ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Bontemps, Christian; Magnac, Thierry; Maurin, Eric

Working Paper

Set identified linear models

cemmap working paper, No. CWP13/11

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Bontemps, Christian; Magnac, Thierry; Maurin, Eric (2011) : Set identified linear models, cemmap working paper, No. CWP13/11, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2011.1311

This Version is available at:
https://hdl.handle.net/10419/64804

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

cemmap
centre for microdata methods and practice

Set identified linear models

Christian Bontemps
Thierry Magnac
Eric Maurin

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP13/11

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Set Identified Linear Models

Christian Bontemps*, Thierry Magnac†, Eric Maurin‡

First version, August 2006
This Version, December 2010

Abstract

We analyze the identification and estimation of parameters β satisfying the *incomplete* linear moment restrictions $E(z^\top(x\beta - y)) = E(z^\top u(z))$ 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 constructed from the identification results. When B is a strictly convex set, we also construct a test of the null hypothesis, $\beta_0 \in B$, whose size is asymptotically correct and which relies on the minimization of the support function of the set $B - \{\beta_0\}$. Results of some Monte Carlo experiments are presented.

Keywords: partial identification; linear prediction; support function; test

*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 we call incomplete linear moment restrictions. Specifically, we consider y, a dependent variable, x, a vector of p variables and z a vector of m variables and assume that parameter β satisfies:

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

where $u(z)$ is any single-dimensional measurable function that takes its values in a given bounded interval $I(z)$. One leading example is the familiar linear projection model $y = x\beta + \varepsilon$, where ε is uncorrelated with z, 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 which satisfy equation (1) for at least one admissible $u(.)$

A general approach to inference when a set only is identified has recently been 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 procedure is valid even when the identified set is neither convex nor bounded. In contrast, the identified set B analysed in this paper is, by construction, bounded and convex. These are key features that we exploit using the concept of a support function whose argument is any direction q of the unit sphere in R^p. The support function sharply characterizes any convex set (Rockafellar, 1970) and Beresteanu and Molinari (2008) used them in the linear projection model in which

[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 insightful comments that significantly improve the paper. We also thank the participants at seminars at PUC-Rio, CEMMAP, CREST Malinvaud seminar, Mannheim, Yale, NYU, Paris I, Cornell, MIT, Northwestern, Toulouse, Cambridge, Bristol and Carlos III 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 Russell Davidson '08, 1st French Econometrics Conference '10) for comments. We thank Mehtap Akguc for her excellent research assistance. The usual disclaimer applies.

variables x and z are identically the same. We show that in the set-up in which there are as many instruments as explanatory variables ($m = p$), the support function of the identified convex set B, is equal to the expectation of a simple random function.

Our first contribution relates to identification. When there are supernumerary instruments ($m > p$), the identified set B might be empty and this would refute the exclusion restrictions as discussed by Manski (2003, Chapter 2). We exhibit a necessary and sufficient condition, a generalization of the usual over-identifying condition *à la* Sargan, under which the identified set is not empty and the exclusion restrictions are acceptable. Set B remains bounded and convex and we show as a second contribution that its support function results from the minimization of the expectation of a simple random function. We also exhibit conditions under which the existence of supernumerary instruments restores point identification.

The next contribution of the paper is to provide a simple estimator of the support function of the identified set. This estimator is the empirical analogue of the expectation of the random function to which the support function is equal. In their closely related contribution, Beresteanu and Molinari (2008) provide an estimation procedure for a class of convex identified sets using the theory of random sets. We directly use the theory of stochastic process from which the theory of random sets is derived because the results are easier to generalize to the endogenous case. Specifically, when the support function is not differentiable, we show that the estimate converges in distribution at a $\sqrt{n}$ rate to the sum of a Gaussian process and of a process that we characterize and whose support comprises the points of non differentiability only. Given the prevalence of discrete regressors that leads to such non differentiability issues, this generalization is worthy of attention. Interestingly enough, our approach also reveals that the asymptotic results of Beresteanu and Molinari (2008) actually simplify to a quite standard linear model format for the covariance matrix in the case in which the support function is differentiable.

Furthermore and more importantly, we develop a new test procedure for null hypotheses concerning parameter values such as $H_0 : \beta_0 \in B$ when the support function is differentiable. This class of hypotheses is particularly adapted to our setting since the generalized Sargan condition developed in the supernumerary moment case can be written as an hypothesis about values. Our test has correct asymptotic size and is very easy to adapt to hypotheses about sub-vectors of the complete parameter. 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 and we exploit

this characteristic to derive the asymptotic distribution of the test statistic even in the case in which the convex set B has kinks. The form of the test statistic is reminiscent of the test statistic proposed by Galichon & Henry (2009) in a more general context since the space over which the minimum of the test statistic is taken in our case is much smaller than theirs because of the convexity assumption.

Finally, the same key feature of convexity allows us to derive asymptotic properties of the estimates in the case in which there are supernumerary moment restrictions and the identified set is a proper set. Estimates are uniformly almost surely consistent and when the support function is differentiable, the inflated difference between the estimated and true functions converges to a Gaussian process whose covariance operator can be characterized and estimated simply.

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 correspond 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). Seminal analysis of the incomplete information case include Fréchet (1951), Hoeffding (1940) and Manski (1989) whereas recent applications include Vazquez-Alvarez, Melenberg and van Soest (2001), Blundell, Gosling, Ichimura and Meghir (2007), Honoré and Lleras-Muney (2006) and Ciliberto and Tamer (2009). Horowitz and Manski (1995) consider the case where the data are corrupted or contaminated while Ridder and Moffitt (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, Galichon and Henry, 2009 among others). Set identification can also be generated by discrete exogeneous variation such as in Chesher (2005) or by structurally missing information such as in the treatment effect literature (see Fan and Park, 2010 or Fan and Wu, 2010 for recent analyses). 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, 2008). In the moment inequality set-up, Rosen (2008) develops simple testing pro-

cedures and Bugni (2010) and Canay (2010) investigate the properties of canonical bootstrap and modifications of it. Andrews and Guggenberger (2009) studies cases that do not fall under the assumptions of Imbens and Manski (2004) or Stoye (2009).

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. Yet, our framework is more restrictive than the popular moment inequality restrictions set up since the unknown function $u(z)$ is single dimensional. The leading examples that we propose are derived from partial observation when outcomes are censored by intervals (Manski & Tamer, 2002, Stoye, 2007, Beresteanu & Molinari, 2008), when the continuous regressor in a binary model is observed by intervals or is discrete (Magnac & Maurin, 2008) or when categorical data on opinions and attitudes are analyzed.

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. Examples of incomplete non linear conditions are offered by Beresteanu, Molchanov and Molinari (2010). In contrast, we believe that estimation is more 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. 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 defines the set up of Incomplete Linear Models and develops examples that are of interest for applied econometricians. Section 3 sharply characterizes the identified set. We analyze the case in which the number of parameters is equal to the number of restrictions as well as the case in which the number of restrictions is larger than the number of parameters. In the latter case, 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 also develop test procedures, construct confidence regions by inversion of the tests and derive asymptotic properties of the estimates in the case in which there are supernumerary restrictions. Section 5 is devoted to Monte Carlo experiments about estimation and 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 of what we call an *incomplete linear model*. In this model, the variables satisfy the *incomplete linear moment* conditions:

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

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[2] in an admissible set $I(z) = [\underline{\Delta}(z), \overline{\Delta}(z)]$. These two bounds $\underline{\Delta}(z)$ and $\overline{\Delta}(z)$ can be constructed using two observable variables $\overline{y}$ and $\underline{y}$ such that $\overline{y} \geq \underline{y}$, $y \in \left[\underline{y}; \overline{y}\right]$, and

$$E(\overline{y} - y \mid z) = \overline{\Delta}(z) > 0 > E(\underline{y} - y \mid z) = \underline{\Delta}(z). \tag{3}$$

The next subsections provide examples, the leading one being that a dependent variable is observed by intervals only so that the lower, $\underline{y}$, and upper bounds, $\overline{y}$, of the interval are explicitly 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 included in $\mathbb{R}^p \times \mathbb{R}^m$. Furthermore, the conditions of full rank, $rank(E(z^\top x)) = p$ and $rank(E(z^\top z)) = m$ hold. Finally, $\Pr\{z = 0\} = 0$.

$R.iii$. The random vector $(\overline{y}, \underline{y}, x, z)$ belongs to the space L^2 of square integrable variables.

Along with equation (2), assumptions $R.i - ii$ defines the linear model where there are p explanatory variables and m instrumental variables (assumption $R.ii$). Assumption $R.ii$, accommodates the standard exogenous case $x = z$ as a particular case and the absence of a mass point at $\{0\}$ is an assumption made for notational convenience and it can be dispensed with (see note 6 further on). Assumption $R.iii$ implies 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]For ease of notation, all statements referring to any value of a random variable should be understood as this random variable almost surely.

2.1 Censored Dependent Variables

The first interesting set of examples corresponds to common 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 that are often reported by interval 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 vector $w = (\overline{y}, \underline{y}, x)$ where $\left[\underline{y}, \overline{y}\right]$ represents the observed interval[4] of a latent variable y^* and x is a vector of p covariates. The observed bounds $\underline{y}$ and $\overline{y}$ are assumed to satisfy *R.iii*.[5] Within this framework, we consider linear latent models :

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

where ε is a random variable uncorrelated with x, $E(x^\top \varepsilon) = 0$. The issue is to characterize the set B of parameters β such that the latent model defined by equation (4) is consistent with the observed bounds. By definition, β belongs to B if and only if there exists 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 $\Delta(x) = E(\frac{\overline{y}-\underline{y}}{2} \mid x)$ the average half-length of the observed interval $\left[\underline{y}, \overline{y}\right]$. Then β belongs to B if and only if there exists a measurable function $u(x)$ which takes values in $I(x) = [-\Delta(x); \Delta(x)]$ such that,*

$$E(x^\top (x\beta - y)) = E(x^\top u(x)). \tag{5}$$

[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 interval is not closed, it is not set B itself but the closure of B which is identified (see Magnac and Maurin, 2008 for more precise statements about closure).

[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).

Proof. See Appendix A.1 ■

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, the latent model is written as that $w^* = x\beta + \varepsilon$ and the semiparametric binary model $w = 1(x\beta + v + \varepsilon > 0)$ is estimated under three assumptions. The random term 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 equation (2) (see Magnac and Maurin, 2008).

Categorical data on individual opinions or attitudes provide 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 equation (2).

3 The Identified Set of Structural Parameters

This section provides a detailed description of B, the set of observationally equivalent parameters, β, that are compatible with the incomplete linear model 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 a particular example). Second we show how the results can be extended to the case in which the number of instruments z is larger than 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 (2) 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:

$$B = \{\beta : \beta = \left(E(z^\top x)\right)^{-1} E\left(z^\top \left(y + u(z)\right)\right), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\}. \tag{6}$$

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.

The key object that we exploit is the *support function* of a convex set defined as:

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

Given that support functions are positive homogenous in q, it is sufficient to define them over the unit sphere of $\mathbb{R}^p$ i.e.$\mathbb{S}_p = \{q \in \mathbb{R}^p; \|q\| = 1\}$. Furthermore, following Rockafellar (1970), set B can be unambiguously characterized as:

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

and identification of B is therefore equivalent to the identification of its support function $\delta^*(. \mid B)$. Beresteanu and Molinari (2008) also use this function in order to apply the theory of random set variables.

We now show that the support function of B can be written as a population moment of two simple random variables. Let β a point which belongs to set B. From (6), there exists some function $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$ such that

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

We can multiply this equation by vector q to express:

$$q^\top \beta = q^\top \left(E(z^\top x)\right)^{-1} E\left(z^\top \left(y + u(z)\right)\right) = E\left(z_q \left(y + u(z)\right)\right), \tag{7}$$

where $z_q = q^\top \Sigma^\top z^\top$ and $\Sigma = \left(E(x^\top z)\right)^{-1}$. As the support function in the direction q is the supremum of $q^\top \beta$ when $\beta \in B$, it is the supremum of (7) over the set of admissible $u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. It reduces to a simple single-dimensional optimization problem whose solution is given by:

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 $\beta_q = \left(E(z^\top x)\right)^{-1} E(z^\top w_q)$ is a frontier point of B such that $\delta^*(q \mid B) = q^\top \beta_q$.

Proof. See Appendix B.1 ■

This proposition sharply characterizes set B. The support function is well defined because assumption *(R.iii)* ensures 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 $\mathbb{S}_p$ except on a set D_f which is composed of directions $q \in \mathbb{S}_p$ such that $\Pr(z_q = 0)$ is positive. The directions in D_f are orthogonal to exposed faces of the identified set.*

Proof. See Appendix B.2 ■

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.[6] Defining this set turns out to be important for asymptotic properties derived in the next section. One Monte Carlo experiment in the Supplementary Appendices analyzes the common case in which one explanatory variable is a dummy variable.

[6] If z has a mass point at 0, the support function would not be differentiable at points such that $\Pr(z_q = 0 \mid z \neq 0) > 0$ as can be seen from the proof of the Lemma. For notational convenience, we choose to exclude this case.

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 ζ^s denotes the vector of residuals of the linear projection of these $m-p$ instruments onto $x(z)$, we assume that rank $\left(E(\zeta^{s\top}\zeta^s)\right) = m-p$. It may very well be the case that other subsets of $m-p$ instruments than z^s satisfy this condition, but, as discussed in Appendix B.3, our results do not depend on the choice of a specific subset.

The parameters of interest β satisfy the incomplete linear moment conditions (2):

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

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 (8) 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 ζ^s of the projection of the supernumerary instruments z^s onto z_F. Analytically, $z_H = \zeta^s E(\zeta^{s\top}\zeta^s)^{-1/2}$ where,

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

After some algebraic manipulations, we have $E(\zeta^{s\top}x) = 0$ (as a matrix) and, consequently, $E(z_H^\top x) = 0$.

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 (8) implies,

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

As $E(z_H^\top x) = 0$, the second set writes $E\left(z_H^\top(y+u(z))\right) = 0$. 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:*

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

$$E\left(z_H^\top(y+u(z))\right) = 0 \tag{10}$$

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

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

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

$$\begin{aligned} B_{\text{Sargan}} &= \left\{\gamma : E\left(z_H^\top(z_H\gamma - y)\right) = E(z_H^\top u(z)), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\right\} \\ &= \left\{\gamma : \gamma = E(z_H^\top(y+u(z))), u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]\right\} \subset \mathbb{R}^{m-p}. \end{aligned} \tag{12}$$

The adapted Sargan condition given by equation (11) 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. $B_{Sargan} \ni O_{m-p}$.

