030	48.8 %	37.6 %	47.7 %	69.2 %	18.2 %	25.7 %	37.2 %	61.9 %
1.040	62.5 %	44.7 %	62.1 %	85.6 %	19.7 %	34.6 %	50.6 %	82.2 %
1.050	73.3 %	51.9 %	73.5 %	94.9 %	22.0 %	44.1 %	64.5 %	92.6 %
1.060	83.5 %	61.6 %	82.5 %	98.3 %	25.4 %	54.3 %	78.4 %	97.5 %
1.070	91.1 %	68.8 %	89.8 %	99.5 %	28.9 %	65.3 %	88.1 %	99.4 %
1.080	95.3 %	78.2 %	94.6 %	99.7 %	32.5 %	73.5 %	93.2 %	99.7 %
1.090	97.3 %	84.3 %	97.8 %	100.0 %	37.5 %	81.3 %	96.4 %	100.0 %
1.100	98.6 %	90.7 %	98.8 %	100.0 %	40.9 %	87.3 %	98.4 %	100.0 %
1.150	100.0 %	99.4 %	100.0 %	100.0 %	63.1 %	99.6 %	100.0 %	100.0 %
1.200	100.0 %	100.0 %	100.0 %	100.0 %	82.7 %	100.0 %	100.0 %	100.0 %
1.300	100.0 %	100.0 %	100.0 %	100.0 %	98.3 %	100.0 %	100.0 %	100.0 %
1.400	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.500	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.600	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.700	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.800	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.900	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
2	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %

SUPPLEMENTARY APPENDIX

Table 6: Results related to the Monte Carlo simulations - nonsmooth set



Set B, $y = \frac{1}{2} + \frac{x}{8} + \varepsilon$, $x \in \{-1, 1\}$

Support function $\delta(q)$ for $q = (0, 1)^T$
True unknown value: 0.375

n	Mean	Q1	Q2	Q3
100	0.374	0.360	0.375	0.390
500	0.375	0.369	0.375	0.382
1000	0.375	0.371	0.375	0.380
2500	0.375	0.372	0.375	0.378

Table 7: Percentage of rejections for the test $H_0 : \beta^r \in B$. Non-smooth set.

	Test with a_n=0				Test with $a_n = \frac{0.5}{n^1/3}$			
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.010	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.050	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.100	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.200	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.300	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.400	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.500	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.600	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.700	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
0.800	0.0 %	0.0 %	0.0 %	0.0 %	0.2 %	0.0 %	0.0 %	0.0 %
0.850	0.0 %	0.0 %	0.0 %	0.0 %	0.9 %	0.0 %	0.0 %	0.0 %
0.900	0.0 %	0.3 %	0.0 %	0.0 %	2.2 %	0.1 %	0.0 %	0.0 %
0.910	0.0 %	0.5 %	0.0 %	0.0 %	2.7 %	0.1 %	0.0 %	0.0 %
0.920	0.0 %	0.6 %	0.1 %	0.0 %	2.9 %	0.1 %	0.0 %	0.0 %
0.930	0.1 %	1.1 %	0.2 %	0.0 %	4.4 %	0.2 %	0.0 %	0.0 %
0.940	0.7 %	2.3 %	0.5 %	0.0 %	5.5 %	0.9 %	0.0 %	0.0 %
0.950	1.5 %	3.5 %	1.1 %	0.0 %	6.1 %	1.2 %	0.2 %	0.0 %
0.960	2.7 %	4.9 %	2.7 %	0.2 %	6.6 %	1.9 %	0.4 %	0.0 %
0.970	5.6 %	7.6 %	5.3 %	1.0 %	8.2 %	3.2 %	1.7 %	0.1 %
0.980	9.8 %	12.2 %	9.6 %	3.2 %	10.3 %	5.7 %	3.4 %	0.7 %
0.990	16.2 %	17.1 %	14.4 %	8.7 %	12.4 %	8.5 %	7.3 %	3.5 %
1	**25.6 %**	**22.7 %**	**22.4 %**	**19.3 %**	**15.0 %**	**12.2 %**	**13.4 %**	**10.7** %
1.010	36.6 %	30.3 %	32.5 %	39.2 %	18.1 %	16.6 %	22.3 %	26.3 %
1.020	47.4 %	39.2 %	47.1 %	60.3 %	19.9 %	25.1 %	33.5 %	48.6 %
1.030	59.6 %	45.4 %	59.3 %	78.1 %	23.3 %	34.6 %	46.9 %	70.6 %
1.040	70.6 %	53.9 %	71.5 %	90.5 %	26.5 %	44.2 %	60.0 %	88.0 %
1.050	82.3 %	63.2 %	80.8 %	97.1 %	29.6 %	54.2 %	73.8 %	95.6 %
1.060	89.7 %	70.1 %	88.4 %	99.3 %	33.7 %	65.1 %	85.7 %	98.8 %
1.070	94.4 %	79.1 %	94.1 %	99.6 %	38.9 %	73.6 %	91.4 %	99.5 %
1.080	96.7 %	84.8 %	97.6 %	99.9 %	42.6 %	81.2 %	95.8 %	99.8 %
1.090	98.6 %	91.2 %	98.5 %	100.0 %	45.8 %	87.4 %	98.2 %	100.0 %
1.100	99.2 %	94.7 %	99.5 %	100.0 %	50.3 %	91.6 %	99.0 %	100.0 %
1.150	100.0 %	99.8 %	100.0 %	100.0 %	72.3 %	100.0 %	100.0 %	100.0 %
1.200	100.0 %	100.0 %	100.0 %	100.0 %	87.9 %	100.0 %	100.0 %	100.0 %
1.300	100.0 %	100.0 %	100.0 %	100.0 %	99.2 %	100.0 %	100.0 %	100.0 %
1.400	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.500	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.600	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.700	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.800	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
1.900	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %
2	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %