*Econometrica Supplementary Material*

# SUPPLEMENT TO "SET IDENTIFIED LINEAR MODELS"

(*Econometrica*, Vol. 80, No. 3, MAY 2012, 1129–1155)

BY CHRISTIAN BONTEMPS, THIERRY MAGNAC, AND ERIC MAURIN

## APPENDIX A: PROOF FOR SECTION 2

### A.1. *Proof of Proposition 1*

*Necessity*. Consider $\beta$ in $\mathbb{R}^K$ and assume that there is a latent random variable $\varepsilon$ uncorrelated with $x$ such that the latent variable $y^* \equiv x\beta + \varepsilon$ lies within the observed bounds, that is, $x\beta + \varepsilon \in [\underline{y}; \overline{y}]$. Denoting $y = (\overline{y} + \underline{y})/2$ and using that $\varepsilon$ 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 $u(x) \equiv E(y^* - y \mid x)$, *Q.E.D.*

$$-E\left(\frac{(\overline{y} - \underline{y})}{2} \,\middle|\, x\right) \leq u(x) \leq E\left(\frac{(\overline{y} - \underline{y})}{2} \,\middle|\, x\right).$$

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 $\varepsilon$ that is uncorrelated with $x$ and is such that $y^* \equiv x\beta + \varepsilon$ lies within the observed bounds.

First, consider $\lambda$ 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 $\varepsilon$ as

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

 DOI: 10.3982/ECTA7637

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

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

where we used that $\lambda$ 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

$$\begin{aligned} E(x^\top \varepsilon) &= E(x^\top (y - x\beta)) + E(x^\top u(x)) \\ &= -E(x^\top u(x)) + E(x^\top u(x)) = 0, \end{aligned}$$

using the moment condition (A.1) involving $y$, $\beta$, and $u(x)$.

## APPENDIX B: PROOFS FOR 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

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

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

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

where $\Delta^*(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]$. Note that $u_q(z)$ is unique (almost everywhere (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.1) 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.2) generate all the points $\beta = (E(z^\top x))^{-1}E(z^\top(y + u_q(z)))$ that 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 $\beta$,

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

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,

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

so that

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

because of equation (3) and $|z_q| > 0$ with positive probability because of the full rank assumption in R.2.

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. *Q.E.D.*

## B.2. *Proof of Lemma 3*

We use the expression derived in Proposition 2:

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

First of all, the support function of a convex set is convex and, therefore, is differentiable except at a countable number of directions $q$ denoted $D_f$. 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 $B$.

### B.2.1. *Characterization of $D_f$*

The first term on the right-hand side (RHS) of equation (B.3) is linear in $q$ since (see the previous proof)

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

and thus is continuously differentiable on $\mathbb{S}_p$. 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 $t \in \mathbb{S}_p$,

$$\begin{aligned}\psi(t) - \psi(q) &= E(z^*(t - q)\mathbf{1}\{z^* q > 0\}) \\ &\quad + E\big(z^* t(\mathbf{1}\{z^* t > 0\} - \mathbf{1}\{z^* q > 0\})\big),\end{aligned}$$

so that

$$\begin{aligned}&\psi(t) - \psi(q) - E(z^* \mathbf{1}\{z^* q > 0\})(t - q) \\ &\quad = E\big(z^* t(\mathbf{1}\{z^* t > 0\} - \mathbf{1}\{z^* q > 0\})\big).\end{aligned} \tag{B.4}$$

Points of nondifferentiability depend on the expression in the RHS. It is the sum of three terms:

$$A_1 = E(z^* t \mathbf{1}\{z^* t > 0, z^* q < 0\}),$$

$$A_2 = -E(z^* t \mathbf{1}\{z^* q > 0, z^* t \leq 0\}),$$
$$A_3 = E(z^* t \mathbf{1}\{z^* q = 0, z^* t > 0\}).$$

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

$$0 \leq |A_1| \leq E(\|z^*\|)\|t - q\| \Pr(z^* t > 0, z^* q < 0),$$
$$0 \leq |A_2| \leq E(\|z^*\|)\|t - q\| \Pr(z^* q > 0, z^* t \leq 0).$$

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

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

since Assumption R.3 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.4),

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

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 $s \in \mathbb{S}_p$, $s^\top q = 0$. We have

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

It follows that $\psi$ has different gradients in different directions $s$, which depend on the term

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

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 Assumption R.2 and, therefore, function $\psi(q)$ is not differentiable.

Overall, the points of nondifferentiability 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.

B.2.2. *Exposed Faces*

Using Lemma 3, we obtain, for any $q$ that does not belong to $D_f$,

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

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 disguised 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 $H_f$ that is not reduced to a singleton. If $q_f$ denotes the vector orthogonal to $H_f$, we have, for any $\beta_f$ in $B_f$,

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

which means (see equation (B.2)) that there exists $\Delta_f^*(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that (recall that $\beta_{q_f} = (E(z^\top x))^{-1} E(z^\top w_{q_f})$)

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

For the set of all $\beta_f$ not to be reduced to the singleton $\{\beta_{q_f}\}$, 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 $\beta_q = (E(z^\top x))^{-1} E(z^\top w_q)$ and let $H_q$ denote the supporting hyperplane at $\beta_q$ orthogonal to $q$. Consider the set $B_f$ of all $\beta_f$ such that there exists $\Delta_f^*(z)$ in $[\underline{\Delta}(z), \overline{\Delta}(z)]$ such that

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

$B_f$ 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 nonzero 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.

### B.2.3. *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), that is, 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 no longer one-to-one. This complicates the construction of testing procedures (as shown in Section 4) and is the reason why it is useful to characterize setups where $B$ has kinks. We have

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

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

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

$$\begin{aligned}
&\beta_q = \beta_r \\
&\quad \Rightarrow \quad E\big(z_q(\bar{y} - \underline{y})(\mathbf{1}\{z_q > 0, z_r < 0\} - \mathbf{1}\{z_q < 0, z_r > 0\})\big) = 0.
\end{aligned}$$

Given that the term within the expectation is necessarily nonnegative, 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. *Q.E.D.*

## B.3. *Proof of Lemma 4*

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

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

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

$$\begin{aligned} z_F &= z(E(z^\top z))^{-1}E(z^\top x)\big[E(x^\top z)(E(z^\top z))^{-1}E(z^\top x)\big]^{-1/2} \\ &= z(E(z^\top z))^{-1/2}Q_F, \end{aligned}$$

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 $A = \binom{0}{I_{m-p}}$ as the $[m, m-p]$ selection matrix, the definition of $z^s$ implies

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

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

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

$$\begin{aligned} \zeta^s &= z^s - z_F E(z_F^\top z^s) = z(E(z^\top z))^{-1/2}(I_m - P_F)A^s \\ &= z(E(z^\top z))^{-1/2}P_H A^s, \end{aligned}$$

which implies

$$\begin{aligned} z_H &= \zeta^s(\zeta^{s\top}\zeta^s)^{-1/2} = z(E(z^\top z))^{-1/2}P_H A^s(A^{s\top}P_H A^s)^{-1/2} \\ &= z(E(z^\top z))^{-1/2}Q_H, \end{aligned}$$

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

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

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

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

which finishes the proof. *Q.E.D.*

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 nonempty because condition (10) is sufficient. Yet, if $B$ is nonempty 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 with condition (10) are invariant to the choice of the specific subset of supernumerary instruments. As discussed above, $z_H$ can be written as $z(E(z^\top z))^{-1/2}Q_H$, where the $m-p$ columns of matrix $Q_H$ are an orthonormal basis of the kernel of the orthogonal projection onto $x(z)$. Changing one specific subset of supernumerary instruments $z_H$ into an alternative subset $z_H^*$ boils down to moving from one orthonormal basis $Q_H$ to an alternative basis $Q_H^*$ (i.e., to $Q_H^* = Q_H R$, 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 subset.

### 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 $\{\gamma = 0\}$, if $x_2^* = (q_2, \lambda_2)$, is

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

Corollary 16.4.1 of Rockafellar (1970, p. 146) 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

$$\delta^*(x^* \mid B_U \cap \{\gamma = 0\}) = \inf_{(x_1^*, x_2^*): x_1^* + x_2^* = x^*} \big(\delta^*(x_1^* \mid B_U) + \delta^*(x_2^* \mid \{\gamma = 0\})\big) \tag{B.5}$$

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

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

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

as the RHS is independent of $\lambda_2$ and $\lambda$. Furthermore, the infimum in $\lambda_1$ 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.5) by its closure even though 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 $B_{\text{Sargan}}$ (see Section 3.2.3). Without loss of generality, suppose that $B_U$ is included in the half-space $\gamma \geq 0$ (i.e., $\gamma$ is a completely positive vector) by changing signs of subparameters of $\gamma$ if necessary. We now consider 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 no longer a proper set. Let $(\beta_I, 0)$ be the intersection point and consider one hyperplane that 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 $(q_I, \lambda_I)$ as its normal vector-oriented outward set $B_U$, where $\lambda_I$ could be infinite (recall that $q_I \in \mathbb{S}_p$). For any value $\lambda \leq \lambda_I$, we have $\forall(\beta, \gamma) \in B_U$,

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

The support function for $(q_I, \lambda)$ is also equal to $\delta((q_I, \lambda_I) \mid B_U)$. If $\lambda_I$ 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. Otherwise, if set $B_U$ is not smooth, the infimum can be attained at a finite $\lambda_I$.

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. *Q.E.D.*

## B.5. *The Construction of $B_U$*

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

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

The support function of $B_U$ is as in Proposition 2,

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

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

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

The last matrix is well defined because of the rank conditions R.2 and Appendix B.3.

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

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

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

## APPENDIX C: PROOFS FOR SECTION 4

We let $M$ denote a generic majorizing constant.

### C.1. *Proof of Proposition 9*

We use that

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

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

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

where

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

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

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

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

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

$$\hat{\Sigma}_n = \begin{cases} \hat{\Sigma}_n^u, & \text{if } \|\hat{\Sigma}_n^u\| \le M, \\ \hat{\Sigma}_n^u\left(\dfrac{M}{\|\hat{\Sigma}_n^u\|}\right), & \text{if not.} \end{cases} \tag{C.2}$$

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

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

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

$$\sqrt{n}(\hat{\Sigma}_n - \hat{\Sigma}_n^u) \underset{n\to\infty}{\overset{P}{\to}} 0, \tag{C.3}$$

and the estimate is asymptotically normal,

$$\sqrt{n}(\operatorname{vec}(\hat{\Sigma}_n^\top - \Sigma^\top)) \implies N(0, W). \tag{C.4}$$

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

### C.1.1. *Consistency and Asymptotic Normality: $\Sigma$ Is Known*

Suppose that $\Sigma$ is known and denote

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

Consider function $f_\theta$ 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 $\theta$ lying in a bounded set $\Theta$.

As the proof of Lemma 3 shows, this function is convex in $\Sigma q$ and, therefore, is Lipschitzian,

$$\begin{aligned} |f_{\theta_1}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta_2}(z_i, \underline{y}_i, \overline{y}_i)| &\le \max(\|z_i^\top \underline{y}_i\|, \|z_i^\top \overline{y}_i\|)\|q_1^\top \Sigma_1^\top - q_2^\top \Sigma_2^\top\| \\ &\le M \max(\|z_i^\top \underline{y}_i\|, \|z_i^\top \overline{y}_i\|)\|\theta_1 - \theta_2\|, \end{aligned} \tag{C.5}$$

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

Under Assumption R.3, we have

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

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

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

Also, under the conditions of Proposition 9, we have

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

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

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

converges in distribution, uniformly in $\Theta$, to a Gaussian process with zero mean and covariance function

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

The second step of the proof of Proposition 9 consists of replacing $\Sigma$ by the almost sure limit $\hat{\Sigma}_n$ defined above. Consistency is proved in Appendix D, since this result was already shown in Beresteanu and Molinari (2008). We rely heavily on Section 19.4 of van der Vaart (1998), where relevant properties are proposed.

C.1.2. *Asymptotic Distribution When $\Sigma$ Is Estimated*

We analyze the asymptotic behavior of $\tau_n(q)$, which is 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 $\theta = (q, \Sigma)$ be the true value and let $\hat{\theta}_n = (q, \hat{\Sigma}_n)$ be its estimate. Let us prove that if $\hat{\theta}_n \overset{P}{\to}_{n\to\infty} \theta$ uniformly in $q$, then

$$E(z_{n,qi}w_{n,qi} - z_{qi}w_{qi})^2 = E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_\theta(z_i, \underline{y}_i, \overline{y}_i))^2 \underset{n\to\infty}{\overset{P}{\to}} 0. \tag{C.6}$$

Using equation (C.5), we have

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

so that

$$\begin{aligned} &E(f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_\theta(z_i, \underline{y}_i, \overline{y}_i))^2 \\ &\quad \leq M^2 E\big(\max(\|z_i^\top \underline{y}_i\|, \|z_i^\top \overline{y}_i\|)^2\big)\|\hat{\theta}_n - \theta\|^2. \end{aligned}$$

Under the conditions of Proposition 9, $E(\max(\|z_i^\top \underline{y}_i\|, \|z_i^\top \overline{y}_i\|)^2) < \infty$ and is independent of $q$. As $\|\hat{\theta}_n - \theta\|^2$ tends in distribution to 0 uniformly in $q \in \mathbb{S}_p$ (equation (C.3)), 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

$$C_n(q) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n z_{qi}w_{qi} - E(z_{qi}w_{qi})\right), \tag{C.7}$$

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

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

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

*and*

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

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

PROOF: For convenience sake, we 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

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

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

$$\begin{aligned} B^2_n(q) &= \sqrt{n}(q^\top (\hat{\Sigma}_n - \Sigma)^\top E(z_i^\top w^*_{qi})) \\ &= \sqrt{n} q^\top (\hat{\Sigma}_n - \Sigma)^\top (\Sigma^\top)^{-1} \beta^*_q \\ &= \sqrt{n} q^\top (\hat{\Sigma}_n^\top (\Sigma^\top)^{-1} - I) \beta^*_q, \end{aligned}$$

using the definition of $\beta^*_q$.

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

$$\begin{aligned} B^1_n(q) &= \sqrt{n} E\bigg( z_{n,qi}(\bar{y}_i - \underline{y}_i) \\ &\qquad \times \bigg( \mathbf{1}\{z_{n,qi} > 0\} - \frac{1}{2}(\mathbf{1}\{z_{qi} > 0\} + \mathbf{1}\{z_{qi} \geq 0\}) \bigg) \bigg) \\ &= \sqrt{n} E\bigg( (z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i) \\ &\qquad \times \bigg( \mathbf{1}\{z_{n,qi} > 0\} - \frac{1}{2}(\mathbf{1}\{z_{qi} > 0\} + \mathbf{1}\{z_{qi} \geq 0\}) \bigg) \bigg) \\ &\qquad + \sqrt{n} E\bigg( z_{qi}(\bar{y}_i - \underline{y}_i) \\ &\qquad \times \bigg( \mathbf{1}\{z_{n,qi} > 0\} - \frac{1}{2}(\mathbf{1}\{z_{qi} > 0\} + \mathbf{1}\{z_{qi} \geq 0\}) \bigg) \bigg). \end{aligned}$$

The first two lines is the sum of two terms,

$$B_n^{11}(q) = \sqrt{n}E\Big((z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i)\Big(\mathbf{1}\{z_{n,qi} > 0\} - \frac{1}{2}\Big)\mathbf{1}\{z_{qi} = 0\}\Big),$$
$$B_n^{12}(q) = \sqrt{n}E\big((z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i) \times (\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\})\mathbf{1}\{z_{qi} \neq 0\}\big),$$

and the last two lines is equal to

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

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–Schwarz inequality and write

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

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

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

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 $(B_n^{13}(q))^2$ as the sum of two terms:

$$\text{(C.8)}\qquad \begin{aligned}&nM^2 E\big(z_{qi}^2 \Pr\big(|\sqrt{n}(z_{n,qi} - z_{qi})| > |\sqrt{n}z_{qi}| \mid z_{qi}\big) \mid z_{qi}^2 \le \alpha\big)\Pr(z_{qi}^2 \le \alpha),\\ &nM^2 E\big(z_{qi}^2 \Pr\big(|\sqrt{n}(z_{n,qi} - z_{qi})| > |\sqrt{n}z_{qi}| \mid z_{qi}\big) \mid z_{qi}^2 > \alpha\big)\Pr(z_{qi}^2 > \alpha).\end{aligned}$$

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

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

Using

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

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

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

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

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

so that the first line of equation (C.8) is bounded by

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

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

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

which is $o_P(1)$.

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

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

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

As for $B_n^{12}(q)$, first we can use the Cauchy–Schwarz inequality to show that

$$\begin{aligned}
|B_n^{12}(q)| &< \big(E[\sqrt{n}(z_{n,qi} - z_{qi})(\bar{y}_i - \underline{y}_i)]^2\big)^{1/2} \\
&\qquad \times \big(E\big(|\mathbf{1}\{z_{n,qi} > 0\} - \mathbf{1}\{z_{qi} > 0\}|\mathbf{1}\{z_{qi} \neq 0\}\big)\big)^{1/2}.
\end{aligned}$$

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

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

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

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

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,

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

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

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

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 $\sqrt{n}(\hat{\Sigma}_n^u - \hat{\Sigma}_n) \xrightarrow{P}_{n\to\infty} 0$ by equation (C.3) and uniform bounds on $q, \Sigma$, and $\beta_q^*$. Moreover, as $\hat{\Sigma}_n$ is bounded and its inverse exists, $\Sigma^\top(\hat{\Sigma}_n^\top)^{-1} \xrightarrow{\text{a.s.}}_{n\to\infty} I$ and we have, uniformly in $q$,

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

*Q.E.D.*

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

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

If there are no exposed faces (i.e., $\Pr(z_i\Sigma q = 0) = 0$), the second term is identically equal to zero, whereas $\varepsilon_{qi}^*$ boils down to the residual of the instrumental variable (IV) regression of $w_q$ onto $x$, 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

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

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

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

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

so that, using the asymptotic normality of the estimate of $\operatorname{vec}(\Sigma^\top)$ in equation (C.4), we have

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

where $\eta$ is a multivariate standard normal random variable of dimension $p^2$ independent of $z_i$. *Q.E.D.*

## C.2. *Proof of Proposition 10*

When $\beta_0$ 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. Therefore, $\sqrt{n}T_n(q_n)$ tends to $-\infty$ (resp. $+\infty$).