Proof. Using the previous developments ■

[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^s, among instruments$z$.

This extends the usual overidentification restrictions. 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, 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:

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

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 moments are in surplus. Consequently, the support function of B_U can be characterized using Proposition 2. The second interesting feature of this construction is that the identified set B is equal to the intersection of B_U and the hyperplane defined by $\gamma = 0$. General results for the support function of intersections 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:*

$$\delta^*(q \mid B) = \inf_{\lambda} \delta^*((q, \lambda) \mid B_U). \tag{13}$$

Proof. Rockafellar (1970) and Appendix B.4 ■

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. 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 $\gamma = 0$,

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

is equal to the set of unconstrained solutions to equation (9). This projection contains set B since parameters in set B are generated by functions $u(z)$ satisfying in addition condition (10). Supernumerary restrictions therefore reduce the size of the identified set and generically they strictly do so. Appendix B.5 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.

3.2.3 Supernumerary Moment Conditions as a Way to Restore Point Identification

The adapted Sargan condition ($O_{m-p} \in B_{\text{Sargan}}$) 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 (11) are not unique and set B is a proper set. More interesting cases arise when O_{m-p} belongs to the frontier of B_{Sargan}. In such cases, using the proof of Proposition 2, we know that there exists q_O in $\mathbb{S}_{m-p}$ such that O_{m-p} is generated by functions $u_{q_O}^{\text{Sargan}}(z)$ defined as,

$$u_{q_O}^{\text{Sargan}}(z) = \underline{\Delta}(z) + (\overline{\Delta}(z) - \underline{\Delta}(z))\mathbf{1}\{z_{Hq_O} > 0\} + \Delta^*(z)\mathbf{1}\{z_{Hq_O} = 0\}$$

where $\Delta^*(z)$ can be any function taking values in $[\underline{\Delta}(z), \overline{\Delta}(z)]$. Hence, if z_H have no mass points, $u_{q_O}^{\text{Sargan}}(z)$is unique and the identified set B is reduced to a point defined as:

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

We summarize this result by,

Proposition 7 *If O_{m-p} belongs to the frontier ∂B_{Sargan} and if instruments z_H have no mass points then parameter β is point-identified.*

More generally, the restoration of point identification only requires that $\Pr\{z_{Hq_O} = 0\} = 0$, that is when $m - p > 1$, O_{m-p} is not on an exposed face of B_{Sargan}.

4 Estimation and Inference

This section describes how we estimate the support function of B and how we test hypotheses of interest. We will deal only with random samples $i = 1, ., n$, in which variables $(\overline{y}_i, \underline{y}_i, y_i, x_i, z_i)$ are independently and identically distributed.[8] We start by analysing the case where there is no supernumerary moment conditions.

4.1 Asymptotic Properties: 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 :

$$\delta^*(q \mid B) = E(z_q w_q). \tag{14}$$

To apply the analogy principle, we first define $\hat{\Sigma}_n$ a bounded estimate of $\left(E(x^\top z)\right)^{-1}$ (see Appendix C) and for any i the variables:

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

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

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

Under usual conditions (White, 1999, p. 35), the estimate $\hat{\delta}_n^*(q \mid B)$ is uniformly consistent.

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

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

Proof. See Beresteanu and Molinari, 2008 and Supplementary Appendix D.3 ■

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 $(q, \Sigma) \in \Theta = \mathbb{S}_p \times \{\|\Sigma\| \leq M\}$. 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

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

surely to $\delta^*(q \mid B)$ 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 reasoning to derive the asymptotic distribution of the estimate by considering the stochastic process defined on $\mathbb{S}_p$:

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

whose asymptotic behavior is characterized under usual conditions (White, 1999, p. 118) in the following.

Proposition 9 *Assume that there exist $M > 0$ and $\gamma > 0$ such that $\|\Sigma\|^{2+\gamma}$, $E(\left\|x^\top z\right\|^{2+\gamma})$, $E(\left\|z^\top \overline{y}\right\|^{2+\gamma})$ and $E(\left\|z^\top \underline{y}\right\|^{2+\gamma})$ are bounded by M. Assume also that the number of mass points in the probability distribution of z_i is finite. The process, $\tau_n(q)$ uniformly converges in distribution when n tends to ∞ to the sum of a Gaussian stochastic process centered at zero and of a point process which is asymptotically equivalent to the following random process:*

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

where W is the asymptotic variance of $vec(\hat{\Sigma}_n^\top)$ and where η is a normally distributed random vector of dimension p^2, independent of $(\overline{y}, \underline{y}, z)$. Vectors q at which this process are non zero are orthogonal to the exposed faces of B (see Lemma 3).

When there are no mass points in the probability distribution of z_i so that set B has no exposed faces, $\tau_n(q)$ uniformly converges in distribution when n tends to ∞ to a Gaussian stochastic process centered at zero. The covariance function of this process for vectors $(q, r) \in \mathbb{S}_p$ is,

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

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

Proof. See Appendix C.1 ■

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 negative values at directions q orthogonal to exposed faces of set B.

4.2 Tests

This section provides testing procedures for null hypothesis such as $H_0 : \beta_0 \in B$, where β_0 represents a potential parameter value. The proposed tests have the correct asymptotic size. Similar approaches could be used for testing assumption about sets, such as $H_0 : B_0 = B$, where B_0 represents a potential value of the identified set. We will focus on the case where 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).[9]

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

One of the hypothesis we have in mind is the generalized Sargan condition. In this case, Assumption D holds if z_H has no mass points. As this variable is constructed by a projection of the original z variables on a subspace of dimension $m - p$, z_H would have generically no mass points if z has more than $m - p$ absolutely continuous components. The result is generic only since the projection could cancel the mixing characteristic of some of these components.

Assumption D does not exclude cases where the derivative of the support function has the same value at different points of the unit sphere. Put differently, it does not exclude that some points of the surface of B may have several different tangent spaces (kinks). In such a case, the relationship between directions of the unit sphere q and points of the surface of B is not one-to-one anymore, which complicates the testing procedure. Set-ups where B has kinks are characterized at the end of the proof of Lemma 3, one leading example being when the density function of some instruments is not positive over the whole real line (for instance an intercept combined with a random variable whose support is bounded).

[9]The difficulty with dealing with the additional point process in the limit is that it is not asymptotically equicontinuous. Some papers derive properties of minimizers of convex functions under more general assumptions than asymptotic equicontinuity (Hjort and Pollard, 1993). We thank Adam Rosen for attracting our attention to this paper but leave the adaptation of these arguments to this subject for future research.

Within this framework, an alternative characterization of $H_0 : \beta_0 \in B$ is:

$$\begin{aligned}\beta_0 \in B &\iff \forall q \in \mathbb{S}_p, T_\infty(q;\beta_0) = \delta^*(q \mid B) - q^\top \beta_0 \geq 0 \\ &\iff \min_{q\in\mathbb{S}_p} T_\infty(q;\beta_0) \geq 0,\end{aligned}$$

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

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

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 $\sqrt{n}T_n(q_0;\beta_0)/\sqrt{V_{q_0}}$ qualifies as a test statistic for H_0.

The two issues that we have to deal with are (i) q_0 is not known (ii) 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 $v_{0,n} \in \mathbb{S}_p$ 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 a single minimizer q_0^* of $T_\infty(q;\beta_0)$. Then,

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

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 special 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 kink, 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 $T_n(q;\beta_0)$ 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:

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

- 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:

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

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 statistic 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 $O_{m-p} \in \partial B_{\text{Sargan}}$ as in Proposition 7).

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 & Romano (2005), 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 D.4 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*

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

where q_n and $\hat{V}_n$ are 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 (see Appendix B.1). As a consequence, the confidence region is not uniformly asymptotically of nominal size equal to $(1-\alpha)$. Uniformity is important however as we might never know in practice how close we are to the point 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 Supplementary Appendix D.5). Indeed, the statistic 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.[10]

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 region (see Supplementary Appendix D.6).

4.4 Asymptotic Properties: The Supernumerary Case

We use the characterization given in equation (13) in Proposition 6. If q is a vector of $\mathbb{R}^p$ and (q,λ) a vector of $\mathbb{R}^m$, we have:

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

On top of assumption D, we assume that:

Condition S: The infimum is attained at values, $\lambda_m(q)$, which belong to a compact set $\Lambda \subset R^{m-p}$.

The last part of the proof of Proposition 6 exhibits the necessary and sufficient condition under which Condition S is obtained and which, combined with assumption D of differentiability, implies that set B is a proper set. This excludes the case in which point identification is recovered.

[10]The diameter of set B is the maximum of $\delta(q \mid B) + \delta(-q \mid B)$ on the compact unit sphere. In our case, the diameter of B is positive, see Appendix B.1.

Let $\hat{\delta}_n^*((q,\lambda) \mid B_U)$ the estimate of $\delta^*((q,\lambda) \mid B_U)$ as derived in Section 4.1 and such that, by Proposition 8:

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

and, by Proposition 9, under assumption D:

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

uniformly converges to a Gaussian process when n tends to infinity.

For any q, define:

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

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:

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

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.

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

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

and:

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

converges uniformly to a Gaussian process when n tends to infinity. The Gaussian process has expectation equal to zero and its covariance operator for two directions $(q,r) \in \mathbb{S}_p$ 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.3 ■

An additional general point is in order about small sample biases arising when minimizing an estimated function as discussed by Manski and Pepper (2009) for instance. A correction approach was recently proposed by Chernozhukov, Lee and Rosen (2009) where they use confidence interval estimates. These are computed as the sum of a minimizer of $\hat{\delta}_n^*((q,\lambda) \mid B_U)$

over a compact set and its estimated standard error times a critical value so that these confidence intervals are exact or conservative at level p. In the Monte Carlo experiments reported below, we did not find in practice sizeable small sample downward biases. Our set-up is arguably different since the variable λ with respect to which the minimum is taken is not random and can take any value. Moreover, the standard errors of estimates varies slowly with λ.

5 Monte-Carlo Experiments

In this section, we develop two 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 explanatory variable is single dimensional. In the first experiment, the explanatory variable is its own instrument whereas we use a single supernumerary restriction in the second experiment. The frontier of the identified set has no kinks and no exposed faces. Additional experiments in dimension two are provided in the Supplementary Appendices D.1. In one of them, the identified set is neither smooth nor strictly convex.

5.1 Smooth and Strictly Convex Set

Consider the model:

$$y^* = \beta^0 x + \varepsilon,$$

where x is uniformly distributed on $[-\sqrt{3}; +\sqrt{3}]$ (and has unit variance), ε is independent of x and uniformly distributed on $[-1/2; 1/2]$. The true value of β^0 is 2. We assume that y^* is interval censored and that the econometrician only observes the lower and upper bounds $\underline{y} = y^* + v\mathbf{1}\{\mathbf{v} < \mathbf{0}\}$ and $\overline{y} = y^* + v\mathbf{1}\{\mathbf{v} \geq \mathbf{0}\}$, where v is a standard normal variable independent of x. This censoring scheme is in line with the example of Section 2.1 and can be viewed as a process implemented by the statistician who conducts the survey to warrant some anonymity. Using the notations of Proposition 1, we have $y = y^* + v/2$ and $\Delta(x) = E(|v|/2|x) = E(|v|)/2$.

The identified interval B can be written as:

$$B = [\beta^{-1}; \beta^1] = [\beta^0 - \frac{1}{2}E(|xv|); \beta^0 + \frac{1}{2}E(|xv|)].$$

We draw 1000 simulations in four different sample size experiments : n =100, 500, 1000 and 2500. The three quartiles as well as the mean of the estimated extreme points β^{-1} and β^1 are displayed in Table 1. Even for small sample sizes, the identified set is well estimated and unsurprisingly, the interquartile interval decreases when the sample size increases.

Regarding the performance of test procedures, let β^r a point such that the algebraic distance between β^0 and β^r is equal to r times the value of the half length of B (positive values refer to points on the side of β^1 and negative values on the side of β^{-1}). Point β^r belongs to B if and only if $|r| \leq 1$. For r varying stepwise from -2 to 2, we computed the rejection frequencies at a 5% level for the interior test $H_0 : \ \beta^r \in B$ developed in Section 4.2 (labeled Test 1). Results are displayed in Table 2 and show that the size of the test is very accurate and remains close to 5% even for $n = 100$ and that its power is very good even in small samples. The frontier test yields similar qualitative results and is not reported here.

5.2 Smooth Set and a single Supernumerary Restriction

The simulated model is the same as before. We now assume that the econometrician is given some additional information on the censoring process. The corresponding supernumerary instrument z is equal, with probability $1 - \pi_0$, to a standard normal variable w independent of v and x and is equal, with probability π_0, to v . Using notations of Section 3.2, we have $z_F = x$ and $z_H = z$ and

$$B_{sargan} = [\frac{-\pi_0 E(|v|)E(|w|))}{2}; 1 - \pi_0 + \frac{\pi_0 E(|v|)E(|w|))}{2}],$$

which implies that B_{sargan} always contains the value zero. Furthermore, the smaller probability π_0 is, the closer to the frontier of B_{sargan} the value zero is. Hence, set B is never empty and shrinks down to the point $\{\beta^0\}$ when π_0 tends to 0. When π_0 is equal to 0, the knowledge of z is implicitly equivalent to that of y^* and point identification is mechanically restored. Finally, the support function of the unconstrained set B_U satisfies,

$$\delta((1, \lambda) \mid B_U) = \beta^0 + \frac{1 - \pi_0}{2}\lambda + E(|v|(x + \lambda z)1(x + \lambda z > 0)).$$

Minimizing $\delta((1, \lambda) \mid B_U)$ (respectively $\delta((-1, \lambda) \mid B_U)$) with respect to λ, yields $\delta(1 \mid B)$ the upper (resp. the opposite of the lower) bound of interval B. Note that the supernumerary restriction unambiguously reduces the size of the identified set since, for instance, the upper bound in the first experiment, $\delta((1, 0) \mid B_U)$, is smaller than $\delta((1, \lambda) \mid B_U)$.

As before, we draw 1000 simulations in four sample size experiments: n =100, 500, 1000 and 2500 with different values of π_0 (from 0.5 to 0). Table 3 displays the rejection frequencies for the Sargan Test proposed in Proposition 5. They are very low when O_{m-p} is inside set B_{Sargan} but converge to the nominal level when O_{m-p} gets closer to the frontier of B_{Sargan}, *i.e.* when π_0 tends to 0.

Table 4 displays the rejection frequencies for the interior test for different points in the same format as in Table 2, for two different values of π_0, 0.5 and 0.1. We only display results for positive r as they are almost symmetric. As B_{sargan} always contains zero, the intersection between set B_U and the hyperplane $\gamma = 0$ is not empty. However, it may happen that for some specific draws the estimate of set B_U has an empty intersection with the hyperplane. We therefore constrain λ in the minimization process to be in a large compact set, $|\lambda| \leq M$ (we set $M = 500$ in the simulations). Asymptotically it does not have any impact on the results.

As expected, the estimated set shrinks down to a point. The rejection frequency tends to the nominal one at the frontier point. However, the interior test can be altered by the fact that the constraint $|\lambda| \leq M$ is binding for some draws when π_0 is very small and the sample size is small. Consequently the estimated variance of the support function is large and points are less likely to be rejected when there are outside the estimated set.

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 because exposed faces are a very common occurrence. 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). Our results might help out with selecting the best criterion in the latter framework.

Finally and more ambitiously, the deep foundation of our approach is a convexity argument. It indeed allows to replace the problem of identifying 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 (see Beresteanu, Molchanov and Molinari, 2010 for a general framework). 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, 2009, "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities", *Econometric Theory,* 25:669-709.

Beresteanu, A. and F., Molinari, 2008, "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica,* 76:763-814.

Beresteanu, A., Molchanov and F., Molinari, 2010, "Sharp Identification Regions in Models with Convex Moment Predictions", CEMMAP WP 25-10.

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.

Bugni, F., 2010, "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities", *Econometrica*, 78:735-754.

Canay, I., 2010, "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156:408-425.

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

Chernozhukov, V., S. Lee and A.M. Rosen, 2009, "Intersection Bounds: Estimation and Inference", CEMMAP WP 19-09.

Chesher, A., 2005, "Non Parametric Identification under Discrete Variation", *Econometrica,* 73:1525-1550.

Ciliberto F. and E. Tamer, 2009, "Market structure and multiple equilibria in airline markets", *Econometrica,* 77:1791-1828.

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

Fan Y. and S., Park, 2010, "Sharp Bounds on the Distribution of the Treatment Effects and Their Statistical Inference", *Econometric Theory*, 26(3):931-951.

Fan Y., and J., Wu, 2010, "Partial Identification of the Distribution of Treatment Effects in Switching Regimes Models and its Confidence Sets", *Review of Economic Studies*, 77(3):1002-1041

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, 2009, "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters", *Journal of Econometrics,* 152:186-196.

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.

Hjort N.L., and D. Pollard, 1993, "Asymptotics for minimizers of convex processes", unpublished manuscript.

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", *Econometrica* 74: 1675-1698.

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

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

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

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

Lehmann, E.L. and J.P Romano, 2005, *Testing Statistical Hypotheses*, Springer: New York.

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", *Review of Economic Studies,* 75:835-864.

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., **and J.V. Pepper,** 2009, "More on monotone instrumental variables", *The Econometrics Journal*, 12:S200-216.

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

McFadden, D., 1994, "Contingent Valuation and Social Choice", *American Journal of Agricultural Economics*, 76:689-708.

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, 2008, "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Statistical Planning and Inference*, 138:2786-2807.

Rosen A.M., 2008, "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146:107-117.

Stoye, J., 2007, "Bounds on Generalized Linear Predictors with Incomplete Outcome Data", *Reliable Computing*, 13: 293–302.

Stoye, J., 2009, "More on Confidence Intervals for Partially Identified Parameters", *Econometrica,* 77:1299-1315.

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

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

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

Appendices

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,

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

We also have :

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

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

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

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

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

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:

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

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,

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

where we used that λ is independent of $\underline{y}$ and $\overline{y}$. Therefore, we have $E(\varepsilon \mid x) = E(y - x\beta \mid x) + u(x)$, 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 (A.1) involving y, β and $u(x)$.

B Proofs in Section 3

B.1 Proof of Proposition 2

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

$$q^\top \beta = E\left(z_q \left(y + u(z)\right)\right), \tag{B.2}$$

where $u(z)$ 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) and by setting $u(z)$ to any value when z_q is equal to 0. It yields a set of "supremum" functions:

$$u_q(z) = \underline{\Delta}(z) + (\overline{\Delta}(z) - \underline{\Delta}(z))\mathbf{1}\{z_q > 0\} + \Delta^*(z)\mathbf{1}\{z_q = 0\} \tag{B.3}$$

where $\Delta^*(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. 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 (3), $E(\overline{y} - y \mid z) = \overline{\Delta}(z)$, $E(\underline{y} - y \mid z) = \underline{\Delta}(z)$, so that the support function or the supremum of (B.2) is equal to:

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

where:

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

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

Furthermore, when $\Pr(z_q = 0) > 0$, since $\Delta^*(z)$ varies in $[\underline{\Delta}(z), \overline{\Delta}(z)]$, the functions $u_q(z)$ defined by equation (B.3) generate all the points $\beta = \left(E(z^\top x)\right)^{-1} E\left(z^\top (y + u_q(z))\right)$ which belong to the tangent space to B whose outer-pointing normal vector is 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 $\Delta^*(z) = 0$, we get the particular value of β:

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

and, by definition:

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

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

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

or equivalently that:

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

Start from consequences of definitions:

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

so that:

$$\begin{aligned}
\delta^*(q \mid B) + \delta^*(-q \mid B) &= E\left(|z_q| \, (\bar{y} - \underline{y})\right) \\
&= E\left(|z_q| \, E((\bar{y} - \underline{y}) \mid z)\right) \\
&= E\left(|z_q| \, (\overline{\Delta}(z) - \underline{\Delta}(z))\right) > 0
\end{aligned}$$

because of equation (3) and $|z_q| > 0$ with positive probability because of the full rank assumption in $R.ii$.

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

$$\min_{q \in \mathbb{S}_p} (\delta^*(q \mid B) + \delta^*(-q \mid B)) > 0$$

since $\mathbb{S}_p$ is compact.

B.2 Proof of Lemma 3

We use the expression derived in Proposition 2:

$$\delta^*(q \mid B) = E(z_q w_q) = E(z_q \underline{y}) + E(z_q \mathbf{1}\{z_q > 0\}(\overline{y} - \underline{y})). \tag{B.4}$$

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. In this proof, we characterize D_f. It corresponds to the set of directions that are orthogonal to the exposed faces of B. We also characterize kink points of set B.

Characterization of D_f The first term on the RHS of equation (B.4) is linear in q since (see the previous proof) :

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

and thus is continuously differentiable on $\mathbb{S}_p$. 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 $q \in \mathbb{S}_p$. For any $t \in \mathbb{S}_p$:

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

so that:

$$\psi(t) - \psi(q) - E(z^* \mathbf{1}\{z^* q > 0\})(t-q) = E(z^* t(\mathbf{1}\{z^* t > 0\} - \mathbf{1}\{z^* q > 0\})). \tag{B.5}$$

Points of non differentiability depends on the expression in the RHS. 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, $\lim_{t\to q}\Pr(z^*q > 0, z^*t \leq 0) = 0$, 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. Its gradient is given by equation (B.5):

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

and is continuous in q.

Consider now the case in which $\Pr(z^*q = 0) > 0$. When $t \to q$, both in $\mathbb{S}_p$, define:

$$t - q = hs + o(h),$$

where $h = \|t-q\|$ and $s \in \mathbb{S}_p$, $s^Tq = 0$. We have

$$\begin{aligned} A_3 &= E(z^*t\mathbf{1}\{z^*q = 0, z^*t > 0\}) \\ &= E(z^*(t-q)\mathbf{1}\{z^*q = 0, z^*(t-q) > 0\}) \\ &= \Pr(z^*q = 0)E(z^*s1\{z^*s \geq 0\} \mid z^*q = 0)h + o(h) \end{aligned}$$

It follows that ψ has different gradients in different directions s, which depend on the term,

$$E(z^*1\{z^*s \geq 0\} \mid z^*q = 0).$$

This vector is constant for any s if and only if (using s and $-s$):

$$E(|z^*s| \mid z^*q = 0) = 0.$$

The support of z^* conditional on $(z^*q = 0)$ boils down to $\{0\}$, that is, if and only if the conditional support of z itself is $\{0\}$. This case is excluded by condition (R.ii) and therefore, function $\psi(q)$ is not differentiable.

Overall, 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 :

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

As $\delta^*(q \mid B) = q^\top \beta_q$, 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 :

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

which means (see equation (B.3)) that there exists $\Delta_f^*(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that (recall that $\beta_{q_f} = \left(E(z^\top x)\right)^{-1} E(z^\top w_{q_f})$):

$$\begin{aligned} \beta_f &= \beta_{q_f} + \left(E(z^\top x)\right)^{-1} E(z^\top \Delta_f^*(z) \mathbf{1}\{z_{q_f} = 0\}), \\ &= \beta_{q_f} + \left(E(z^\top x)\right)^{-1} E(z^\top \Delta_f^*(z) \mid z_{q_f} = 0) \Pr(z_{q_f} = 0). \end{aligned}$$

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

Conversely, suppose that there exists a direction q such that $\Pr(z_q = 0) > 0$ and suppose that the conditional support of z 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 $\Delta_f^*(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that:

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

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

Kinks Assume that $\Pr(z_q = 0) = 0$ so that the support function is differentiable and B is strictly convex. Even in this case, it is still possible to observe points $\beta_k \in \partial B$ where the tangent space is not unique (kinks), *i.e.*, points of the surface such that there exist at least two distinct vectors q and r ($r \neq q$) satisfying $\beta_k = \beta_q = \beta_r$. When there exist such points, the relationship between directions of the unit sphere and points of the frontier of *B* is not one-to-one anymore. It

complicates the construction of testing procedures (as shown in Section 4) and this is the reason why it is useful to characterize set-ups where B has kinks. We have,

$$\begin{aligned}
\beta_q &= \beta_r, \\
&\Leftrightarrow E(z^\top w_q) = E(z^\top w_r) \\
&\Leftrightarrow E(z^\top(\bar{y} - \underline{y})(\mathbf{1}\{z_q > 0\} - \mathbf{1}\{z_r > 0\})) = 0 \\
&\Leftrightarrow E(z^\top(\bar{y} - \underline{y})\left(\mathbf{1}\{z_q > 0, z_r < 0\} - \mathbf{1}\{z_q < 0, z_r > 0\}\right)) = 0,
\end{aligned}$$

the last equation holding because we have assumed that $\Pr(z_q = 0) = 0$.

Premultiplying last equation by $q^\top \left(E(z^\top x)\right)^{-1}$, we get,

$$\beta_q = \beta_r \Rightarrow E(z_q(\bar{y} - \underline{y})\left(\mathbf{1}\{z_q > 0, z_r < 0\} - \mathbf{1}\{z_q < 0, z_r > 0\}\right)) = 0.$$

Given that the term within the expectation is necessarily non negative, the fact that the expectation is zero implies necessarily

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

It follows that the existence of q and r ($r \neq q$) satisfying the latter condition is not only sufficient, but also necessary for the existence of kinks.

B.3 Proof of Lemma 4

We have already proven that conditions (9) and (10) are necessary, we want to prove that they are sufficient. Specifically, we suppose that conditions (9) and (10) hold true and we want to prove that

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

To prove this, we are going to show that z can be written as a linear combination of z_F and z_H. Note first that :

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

where Q_F is a $[m, p]$ matrix of rank p satisfying $Q_F^\top Q_F = I_p$ (where I_p is the identity matrix of dimension p). Second, denoting $A = \begin{pmatrix} 0 \\ I_{m-p} \end{pmatrix}$ the $[m, m-p]$ selection matrix, the definition of z^s implies,

$$z^s = zA = z\left(E(z^\top z)\right)^{-1/2} A^s,$$

where $A^s = \left(E(z^\top z)\right)^{1/2} A$.

Denoting $P_F = Q_F Q_F^\top$ and $P_H = I_m - P_F$, P_F and P_H are two orthogonal projections and we have,

$$\zeta^s = z^s - z_F E(z_F^\top z^s) = z\left(E(z^\top z)\right)^{-1/2}(I_m - P_F)A^s = z\left(E(z^\top z)\right)^{-1/2} P_H A^s,$$

which implies,

$$z_H = \zeta^s(\zeta^{s\top}\zeta^s)^{-1/2} = z\left(E(z^\top z)\right)^{-1/2} P_H A^s (A^{s\top} P_H A^s)^{-1/2} = z\left(E(z^\top z)\right)^{-1/2} Q_H,$$

where $Q_H = P_H A^s (A^{s\top} P_H A^s)^{-1/2}$ is a matrix of dimension $[m, m-p]$ of rank $(m-p)$ satisfying $Q_H^\top Q_H = I_{m-p}$ and $Q_F^\top Q_H = 0$ (as a matrix).

Overall, the relationship between (z_F, z_H) and z boils down to,

$$(z_F, z_H) = z\left(E(z^\top z)\right)^{-1/2}(Q_F, Q_H) = z\left(E(z^\top z)\right)^{-1/2} Q,$$

where the $[m, m]$ matrix $Q = (Q_F, Q_H)$ satisfies $Q^\top Q = I_m$ and hence has full rank. Hence z may be written $(z_F, z_H)Q^{-1}\left(E(z^\top z)\right)^{1/2}$ *i.e.*, a linear combination of z_F and z_H. In such a case, conditions (9) and (10) implies

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

which finishes the proof.

We can now show that the choice of z^s among z is without loss of generality. Suppose that z_H associated with a given subset of supernumerary instruments z^s satisfies condition (10). Then, B is non empty because condition (10) is sufficient. Yet, if B is non empty and since condition (10) is necessary, condition (10) 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 (10) is both necessary and sufficient for the condition that B is not empty, when it is satisfied by a given subset of supernumerary instruments, it is necessarily satisfied by any alternative subsets.