Consider now the case $\beta_0 \in \partial B$ and let $\mathcal{Q}(\beta_0)$ be the set of all $q_0 \in \mathbb{S}_p$ that minimize $T_\infty(q; \beta_0)$, that is, the set of all $q_0 \in \mathbb{S}_p$ that satisfy $\delta^*(q_0 \mid B) = q_0^\top \beta_0$. Set $\mathcal{Q}(\beta_0)$ is a nonempty 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

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

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 $\tau_n(\cdot)$ is stochastically equicontinuous (for instance, Andrews (1994, p. 2251)).

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

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

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

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

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

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 that tend to $q_0^*$.
3. We show that any sequence of minimizers of the empirical programs converges to $q_0^*$.

*Step 1. The selection of a single* $q_0^* \in \mathcal{Q}(\beta_0)$. For this, we select a vector oriented outward set $B$ and consider its projection on the smallest convex cone that 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 $\beta_0$, which is a frontier point of $B$, and any interior point $\beta^*$ of $B$. For instance the "center" of $B$ obtained by setting $u(z) = \frac{\underline{\Delta}(z)+\overline{\Delta}(z)}{2}$ is interior and

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

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

$$\delta^*(q_0 \mid B) - q_0^\top \beta^* > 0 \quad \Longrightarrow \quad q_0^\top v_0 > 0. \tag{C.9}$$

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

$$\min_{v, \delta^*(v|B) - v^\top \beta_0 \leq 0} \frac{(v_0 - v)^\top (v_0 - v)}{2}. \tag{C.10}$$

This projection is unique and defined by $v_0^* = \lambda^* q_0^*$, where $(\lambda^*, q_0^*)$ 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.9)) 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 $\mathcal{Q}(\beta_0)$.

*Step 2. Minimization programs whose well separated solutions converges to* $q_0^*$. The estimation of $q_0^*$ cannot proceed directly from program (C.10) since we do not know the set of constraints, $\mathcal{Q}(\beta_0)$. Consider the generalization of (C.10) for any $\alpha \geq 0$,

$$b(\alpha) \equiv \min_{v, \delta^*(v|B) - v^\top \beta_0 \leq \alpha} \frac{(v_0 - v)^\top (v_0 - v)}{2}, \tag{C.11}$$

where $b(\alpha)$ is continuous and nonincreasing in $\alpha$ because the constraint is continuous. The unique solution of this program, denoted $v_\alpha^*$, 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 Appendix C.2.3.

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

LEMMA 14: *For any $\alpha > 0$, the strictly convex program* (C.11) *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 $\alpha$.*

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:

$$\forall \varepsilon > 0, \exists a_0 > 0, \exists \eta > 0 \quad \text{such that} \quad \inf_{0<a\le a_0, \|q-q_0^*\|\ge\varepsilon} \frac{\Psi_a(q) - \Psi_a(q_a)}{a} > \eta. \tag{C.12}$$

*Step 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 $\hat{\beta}_n^* = \frac{1}{n}\sum_{i=1}^n \hat{\Sigma}_n^\top z_i \frac{\bar{y}_i + \underline{y}_i}{2}$.

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

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

We have

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

whereas the square-root uniform convergence of $\Psi_{n,a}$ to $\Psi_a$ 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 can write

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

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 $a_n = O(n^{-\alpha})$ be 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

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

Then

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

and, therefore,

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

By condition (C.12), for any $\varepsilon > 0$, there exist $n_0$ and $\eta > 0$ such that, for $n \geq n_0$, the event

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

Therefore,

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

To finish the proof of Proposition 10 we can now use the same argument as in Appendix C.2.1 so that

$$
\sqrt{n}\big(T_n\big(q_{n,a_n}; \beta_0\big) - T_n(q_0^*; \beta_0)\big) \underset{n\to\infty}{\overset{P}{\to}} 0.
$$

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

*Q.E.D.*

C.2.3. *Proofs of Lemmas 14–16*

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 $v_\alpha^*$ 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

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

When $\frac{v_0}{\|v_0\|} \notin \mathcal{Q}(\beta_0)$ and $\alpha$ runs from 0 to $\alpha_0$, then $v_\alpha^*$ 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))

$$\alpha = \min_{v, (v_0 - v)^\top (v_0 - v)/2 \leq b(\alpha)} \left(\delta^*(v \mid B) - v^\top \beta_0\right), \tag{C.13}$$

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.13) as the Lagrangian where $a > 0$,

$$\begin{aligned} L(v, a) = {} & \delta^*(v \mid B) - v^\top \beta_0 \\ & + a\left(\frac{(v_0 - v)^\top (v_0 - v)}{2} - b(\alpha)\right), \end{aligned} \tag{C.14}$$

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 $\beta_{q_\alpha} = \frac{\partial \delta^*(v|B)}{\partial v}|_{v_\alpha^*}$. 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 $\alpha = 0$, then $\beta_{q_\alpha} = \beta_0$ 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 $\alpha$ for any $\alpha < \alpha_0$ since all objects in the expression are continuous.

We now prove that $a(\alpha)$ is increasing with $\alpha$. Consider $0 < \alpha < \alpha' < \alpha_0$ and the optimal solutions $v_\alpha^*$ 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,

$$\begin{aligned} L(v_\alpha^*, a(\alpha)) &= \delta^*(v_\alpha^* \mid B) - v_\alpha^{*\top} \beta_0 \\ &< \delta^*(v_{\alpha'}^* \mid B) - v_{\alpha'}^{*\top} \beta_0 \\ &\quad + a(\alpha)\left(\frac{(v_0 - v_{\alpha'}^*)^\top (v_0 - v_{\alpha'}^*)}{2} - b(\alpha)\right), \\ L(v_\alpha^*, a(\alpha')) &= \delta^*(v_\alpha^* \mid B) - v_\alpha^{*\top} \beta_0 \\ &\quad + a(\alpha')\left(\frac{(v_0 - v_\alpha^*)^\top (v_0 - v_\alpha^*)}{2} - b(\alpha')\right) \\ &> \delta^*(v_{\alpha'}^* \mid B) - v_{\alpha'}^{*\top} \beta_0, \end{aligned}$$

so that by differencing,

$$\begin{aligned} &a(\alpha')(b(\alpha) - b(\alpha')) > -a(\alpha)(b(\alpha') - b(\alpha)) \\ &\quad \Rightarrow \quad (a(\alpha') - a(\alpha))(b(\alpha) - b(\alpha')) > 0. \end{aligned}$$