There is another interesting way to see why restrictions involved by condition (10) are invariant to the choice of the specific subset of supernumerary instruments. As discussed above z_H can be written as $z\left(E(z^\top z)\right)^{-1/2} Q_H$ 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 $Q_H^* = Q_H R$, where R is an orthogonal matrix). In other words, for any z_H^* satisfying the same conditions as z_H, there exists necessarily an orthogonal matrix R (with $R = Q_H^\top Q_H^*$) such that $z_H^* = z_H R$. This basic linear relationship between all possible subsets of supernumerary instruments implies that when linear moment condition (10) 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 $\delta^*(x_1^* \mid B_U)$ where $x_1^* = (q_1, \lambda_1)$. The support function of $\{\gamma = 0\}$ is as follows if $x_2^* = (q_2, \lambda_2)$:

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

Corollary 16.4.1 page 146 of Rockafellar (1970) states that the support function $\delta^*(x^*)$, where $x^* = (q, \lambda)$, of the intersection of two convex sets such that their relative interiors[11] have one point in common, can be written:

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

and the infimum is attained.

Therefore, when the hyperplane $\{\gamma = 0\}$ is not tangent to B_U and their intersection is not empty, their relative interiors have all the points of the relative interior of their intersection in common, and we have:

$$\delta^*((q, \lambda) \mid B) = \inf_{(\lambda_1,\lambda_2):\lambda_1+\lambda_2=\lambda} \delta^*((q, \lambda_1) \mid B_U)) = \inf_{\lambda_1} \delta^*((q, \lambda_1) \mid B_U))$$

as the RHS is independent of λ_2 and λ. Furthermore, the infimum in λ_1 is attained.

On the other hand, when the hyperplane $\{\gamma = 0\}$ is tangent to B_U, the relative interiors have no points in common since all intersection points belong to the closure of B_U. The same corollary 16.4.1 of Rockafellar (1970) nonetheless states that we should replace equation (B.6) by its closure although the infimum is not necessarily attained.

Specifically, the condition under which the hyperplane $\{\gamma = 0\}$ is tangent to B_U, is obtained when the origin point belongs to the frontier of B_{Sargan} (see Section 3.2.3). Without loss of

[11] 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:

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

generality, suppose that B_U is included in the half-space $\gamma \geq 0$ (*i.e.* γ is a completely positive vector) by changing signs of sub-parameters of γ if necessary. We consider now two cases.

In the first case in which the support function is differentiable, there are no exposed faces and the tangency of the hyperplane $\{\gamma = 0\}$ to B_U results in a single intersection point. Set B is reduced to a point and is not a proper set any longer. Let $(\beta_I, 0)$ be the intersection point and consider one hyperplane which is tangent to set B_U at this point. If there is a kink of set B_U at this point, there exist many hyperplanes tangent to set B_U. In any case, choose one and denote (q_I, λ_I) its normal vector oriented outward set B_U and where λ_I could be infinite (recall that $q_I \in \mathbb{S}_p$). For any value $\lambda \leq \lambda_I$ we have $\forall(\beta, \gamma) \in B_U$:

$$\begin{aligned} q_I\beta + \lambda\gamma &\leq q_I\beta + \lambda_I\gamma \ \text{ (as } \gamma \geq 0), \\ &\leq q_I\beta + \lambda_I 0 \text{ as } \delta(q_I, \lambda_I) = (q_I, \lambda_I)(\beta_I, 0)^\top, \\ &= q_I\beta + \lambda 0. \end{aligned}$$

The support function for (q_I, λ) is also equal to $\delta((q_I, \lambda_I) \mid B_U)$.If λ_I is finite, the minimum is attained for any $\lambda \leq \lambda_I$.

If set B_U is smooth at $(\beta_I, 0)$, there is only one tangent space at $(\beta_I, 0)$ and this is only possible if $\lambda_I = -\infty$. Consequently, the infimum of $\delta^*((q_I, \lambda) \mid B_U)$ is not attained. Else if set B_U is not smooth, the infimum can be attained at a finite λ_I.

In the second case in which the support function is not differentiable, there are exposed faces and set B is a proper set. Depending on the smoothness of B_U at the frontier points of the intersection with the hyperplane $\gamma = 0$, the previous discussion can be extended to see whether the infimum of $\delta^*((q_I, \lambda) \mid B_U)$ 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

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

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

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

The last matrix 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:

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

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 $Q^\top \lambda$ is equivalent.

C Proofs in Section 4

We denote M a generic majorizing constant.

C.1 Proof of Proposition 9

We use that:

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

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

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

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:

$$\hat{\Sigma}_n^u = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^\top z_i\right)^{-1}, \tag{C.7}$$

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

$$\begin{cases} \hat{\Sigma}_n = \hat{\Sigma}_n^u & \text{if } \left\|\hat{\Sigma}_n^u\right\| \leq M, \\ \hat{\Sigma}_n = \hat{\Sigma}_n^u \left(\frac{M}{\left\|\hat{\Sigma}_n^u\right\|}\right) & \text{if not.} \end{cases} \tag{C.8}$$

The element $(q, \hat{\Sigma}_n)$ always belongs to the bounded set $\Theta = \mathbb{S}_p \times \{\|\Sigma\| \leq M\}$. Under the conditions of Proposition 8 , $\hat{\Sigma}_n$ is almost surely consistent to Σ:

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

Under the conditions of Proposition 9, $\hat{\Sigma}_n^u$ and $\hat{\Sigma}_n$ are asymptotically equivalent:

$$\sqrt{n}\left(\hat{\Sigma}_n - \hat{\Sigma}_n^u\right) \underset{n \to \infty}{\overset{P}{\to}} 0, \tag{C.9}$$

and the estimate is asymptotically normal:

$$\sqrt{n}\left(vec(\hat{\Sigma}_n^\top - \Sigma^\top)\right) \Longrightarrow N(0, W). \tag{C.10}$$

We proceed in two steps. As the first step is simple, we give the proof 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 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, this function is convex in Σq and therefore Lipschitzian:

$$\begin{aligned} \left| f_{\theta_1}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_2}(z_i, \underline{y}_i, \overline{y}_i) \right| &\leq \max(\left\| z_i^\top \underline{y}_i \right\|, \left\| z_i^\top \overline{y}_i \right\|) \left\| q_1^\top \Sigma_1^\top - q_2^\top \Sigma_2^\top \right\|, \\ &\leq M \max(\left\| z_i^\top \underline{y}_i \right\|, \left\| z_i^\top \overline{y}_i \right\|) \left\| \theta_1 - \theta_2 \right\|. \end{aligned} \tag{C.11}$$

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

Under conditions $R.iii$, we have:

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

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:

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

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, uniformly in Θ, 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 Supplementary Appendix since this result is already shown in Beresteanu and Molinari (2008). We will rely heavily on Section 19.4 of van der Vaart (1998) where relevant properties are proposed.

C.1.2 Asymptotic distribution when Σ is estimated

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

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

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

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

where $E(z_{n,qi}w_{n,qi})$ is evaluated along a specific sequence $\{\hat{\Sigma}_n\}_{n\geq 1}$ and the expectation operator is taken with respect to the probability measure of z_i, $\underline{y}_i$ and $\overline{y}_i$ (see Section 19.4 of van der Vaart, 1998).

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

$$E(z_{n,qi}w_{n,qi} - z_{qi}w_{qi})^2 = E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_\theta(z_i, \underline{y}_i, \overline{y}_i))^2 \underset{n\to\infty}{\overset{P}{\to}} 0. \tag{C.12}$$

Using equation (C.11), we have:

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

so that:

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

Under the conditions of Proposition 9, $E\left(\max(\left\|z_i^\top \underline{y}_i\right\|, \left\|z_i^\top \overline{y}_i\right\|)^2\right) < \infty$ and is independent of q. As $\left\|\hat{\theta}_n - \theta\right\|^2$ tends in distribution to 0 uniformly in $q \in \mathbb{S}_p$ (equation (C.9)), it tends also in probability to 0, uniformly in $q \in \mathbb{S}_p$, 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:

$$C_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} z_{qi}w_{qi} - E(z_{qi}w_{qi})\right). \tag{C.13}$$

uniformly in $q \in S$. Therefore the problem boils down to compute the limit of processes $B_n(q)$ and $C_n(q)$ as given in the following lemma:

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

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

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

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

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

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

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 $z_{qi} = q^\top \Sigma^\top z_i^\top$ and as we are evaluating these expressions along a specific sequence $\{\hat{\Sigma}_n\}_{n\geq 1}$, the second term on the RHS is equal to:

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

using the definition of β_q^*.

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

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

The first line is the sum of two terms:

$$\begin{aligned} B_n^{11}(q) &= \sqrt{n}E\left((z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i)\left(\mathbf{1}\{z_{n,qi} > 0\} - \frac{1}{2}\right)\mathbf{1}\{z_{qi} = 0\}\right), \\ B_n^{12}(q) &= \sqrt{n}E\left((z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i)\left(\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\}\right)\mathbf{1}\{z_{qi} \neq 0\}\right). \end{aligned}$$

and the second line is equal to:

$$B_n^{13}(q) = \sqrt{n}E\left(z_{qi}(\bar{y}_i - \underline{y}_i)(\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\})\right).$$

We shall prove that $B_n^{12}(q)$ and $B_n^{13}(q)$ are bounded from above by $o_P(1)$ terms.

Considering $B_n^{13}(q)$ first, use the Cauchy-Schwartz inequality and write:

$$\left|B_n^{13}(q)\right| \leq \sqrt{n}\left[E(\bar{y}_i - \underline{y}_i)^2\right]^{1/2}\left[E\left(z_{qi}^2\left|\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\}\right|\right)\right]^{1/2},$$

since squares of dummy variables are equal to themselves. Denote generic $M = \left[E(\bar{y}_i - \underline{y}_i)^2\right]^{1/2}$ and write:

$$\begin{aligned} \left|B_n^{13}(q)\right| &\leq \sqrt{n}M\left[E\left(z_{qi}^2\left|\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\}\right|\right)\right]^{1/2}, \\ &\leq \sqrt{n}M\left[E\left(z_{qi}^2 E(\left|\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\}\right| \mid z_{qi})\right)\right] \\ &\leq \sqrt{n}M\left[E\left(z_{qi}^2 \Pr(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right| \geq \left|\sqrt{n}z_{qi} \mid z_{qi}\right|)\right)\right]^{1/2}, \end{aligned}$$

since the alternation in signs between $z_{n,qi}$ and z_{qi} means that $\sqrt{n}(z_{n,qi} - z_{qi})$ is bounded further away from zero when n increases.

As the number of mass points is finite, there exists a finite $\alpha > 0$ such that there is no mass point between $z_q = 0$ (excluded) and $(z_q)^2 = \alpha$ and such that the density function of $(z_q)^2$ between these two values is bounded. Write the upper bound on $\left(B_n^{13}(q)\right)^2$ as the sum of two terms:

$$\begin{aligned} &nM^2E\left(z_{qi}^2 \Pr(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right| > \left|\sqrt{n}z_{qi}\right| \mid z_{qi}) \mid z_{qi}^2 \leq \alpha)\right) \Pr(z_{qi}^2 \leq \alpha), \\ &nM^2E\left(z_{qi}^2 \Pr(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right| > \left|\sqrt{n}z_{qi}\right| \mid z_{qi}) \mid z_{qi}^2 > \alpha)\right) \Pr(z_{qi}^2 > \alpha). \end{aligned} \tag{C.14}$$

Using the conditions in Proposition 9, consider $0 < \mu < \min(2, \gamma)$ so that $E(\left\|x^\top z\right\|^{2+\mu}) < \infty$. We also have,

$$\left\|\sqrt{n}(\hat{\Sigma}_n - \Sigma)\right\|^{2+\mu} = O_P(1).$$

Using:

$$\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu} \leq \left\|\sqrt{n}(\hat{\Sigma}_n - \Sigma)\right\|^{2+\mu} \|z_i\|^{2+\mu},$$

and the same conditions in Proposition 9, there exists $M_n = O_P(1)$ such that:

$$\sup_{z_{qi}^2 \leq \alpha} E(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu} \mid z_{qi}) \leq M_n \text{ and } E(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu}) \leq M_n.$$

Use Markov inequality with exponent $2 + \mu$ to write:

$$\Pr(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right| > \left|\sqrt{n}z_{qi}\right| \mid z_{qi}) \leq \frac{E(|\sqrt{n}(z_{n,qi} - z_{qi})|^{2+\mu} \mid z_{qi})}{|\sqrt{n}z_{qi}|^{2+\mu}}$$

so that the first line of equation (C.14) is bounded by:

$$\begin{aligned} &(\sqrt{n})^{-\mu}M^2 \int_0^\alpha E(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu} \mid z_{qi}) \left|z_{qi}\right|^{-\mu} \Pr(d\left(z_{qi}\right)^2) \\ \leq &(\sqrt{n})^{-\mu}M^2 \sup_{(z_{qi})^2 \leq \alpha} E(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu} \mid z_{qi}) \int_0^\alpha \left|z_{qi}\right|^{-\mu} d\left(z_{qi}\right)^2, \end{aligned}$$

as the density of $(z_{qi})^2$ is bounded on $(0, \alpha]$. The last term can then be written as:

$$(\sqrt{n})^{-\mu}M^2M_n \left[\frac{(z_{qi})^{2-\mu}}{1 - \mu/2}\right]_0^\alpha = (\sqrt{n})^{-\mu}M^2M_n \frac{\alpha^{1-\mu/2}}{1 - \mu/2},$$

which is $o_P(1)$.

Moreover, using the same Markov inequality, the second line is bounded by:

$$\begin{aligned} &(\sqrt{n})^{-\mu}M^2 \int_\alpha^{+\infty} E(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu} \mid z_{qi}) \left|z_{qi}\right|^{-\mu} \Pr(d\left(z_{qi}\right)^2) \\ \leq &(\sqrt{n})^{-\mu} \frac{M^2}{\alpha^{\mu/2}} E(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right|^{2+\mu} \mid z_{qi}^2 > \alpha) \end{aligned}$$

which is $o_P(1)$. This proves that $\left(B_n^{13}(q)\right)^2$ is bounded by a $o_P(1)$ term.

As for $B_n^{12}(q)$ first we can use the Cauchy-Schwartz inequality to show that:

$$\left|B_n^{12}(q)\right| < \left(E\left[\sqrt{n}(z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i)\right]^2\right)^{1/2} \left(E(|\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\}|\,\mathbf{1}\{z_{qi} \neq 0\}\right)^{1/2}.$$

Since $z_{n,qi} - z_{qi} = q^T(\Sigma_n^T - \Sigma^T)z_i^T$, the first term in the product is bounded by:

$$\left\|\sqrt{n}(\Sigma_n^T - \Sigma^T)\right\| \left(E\left[z_i^T(\bar{y}_i - \underline{y}_i)\right]^2\right)^{1/2} = O_P(1),$$

because all variables are in L^2. The second term is bounded by:

$$\Pr(\left|\sqrt{n}(z_{n,qi} - z_{qi})\right| \geq \left|\sqrt{n}z_{qi}\right|, z_{qi} \neq 0) = o_P(1)$$

using a similar proof as in the above proof for $B_n^{13}(q)$. We thus have $B_n^{12}(q) = o_P(1)$.

Therefore:

$$\begin{aligned} B_n^1(q) &= B_n^{11}(q) + o_P(1) \\ &= \sqrt{n}E\left(\frac{|z_{n,qi} - z_{qi}|}{2}(\bar{y}_i - \underline{y}_i)\mathbf{1}\{z_{qi} = 0\}\right) + o_P(1) \\ &= \sqrt{n}E\left(\left|q^\top(\Sigma_n^\top - \Sigma^\top)z_i^\top\right|(\bar{y}_i - \underline{y}_i)\mathbf{1}\{z_{qi} = 0\}\right)/2 + o_P(1). \end{aligned}$$

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

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

The third line uses that $\sqrt{n}(\hat{\Sigma}_n^u - \hat{\Sigma}_n) \xrightarrow[n\to\infty]{P} 0$ by equation (C.9) and uniform bounds on q, Σ and β_q^*. Moreover, as $\hat{\Sigma}_n$ is bounded and its inverse exists, $\Sigma^\top(\hat{\Sigma}_n^\top)^{-1} \xrightarrow[n\to\infty]{a.s} I$, and we have, uniformly in q:

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

■

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

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

If there are no exposed faces (*i.e.*, $\Pr(z_i\Sigma q = 0) = 0$), the second term is identically equal to zero whereas ε^*_{qi} boils down to the residual of the IV regression of w_q on to x using instruments z so that $\tau_n(q)$ converges in distribution, uniformly in q, to a Gaussian process centered at zero and of covariance function:

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

with $\varepsilon_{qi} = w_{qi} - x_i\beta_q$.

Suppose that there exist exposed faces ($\Pr(z_{qi}=0) > 0$). Write:

$$\Sigma q = (I_p \otimes q^\top)vec(\Sigma^\top)$$

so that, using the asymptotic normality of the estimate of $vec(\Sigma^\top)$ in equation (C.10) we have:

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

where η is a multivariate standard normal random variable of dimension p^2 independent of z_i.

C.2 Proof of Proposition 10

When β_0 is outside (resp. inside) set B but not on the frontier, we know that $\inf_q T_\infty(q)$ is strictly negative (resp. positive). As $T_n(q)$ converges uniformly in q to $T_\infty(q)$, $\min_q T_n(q)$ is negative (resp. positive) and bounded away from zero for n sufficiently large. $\sqrt{n}T_n(q_n)$ tends therefore to $-\infty$ (resp. $+\infty$).

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

C.2.1 $\mathcal{Q}(\beta_0)$ is a singleton: $\mathcal{Q}(\beta_0) = \{q_0\}$

As $\delta^*(q \mid B)$ is differentiable (assumption D), the empirical stochastic process defined for $q \in \mathbb{S}_p$ as,

$$\sqrt{n}\left(T_n(q;\beta_0) - T_\infty(q;\beta_0)\right) = \sqrt{n}(\hat{\delta}^*_n(q \mid B) - \delta^*(q \mid B)) = \tau_n(q),$$

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

Let $q_n \in \mathbb{S}_p$ 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 well separated minimum, imply that:

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

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

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

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

C.2.2 $\mathcal{Q}(\beta_0)$ is not a singleton

The proof proceeds in various steps:

1. We select and characterize a unique q_0^* from $\mathcal{Q}(\beta_0)$.
2. We construct a sequence of well separated minima of minimization programs which tends to q_0^*.
3. We show that any sequence of minimizers of the empirical programs converge to q_0^*.

1. The selection of a single $q_0^* \in \mathcal{Q}(\beta_0)$ For this, we select a vector oriented outwards B and consider its projection on the smallest convex cone which includes $\mathcal{Q}(\beta_0)$:

$$\mathcal{C}(\beta_0) = \{\lambda q_0; q_0 \in \mathcal{Q}(\beta_0), \lambda \geq 0\} = \{v; \delta^*(v \mid B) - v^\top \beta_0 \leq 0\}.$$

The vector oriented outward set B can be constructed as the difference between β_0 which is a frontier point of B and any interior point β^* of B. For instance the "center" of B obtained by setting $u(z) = \frac{\underline{\Delta}(z)+\overline{\Delta}(z)}{2}$) is interior and

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

Denote $v_0 = \beta_0 - \beta^* \neq 0$ and note that, as $\beta^* \in int(B)$, we have for all q_0 in $\mathcal{Q}(\beta_0)$:

$$\delta^*(q_0 \mid B) - q_0^\top \beta^* > 0 \Longrightarrow\ q_0^\top v_0 > 0. \tag{C.15}$$

The projection of v_0 on the convex cone $\mathcal{C}(\beta_0)$ is given by:

$$\min_{v,\delta^*(v|B)-v^\top\beta_0\leq 0} \frac{(v_0-v)^\top(v_0-v)}{2} \tag{C.16}$$

This projection is unique and defined by $v_0^* = \lambda^* q_0^*$ where (λ^*, q_0^*) is the argument of the minimum:

$$\min_{(\lambda\geq 0, q\in\mathcal{Q}(\beta_0))} (v_0-\lambda q)^\top(v_0-\lambda q) \propto \min_{(\lambda\geq 0, q\in\mathcal{Q}(\beta_0))} \{-2\lambda q^\top v_0 + \lambda^2\}$$

which yields $\lambda^* = q_0^{*\top} v_0 > 0$ (see equation C.15) whereas q_0^* is the argument of:

$$\max_{q\in\mathcal{Q}(\beta_0)} q^\top v_0.$$

Vector q_0^* is unique because it is a (normalized) projection. Furthermore, when $\frac{v_0}{\|v_0\|} \in \mathcal{Q}(\beta_0)$ (or equivalently $v_0 \in \mathcal{C}(\beta_0)$), we have $q_0^* = \frac{v_0}{\|v_0\|}$ whereas in other cases q_0^* belongs to the frontier of $\mathcal{Q}(\beta_0)$.

2. Minimization programs whose well separated solutions converges to q_0^* The estimation of q_0^* cannot proceed directly from program (C.16) since we do not know the set of constraints, $\mathcal{Q}(\beta_0)$. Consider the generalization of (C.16) for any $\alpha \geq 0$:

$$b(\alpha) \equiv \min_{v, \delta^*(v|B) - v^\top \beta_0 \leq \alpha} \frac{(v_0 - v)^\top (v_0 - v)}{2} \tag{C.17}$$

where $b(\alpha)$ is continuous and non increasing in α because the constraint is continuous. The unique solution of this program, denoted v_α^*, is the projection of $v_0 = \beta_0 - \beta^*$ on the convex cone $\{v \in \mathbb{R}^p, \delta^*(v \mid B) - v^\top \beta_0 \leq \alpha\}$.

We state a sequence of Lemmas that are proved below in Section C.2.3.

It turns out that the following equivalent characterization to this program will be more amenable to estimation.

Lemma 14 *For any $\alpha > 0$, the strictly convex program (C.17) is equivalent to the minimization of:*

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

where a is an increasing function of α.

This equivalence covers the case where $a > 0$. We need to complete this result by showing how the minimizer q_a of $\Psi_a(q)$ converges to q_0^* when $a \to 0$.

Lemma 15 *The limit of the sequence $\{q_a\}_{a>0}$ exists when $a \to 0$ and is equal to q_0^*. Furthermore:*

$$\Psi_a(q_0^*) - \Psi_a(q_a) = o(a).$$

Moreover, we have the following uniform result:

Lemma 16

$$\forall \varepsilon > 0, \exists a_0 > 0, \ \exists \eta > 0 \text{ such that } \inf_{0 < a \leq a_0, \left\| q - q_0^* \right\| \geq \varepsilon} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} > \eta. \tag{C.18}$$

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

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

where $v_{0,n} = \beta_0 - \hat{\beta}_n^*$ and $\hat{\beta}_n^* = \dfrac{1}{n} \sum_{i=1}^n \hat{\Sigma}_n^\top z_i \dfrac{\bar{y}_i + \underline{y}_i}{2}$.

Define $q_{n,a}$ as a near minimizer of $\Psi_{n,a}$:

$$\Psi_{n,a}(q_{n,a}) \leq \inf_q \Psi_{n,a}(q) + O_P(n^{-1/2}).$$

We have

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

whereas the square-root uniform convergence of $\Psi_{n,a}$ to Ψ_a ensures that:

$$\Psi_{n,a}(q_{n,a}) = \Psi_a(q_{n,a}) + O_P(n^{-1/2}).$$

Using successively the last equality and the previous inequality, we canwrite,

$$\begin{aligned} 0 \leq \Psi_a(q_{n,a}) - \Psi_a(q_a) &= \Psi_{n,a}(q_{n,a}) - \Psi_a(q_a) + O_P(n^{-1/2}), \\ &\leq \Psi_{n,a}(q_a) - \Psi_a(q_a) + O_P(n^{-1/2}), \\ &\leq \sup_q |\Psi_a(q) - \Psi_{n,a}(q)| + O_P(n^{-1/2}). \end{aligned}$$

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 sequence 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:

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

Then:

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

and therefore:

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

By condition (C.18), for any $\varepsilon > 0$, there exist n_0 and $\eta > 0$ such that, for $n \geq n_0$ the event:

$$\{d(q_{n,a_n}, q_0^*) \geq \varepsilon\} \subset \{\frac{\Psi_{a_n}(q_{n,a_n}) - \Psi_{a_n}(q_{a_n})}{a_n} > \eta\}.$$

Therefore:

$$\forall \varepsilon > 0, \lim_{n\to\infty} \Pr(d(q_{n,a_n}, q_0^*) \geq \varepsilon) = 0 \Longrightarrow q_{n,a_n} - q_0^* \xrightarrow[n\to\infty]{P} 0.$$

To finish the proof of Proposition 10 we can now use the same argument than in Section C.2.1 so that:

$$\sqrt{n}\left(T_n(q_{n,a_n}; \beta_0) - T_n(q_0^*; \beta_0)\right) \xrightarrow[n\to\infty]{P} 0.$$

The variance of $T_n(q_0^*; \beta_0)$ is estimated as the variance of $T_n(q_{n,a_n}; \beta_0)$.

C.2.3 Proofs of Lemma 14 to 16

Proof of Lemma 14: Let $\alpha_0 = \delta^*(v_0 \mid B) - v_0^\top \beta_0$, we have $v_\alpha^* = v_0$ for any $\alpha \geq \alpha_0$ whereas in other cases the optimal solution v_α^* is such that the constraint is binding, $\delta^*(v_\alpha^* \mid B) - v_\alpha^{*\top} \beta_0 = \alpha$. If $\frac{v_0}{\|v_0\|} \in \mathcal{Q}(\beta_0)$, we have that $\alpha_0 = 0$ and

$$q_0^* = \frac{v_\alpha^*}{\|v_\alpha^*\|} = \frac{v_0}{\|v_0\|}, \ \forall \alpha \geq 0.$$

When $\frac{v_0}{\|v_0\|} \notin \mathcal{Q}(\beta_0)$ and α runs from 0 to α_0, v_α^* describes a trajectory between v_0^* and v_0. We now characterize this trajectory.

It is easier to work with the equivalent dual program (Rockafellar, 1970):

$$\alpha = \min_{v, \frac{(v_0-v)^\top(v_0-v)}{2} \leq b(\alpha)} (\delta^*(v \mid B) - v^\top \beta_0), \tag{C.19}$$

where $b(\alpha)$ runs from $\frac{(v_0-v_0^*)^\top(v_0-v_0^*)}{2}$ to 0 to generate the same trajectory $\{v_\alpha^*\}_{\alpha \geq 0}$. Writing the program (C.19) as the Lagrangian where $a > 0$:

$$L(v, a) = \delta^*(v \mid B) - v^\top \beta_0 + a(\frac{(v_0 - v)^\top (v_0 - v)}{2} - b(\alpha)) \tag{C.20}$$

we obtain the first order condition (by assumption D, $\delta^*(v \mid B)$ is differentiable):

$$\beta_{q_\alpha} - \beta_0 - a(\alpha)(v_0 - v_\alpha^*) = 0,$$

where $q_\alpha = \frac{v_\alpha^*}{\|v_\alpha^*\|} \in \mathbb{S}_p$ and $\beta_{q_\alpha} = \left.\frac{\partial \delta^*(v|B)}{\partial v}\right|_{v_\alpha^*}$. To obtain a, multiply the equation by $(v_0 - v_\alpha^*)^\top$:

$$2a(\alpha)b(\alpha) = (v_0 - v_\alpha^*)^\top (\beta_{q_\alpha} - \beta_0).$$

When $\alpha = 0$, $\beta_{q_\alpha} = \beta_0$ and therefore $a(\alpha) = 0$ since $\frac{v_0}{\|v_0\|} \notin \mathcal{Q}(\beta_0)$ and $b(\alpha) > 0$. Furthermore, $a(\alpha)$ is continuous in α for any $\alpha < \alpha_0$ since all objects in the expression are continuous.