As $b(\alpha)$ is nonincreasing, it implies that $a(\alpha)$ is increasing with $\alpha$ 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 $\infty$. Let us rewrite the minimization program (C.14) so as to consider vectors on $\mathbb{S}_p$, since estimates are defined on $\mathbb{S}_p$ only:

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

Minimizing with respect to $\lambda$ yields the first order condition for the optimal solution $\lambda_q$,

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

which implies that

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

When $a > 0$, minimizing $L(\lambda_q q, a)$ is equivalent to maximizing $\lambda_q$ and, thus, is 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. *Q.E.D.*

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 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

$$0 \leq \delta^*(q_a \mid B) - q_a^\top \beta_0 \leq a(q_a - q_0^*)^\top v_0 \leq 2a\|v_0\|, \tag{C.15}$$

since $\beta_0 \in B$ (the left-hand side, $\delta^*(q_a \mid B) - q_a^\top \beta_0$, is nonnegative) 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 $q_m$ to be any accumulation point of the sequence $q_a$, that is, 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

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

where the first inequality holds true because $q_a$ minimizes $\Psi_a$ 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

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

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

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 $\max_{q \in \mathcal{Q}(\beta_0)} q^\top v_0$. But this argument is unique and is precisely $q_0^*$. Hence, we have necessarily $q_m = q_0^*$ and therefore,

$$\text{(C.16)} \quad \lim_{a \to 0} \|q_a - q_0^*\| = 0.$$

Furthermore, as

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

we have

$$\text{(C.17)} \quad \Psi_a(q_0^*) - \Psi_a(q_a) = o(a). \qquad \textit{Q.E.D.}$$

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,

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

the last equality resulting from the expansion

$$\text{(C.18)} \quad \|q\|^2 = 1 = \left\| q - \frac{v_0}{\|v_0\|} + \frac{v_0}{\|v_0\|} \right\|^2$$

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

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

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

- We have $q \in \mathcal{Q}(\beta_0)$ 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 $\|q - q_0^*\| \geq \varepsilon$.

• We have $q \notin \mathcal{Q}(\beta_0)$. 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 nonpositive infimum can only be obtained at some $a > 0$. This 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 $q$s is thus positive also. *Q.E.D.*

## 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 the end of Appendix B.4). As $\mathbb{S}_p$ is compact, let $\Lambda$ denote a compact set of $\mathbb{R}^m$ such that for all $q \in \mathbb{S}_p$, $\lambda_0(q) \in \text{int}(\Lambda)$.

The proof consists of 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 $\lambda$ 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 $\lambda$ 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 is 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, p. 2251), 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 $\tau_n^U(\cdot)$, defined for $s = (q, \lambda) \in \mathbb{S}_m$, the unit sphere in $\mathbb{R}^m$, as

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

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

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

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

$\mathcal{S}(q)$ is a nonempty 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,

$$\delta^*(s \mid B_U) = \|s\| \delta^*\left(\frac{s}{\|s\|} \;\middle|\; B_U\right),$$

where $\frac{s}{\|s\|} \in \mathbb{S}_m$. In the following discussion, we 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, $\mathcal{S}(q) = \{s_0 = (q, \lambda_0)\} \subset \mathrm{int}(S_B)$. 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:

$$\hat{\delta}^*_n(q \mid B) = \hat{\delta}^*_n(s_n \mid B_U). \tag{C.19}$$

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

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

Second, because (i) $\tau^U_n(\cdot)$ is stochastically equicontinuous, (ii) $s_n \in \mathbb{S}_p \times \Lambda$, and (iii) $\mathrm{plim}_{n \to \infty} s_n = s_0$, Andrews (1994, equation (3.36), p. 2265) showed that

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

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

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

which implies that

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

using equation (C.19) and where $V_{s_0}$ is consistently estimated by $V_{s_n}$.

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

(b) Suppose now that $\mathcal{S}(q)$ is not a singleton because there are various minimizers of $\delta^*(s \mid B_U)$ in $\lambda$. Note first that set $\mathcal{S}(q) \subset \text{int}(S_B)$ is convex and compact because $\delta^*(s \mid B_U)$ is convex and continuous. We first select and characterize a unique $(q, \lambda_0^*)$ from $\mathcal{S}(q)$. Consider the smallest convex cone that 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 is defined by $c^* s_0^*$, where $(c^*, s_0^*)$ is the argument of the minimum,

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

since $q^\top q = 1$. It yields $c^* = \frac{1}{1+\lambda^\top\lambda} > 0$, whereas $\lambda_0^*$ is the argument of

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

Vector $\lambda_0^*$ 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 $\lambda$ and $\lambda^\top\lambda$ is strictly convex in $\lambda$, then $\Psi_a(s)$ is a strictly convex function in $\lambda$.

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 $\lambda_a$ tends to $\lambda_0^*$ when $a \to 0$.

LEMMA 17: *The limit of the sequence* $\{\lambda_a\}_{a>0}$ *exists when* $a \to 0$ *and is equal to* $\lambda_0^*$.

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 $\lambda_a$ 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 $\lambda_a$ as a minimum,

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

It implies that

$$0 \le \delta^*(s_a \mid B_U) - \delta^*(s_0^* \mid B_U) \le a(\lambda_0^{*\top}\lambda_0^* - \lambda_a^\top \lambda_a) \le a\lambda_0^{*\top}\lambda_0^*, \tag{C.20}$$

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 $\lambda_m$ to be any accumulation point of the sequence $\lambda_a$, that is, 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 $\lambda_a$ and $\lambda_0^*$, we have

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

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

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

Combining the two equations gives

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

By taking limits and using that $\lambda_a$ tends to $\lambda_m$ 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

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

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

$$\Psi_a(s_0^*) - \Psi_a(s_a) = o(a). \tag{C.21}$$

*Q.E.D.*

The next step is to construct an estimate of $\lambda_a$. Before moving on to this step, we prove a lemma that will be useful for showing that the estimate of $\lambda_a$ actually converges to $\lambda_0^*$.

LEMMA 18: *We have*,

$$\forall \varepsilon > 0, \exists a_0 > 0, \exists \eta > 0, \forall a \quad \textit{such that} \quad \inf_{0<a\leq a_0, \|\lambda-\lambda_0^*\|\geq\varepsilon} \frac{\Psi_a(s) - \Psi_a(s_a)}{a} > \eta. \tag{C.22}$$

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

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

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

Assume that $(q, 0) \notin \mathcal{S}(q)$. As in Lemma 16, we 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 Lemma 17 that $\lambda \to \lambda_0^*$ when $a$ tends to 0.

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

As $\lambda_a \to \lambda_0^*$ when $a$ tends to zero, there exist some positive $a_0$ such that the joint events $\{0 < a < a_0\}$ and $\{\|\lambda - \lambda_0^*\| \geq \varepsilon\}$ imply 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 nonpositive infimum can only be obtained at some $a > 0$ and $s_a \in S_B$. This 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 $s$s in this set is positive also. *Q.E.D.*

Finally, we construct the estimate of $\lambda_a$. 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 used in the proof of Proposition 10 to show that

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

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

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

Then

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

We thus have

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

By condition (C.22), for any $\varepsilon$ there exist $\eta > 0$ and $n_0$ such that, for any $n \geq n_0$, the event

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

Therefore,

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

Then the same argument in part (i) applies and

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

We can then use the asymptotic distribution of $\hat{\delta}_n^*(s_0^* \mid B_U)$ in place of $\hat{\delta}_n^*(s_{a_n,n} \mid B_U)$. By the same development, it applies to a finite vector of such estimates defined at values $q_1, q_2, \ldots, q_J$.

*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 $\delta$ such that

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

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

$$\forall \varepsilon, \forall \eta, \exists \delta > 0, \quad \lim_{n\to\infty} \Pr\Big[\sup_{s_1,s_2 \in S_B, \|s_1 - s_2\| < \delta} |\tau_n^U(s_1) - \tau_n^U(s_2)| > \eta\Big] < \varepsilon. \tag{C.23}$$

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 defined by the argument used in Step 2(b) if they are not. Consider the difference

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

Alternatively, consider

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

Consequently,

$$\begin{aligned}
&|\hat{\delta}_n^*(s_{1n} \mid B_U) - \hat{\delta}_n^*(s_{2n} \mid B_U)| \\
&\quad \leq \max_{\lambda \in \Lambda} \big| \hat{\delta}_n^*((q_1, \lambda) \mid B_U) - \hat{\delta}_n^*((q_2, \lambda) \mid B_U) \big|.
\end{aligned}$$

By definition,

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

so that for any $(\varepsilon, \eta)$and for any $\delta$ satisfying equation (C.23),

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

This proves that the process $\tau_n(q)$ is equicontinuous by equation (C.23).

The proof when the minimizers are replaced by near minimizers can be adapted in a straightforward way. *Q.E.D.*

## APPENDIX D: ADDITIONAL EXPERIMENTS AND ANALYSIS

### 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 2. 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 $\varepsilon$ is independent of $x$ and uniformly distributed on $[-1/2, 1/2]$. The true value of $\beta$ is $(0, 0)^\top$. We

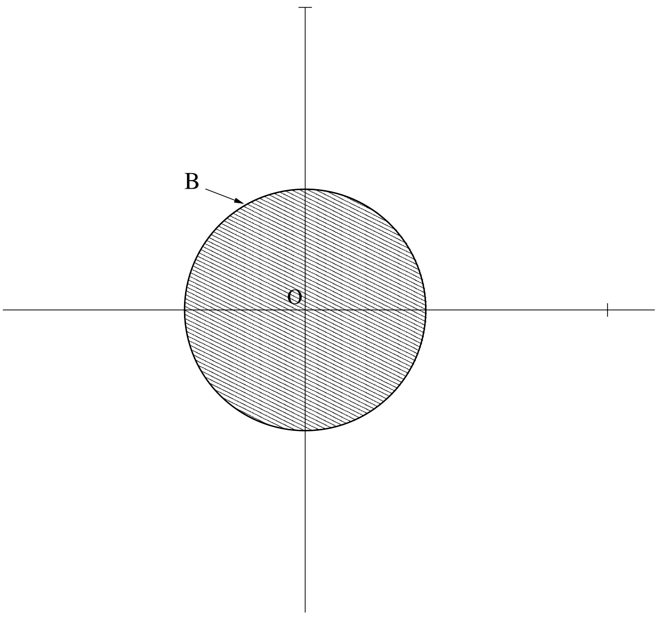


FIGURE D.1.—Set $B$, $y = 0.x_1 + 0.x_2 + \varepsilon$, $(x_1, x_2)^\top \sim N(0, I_2)$.

assume that $y^*$ is observed by intervals defined as ($I_k = [-1/2 + k/K; -1/2 + (k+1)/K]$, $k = 0, \dots, K-1$).

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 Figure D.1).

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 V, 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 $\beta^r$ a point on a ray such that the distance between 0 and $\beta^r$ 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 $\beta^r$ belongs to $B$ if and only if $r \leq 1$ and $\beta^1$ belongs to the frontier. For $r$ varying stepwise from

TABLE V

RESULTS RELATED TO THE ADDITIONAL MONTE CARLO SIMULATIONS IN APPENDIX D.1.1 FOR THE SET $B$ IN FIGURE D.1[a]

| $n$ | 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 |

[a] Support function $\delta(q)$ for $q = (0, 1)^\top$. True unknown value 0.199.

0 to 3, we computed the rejection frequencies at a 5% level for the two tests developed in Section 4.2: whether $\beta^r$ 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 VI. 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 independent and identically distributed (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 $\pi$ (respectively $\nu$) 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 $\pi$ and $\nu$ are positive and strictly lower than 1, there is some information loss due to the use of $e_2$ and $w$ instead of $x_2$, and set $B$ is stretched along the second axis (see Figure D.2).

TABLE VI

PERCENTAGE OF REJECTIONS FOR THE TWO TESTS IN THE EXAMPLE IN APPENDIX D.1.1[a]

| | 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% |

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

As before, we draw 1000 simulations in four sample size experiments: $n$ =100, 500, 1000, and 2500 with $\pi = \nu = 0.9$. Table VII displays descriptive statistics (mean and quartiles) related to the distribution of the estimated support function at one angle. Table VIII displays the percentage of rejections for the tests for different points along the $x$-axis. The line that corresponds to the frontier point ($r = 1$) is reported in bold. As before, there is no significant distortion when using supernumerary instruments in the estimation and test procedures.

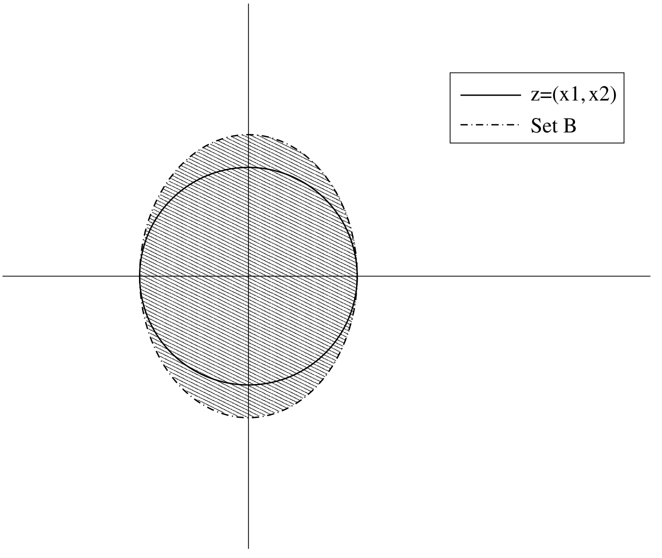


FIGURE D.2.—Set $B$, $y = 0.x_1 + 0.x_2 + \varepsilon$, $z = (x_1, e_2, w)$.

### 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 $\varepsilon$ is independent of $x$ and is uniformly distributed on $[-\frac{1}{4}, \frac{1}{4}]$. The true value of $\beta$ is $(\frac{1}{2}, \frac{1}{8})^\top$. As before, we only observe $y^*$ by intervals ($I_1 = [0, \frac{1}{2}]$

TABLE VII

RESULTS RELATED TO THE ADDITIONAL MONTE CARLO SIMULATIONS IN APPENDIX D.1.2 WITH SUPERNUMERARY INSTRUMENTS FOR THE SET $B$ IN FIGURE D.2[a]

| $n$ | 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 |

[a]Support function $\delta(q)$ for $q = (0, 1)^\top$. True unknown value 0.243.

TABLE VIII

PERCENTAGE OF REJECTIONS FOR THE TWO TESTS IN THE EXAMPLE IN APPENDIX D.1.2[a]

| | 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% |

[a]The point tested $\beta^r$ is located on the $x$-axis. $r$ is the fraction of the distance from the origin with respect to the distance origin–frontier point on this axis. $r = 1$ is the frontier point (results in bold), $r = 0$ to the origin.

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 four sample sizes: 100, 500, 1000, and 2500. The same conclusions concerning the estimation of the set remain valid here (see Figure D.3 and Table IX).

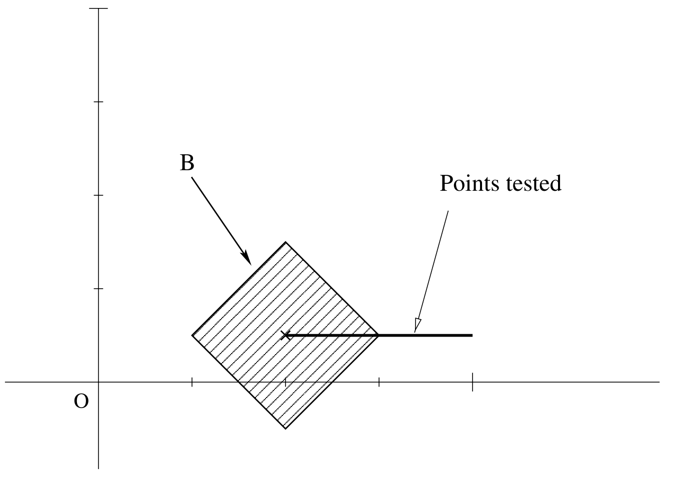


FIGURE D.3.—Set $B$, $y = \frac{1}{2} + \frac{x}{8} + \varepsilon$, $x \in \{-1, 1\}$.

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. As 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 X 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 XI reports the results for the test for the interior. In the columns labeled $a_n = 0$, we display results that ignore 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 do overreject for the interior test. In the panels

TABLE IX

RESULTS RELATED TO THE ADDITIONAL MONTE CARLO SIMULATIONS IN APPENDIX D.1.2 FOR THE NONSMOOTH SET SHOWN IN FIGURE D.3[a]

| $n$ | 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 |

[a]Support function $\delta(q)$ for $q = (0, 1)^\top$. True unknown value 0.375.

TABLE X

PERCENTAGE OF REJECTIONS FOR THE TEST $H_0 : \beta^r \in \partial B$ FOR THE NONSMOOTH SET OF APPENDIX D.1.3

| | 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% |
| **1.000** | **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 XI

PERCENTAGE OF REJECTIONS FOR THE TEST $H_0 : \beta^r \in B$ FOR THE NONSMOOTH 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% |

labeled $a_n = \frac{0.5}{n^{1/3}}$, we display the rejection rates using the perturbed program defined in Proposition 10. Rejection rates are pretty close to the nominal size for both the frontier test and the interior test. Perturbing the program leads to quite efficient correction 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 Appendix D.1*

### D.2.1. *Example in Appendix 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

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

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

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

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 in Appendix D.1.2*

The simulated model is

$$y^* = 0x_1 + 0x_2 + \varepsilon, \quad x_2 = \pi e_2 + \sqrt{1-\pi^2} e_3, \quad w = \nu e_3 + \sqrt{1-\nu^2} e_4,$$

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

$$E(x^\top z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \pi & \mu \end{pmatrix},$$

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

and

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

which is standard unit bivariate normally distributed. Moreover, as

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

the normalized vector $(0 \quad \frac{\mu}{a} \quad -\frac{\pi}{a})^\top$ belongs to the kernel of this operator and, consequently, $z_H = \frac{\mu e_2 - \pi w}{a}$.

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

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

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

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

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

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

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. Consequently, 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 in Appendix D.1.3*

The simulated model is

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

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

$$z_q = z\Sigma q = \cos\theta + x\sin\theta,$$

$$w_q = \underline{y} + \frac{1}{2}\mathbf{1}\{z_q > 0\},$$

$$\underline{y} = \frac{1}{2}\mathbf{1}\{y^* \geq 0.5\}.$$

As $E(\underline{y}) = \frac{1}{4}$ and $E(x\underline{y}) = \frac{1}{8}$, we can derive the frontier points

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

Let $\theta_0 = \pi/4$. For $\theta$ between $-\theta_0$ and $\theta_0$, $z_q$ is always positive regardless of the value of $x$:

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

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

For $\theta$ between $\theta_0$ and $-\theta_0 + \pi$, $z_q$ is negative when $x = -1$; otherwise it is positive:

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

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

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

The term $\tau_1(q)$ defined in Proposition 9 is equal to zero when $P(z_q = 0) = 0$, that is, 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 $\theta$.

### D.3. *Proof of Proposition 8*

We denote by $M$ 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 $\Theta$. 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

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

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

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

we have

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

Use equation (C.5),

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

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

$$|f_{\hat{\theta}_n}(z_i, \underline{y}_i, \overline{y}_i) - f_{\theta}(z_i, \underline{y}_i, \overline{y}_i)| \underset{n\to\infty}{\overset{\text{a.s.}}{\to}} 0. \tag{D.2}$$

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

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

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

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

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

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

Thus, equation (D.2) implies that, by the dominated convergence theorem, uniformly over $q$,

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

From the latter equation, equation (D.1), and the triangular inequality, we thus conclude that, uniformly for $q \in \mathbb{S}$,

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

*Q.E.D.*

### D.4. *Construction of the Confidence Region in Proposition 11*

As before, for simplicity of exposition, we focus on the case where $B$ is strictly convex and smooth. We here provide a simple way to construct $\mathrm{CI}_\alpha^n$ when $\alpha < 1/2$; that is,

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

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

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

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

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

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

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

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

We have

$$T_n(q; \beta_f(\lambda)) = T_n(q; \beta_f) + q^\top(\beta_f - \beta_f(\lambda))$$
$$= T_n(q; \beta_f) - \lambda q^\top q_f,$$

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 $T_n(q;\beta_f)$, also minimizes $T_n(q;\beta_f(\lambda))$.

We can therefore characterize the points of the half-line that belongs to $\mathrm{CI}_\alpha^n$. Given that $\lambda$ is positive,

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

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

$$\hat{B}_n \cup \left\{ \bigcup_{\beta_f \in \partial B_n} \bigcup_{q_f \in \mathcal{C}(\beta_f)} \left(\beta_f, \beta_f - \frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha q_f\right)\right\} \subset \mathrm{CI}_\alpha^n, \tag{D.3}$$

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

Conversely, let us prove that $\mathrm{CI}_\alpha^n$ is included in the set on the left-hand side (LHS). Let $\beta_c$ be a point in $\mathrm{CI}_\alpha^n$. If $\beta_c$ belongs to $\hat{B}_n$, the inclusion is proved. Assume that $\beta_c$ is outside the estimated set and let $\beta_f$ be the point on the frontier of $\hat{B}_n$ that is the projection of $\beta_c$ 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 $\beta_f$ is the projection of $\beta_c$ 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 $\hat{\delta}_n^*(q_f \mid B) = q_f^\top \beta_f$. The pair $(\beta_f, q_f)$ satisfies the condition of the previous paragraphs.

As $\beta_c$ is a point of $\mathrm{CI}_\alpha^n$, then $\lambda$ is necessarily less than or equal to the value $-\frac{\sqrt{\hat{V}_{q_f}}}{\sqrt{n}}\mathcal{N}_\alpha$. Thus it belongs to the LHS of equation (D.3). As a consequence, equation (D.3) is an equality. *Q.E.D.*

### D.5. *Behavior of $\xi_n(\beta)$ When the Set Is a Singleton*

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

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

where $V$ is some positive definite matrix.

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

When $\beta = \beta_0$,

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

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

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, that is, $\mathcal{N}_\alpha$, 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 Appendix B.1, the width of set $B$ for direction $q$ is equal to

$$\Delta(q) = \delta^*(q \mid B) + \delta^*(-q \mid B) = E(|z_{qi}|(\bar{y}_i - \underline{y_i})).$$

As by assumption, $\bar{y}_i - \underline{y_i} > 0$ and $\Pr(z_{qi} = 0) < 1$ because of the rank condition in Assumption R.2, the limit point $\Delta = 0$ is outside the range of data that we consider.

Its empirical counterpart is

$$\hat{\Delta}_n(q) = \left(\frac{1}{n}\sum_{i=1}^{n} |z_{n,qi}|(\bar{y}_i - \underline{y}_i)\right).$$

Therefore, if we extend our setting to include the case $\Delta = 0$ because $|z_{qi}|(\bar{y}_i - \underline{y_i}) = 0$ almost surely $z_i$, we have

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

so that trivially

$$\sqrt{n}(\hat{\Delta}_n(q) - \Delta(q)) \overset{P}{\underset{n\to\infty,\Delta(q)=0}{\longrightarrow}} 0.$$

More generally, consider a sequence of experiments indexed by $\varepsilon \downarrow 0$ such that $\Pr(|z_{qi}|(\bar{y}_i - \underline{y_i}) < \varepsilon) = 1$ and $\Pr(|z_{qi}|(\bar{y}_i - \underline{y_i}) > \varepsilon/2) > 0$. Therefore, $\Delta(q) = E(|z_{qi}|(\bar{y}_i - \underline{y_i}))$ and $\varepsilon$ go to zero at the same rate.

We have

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

which has a variance approximately equal to $V(|z_{qi}|(\bar{y}_i - \underline{y}_i))$ which is bounded by a term $O_P(\varepsilon^2)$ and, therefore, $O_P(\Delta^2)$. We thus have

$$\sqrt{n}|\hat{\Delta}_n(q) - \Delta(q)| \leq O_P(\Delta(q)).$$

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*

$$\hat{\sigma}_q = \sqrt{\hat{V}_q} = \sqrt{q^\top \hat{\Sigma}_n \hat{V}(z^\top \varepsilon_q)\hat{\Sigma}_n q}.$$

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

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

The overall confidence region $\tilde{\mathrm{CI}}_\alpha^n$, which is the union of the previous sets, is then characterized by

$$\lim_{n \to +\infty} \inf_{\beta \in B, P \in \mathbb{P}} \Pr(\beta \in \tilde{\mathrm{CI}}_\alpha^n) = 1 - \alpha,$$

in which $\mathbb{P}$ is the set of probability distributions that satisfy the condition

$$\mathbb{P} = \{P(\bar{y}_i, \underline{y}_i, z_i) \text{ such that } \forall q; \Pr(z_{qi} = 0) = 0 \text{ and Assumption R}\}.$$

## REFERENCE

ANDREWS, D. W. K. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: North Holland, 2247–2294. [20,28,29]