We now prove that $a(\alpha)$ is increasing with α. Consider $0 < \alpha < \alpha' < \alpha_0$ and the optimal solutions v_α^* and $v_{\alpha'}^*$, where $v_\alpha^* \neq v_{\alpha'}^*$ because:

$$\delta^*(v_\alpha^* \mid B) - v_\alpha^{*\top} \beta_0 = \alpha < \delta^*(v_{\alpha'}^* \mid B) - v_{\alpha'}^{*\top} \beta_0 = \alpha'.$$

Note that by optimality:

$$L(v_\alpha^*, a(\alpha)) = \delta^*(v_\alpha^* \mid B) - v_\alpha^{*\top} \beta_0 < \delta^*(v_{\alpha'}^* \mid B) - v_{\alpha'}^{*\top} \beta_0 + a(\alpha)(\frac{(v_0 - v_{\alpha'}^*)^\top (v_0 - v_{\alpha'}^*)}{2} - b(\alpha)),$$

$$L(v_\alpha^*, a(\alpha')) = \delta^*(v_\alpha^* \mid B) - v_\alpha^{*\top} \beta_0 + a(\alpha')(\frac{(v_0 - v_\alpha^*)^\top (v_0 - v_\alpha^*)}{2} - b(\alpha')) > \delta^*(v_{\alpha'}^* \mid B) - v_{\alpha'}^{*\top} \beta_0.$$

so that by differencing:

$$a(\alpha')(b(\alpha) - b(\alpha')) > -a(\alpha)(b(\alpha') - b(\alpha)) \Rightarrow (a(\alpha') - a(\alpha))\,(b(\alpha) - b(\alpha')) > 0.$$

As $b(\alpha)$ is non increasing, it implies that $a(\alpha)$ is increasing with α from $a(0) = 0$ to $\lim_{\alpha \to \alpha_0} a(\alpha) = +\infty$.

We can thus generates the arc $\{v_\alpha^*\}_{\alpha > 0}$ equivalently by making a varies between 0 and ∞. Let us rewrite the minimization program (C.20) in order to consider vectors on $\mathbb{S}_p$ since estimates are defined on $\mathbb{S}_p$ only:

$$\begin{aligned} L(\lambda q, a) &= \delta^*(\lambda q \mid B) - (\lambda q)^\top \beta_0 + a(\frac{(v_0 - \lambda q)^\top (v_0 - \lambda q)}{2} - b(\alpha)), \\ &= \lambda(\delta^*(q \mid B) - q^\top \beta_0) + a(\frac{(v_0 - \lambda q)^\top (v_0 - \lambda q)}{2} - b(\alpha)). \end{aligned}$$

Minimizing wrt λ yields the FOC for the optimal solution λ_q:

$$\delta^*(q \mid B) - q^\top \beta_0 + a(\lambda_q - q^\top v_0) = 0,$$

which implies that:

$$\begin{aligned} L(\lambda_q q, a) &= a(-\frac{\lambda_q^2}{2} - b(\alpha)), \\ -a\lambda_q &= \delta^*(q \mid B) - q^\top \beta_0 - aq^\top v_0 \equiv \Psi_a(q). \end{aligned}$$

When $a > 0$, minimizing $L(\lambda_q q, a)$ is equivalent to maximizing λ_q and thus equivalent to minimizing $\Psi_a(q)$. As $L(\lambda_q q, a)$ is a strictly convex program, the minimizer of $\Psi_a(q)$ is unique and well separated.

Proof of Lemma 15: To begin with, it is useful to note that $-a \|v_0\|$ provides a lower bound of $\Psi_a(q)$,

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

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

We are going to consider in turn two cases:

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

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

  $$\Psi_a(q_a) = \delta^*(q_a \mid B) - q_a^\top \beta_0 - aq_a^\top v_0 \leq \Psi_a(q_0^*) = -aq_0^{*\top} v_0,$$

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

  $$0 \leq \delta^*(q_a \mid B) - q_a^\top \beta_0 \leq a(q_a - q_0^*)^\top v_0 \leq 2a \|v_0\|, \tag{C.21}$$

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

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

  and the distance between set $\mathcal{Q}(\beta_0)$ and q_a tends to zero by continuity of the 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, $\forall \eta > 0$, $\exists a_0 > 0$ such that $\forall a < a_0$, $\|q_a - q_m\| < \eta$.[12] 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

  $$\frac{\Psi_a(q_a)}{a} \leq \frac{\Psi_a(q_0^*)}{a} = -q_0^{*\top} v_0 \leq -q_m^\top v_0.$$

[12] Such a sequence exists as the distance between q_a and $\mathcal{Q}(\beta_0)$, a compact set, tends to zero. In the following we will work with a instead of working with a sequence indexed by a without loss of generality.

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 $q^\top v_0$ 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,

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

By taking limits and using that q_a tends to q_m when a tends to zero, we obtain that $q_m^\top v_0 = q_0^{*\top} v_0$. Given the definition of q_0^*, it means that q_m is the argument of $\max_{q \in Q(\beta_0)} q^\top v_0$. But this argument is unique and is precisely q_0^*. Hence, we have necessarily $q_m = q_0^*$ and therefore:

$$\lim_{a \to 0} \|q_a - q_0^*\| = 0. \tag{C.22}$$

Furthermore, as:

$$0 \leq \frac{\Psi_a(q_0^*) - \Psi_a(q_a)}{a} \leq (q_a - q_0^*)^\top v_0$$

we have:

$$\Psi_a(q_0^*) - \Psi_a(q_a) = o(a). \tag{C.23}$$

Proof of Lemma 16: First, the Lemma is trivially satisfied when $\frac{v_0}{\|v_0\|} \in \mathcal{Q}(\beta_0)$ since $q_a = q_0^* = \frac{v_0}{\|v_0\|}$ and therefore:

$$\frac{\Psi_a(q) - \Psi_a(q_a)}{a} \geq -(q - \frac{v_0}{\|v_0\|})^\top v_0 = \frac{1}{2}\|v_0\| \left\| q - \frac{v_0}{\|v_0\|} \right\|^2,$$

the last equality resulting from the following expansion:

$$\|q\|^2 = 1 = \left\| q - \frac{v_0}{\|v_0\|} + \frac{v_0}{\|v_0\|} \right\|^2 = 1 + 2(q - \frac{v_0}{\|v_0\|})^\top \frac{v_0}{\|v_0\|} + \left\| q - \frac{v_0}{\|v_0\|} \right\|^2. \tag{C.24}$$

Consequently, this quantity is bounded from below by a positive number when $\left\| q - \frac{v_0}{\|v_0\|} \right\| \geq \varepsilon$.

Assume now that $\frac{v_0}{\|v_0\|} \notin \mathcal{Q}(\beta_0)$. We will first show that, for a given q, the infimum if it is attained when a tends to zero is strictly positive. Using the results of Lemma (15), we know that when $a \to 0$, $q_a \to q_0^*$ and $\frac{\Psi_a(q_a)}{a} \to -q_0^{*\top} v_0$.

- Either $q \in \mathcal{Q}(\beta_0)$ and $\frac{\Psi_a(q)}{a} = -q^\top v_0 \geq -q_0^{*\top} v_0$, by construction of q_0^*. Consequently,

  $$\frac{\Psi_a(q) - \Psi_a(q_a)}{a} \underset{a \to 0}{\to} -(q - q_0^*)^\top v_0,$$

  which is strictly positive when $\|q - q_0^*\| \geq \varepsilon$.

- Or $q \notin \mathcal{Q}(\beta_0)$. In this case $\frac{\Psi_a(q)}{a} \to +\infty$ and cannot deliver the infimum.

As q_a tends to q_0^* when a tends to zero, there exists some a_0 for which the joint events $\{0 < a \leq a_0\}$ and $\{\|q - q_0^*\| \geq \varepsilon\}$ imply that $\|q - q_a\| \geq \frac{\varepsilon}{2}$. Assume now by contradiction that the infimum over $0 < a \leq a_0$ is not positive. By continuity of function $\frac{\Psi_a(q) - \Psi_a(q_a)}{a}$ in a and q when $a > 0$ (see Lemma 14), and as the infimum is positive at the limit $a \to 0$, a non-positive infimum can only be obtained at some $a > 0$. It is a contradiction because q_a is a well separated minimum for any $a > 0$ (Lemma 14).

The infimum in $0 \leq a \leq a_0$ is therefore positive for any q such that $q \in \mathbb{S}_p \cap \{\|q - q_0^*\| \geq \varepsilon\}$. The last set is a compact set in q. The infimum over such qs is thus positive also.

C.3 Proof of Proposition 12

By condition S, the relative interiors of sets B_U and $\{\gamma = 0\}$ have points in common and the infimum is attained at $\lambda_0(q)$ (see end of proof, Section B.4). As $\mathbb{S}_p$ is compact, denote Λ a compact set of $\mathbb{R}^m$ such that for all $q \in \mathbb{S}_p$, $\lambda_0(q) \in int(\Lambda)$.

The proof consists in three steps:

1. Under assumption D that the unconstrained set B_U has no faces, the estimate of the unconstrained support function is a consistent and asymptotically Gaussian random process (Proposition 9).

2. The minimization of the estimate $\hat{\delta}_n^*((q, \lambda) \mid B_U)$ with respect to λ holding q constant for any q can be analyzed as in Proposition 10.

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

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

   In both cases, this argument is valid for any finite list of q and the vector of those estimates are jointly asymptotically normal.

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

Using Andrews (1994, p2251), the three steps prove that $\tau_n(q)$ is a consistent and asymptotically Gaussian random process.

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

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

converges to a Gaussian process whose sample paths are uniformly continuous on the unit sphere $\mathbb{S}_m$, using the usual Euclidean norm. Hence $\tau_n^U(.)$ 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.*:

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

$\mathcal{S}(q)$ is a non-empty subset included in the interior of the compact set $S_B = \mathbb{S}_p \times \Lambda \subset \mathbb{R}_m$ by the above. Note also that to obtain the standard evaluation on the unit sphere some renormalization is necessary since $\|s\| \geq \|q\| = 1$, and this is done using the positive homogeneity of support functions:

$$\delta^*(s \mid B_U) = \|s\| \, \delta^*(\frac{s}{\|s\|} \mid B_U).$$

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

We first consider the case where $\mathcal{S}(q)$ is a singleton. In the second part of the proof, $\mathcal{S}(q)$ 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, $\mathcal{S}(q) = \{s_0 = (q, \lambda_0)\} \subset int(S_B)$.

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

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

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

$$\hat{\delta}_n^*(q \mid B) = \hat{\delta}_n^*(s_n \mid B_U). \tag{C.25}$$

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

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

The step 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:

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

using equation (C.25) and 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 because there are various minimizers of $\delta^*(s \mid B_U)$ in λ. Note first that set $\mathcal{S}(q) \subset int(S_B)$ is convex and compact because $\delta^*(s \mid B_U)$ is convex and continuous. We are first going to select and characterize a unique (q, λ_0^*) from $\mathcal{S}(q)$. Consider the smallest convex cone which includes $\mathcal{S}(q)$:

$$\mathcal{CS}(q) = \{cs_0; s_0 \in \mathcal{S}(q), c \geq 0\}.$$

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:

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

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 in λ.

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

Lemma 17 *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 for $\Psi_a(s)$,

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

because s_0^* is a minimizer of $\delta^*(s \mid B_U)$.

We consider 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,

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

It implies that:

$$0 \leq \delta^*(s_a \mid B_U) - \delta^*(s_0^* \mid B_U) \leq a(\lambda_0^{*\top}\lambda_0^* - \lambda_a^\top \lambda_a) \leq a\lambda_0^{*\top}\lambda_0^*, \tag{C.26}$$

so that the distance between s_a and the set $\mathcal{S}(q) = \{s; \delta^*(s \mid B_U) = \delta^*(s_0^* \mid B_U)\}$ tends to zero when a tends to zero by the continuity of 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 $\mathcal{S}(q)$ is compact, $s_m = (q, \lambda_m) \in \mathcal{S}(q)$. We are going to show that $s_m = s_0^*$. By definition of λ_a and λ_0^*, we have

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

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

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

Combining the two equations,

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

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:

$$\lim_{a \to 0} \|\lambda_a - \lambda_0^*\| = 0.$$

Furthermore, we check:

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

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

$$\Psi_a(s_0^*) - \Psi_a(s_a) = o(a). \tag{C.27}$$

■

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 18

$$\forall \varepsilon > 0, \exists a_0 > 0, \exists \eta > 0, \forall a \textit{ such that } \inf_{0 < a \leq a_0, \left\|\lambda - \lambda_0^*\right\| \geq \varepsilon} \frac{\Psi_a(s) - \Psi_a(s_a)}{a} > \eta. \tag{C.28}$$

Proof. First, the Lemma is trivially satisfied when $s_0^* = (q, 0)$ since $\lambda_a = \lambda_0^* = 0$ for any a and therefore:

$$\frac{\Psi_a(s) - \Psi_a(s_a)}{a} = \lambda^T \lambda$$

which is bounded from below by ε^2 when $\|\lambda\| \geq \varepsilon$.

Assume that $(q, 0) \notin \mathcal{S}(q)$. Like for Lemma 16, we will first show that the infimum is positive for a given q and then use the compactness of the space where q evolves to conclude. We know from the previous Lemma that $\lambda \to \lambda_0^*$ when a tends to 0.

- When $s = (q, \lambda) \in \mathcal{S}(q)$, $\frac{\Psi_a(s) - \Psi_a(s_a)}{a} \to \lambda^T \lambda - \lambda_0^{*\top} \lambda_0$ when a tends to zero; which is strictly positive when $\|\lambda - \lambda_0^*\| \geq \varepsilon$.
- When $s = (q, \lambda) \notin \mathcal{S}(q)$, $\frac{\Psi_a(s) - \Psi_a(s_a)}{a} \to +\infty$ when a tends to zero and cannot deliver the infimum.

As $\lambda_a \to \lambda_0^*$ when a tends to zero, there exists some positive a_0 such that the joint events $\{0 < a < a_0\}$ and $\{\|\lambda - \lambda_0^*\| \geq \varepsilon\}$ implies that $\|\lambda - \lambda_a\| \geq \varepsilon/2$. Assume now by contradiction that the infimum over $0 < a < a_0$ is not positive. By continuity of function $\frac{\Psi_a(s) - \Psi_a(s_a)}{a}$ in a and s when $a > 0$, and as the infimum is positive at the limit $a \to 0$, a non-positive infimum can only be obtained at some $a > 0$ and $s_a \in S_B$. It is a contradiction because s_a is a well separated minimum for any $a > 0$. The compactness of $S_B \cap \{s \in S_B,\ \|\lambda - \lambda_0^*\| \geq \varepsilon\}$ ensures that the infimum over such ss in this set is positive also. ■

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 $s_{n,a}$ as a near minimizer of $\Psi_{n,a}$ over S_B. We can adapt the same kind of argument than the one used in the proof of Proposition 10 to show that

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

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 $\delta^*(s \mid B_U)$, when $s \in S_B$, we have that:

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

Then:

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

We thus have:

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

By condition (C.28), for any ε, there exist $\eta > 0$ and n_0 such that for any $n \geq n_0$ the event:

$$\{d(s_{n,a_n}, s_0^*) \geq \varepsilon\} \subset \{\frac{\Psi_{a_n}(s_{n,a_n}) - \Psi_{a_n}(s_{a_n})}{a_n} > \eta\}.$$

Therefore:

$$\forall \varepsilon, \lim_{n\to\infty} \Pr(d(s_{n,a_n}, s_0^*) > \varepsilon) = 0 \Longrightarrow s_{n,a_n} - s_0^* \xrightarrow[n\to\infty]{P} 0.$$

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

$$\sqrt{n}\left(\hat{\delta}_n^*(s_{a_n,n} \mid B_U) - \hat{\delta}_n^*(s_0^* \mid B_U)\right) \xrightarrow[n\to\infty]{P} 0.$$

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

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 $\tau_n^U(s)$ is equicontinuous, we know that for any $\varepsilon > 0$ and $\eta > 0$ there exists δ such that:

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

It is straightforward to extend this result to the compact set $S_B = \mathbb{S}_p \times \Lambda$ so that:

$$\forall \varepsilon, \forall \eta, \exists \delta > 0, \lim_{n\to\infty} \Pr\left[\sup_{s_1,s_2\in S_B, \|s_1 - s_2\| < \delta} \left|\tau_n^U(s_1) - \tau_n^U(s_2)\right| > \eta\right] < \varepsilon. \tag{C.29}$$

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:

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

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

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

Or alternatively:

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

In consequence:

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

By definition:

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

so that for any (ε, η) and for any δ satisfying equation (C.29) :

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

that proves that the process $\tau_n(q)$ is equicontinuous by equation (C.29).

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

Figures and Tables

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

Estimation of the interval boundaries $[\beta^{-1}; \beta^{1}]$.

	β^{-1}			β^{1}		
True value	1.6545			2.3455		
n	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
100	1.6089	1.6543	1.7000	2.2990	2.3426	2.3887
500	1.6333	1.6540	1.6739	2.3242	2.3438	2.3653
1000	1.6410	1.6554	1.6697	2.3313	2.3458	2.3601
2500	1.6463	1.6551	1.6640	2.3364	2.3457	2.3544

Q_1, Q_2, Q_3 are the three quantiles of the distribution of the estimated interval boundaries across simulations.

Table 2: Percentage of rejections for the interior test - first example

	$H_0 : \beta^r \in B, H_a : \beta^r \notin B.$			
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
-2.00	99.8%	100.0%	100.0%	100.0%
-1.70	95.1%	100.0%	100.0%	100.0%
-1.50	79.9%	100.0%	100.0%	100.0%
-1.30	48.0%	95.5%	99.9%	100.0%
-1.20	30.7%	73.1%	95.0%	100.0%
-1.10	16.5%	32.6%	52.0%	83.2%
-1.00	**6.7%**	**6.0%**	**6.2%**	**5.9%**
-0.90	2.5%	0.6%	0.1%	0.0%
-0.80	0.7%	0.0%	0.0%	0.0%
-0.70	0.1%	0.0%	0.0%	0.0%
-0.60	0.0%	0.0%	0.0%	0.0%
to				
0.60	0.0%	0.0%	0.0%	0.0%
0.70	0.2%	0.0%	0.0%	0.0%
0.80	0.5%	0.0%	0.0%	0.0%
0.90	2.3%	0.5%	0.1%	0.0%
1.00	**7.3%**	**6.5%**	**5.9%**	**5.5%**
1.10	17.0%	34.4%	49.9%	82.3%
1.20	32.1%	74.6%	93.5%	100.0%
1.30	49.5%	95.3%	99.9%	100.0%
1.50	80.8%	100.0%	100.0%	100.0%
1.70	95.6%	100.0%	100.0%	100.0%
2.00	99.9%	100.0%	100.0%	100.0%

The point tested is $\beta^r = \beta^0 + r(\beta^1 - \beta^0)$.
β^1 and β^{-1} are on the frontier of B.

Table 3: Percentage of rejections for the Sargan test - second example

	$H_0 : 0 \in B_{Sargan}, H_a : 0 \notin B_{Sargan}$			
π_0	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
1	0.0%	0.0%	0.0%	0.0%
to				
0.3	0.0%	0.0%	0.0%	0.0%
0.2	0.1%	0.0%	0.0%	0.0%
0.1	0.8%	0.0%	0.0%	0.0%
0.0	1.5%	0.0%	0.0%	0.0%
0.06	2.6%	0.0%	0.0%	0.0%
0.04	3.2%	0.4%	0.1%	0.0%
0.02	4.1%	1.5%	0.9%	0.8%
0.01	5.3%	3.4%	1.9%	1.8%
0.005	6.1%	4.2%	2.9%	2.8%
0.002	5.9%	4.9%	3.8%	4.1%
0.001	5.9%	4.9%	4.3%	4.2%
0.000	6.3%	4.9%	4.6%	5.1%

Table 4: Percentage of rejections for the interior test - second example

$\pi_0 = 0.5$					$\pi_0 = 0.1$				
r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$	r	$n = 100$	$n = 500$	$n = 1000$	$n = 2500$
0.000	0.0%	0.0%	0.0%	0.0%	0.000	0.0%	0.0%	0.0%	0.0%
0.100	0.0%	0.0%	0.0%	0.0%	0.100	0.0%	0.0%	0.0%	0.0%
0.150	0.0%	0.0%	0.0%	0.0%	0.150	0.0%	0.0%	0.0%	0.0%
0.200	0.0%	0.0%	0.0%	0.0%	0.200	0.1%	0.1%	0.1%	0.1%
0.250	0.0%	0.0%	0.0%	0.0%	0.250	0.3%	1.6%	3.1%	5.1%
					0.251	**0.3%**	**1.6%**	**3.5%**	**5.2%**
0.300	0.0%	0.0%	0.0%	0.0%	0.300	1.1%	6.0%	15.9%	30.5%
0.400	0.2%	0.0%	0.0%	0.0%	0.400	7.3%	44.9%	73.6%	97.5%
0.500	0.3%	0.0%	0.0%	0.0%	0.500	19.4%	83.5%	98.1%	100.0%
0.600	1.6%	0.0%	0.0%	0.0%	0.600	37.5%	95.3%	99.8%	100.0%
0.700	5.6%	2.4%	2.1%	0.7%	0.700	55.9%	97.4%	99.8%	100.0%
0.733	**8.1%**	**5.4%**	**6.7%**	**5.3%**					
0.750	9.5%	8.2%	11.3%	10.6%	0.750	62.9%	97.4%	99.8%	100.0%
0.800	14.0%	19.4%	28.7%	47.8%	0.800	68.2%	97.4%	99.8%	100.0%
0.900	26.4%	57.8%	81.7%	98.9%	0.900	75.8%	97.4%	99.8%	100.0%
1.000	43.8%	88.1%	98.6%	100.0%	1.000	80.2%	97.4%	99.8%	100.0%
1.400	94.2%	100.0%	100.0%	100.0%	1.400	83.1%	97.4%	99.8%	100.0%
1.800	99.8%	100.0%	100.0%	100.0%	1.800	83.4%	97.4%	99.8%	100.0%
2.000	99.9%	100.0%	100.0%	100.0%	2.000	83.4%	97.4%	99.8%	100.0%

The points tested are the ones considered in example 1. β^1 is on the frontier of B when there are no supernumerary instruments. The points in bold are on the frontier of the new set B using z as a supernumerary instrument.

D Supplementary Appendix

D.1 Additional Monte Carlo experiments

We report three additional 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 two. 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 and serves as a benchmark 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.

D.1.1 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 $[-1/2, 1/2]$. 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 constant (see Appendix D.2.1):

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

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 5).

We draw 1000 simulations in four different sample size experiments: 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 increases. The three quartiles as well as the mean of the distribution of the estimated support function at one angle are displayed in Table 5 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.

With respect to 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 6. 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.

D.1.2 Smooth set with one supernumerary instrument

The simulated model is as before except that the second explanatory variable x_2 is now generated by:

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

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 observed variable where e_4 is i.i.d. standard normal. Using general notations, we have $x = (x_1, x_2)$ and $z = (x_1, e_2, w)$. 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.2.2):

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

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 the figure of Table 7).

As before, we draw 1000 simulations in four sample size experiments : n =100, 500, 1000 and 2500 with $\pi = \nu = 0.9$. Table 7 displays descriptive statistics (Mean, quartiles) related to the distribution of the estimated support function at one angle. Table 8 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.

D.1.3 A set with kinks and faces

In this experiment, the explanatory variable has mass points so that the identified set has exposed faces. Also 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.2.3). 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 9).

One 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.2.3) 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 10 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 and Table 11 reports the 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 ($a_n > 0$). Surprisingly, for the frontier test we do not overreject too much but we overreject for the interior test. In the second panel, we display the rejection rates using the perturbed program defined in Proposition 10 with $a_n = \frac{0.5}{n^{1/3}}$. Rejection rates are pretty close to the nominal size for both the frontier test and the interior test. Perturbing the program leads to correct quite efficiently for the presence of kinks except perhaps for very small sample sizes. Sample size properties can also be improved while estimating the variance with i.i.d. bootstrap techniques.

D.2 Computations of Section D.1

D.2.1 Example of Section D.1.1

The simulated model is:

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

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

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

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.2.2 Example of Section D.1.2

The simulated model is:

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

$x_2 = \pi e_2 + \sqrt{1-\pi^2}e_3$, $w = \nu e_3 + \sqrt{1-\nu^2}e_4$ 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:

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

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

$$\begin{aligned} z_{q,\lambda} &= \begin{pmatrix} q^\top & \lambda \end{pmatrix}\begin{pmatrix} x_1 \\ a^{-2}(\pi e_2 + \mu w) \\ z_H \end{pmatrix} \\ &= x_1 q_1 + (a^{-2}(\pi e_2 + \mu w))q_2 + z_H\lambda. \end{aligned}$$

The variance of $z_{q,\lambda}$ is therefore

$$V_{q,\lambda} = q_1^2 + \frac{q_2^2}{a^2} + \lambda^2.$$

As in the previous example,

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

The covariances 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,

$$Ex_1\mathbf{1}_{z_{q,\lambda}>0} = \frac{1}{\sqrt{2\pi}}\frac{q_1}{\sqrt{V_{q,\lambda}}},$$

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

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

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.2.3 Example of Section D.1.3

The simulated model is:

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

and variable $z \equiv (1, x_1)^\top$ are the instruments. As $\Sigma = \left(E(z^\top z)\right)^{-1} = I_2$,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 frontier 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:

$$\begin{aligned} E\mathbf{1}\{z_q > 0\} &= \frac{1}{2} \\ Ex\mathbf{1}\{z_q > 0\} &= \frac{1}{2}, \end{aligned}$$

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 $q^\top(\hat{\Sigma}_n - \Sigma)z^\top$ is equal to $\frac{1}{\sqrt{2}}(1+x)\frac{1}{n}\sum_{i=1}^n x_i$ which is equal to zero when $x = -1$. The additional term in the asymptotic distribution $\tau_1(q)$ is therefore equal to zero. The proof is similar for other values of θ.

D.3 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 Appendix C.1 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:

$$\left| \frac{1}{n} \sum_{i=1}^{n} f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i)) \right| \underset{n \to \infty}{\overset{a.s}{\to}} 0. \tag{D.30}$$

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.11):

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

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

$$\left| f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta}(z_i, \underline{y}_i, \overline{y}_i) \right| \underset{n \to \infty}{\overset{a.s}{\to}} 0. \tag{D.31}$$

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 (D.31) 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 (D.30) and the triangular inequality, we thus conclude that, uniformly for $q \in \mathbb{S}$:

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

D.4 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 and smooth. We here provide a simple way to construct CI^n_α when $\alpha < 1/2$:

$$CI^n_\alpha = \{\beta; \frac{\sqrt{n}}{\sqrt{\hat{V}_{q_n}}}(T_n(q_n;\beta)) > \mathcal{N}_\alpha\}$$

where

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

and where q_n is any sequence of local minimizers of $T_n(q;\beta)$ over the unit sphere (and therefore depends on β). Therefore, the confidence region is also given by $CI^n_\alpha = \{\beta; \min_{q\in\mathbb{S}}(T_n(q;\beta)) > \frac{\sqrt{\hat{V}_{q_n}}}{\sqrt{n}}\mathcal{N}_\alpha\}$

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

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

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

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

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

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

We have:

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

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^n_α. Given that λ is positive,

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

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

$$\hat{B}_n \cup \{\cup_{\beta_f \in \partial B_n} \cup_{q_f \in \mathcal{C}(\beta_f)} (\beta_f, \beta_f - \frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha q_f)\} \subset CI^n_\alpha, \tag{D.32}$$

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

Conversely, let us prove that CI^n_α is included in the set on the LHS. Let β_c a point in CI^n_α. 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^n_α, 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 (D.32). As a consequence, equation (D.32) is an equality.

D.5 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 $1 - 2\alpha$.

D.6 Uniform confidence regions

Starting from the end of Section B.1, the diameter of set B is equal to:

$$\Delta = \max_{q \in \mathbb{S}_p} (\delta^*(q|B) + \delta^*(-q|B)) = \max_{q \in \mathbb{S}_p} (E\left(|z_{qi}|\left(\bar{y}_i - \underline{y_i}\right)\right)).$$

As by assumption, $\bar{y}_i - \underline{y_i} > 0$, the limit point $\Delta = 0$ is outside the range of data that we consider since it is obtained if and only if $\bar{y}_i - \underline{y_i} = 0$ almost surely $|z_{qi}| \neq 0$ for any q and thus almost surely $z_i \neq 0$. Using the regularity condition $R.ii$, $\Pr(z_i = 0) = 0$ and the almost sure statement extends to all z_i. This simplifies notations and notations only.

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

$$\begin{aligned} \hat{\Delta}_n &= \max_{q \in \mathbb{S}_p} \left(\hat{\delta}_n^*(q|B) + \hat{\delta}_n^*(-q|B) \right), \\ &= \max_{q \in \mathbb{S}_p} \left(\frac{1}{n} \sum_{i=1}^{n} |z_{n,qi}| \left(\bar{y}_i - \underline{y}_i\right) \right) \end{aligned}$$

Therefore, if we extend our setting to include the case $\Delta = 0$, $\bar{y}_i - \underline{y}_i = 0$ almost surely z_i and we have:

$$\Pr(\hat{\Delta}_n = 0; \Delta = 0) = 1,$$

so that trivially:

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

More generally, consider a sequence of experiments indexed by $\alpha \downarrow 0$ such that $\Pr(\bar{y}_i - \underline{y}_i < \alpha) = 1$ and $\Pr(\bar{y}_i - \underline{y}_i > \alpha/2) > 0$. By Cauchy-Schwartz and since $E((\bar{y}_i - \underline{y}_i)^2)^{1/2} < \alpha$, the diameter Δ is such that $\Delta < \alpha \max_{q \in S_p} (E(z_{qi}^2))^{1/2} \leq \alpha . \|\Sigma\| \|z_i\|_{L^2} = \alpha M$. Therefore Δ is at most $O(\alpha)$. Using minima instead of maxima and the other bound on $\bar{y}_i - \underline{y}_i$ yields that Δ and α go to zero at the same rate.

We have:

$$\sqrt{n}(\Delta_n - \Delta) = \sqrt{n} \left(\max_q \frac{1}{n} \sum_{i=1}^{n} |z_{n,qi}| \left(\bar{y}_i - \underline{y}_i\right) - \max_q E(|z_{qi}| \left(\bar{y}_i - \underline{y}_i\right)) \right).$$

Denote q_α any maximizer of $E(|z_{qi}| \left(\bar{y}_i - \underline{y}_i\right))$ in experiment α and construct q_n as a sequence of local maximizers of the first maximum term (perturbing the objective function in the case in which there are kinks as in Proposition 10). Then, as in Proposition 10:

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} |z_{n,q_n i}| \left(\bar{y}_i - \underline{y}_i\right) - \frac{1}{n} \sum_{i=1}^{n} |z_{n,q_\alpha i}| \left(\bar{y}_i - \underline{y}_i\right) \right) = o_P(1),$$

so that:

$$\sqrt{n}(\Delta_n - \Delta) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} |z_{n,q_\alpha i}| \left(\bar{y}_i - \underline{y}_i\right) - E(|z_{q_\alpha i}| \left(\bar{y}_i - \underline{y}_i\right)) \right) + o_P(1).$$

Furthermore, using the same argument as in Proposition 9, write:

$$\begin{aligned}\sqrt{n}(\Delta_n - \Delta) =& \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|z_{n,q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right) - E(|z_{n,q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right))\right) \\ &+ \sqrt{n}\left(E(|z_{n,q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right)) - E(|z_{q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right)\right) + o_P(1).\end{aligned}$$

The first term is asymptotically equivalent to $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|z_{q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right) - E(|z_{q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right))\right)$ and has therefore a variance approximately equal to $V(|z_{q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right))$ which is bounded by a term $O_P(\alpha^2)$ and therefore $O_P(\Delta^2)$. The second term

$$\sqrt{n}\left(E(|z_{n,q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right)) - E(|z_{q_\alpha i}|\left(\bar{y}_i - \underline{y}_i\right)\right),$$

can also be easily bounded by a term $O_P(\alpha)$ so that:

$$\left|\sqrt{n}(\Delta_n - \Delta)\right| \leq O_P(\Delta).$$

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

Proposition 19 *Let:*

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

where q_n is a sequence of local minimizers of a (possibly perturbed) program defined in Proposition 10.

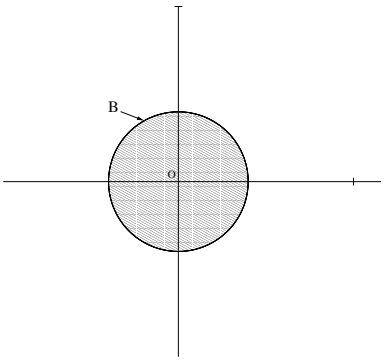
A confidence region $\tilde{CI}_\alpha^n$ of asymptotic level equal to $1-\alpha$ is defined by the collection of the points such that $\xi(\beta) \geq \tilde{\mathcal{N}}_\alpha$ where $\tilde{\mathcal{N}}_\alpha$ satisfies the equation

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

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

D.7 Figures an Tables for the additional Monte Carlo Experiments of Section D.1

Table 5: Results related to the additional 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

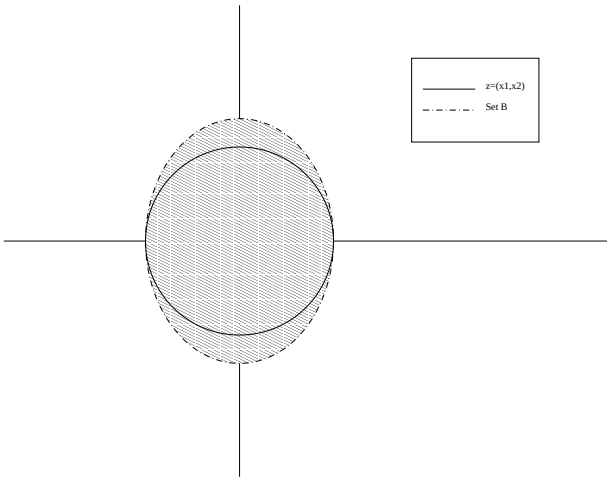
n	Mean	Q1	Q2	Q3
100	0.198	0.178	0.197	0.216
500	0.199	0.190	0.199	0.208
1000	0.199	0.193	0.199	0.206
2500	0.199	0.196	0.199	0.203

Table 6: Percentage of rejections for the two tests - first example of Section D.1

	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 7: Results related to the additional 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

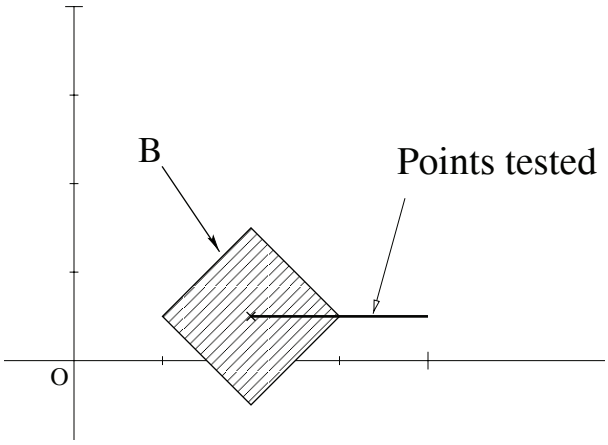
n	Mean	Q1	Q2	Q3
100	0.244	0.216	0.242	0.268
500	0.244	0.232	0.244	0.256
1000	0.243	0.234	0.243	0.252
2500	0.243	0.238	0.243	0.248

Table 8: Percentage of rejections for the two tests - example 2 of Section D.1

	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 9: Results related to the additional 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 10: Percentage of rejections for the test $H_0 : \beta^r \in \partial B$. Non-smooth set of Section D.1

	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%	100%	100%	100%	100%	100%	100%	100%
0.050	100%	100%	100%	100%	100%	100%	100%	100%
0.100	99.8 %	100%	100%	100%	100%	100%	100%	100%
0.200	100%	100%	100%	100%	100%	100%	100%	100%
0.300	99.8 %	100%	100%	100%	99.8 %	100%	100%	100%
0.400	98.5 %	100%	100%	100%	98.5 %	100%	100%	100%
0.500	94.1 %	100%	100%	100%	95%	100%	100%	100%
0.600	78.2 %	100%	100%	100%	84%	100%	100%	100%
0.700	41.9 %	99.8 %	100%	100%	61.3 %	99.8 %	100%	100%
0.800	9.9 %	90.9 %	100%	100%	29.1 %	96.2 %	100%	100%
0.850	4.2 %	60.6 %	94.5 %	100%	15.4 %	85.7 %	99.2 %	100%
0.900	1.8 %	20%	52.6 %	97.7 %	7.6 %	55.3 %	87.3 %	99.9 %
0.910	1.6 %	14.7 %	40.4 %	92.9 %	6.4 %	48.3 %	79.6 %	99.8 %
0.920	1.8 %	9%	28.5 %	84.2 %	5.8 %	40.1 %	70.2 %	98.8 %
0.930	2.3 %	5.6 %	18.8 %	66.4 %	4.8 %	30.8 %	59.8 %	96.2 %
0.940	2.3 %	3.4 %	11%	44.1 %	4.3 %	23.6 %	47%	90.1 %
0.950	2.7 %	2.1 %	6.6 %	25.4 %	3.9 %	17%	33.5 %	76.7 %
0.960	2.9 %	1.7 %	3.4 %	13.1 %	3.7 %	11.9 %	20.9 %	54.4 %
0.970	3.3 %	1.6 %	2.8 %	5.5 %	3.8 %	7.9 %	12%	32.1 %
0.980	3.8 %	2.4 %	3.4 %	2.4 %	3.8 %	4.9 %	8%	14.9 %
0.990	4.9 %	3.9 %	3.9 %	1.9 %	4.7 %	4.2 %	5.3 %	6.2 %
1 %	**6.4** %	**6.5** %	**6.4** %	**4.7** %	**5.3** %	**5.2** %	**5.2** %	**4.9** %
1.010	7.9 %	9.3 %	11.7 %	12.3 %	5.7 %	7.3 %	9.2 %	9.2 %
1.020	9%	13.1 %	19.7 %	30.4 %	6.4 %	9.9 %	15.8 %	22.6 %
1.030	10.1 %	16.9 %	29.4 %	53.2 %	7.9 %	13.6 %	23.6 %	44.4 %
1.040	12.3 %	25.6 %	41%	73.5 %	9.8 %	19.1 %	34.8 %	67.4 %
1.050	14.1 %	33.7 %	54.3 %	88.2 %	11.9 %	27.8 %	48.6 %	86.4 %
1.060	16.4 %	44%	67.4 %	95.7 %	13.6 %	37.8 %	62.5 %	94.7 %
1.070	19.3 %	52.9 %	80.3 %	98.9 %	15.7 %	46.5 %	76.7 %	98.7 %
1.080	21.8 %	63.3 %	88.7 %	99.5 %	18.1 %	59.1 %	86.8 %	99.4 %
1.090	24.2 %	71%	93.9 %	99.9 %	21.2 %	67.7 %	92.2 %	99.8 %
1.100	27.7 %	79.2 %	97%	100%	23.6 %	76.6 %	96%	100%
1.150	48.4 %	98.5 %	100%	100%	44.6 %	98.3 %	100%	100%
1.200	68.6 %	100%	100%	100%	66%	100%	100%	100%
1.300	95.1 %	100%	100%	100%	94.7 %	100%	100%	100%
1.400	99.9 %	100%	100%	100%	99.8 %	100%	100%	100%
1.500	100%	100%	100%	100%	100%	100%	100%	100%
1.600	100%	100%	100%	100%	100%	100%	100%	100%
1.700	100%	100%	100%	100%	100%	100%	100%	100%
1.800	100%	100%	100%	100%	100%	100%	100%	100%
1.900	100%	100%	100%	100%	100%	100%	100%	100%
2.000	100%	100%	100%	100%	100%	100%	100%	100%

Table 11: 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 %
0.050	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.100	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.200	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.300	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.400	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.500	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.600	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.700	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.800	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
0.850	0.5 %	0 %	0 %	0 %	0.2 %	0 %	0 %	0 %
0.900	1.9 %	0.1 %	0 %	0 %	0.9 %	0 %	0 %	0 %
0.910	2.0 %	0.1 %	0 %	0 %	1.2 %	0 %	0 %	0 %
0.920	2.5 %	0.3 %	0 %	0 %	1.5 %	0.1 %	0 %	0 %
0.930	3.1 %	0.3 %	0 %	0 %	2.0 %	0.1 %	0 %	0 %
0.940	3.3 %	0.4 %	0 %	0 %	2.4 %	0.1 %	0 %	0 %
0.950	4.0 %	1.1 %	0.2 %	0 %	2.9 %	0.6 %	0 %	0 %
0.960	5.5 %	1.8 %	1.1 %	0 %	3.8 %	0.9 %	0.1 %	0 %
0.970	6.6 %	3.3 %	2.3 %	0.2 %	4.8 %	1.5 %	0.6 %	0 %
0.980	8.2 %	5.1 %	3.6 %	1.3 %	5.6 %	2.6 %	1.8 %	0.3 %
0.990	9.1 %	7.4 %	6.2 %	3.3 %	6.3 %	4.6 %	3.3 %	1.4 %
1.000	**10.8 %**	**10.6 %**	**11.7 %**	**9.3 %**	**7.9 %**	**7.6 %**	**7.4 %**	**5.3** %
1.010	12.9 %	14.3 %	20.3 %	22.1 %	9.5 %	10.8 %	14.0 %	15.4 %
1.020	14.8 %	21.4 %	29.5 %	44.7 %	11.5 %	14.7 %	22.6 %	34.3 %
1.030	16.9 %	28.6 %	40.9 %	65.9 %	13.0 %	20.9 %	33.7 %	57.4 %
1.040	20.1 %	38.2 %	53.9 %	83.4 %	15.9 %	30.5 %	47.2 %	79.0 %
1.050	22.5 %	48.8 %	66.8 %	93.2 %	18.7 %	41.0 %	60.6 %	91.6 %
1.060	25.6 %	56.9 %	79.7 %	97.4 %	21.3 %	49.3 %	74.7 %	96.9 %
1.070	29.4 %	65.9 %	88.2 %	99.4 %	23.7 %	61.0 %	85.9 %	98.9 %
1.080	33.1 %	74.7 %	93.3 %	99.8 %	28.0 %	70 %	91.6 %	99.7 %
1.090	37.6 %	81.2 %	96.8 %	100 %	32.5 %	78.2 %	95.9 %	100 %
1.100	43.4 %	87.0 %	98.5 %	100 %	37.3 %	85.0 %	98.2 %	100 %
1.150	61.8 %	99.4 %	100 %	100 %	57.7 %	99.2 %	100 %	100 %
1.200	81.0 %	100 %	100 %	100 %	77.5 %	100 %	100 %	100 %
1.300	97.8 %	100 %	100 %	100 %	97.6 %	100 %	100 %	100 %
1.400	99.9 %	100 %	100 %	100 %	99.9 %	100 %	100 %	100 %
1.500	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
1.600	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
1.700	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
1.800	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
1.900	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %
2.000	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %