APPENDIX A: MAIN RESULTS

A.1. *Lemma 1*

We now show that Assumptions 1 and 2 imply that given a collection $r \in \mathbb{R}^{n\times m}$ of 'active' constraints ($m \leq n-1$) the maximum response is determined in closed-form (and up to sign) by the Karush-Kuhn-Tucker conditions of programs (2.5) and (2.6). The following Lemma constitutes the basis of Theorem 1.

Lemma 1 *Suppose that Assumptions 1 and 2 hold. Let r be a matrix of dimension $n \times m$ collecting the gradients of the 'active' (binding) constraints at a solution $x^*(\mu)$ of the mathematical program (2.5), then :*

a) $\overline{v}_{k,i,j}(\mu)$ is given by either plus or minus the norm of the residual of the projection of $\Sigma^{1/2}C_k(A)'e_i$ into the space spanned by the columns of $\Sigma^{1/2}r$; that is

$$\overline{v}_{k,i,j}(\mu) = \Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}, \tag{A.1}$$

or

$$\overline{v}_{k,i,j}(\mu) = -\Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}, \tag{A.2}$$

where

$$M_{\Sigma^{1/2}r} \equiv \mathbb{I}_n - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma^{1/2}.$$

b) If in addition $\overline{v}_{k,i,j}(\mu) \neq 0$, then $x^*(\mu)$ is given by

$$x^*(\mu) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \overline{v}_{k,i,j}(\mu).$$

Consequently, the sign of $\overline{v}_{k,i,j}(\mu)$ depends on which of the two values of $x^*(\mu)$ in the equation above (the one with (A.1) in the denominator or the one with (A.2)) satisfies the sign restrictions that are not in r.

Proof: Let $S(\mu)$ denote the $n \times m_s$ matrix of m_s 'sign' restrictions and let $Z(\mu)$ denote the $n \times m_z$ matrix of 'zero' restrictions. For notational simplicity, we deliberately ignore the dependence of the equality/inequality restrictions on μ. The problem in equation (2.5) is equivalent to

$$\overline{v}_{k,i,j}(\mu) \equiv \max_{x\in\mathbb{R}^n} e_i'C_k(A)x \quad \text{subject to} \;\; x'\Sigma^{-1}x = 1, \quad S'x \geq \mathbf{0}_{m_s\times 1}, \quad Z'x = \mathbf{0}_{m_z\times 1}. \tag{A.3}$$

The auxiliary Lagrangian function is given by

$$\mathcal{L}(x,\lambda,w_1,w_2;\mu) = e_i'C_kx - \lambda(x'\Sigma^{-1}x-1) - w_1'(S'x) - w_2'(Z'x).$$

Assumptions 1–2 imply that we can characterize the maximum response using the Karush-Kuhn-Tucker conditions for the mathematical program in (2.5). The Karush-Kuhn-Tucker necessary conditions for this problem are as follows:

$$\begin{aligned}
\text{Stationarity} \quad &: \quad C_k'(A)e_i - 2\lambda\Sigma^{-1}x - Sw_1 - Zw_2 = \mathbf{0}_{n\times 1},\\
\text{Primal Feasibility} \quad &: \quad x'\Sigma^{-1}x = 1,\\
&\quad\;\; S'x \geq \mathbf{0}_{m_s\times 1},\\
&\quad\;\; Z'x = \mathbf{0}_{m_z\times 1},\\
\text{Complementary Slackness} \quad &: \quad w_{1i}(e_i'Sx) = 0 \quad \forall \quad i = 1\ldots m_s,
\end{aligned}$$

plus the additional dual feasibility constraint requiring the Lagrange multipliers, w_{1i}, to be smaller than or equal to zero.

Let $x^*(\mu)$ be one (out of possibly many) maximizers of the program of interest and suppose that the $n \times m$ matrix r collects all the restrictions that are active (binding). Because of Assumption 1 and 2 , the matrix r is of full rank m and m must be smaller than or equal $n-1$. Using Stationarity, Primal Feasibility, and Complementary Slackness at x^* we get

$$\begin{aligned}
0 = x^{*\prime}[C_k(A)'e_i - 2\lambda^*\Sigma^{-1}x^* - Sw_1 - Zw_2] &= x^{*\prime}C_k(A)'e_i - 2\lambda^* x^{*\prime}\Sigma^{-1}x^* - x^{*\prime}Sw_1 - x^{*\prime}Zw_2 \\
&= x^{*\prime}C_k(A)'e_i - 2\lambda^* - x^{*\prime}Sw_1 - x^{*\prime}Zw_2 \\
&\quad \text{(where we have used } x^{*\prime}\Sigma^{-1}x^* = 1) \\
&= x^{*\prime}C_k(A)'e_i - 2\lambda^* \\
&\quad \text{(where we have used } x^{*\prime}Z = \mathbf{0}_{m_z \times 1} \text{ and complementary slackness)} \\
&= \overline{v}_{k,i,j}(\mu) - 2\lambda^*,
\end{aligned}$$

where $\overline{v}_{k,i,j}(\mu)$ denotes the value of the maximum response when the constraints in r are active. Thus, the Lagrange multiplier λ^* is unique and given by

$$\lambda^* = \frac{1}{2}\overline{v}_{k,i,j}(\mu).$$

Note also that $\lambda^* \neq 0$ if and only if $\overline{v}_{k,ij}(\mu; r) \neq 0$. We now show that there are unique w_1^* and w_2^* that satisfy the Karush-Kuhn Tucker conditions. Let w^* denote the nonzero components of w_1^* and all the components of w_2^*. Note that left multiplying the stationarity condition by Σ and rearranging the terms we have :

$$2\lambda^* x^{*\prime} = \Big(C_k(A)'e_i - rw^*\Big)'\Sigma,$$

$$\begin{aligned}
\text{(A.4)} \qquad \Big(C_k(A)'e_i - rw^*\Big)'\Sigma\Big(C_k(A)'e_i - rw^*\Big) &= 4(\lambda^*)^2 x'\Sigma^{-1}x \\
&= 4(\lambda^*)^2 \\
&\quad \text{(where we have used} \quad x^{*\prime}\Sigma^{-1}x^* = 1) \\
&= 4\Big(\frac{1}{2}\overline{v}_{k,i,j}(\mu)\Big)^2 \\
&= \overline{v}_{k,i,j}(\mu)^2.
\end{aligned}$$

Consequently the value function given active constraints r is given by either

$$\overline{v}_{k,i,j}(\mu) = \Big[\Big(C_k(A)'e_i - rw^*\Big)'\Sigma\Big(C_k(A)'e_i - rw^*\Big)\Big]^{1/2},$$

or

$$\overline{v}_{k,i,j}(\mu) = -\Big[\Big(C_k(A)'e_i - rw^*\Big)'\Sigma\Big(C_k(A)'e_i - rw^*\Big)\Big]^{1/2}.$$

We will use the first order conditions to find the vector of Lagrange multipliers w^* and show that they are unique. Note that

$$\begin{aligned}
0 = 2\lambda^* r'x^* &= \Big[r'\Sigma(C_k(A)'e_i - rw^*)\Big] \\
&= \Big[r'\Sigma C_k(A)'e_i - r'\Sigma rw^*\Big].
\end{aligned}$$

Under the assumptions of the lemma, r is of rank m. The equation above holds if and only if

$$w^* = (r'\Sigma r)^{-1} r'\Sigma C_k(A)'e_i.$$

Consequently, the Lagrange multipliers for the active restrictions are unique. To conclude the proof, we get an explicit expression of the value function in terms of μ. To do so, note that

$$\begin{aligned}
\Sigma^{1/2}\Big(C_k(A)'e_i - rw^*\Big) &= \Sigma^{1/2}C_k(A)'e_i - \Sigma^{1/2}rw^* \\
&= \Sigma^{1/2}C_k(A)'e_i - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma C_k(A)'e_i \\
&= \Big(\mathbb{I}_n - \Sigma^{1/2}r(r'\Sigma r)^{-1}r'\Sigma^{1/2}\Big)\Sigma^{1/2}C_k(A)'e_i \\
&= \Big(\mathbb{I}_n - P_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \\
&= M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i.
\end{aligned}$$

Therefore, the equation above and (A.4) imply that either

$$\overline{v}_{k,i,j}(\mu) = \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}$$

or

$$\overline{v}_{k,i,j}(\mu) = -\Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}.$$

Furthermore, since any solution for which r is the set of binding constraints satisfies $2\lambda^* x^{*\prime} = (C_k(A)'e_i - rw^*)'\Sigma$, then for any $\overline{v}_{k,i,j}(\mu) \neq 0$ the solution x^* should be given by either

$$x^* = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2},$$

or

$$x^* = -\Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big[e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big]^{1/2}.$$

In any case the Lagrange multipliers for the active constraints are given (as shown above) by,

$$w^* = (r'\Sigma r)^{-1} r'\Sigma C_k(A)'e_i.$$

A.2. *Proof of Theorem 1*

The choice set of program (2.5) is non-empty (by Assumption 1) and compact (because of the ellipsoid constraint $BB' = \Sigma$). Hence, the maximum exists. Let $x^* \in \mathbb{R}^n$ be a solution and let r^* be the set of constraints that are active at x^*.

Step 1: We show first that

$$\overline{v}_{k,i,j}(\mu) \geq \max_{r\in R}\Big(\max\{f^+_{max}(\mu;r), f^-_{max}(\mu;r)\}\Big).$$

We do so by considering two different cases.

Case 1.1: Take any $r \in R$, and assume first that $v_{k,i,j}(\mu;r) \neq 0$. If $\mathbf{1}_{m_s}(x^*_+(\mu;r)) = 0$, then

$$f^+_{max}(\mu;r) = v_{k,i,j}(\mu;r) - 2c \leq c - 2c = -c < \overline{v}_{k,i,j}(\mu),$$

where the first equality above follows from the definition of f^+_{max} and the two remaining inequalities follow from the definition of the penalty term c.

Note, however, that if $r \in R$ is such that $\mathbf{1}_{m_s}(x^*_+(\mu;r)) = 1$, then $x^*_+(\mu;r)$ satisfies all the equality and inequality restrictions in (2.5) and, by construction, also satisfies the ellipsoid constraint

$$x^*_+(\mu;r)'\Sigma^{-1}x^*_+(\mu;r) = 1.$$

Consequently, $\overline{v}_{k,i,j}(\mu) \geq f^+_{max}(\mu;r)$ for all $r \in R$. An analogous argument shows that $\overline{v}_{k,i,j}(\mu) \geq f^-_{max}(\mu;r)$. This implies that

$$\overline{v}_{k,i,j}(\mu) \geq \max\{f^+_{max}(\mu;r), f^-_{max}(\mu;r)\},$$

for all $r \in \mathbb{R}$ such that $v_{k,i,j}(\mu;r) \neq 0$.

Case 1.2: Consider now any r such that $v_{k,i,j}(\mu;r) = 0$. If there is no feasible point x^* that gives such a value, then $f^+_{max}(\mu;r) = f^-_{max}(\mu;r) = -2c < \overline{v}_{k,i,j}(\mu)$. If there is such a feasible point $x^* \neq 0$ then $f^+_{max}(\mu;r) = f^-_{max}(\mu;r) = 0$. Since x^* is in the choice set of the program (2.5), then $f^+_{max}(\mu;r) = f^-_{max}(\mu;r) = 0 \leq \overline{v}_{k,i,j}(\mu)$.

Therefore, Case 1.1 and 1.2 imply that

$$\overline{v}_{k,i,j}(\mu) \geq \max\{f^+_{max}(\mu;r), f^-_{max}(\mu;r)\} \quad \text{for all } r \in R.$$

Step 2: We now show that

$$\overline{v}_{k,i,j}(\mu) \leq \max_{r\in R}\Big(\max\{f^+_{max}(\mu;r), f^-_{max}(\mu;r)\}\Big).$$

Again, we consider two cases.

Case 2.1: Assume first that $\overline{v}_{k,i,j}(\mu) \neq 0$. Without loss of generality, let us assume that $\overline{v}_{k,i,j}(\mu) > 0$ (the case in which $\overline{v}_{k,i,j}(\mu) < 0$ is completely analogous). Let $r^* \in R$ denote the set of active restrictions (which by Assumptions 1 and 2 has at most $n-1$ columns) at the solution x^* (this is one out of the potentially many solutions to the program). By Lemma 1 we know that

$$\overline{v}_{k,i,j}(\mu) = \Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r^*}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2},$$

and

$$x^*(\mu;r^*) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r^*}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ \Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r^*}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}.$$

Since this point satisfies the sign restrictions not in r^* (because it is a solution), then

$$\Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r^*}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2} = f^+_{max}(\mu;r^*).$$

Consequently,

$$\overline{v}_{k,i,j}(\mu) = f^+_{max}(\mu;r^*) \leq \max_{r\in R}\Big(\max\{f^+_{max}(\mu;r), f^-_{max}(\mu;r)\}\Big).$$

Case 2.2: If $\overline{v}_{k,i,j}(\mu) = 0$, there is an $x^* \neq 0$ in the choice set. Hence, the Karush-Kuhn-Tucker conditions imply that $C_k(A)'e_i$ is a linear combination of the active constraints that generate the value of zero (which means, by definition of the algorithm, that there is an r^* such that $f^+_{max}(\mu;r^*) = f^-_{max}(\mu;r^*) = 0$). Therefore, $\overline{v}_{k,i,j}(\mu) = f(\mu;r^*) \leq \max_{r\in R}\Big(\max\{f^+_{max}(\mu;r), f^-_{max}(\mu;r)\}\Big)$.

As the result, the value function $\overline{v}_{k,i,j}(\mu)$ is obtained by computing the Karush-Kuhn-Tucker points in Lemma 1 for each r, penalizing the value $\overline{v}_{k,i,j}(\mu;r)$ if not feasible, and maximizing over all the possible values of r.

The proof for the lower bound is analogous;

$$\underline{v}_{k,i,j}(\mu) = \min_{r\in R}\Big(\min\{f^{+}_{min}(\mu;r), f^{-}_{min}(\mu;r)\}\Big),$$

with:

$$\begin{aligned} f^{+}_{min}(\mu;r) &\equiv v_{k,i,j}(\mu;r) + 2(1-\mathbf{1}_{m_s}(x^*_+(\mu;r)))c, \\ f^{-}_{min}(\mu;r) &\equiv -v_{k,i,j}(\mu;r) + 2(1-\mathbf{1}_{m_s}(x^*_-(\mu;r)))c. \end{aligned}$$

A.3. *Lemma 2*

Lemma 2 *Suppose that Assumptions 1-3 hold. Let $r(\mu)$ be a matrix of dimension $n\times l$ collecting the gradients of the 'active' (binding) constraints at a solution $x^*(\mu)$ of the mathematical program (2.5) such that $v_{k,i,j}(\mu;r(\mu))\neq 0$. Then $v_{k,i,j}(\mu;r(\mu))$ is differentiable with respect to μ with the derivative $\dot{v}_{k,i,j}(\mu;r(\mu))$ given by*

$$\begin{bmatrix} \frac{\partial v_{k,i,j}(\mu;r(\mu))}{\partial vec(A)} \\ \\ \frac{\partial v_{k,i,j}(\mu;r(\mu))}{\partial vec(\Sigma)} \end{bmatrix} = \begin{bmatrix} \frac{\partial vec(C_k(A))}{\partial vec(A)}(x^*(\mu;r(\mu))\otimes e_i) - \sum_{k=1}^{l} w^*_k \frac{\partial vec(r_k(\mu))}{\partial vec(A)} x^*(\mu;r(\mu)) \\ \\ \lambda^*(\Sigma^{-1}x^*(\mu;r(\mu))\otimes\Sigma^{-1}x^*(\mu;r(\mu))) - \sum_{k=1}^{l} w^*_k \frac{\partial vec(r_k(\mu))}{\partial vec(\Sigma)} x^*(\mu;r(\mu)) \end{bmatrix},$$

where $r_k(\mu)$ denotes the k-th column of $r(\mu)$,

$$x^*(\mu;r(\mu)) = \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r(\mu)}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu;r(\mu)),$$

$$\lambda^* \equiv \frac{1}{2}v_{k,i,j}(\mu;r(\mu)), \quad w^* \equiv [r(\mu)'\Sigma r(\mu)]^{-1}r(\mu)'\Sigma C_k(A)e_i,$$

and w^*_k is the k-th component of the vector w^*.

Proof: Note first that Assumption 3 implies that $r\equiv r(\mu)$ is differentiable with respect to μ. Moreover, since $v_{k,i,j}(\mu;r)\neq 0$ the function

$$v_{k,i,j}(\mu;r) = \Big(e_i'C_k(A)\Sigma^{1/2}M_{\Sigma^{1/2}r}\Sigma^{1/2}C_k(A)'e_i\Big)^{1/2}$$

is differentiable as well. Moreover, the function

$$x^*(\mu;r) \equiv \Sigma^{1/2}\Big(M_{\Sigma^{1/2}r}\Big)\Sigma^{1/2}C_k(A)'e_i \Big/ v_{k,i,j}(\mu;r)$$

is also differentiable. Therefore,

$$\begin{aligned} \frac{dv_{k,i,j}(\mu;r)}{d\mu} &= \frac{d[e_i'C_k(A)x^*(\mu;r)]}{d\mu} \\ &\quad \text{(since } v_{k,i,j}(\mu;r) = e_i'C_k(A)x^*(\mu;r)\text{)} \\ &= \frac{dx^*(\mu;r)}{d\mu}C_k'(A)e_i + \frac{d(x^*(\mu;r)'\otimes e_i')\text{vec}(C_k(A))}{d\mu}, \\ &\quad \text{(where we have re-written } e_i'C_k(A)x^* \text{ as } (x^{*\prime}\otimes e_i')\text{vec}(C_k(A))\text{)} \\ &= \frac{dx^*(\mu;r)}{d\mu}C_k'(A)e_i + \frac{d\text{vec}(C_k(A))}{d\mu}(x^*(\mu;r)\otimes e_i) \\ &\quad \text{(where we have applied the chain rule for matrix derivatives).} \end{aligned}$$

We now use the envelope theorem to compute this derivative. Note that —using Assumptions 1 and 2— Lemma 1 shows the existence of unique multipliers $\lambda^* \in \mathbb{R}$ and $w^* \in \mathbb{R}^l$ such that

$$C_k(A)'e_i = \lambda^* 2\Sigma^{-1}x^*(\mu;r) + rw^*.$$

Therefore,

$$\frac{dv_{k,i,j}(\mu;r)}{\text{dvec}(A)} = \frac{dx^*(\mu;r)}{\text{dvec}(A)}\left[\lambda^* 2\Sigma^{-1}x^*(\mu;r) + rw^*\right] + \frac{\text{dvec}(C_k(A))}{\text{dvec}(A)}(x^*(\mu;r)\otimes e_i)$$

and

$$\frac{dv_{k,i,j}(\mu;r)}{\text{dvec}(\Sigma)} = \frac{dx^*(\mu;r)}{\text{dvec}(\Sigma)}\left[\lambda^* 2\Sigma^{-1}x^*(\mu;r) + rw^*\right] + \frac{\text{dvec}(C_k(A))}{\text{dvec}(\Sigma)}(x^*(\mu;r)\otimes e_i).$$

Note also that because $x^*(\mu,r)$ satisfies the ellipsoid constraint

$$0 = \frac{dx^*(\mu;r)'\Sigma^{-1}x^*(\mu;r)}{\text{dvec}(A)} = 2\frac{dx^*(\mu;r)}{\text{dvec}(A)}\Sigma^{-1}x^*(\mu;r)$$

and, also, since the equality constraints are met,

$$\begin{aligned} 0 &= \frac{dr(\mu)'x^*(\mu;r(\mu))}{\text{dvec}(A)} \\ &= \frac{dx^*(\mu;r)}{\text{dvec}(A)}r(\mu) + \left(\tfrac{dr_1(\mu)}{\text{dvec}(A)}x^*(\mu;r(\mu)), \quad \cdots \quad, \tfrac{dr_l(\mu)}{\text{dvec}(A)}x^*(\mu;r(\mu))\right), \end{aligned}$$

where $r_k(\mu)$ denotes the k-th column of $r(\mu)$. Consequently,

$$\frac{dv_{k,i,j}(\mu;r)}{\text{dvec}(A)} = \frac{\text{dvec}(C_k(A))}{\text{dvec}(A)}(x^*(\mu;r)\otimes e_i) - \sum_{k=1}^{l} w_k^* \frac{\text{dvec}(r_k(\mu))}{\text{dvec}(A)}x^*(\mu;r),$$

where w_k^* is the k-th entry of the vector of lagrange multipliers w^*. This gives the partial derivative of $v_{k,i,j}(\mu;r_l(\mu))$ with respect to vec(A). We note that this derivative can also be written as

$$\frac{dv_{k,i,j}(\mu;r)}{\text{dvec}(A)} = \frac{\text{dvec}(C_k(A))}{\text{dvec}(A)}(x^*(\mu;r)\otimes e_i) - \frac{\text{dvec}(r(\mu)')}{\text{dvec}(A)}(x^*(\mu;r)\otimes \mathbb{I}_l)w^*,$$

which is the expression given in the overview. Finally, to get the derivative with respect to vec(Σ) we note that

$$0 = \frac{dx^*(\mu;r)'\Sigma^{-1}x^*(\mu;r)}{\text{dvec}(\Sigma)} = 2\frac{dx^*(\mu;r)}{\text{dvec}(\Sigma)}\Sigma^{-1}x^*(\mu;r) - (\Sigma^{-1}x^*(\mu;r)\otimes\Sigma^{-1}x^*(\mu;r)),$$

and

$$\begin{aligned} 0 &= \frac{dr(\mu)'x^*(\mu;r(\mu))}{\text{dvec}(\Sigma)} \\ &= \frac{dx^*(\mu;r(\mu))}{\text{dvec}(\Sigma)}r(\mu) + \left(\tfrac{dr_1(\mu)}{\text{dvec}(\Sigma)}x^*(\mu;r(\mu)), \quad \cdots \quad, \tfrac{dr_l(\mu)}{\text{dvec}(\Sigma)}x^*(\mu;r(\mu))\right). \end{aligned}$$

Consequently,

$$\frac{dv_{k,i,j}(\mu;r)}{\text{dvec}(\Sigma)} = \lambda^*(\Sigma^{-1}x^*(\mu;r)\otimes\Sigma^{-1}x^*(\mu;r)) - \sum_{k=1}^{l} w_k^* \frac{\text{dvec}(r_{k,l}(\mu))}{\text{dvec}(\Sigma)}x^*(\mu;r).$$

A.4. *Proof of Theorem 2*

Structure of the proof: The proof proceeds in five steps. First, we show that Assumptions 1 and 2 imply that the choice set of program (2.5) is non-empty for any $\tilde{\mu}$ in a neighborhood of μ. Second, we show that the choice set of program (2.5) is both lower and upper-hemicontinuous correspondence at μ. Third, we use the continuity of the choice set and the Maximum theorem to establish continuity of $\overline{v}_{k,i,j}(\cdot)$ at μ. Fourth, we use Lemma 1 and the continuity of $\overline{v}_{k,i,j}(\cdot)$ to show that

$$\max_{r_l^*(\mu)\in R^*(\mu)}\{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\} \leq \liminf_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

Finally, we use Lemma 1, Theorem 1, and Lemma 2 to show (by contradiction) that

$$\limsup_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) \leq \max_{r_l^*(\mu)\in R^*(\mu)}\{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}.$$

Step 1 : By Assumption 1 there is a point $x^* \in \mathbb{R}^n$ that belongs to the choice set of program (2.5). Let $Z^*(\mu) \in \mathbb{R}^{n\times m_e}$ denote the restrictions in program (2.5) that are active at x^*. By Assumption 2, we know that $m_e \leq n-1$. Let $S^*(\mu) \in \mathbb{R}^{n\times m_i}$ denote all the other restrictions that are not in $Z^*(\mu)$. This means that $S^*(\mu)'x^* > 0_{m_i\times 1}$ (since these restrictions are not in $Z^*(\mu)$). Note first that Assumption 2 implies there is $\epsilon_1 > 0$ such that $\lambda_{min}(\mu) \equiv \min \text{eig}(Z^*(\mu)'Z^*(\mu)) > \epsilon_1$. Since x^* is feasible we can also pick ϵ_2 such that $(s_m^*(\mu)/||s_m^*(\mu)||)'x^*(\mu)$ is larger than ϵ_2 for each $m \in \{1, 2, \ldots, m_i\}$. Define

$$U(\mu) \equiv \{\tilde{\mu} \quad | \quad \lambda_{min}(\tilde{\mu}) > \epsilon_1, \quad (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'x^* > \epsilon_2 \;\; \forall\, m, \quad ||Z^*(\tilde{\mu})'x^*|| < \sqrt{\epsilon_1}\epsilon_2/2\} \cap \mathcal{M}.$$

By construction $\mu \in U(\mu)$. Moreover, the continuity of $Z(\cdot)$ and $S(\cdot)$ and openness of $\mathcal{M}$ implies that $Z^*(\cdot)$ and $S^*(\cdot)$ are continuous and therefore $U(\mu)$ is open. We now show that for every $\tilde{\mu} \in U(\mu)$ there is $\tilde{x} \in \mathbb{R}^d$ that satisfies the equality restrictions in $Z^*(\tilde{\mu})$ and also the inequalities in $S^*(\tilde{\mu})$ with slack. To formalize this point, define

$$\tilde{x} \equiv \tilde{x}(\tilde{\mu}, \mu) \equiv x^* - N_{Z^*(\tilde{\mu})}x^*, \tag{A.5}$$

where $N_{Z^*(\tilde{\mu})} \equiv Z^*(\tilde{\mu})(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}Z^*(\tilde{\mu})'$ is well defined because $\lambda_{min}(\tilde{\mu}) > \epsilon_1$. Note first that, by construction,

$$Z^*(\tilde{\mu})'\tilde{x} = Z^*(\tilde{\mu})'x^* - Z^*(\tilde{\mu})'N_{Z^*(\tilde{\mu})}x^* = Z^*(\tilde{\mu})'x^* - Z^*(\tilde{\mu})'Z^*(\tilde{\mu})(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}Z^*(\tilde{\mu})'x^* = 0_{m_e\times 1},$$

implying that the equality restrictions at $Z^*(\tilde{\mu})$ are satisfied by $\tilde{x}$. Thus, we only need to show that the inequalities in $s_m^*(\tilde{\mu})$ are satisfied with slack (after normalizing by its norm). To see this, simply note that

$$\begin{aligned}
(s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'\tilde{x} &= (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'(\tilde{x} - x^*) + (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'x^* \\
&> (s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'(\tilde{x} - x^*) + \epsilon_2 \\
&\geq -|s_m^*(\tilde{\mu})/||s_m^*(\tilde{\mu})||)'(\tilde{x} - x^*)| + \epsilon_2 \\
&\geq -||(\tilde{x} - x^*)|| + \epsilon_2.
\end{aligned}$$

But

$$\begin{aligned}
||\tilde{x} - x^*|| &= (x^{*\prime}Z^*(\tilde{\mu})(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}Z^*(\tilde{\mu})x^*)^{1/2} \\
&\leq \sup_{\omega \text{ s.t. } ||\omega||=1} (\omega(Z^*(\tilde{\mu})'Z^*(\tilde{\mu}))^{-1}\omega)^{1/2}||Z^*(\tilde{\mu})x^*|| \\
&= ||Z^*(\tilde{\mu})x^*||\sqrt{\lambda_{\min}(\tilde{\mu})} \\
&\leq (\sqrt{\epsilon_1}\epsilon_2/2\sqrt{\epsilon_1}) \\
&= \epsilon_2/2.
\end{aligned}$$

This implies that $s_m^*(\tilde{\mu})'\tilde{x} > 0$ for every $m \in \{1, 2, \ldots, m_i\}$. This shows that for every $\tilde{\mu} \in U(\mu)$, $\tilde{x} \in \mathcal{R}(\tilde{\mu})$. To complete Step 1, notice that $x^\dagger \equiv \tilde{x}/(\tilde{x}'\tilde{\Sigma}^{-1}\tilde{x}) \in \mathcal{R}(\tilde{\mu})$. By construction, $x^\dagger$ is in the choice set of program 2.5 evaluated at $\tilde{\mu}$.

Step 2 : Let the multivalued correspondence $\Gamma(\cdot) : \mathcal{M} \to \mathbb{R}^n$ be defined as the choice set of program (2.5). We show continuity of this correspondence at μ by showing that it is both lower and upper hemicontinuous.

To establish upper hemicontinuity, pick any sequence $\mu_N \in \mathcal{M}$ s.t. $\mu_N \to \mu$ and any converging sequence $x_N \in \Gamma(\mu_N)$ s.t. $x_N \to x^*$. Consider any sign restriction $s(\mu_N)$. By construction, $s(\mu_N)'x_N \geq 0$. By continuity of $s(\mu_n)$, we get in the limit $s(\mu)'x^* \geq 0$. Similarly, $(x^*)'\Sigma^{-1}(x^*) = 1$ and for any zero restriction $z(\mu)'x^* = 0$. The set $\Gamma(\mu)$ is compact, so by Theorem 2 on p. 218 in Ok (2007), $\Gamma(\cdot)$ is upper hemicontinuous at μ.

To establish lower hemicontinuity, consider any sequence $\mu_N \in \mathcal{M}$ s.t. $\mu_N \to \mu$ and any point $x^* \in \Gamma(\mu)$. Then, by Step 1, the elements of the sequence defined as

$$x_N \equiv \tilde{x}(\mu_N, \mu)/(\tilde{x}(\mu_N, \mu)'\Sigma^{-1}\tilde{x}(\mu_N, \mu))$$

belong to $\Gamma(\mu_N)$. By continuity of $Z^*(\cdot)$ and Σ^{-1} at $\mu \in \mathcal{M}$ (implied by Assumption 3) and using the invertibility of the matrices $(Z^*(\tilde{\mu_N})'Z^*(\tilde{\mu_N}))$ for N large enough (implied by Assumption 2) we have $x_N \to x^*$. By Proposition 4 on p. 224 in Ok (2007), $\Gamma(\mu)$ is lower hemicontinuous. By definition, it is continuous at μ.

Step 3 : Let $(\Theta, \rho) \equiv (U(\mu), \rho)$ be a metric space with Euclidean metric $\rho(\cdot)$. By Steps 1 and 2, the choice set of the program in (2.5) is a non-empty, compact-valued, continuous correspondence at μ. By the Maximum theorem, see p. 229 in Ok (2007), $\overline{v}_{k,i,j}(\cdot)$ is continuous at μ.

Additional Notation: Consider any sequence $\mu_N = (\text{vec}A_N{}', \text{vec}\Sigma_N{}')'$ such that

$$\mu_N = \mu + h_N/t_N,$$

where $h_N \to h \in \mathbb{R}^d$, $t_N \to \infty$ and such that μ_N belongs to the parameter space $\mathcal{M}$ for N large enough. By Step 1 there exists N^* large enough such that the choice set of the program in (2.5) at μ_N is non-empty for every $N \geq N^*$. Thus, $\overline{v}_{k,i,j}(\mu_N)$ is well-defined for N large enough. Moreover, the continuity of the value function established in Step 3 implies that we can assume that $\overline{v}_{k,i,j}(\mu_N) \neq 0$ for N large enough. In fact, it is without loss of generality to assume that $\overline{v}_{k,i,j}(\mu_N) > 0$ for N large enough.

Let $X^*(\mu)$ denote the argmax of program (2.5) at μ. By Theorem 1—and using the fact that $\overline{v}_{k,i,j}(\mu) \neq 0$—$X^*(\mu)$ has a finite number of elements. Assume then that the argmax has L elements and denote them as $x_1^*(\mu), x_2^*(\mu), \ldots, x_L^*(\mu)$.

For each $l \in \{1, 2, \ldots, L\}$, let $r_l^*(\mu)$ denote the $n \times m_{z_l}$ matrix of *all* active restrictions at a solution $x_l^*(\mu)$. Likewise, let $S_l^*(\mu)$ be the matrix of dimension $n \times m_{s_l}$ that collects *all* slack restrictions at $x_l^*(\mu)$. Consequently, for each solution $x_l^*(\mu)$ there are unique matrices $r_l^*(\mu)$ and $S_l^*(\mu)$ such that

$$r_l^*(\mu)'x_l^*(\mu) = \mathbf{0}_{m_{z_l} \times 1}, \quad S_l^*(\mu)'x_l^*(\mu) > \mathbf{0}_{m_{s_l} \times 1}.$$

Define

$$R^*(\mu) \equiv \{r_1^*(\mu), r_2^*(\mu), \ldots, r_l^*(\mu)\}.$$

Proof of differentiability: We establish the differentiability of the value function in two sub-steps.

Step 4: First, we show that

$$\max_{r_l^*(\mu) \in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\} \leq \liminf_{N \to \infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)). \tag{A.6}$$

Proof: Take any $r_l^*(\mu) \in R^*(\mu)$. By definition of $r_l^*(\mu)$ all the columns of $S(\mu)$ that are not contained in $r_l^*(\mu)$ are slack (at μ). Consider then the candidate solution $x_+^*(\mu, r_l^*(\mu))$. This candidate solution is continuous at μ (which follows from the formula in Lemma 1 and the fact that $v_{k,i,j}(\mu, r_l^*(\mu)) = \overline{v}_{k,i,j}(\mu) > 0$). Therefore, for N large enough this candidate solution $x_+^*(\mu_N, r_l^*(\mu_N))$ is in the choice set of program (2.5) at μ_N, which implies that

$$v_{k,i,j}(\mu_N, r_l^*(\mu_N)) \leq \overline{v}_{k,i,j}(\mu_N).$$

Hence, the inequality above implies that for any $r_l^*(\mu) \in R^*(\mu)$ we have that

$$t_N(v_{k,i,j}(\mu_N, r_l^*(\mu_N)) - v_{k,i,j}(\mu, r_l^*(\mu))) \leq t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

Lemma 2 thus implies that for any $r_l^*(\mu) \in R^*(\mu)$,

$$\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h \leq \liminf_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)),$$

which establishes equation (A.6).

Step 5: Now we show that

$$\limsup_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) \leq \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}. \tag{A.7}$$

Proof: We prove the statement above by contradiction. So, suppose that (A.7) does not hold. Then, there exists $\epsilon_0 > 0$ and a subsequence μ_{N_k} such that for every $r_l^*(\mu) \in R^*(\mu)$,

$$\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h + \epsilon_0 \leq t_N(\overline{v}_{k,i,j}(\mu_{N_k}) - \overline{v}_{k,i,j}(\mu)). \tag{A.8}$$

We will show that assuming the existence of $\epsilon_0 > 0$ and a subsequence μ_{N_k} will lead to a contradiction.

Additional Notation: Let $x_{N_k}^*$ be any element in the argmax of program (2.5) at μ_{N_k}. Let $r_{N_k}^*(\mu_{N_k})$ be the matrix that collects all active restrictions at $x_{N_k}^*$ and let $S_{N_k}^*(\mu_{N_k})$ be the matrix that collects all of the sign restrictions that are slack at $x_{N_k}^*$; i.e, $S_{N_k}^*(\mu_{N_k})'x_{N_k}^* > \mathbf{0}$. Let

$$R_+(\mu) \equiv \{r \in R(\mu) \mid v_{k,i,j}(\mu; r(\mu)) > 0\}.$$

Partition the set $R_+(\mu)$ into the following four disjoint sets:

i) $R^*(\mu)$,

ii) The restrictions $r(\mu) \in R_+(\mu)/R^*(\mu)$ for which $x_+(\mu; r(\mu))$ belongs to $X^*(\mu)$,

iii) The restrictions $r(\mu) \in R_+(\mu)$ that do not fall in neither i) nor ii) and for which some sign restriction not included in $r(\mu)$ is violated,

iv) The restrictions $r(\mu) \in R(\mu)$ that do not fall in i), ii), iii) for which $x_+(\mu, r(\mu))$ is feasible but $v_{k,i,j}(\mu, r(\mu)) < \overline{v}_{k,i,j}(\mu)$.

Proof of A.7): Note that the restrictions of Type i) cannot be satisfied by $x_{N_k}^*$ infinitely often. In other words, there is no $l = 1, \dots L$ such that

$$r_{N_k}^*(\mu_{N_k}) = r_l^*(\mu_{N_k}), \text{ and } S_{N_k}^*(\mu_{N_k}) = S_l^*(\mu_{N_k})$$

for infinitely many values of k. If this were the case, there would be a further subsequence N_{K_T} for which $\overline{v}_{k,i,j}(\mu_{N_{K_T}}) = v_{k,i,j}(\mu_{N_{K_T}}, r_l(\mu_{N_{K_T}}))$. Thus, equation (A.8) would contradict the differentiability of $v_{k,i,j}(\mu, r_l(\mu))$.

Restrictions of Type iii) cannot be satisfied infinitely often by $x_{N_k}^*$. This follows from the fact that if $r_{N_k}^*(\mu_{N_K})$ were equal to some $r(\mu_{N_K})$ for $r(\mu)$ of type iii) infinitely often, then we could always find some large k for which $x_{N_k}^*$ is the form $x_+(\mu_{N_k}, r(\mu_{N_K}))$. Such candidate solution will eventually violate a sign restriction, contradicting the fact that $x_{N_k}^*$ is in fact a solution.

Restrictions of Type iv) cannot be satisfied infinitely often by $x_{N_k}^*$. If this were the case, then we could always find some large k for which

$$\begin{aligned}\dot{v}_{k,i,j}(\mu,r_l^*(\mu))'h+\epsilon_0 &\leq t_{N_k}(\overline{v}_{k,i,j}(\mu_{N_k})-\overline{v}_{k,i,j}(\mu))\\ &= t_{N_k}(v_{k,i,j}(\mu_{N_k},r_l(\mu_{N_k}))-\overline{v}_{k,i,j}(\mu))\\ &= t_{N_k}(v_{k,i,j}(\mu_{N_k},r_l(\mu_{N_k}))-v_{k,i,j}(\mu,r_l(\mu)))\\ &+ t_{N_k}(v_{k,i,j}(\mu_{N_k},r_l(\mu))-\overline{v}_{k,i,j}(\mu))\end{aligned}$$

(where $r_l(\cdot)$ is some set of restrictions of type iv)). But the fact that $(v_{k,i,j}(\mu_{N_k};r_l(\mu))-\overline{v}_{k,i,j}(\mu)<0)$ contradicts the definition of the subsequence μ_{N_k}.

Finally, we show that if r is a restriction of Type ii) it cannot be the case that

$$r^*_{\mu_{N_k}}(\mu_{N_p})=r(\mu_{N_p})$$

infinitely often. To establish this claim, suppose that there is a restriction r of Type ii) such that

$$r(\mu_{N_p})'x^*(\mu_{N_p})=\mathbf{0}$$

infinitely often. This means we can construct a further subsequence $\mu_{N_{p_q}}$ for which (by Lemma 1)

$$\overline{v}_{k,i,j}(\mu_{N_{p_q}})=v_{k,i,j}(\mu_{N_{p_q}},r(\mu_{N_{p_q}})).$$

Therefore, by equation (A.8) we must have that for every $r_l^*(\mu)\in R^*(\mu)$,

$$\begin{aligned}\dot{v}_{k,i,j}(\mu,r_l^*(\mu))'h+\epsilon_0 &\leq t_{N_{p_q}}(v_{k,i,j}(\mu_{N_{p_q}},r(\mu_{N_{p_q}}))-\overline{v}_{k,i,j}(\mu))\\ &= t_{N_{p_q}}(v_{k,i,j}(\mu_{N_{p_q}},r(\mu_{N_{p_q}}))-v_{k,i,j}(\mu,r(\mu)))\\ &+ t_{N_{p_q}}(v_{k,i,j}(\mu,r(\mu))-\overline{v}_{k,i,j}(\mu))\\ &= t_{N_{p_q}}(v_{k,i,j}(\mu_{N_{p_q}},r(\mu_{N_{pp_q}}))-v_{k,i,j}(\mu,r(\mu))),\end{aligned}$$

where the last line follows from the fact that $r(\mu)$ is of Type ii) and, hence, leads to a candidate solution $x_+(\mu;r(\mu))$ that equals $x_l^*(\mu)$ for some l, which we will assume (without loss of generality) to be equal to 1. The differentiability result in Lemma 2 implies that for every $l=1,\dots,L$,

$$\dot{v}_{k,i,j}(\mu;r_l^*(\mu))'h+\epsilon_0\leq\dot{v}_{k,i,j}(\mu,r(\mu))'h.$$

We show that this last inequality leads to a contradiction as we must have

$$\dot{v}_{k,i,j}(\mu,r_1^*(\mu))'h=\dot{v}_{k,i,j}(\mu;r(\mu))'h. \tag{A.9}$$

To see this, note first that $r_1^*(\mu)$ must contain all the columns of $r(\mu)$ as

$$r(\mu)'x_1^*(\mu)=\mathbf{0},$$

and, by definition, $r_1^*(\mu)$ contains all the constraints that are active at $x_1^*(\mu)$. Thus, we can write $r_1^*(\mu)$ as

$$r_1^*(\mu)=[r(\mu),\tilde{r}(\mu)],$$

where $r(\mu)$ and $\tilde{r}(\mu)$ are linearly independent. Our formula for $\dot{v}_{k,i,j}$ in Lemma 2 implies that (A.9) will hold if the Lagrange multipliers corresponding to the constraints in $\tilde{r}(\mu)$ are zero. To see that this is indeed the case, note that by the argument used in the proof of Lemma 2, the Karush-Kuhn-Tucker conditions for the program that only imposes $r(\mu)$ as equality conditions (along with the ellipsoid constraint) imply that

$$C_k'(A)e_i=v_{k,i,j}(\mu;r(\mu))\Sigma^{-1}x_+(\mu;r(\mu))+r(\mu)w_1.$$

The analogous conditions for the program that imposes $r_1^*(\mu)$ as constraints imply that

$$C_k'(A)e_i = v_{k,i,j}(\mu; r_1^*(\mu))\Sigma^{-1}x_+(\mu; r_1^*(\mu)) + r_1^*(\mu)w_1^*.$$

Therefore—since by assumption $x_+(\mu; r_1^*(\mu)) = x_+(\mu; r(\mu))$—it has to be the case that

$$r(\mu)w_1 - r_1^*(\mu)w_1^* = \mathbf{0}_{n\times 1}.$$

Partitioning $w_1^* = [w_{1,1}^{*}{}', w^{*}{}'_{1,2}]'$ according to $r(\mu) = [r(\mu), \tilde{r}(\mu)]$, we have that

$$r(\mu)(w_1 - w_{1,1}^*) + \tilde{r}(\mu)w_{1,2}^* = \mathbf{0}_{n\times 1}.$$

Assumption 2 implies that the latter equality holds if and only if $w_1 = w_{1,1}^*$ and $w_{1,2}^* = \mathbf{0}$. Therefore we conclude that equation (A.9) must holds. This leads to a contradiction as $\epsilon_0 > 0$ and

$$\dot{v}_{k,i,j}(\mu; r_1^*(\mu))'h + \epsilon_0 \leq \dot{v}_{k,i,j}(\mu; r_1^*(\mu))'h.$$

Summary of Step 4: Step 4.1 showed that

$$\max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu; r_l^*(\mu))'h\} \leq \liminf_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)).$$

Step 4.2 showed that

$$\limsup_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) \leq \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}.$$

We conclude that

$$\lim_{N\to\infty} t_N(\overline{v}_{k,i,j}(\mu_N) - \overline{v}_{k,i,j}(\mu)) = \max_{r_l^*(\mu)\in R^*(\mu)} \{\dot{v}_{k,i,j}(\mu, r_l^*(\mu))'h\}.$$

A.5. *Proof of Theorem 3 Part a)*

Let P denote the data generating process. For notational simplicity we write μ instead of $\mu(P)$ and Ω instead of $\Omega(P)$ whenever convenient. Note first that

$$P\left(\lambda \in \left[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T}/\sqrt{T}\right]\right) \tag{A.10}$$

is bounded from below by

$$P\left(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T) - \underline{v}_{k,i,j}(\mu)) \leq z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T} \text{ and } \; -z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \leq \sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T) - \overline{v}_{k,i,j}(\mu))\right),$$

which is itself bounded from below by

$$P\left(\sqrt{T}(\underline{v}_{k,i,j}(\widehat{\mu}_T) - \underline{v}_{k,i,j}(\mu)) \leq z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \text{ and } -z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \leq \sqrt{T}(\overline{v}_{k,i,j}(\widehat{\mu}_T) - \overline{v}_{k,i,j}(\mu)), \text{ and } ||\sqrt{T}(\widehat{\mu}_T - \mu)|| \leq M_\epsilon\right),$$

where M_ϵ is such that

$$P\Big(||\zeta(P)|| > M_\epsilon\Big) \leq \epsilon.$$

By Theorem 2, both $\underline{v}_{k,i,j}(\cdot)$ and $\overline{v}_{k,i,j}(\mu)$ are directionally differentiable function with directional derivatives denoted by $\underline{\dot{v}}_{k,i,j;\mu}(\cdot), \dot{\overline{v}}_{k,i,j;\mu}(\cdot)$. The directional differentiability implies that for any $\delta > 0$ there is T large enough such that for any $h \in \mathbb{R}^d$ such that $||h|| \leq M_\epsilon$,

$$-\delta \leq \sqrt{T}(\underline{v}_{k,i,j}(\mu + h/\sqrt{T}) - \underline{v}_{k,i,j}(\mu)) - \underline{\dot{v}}_{k,i,j;\mu}(h) \leq \delta$$

and

$$-\delta \leq \sqrt{T}(\overline{v}_{k,i,j}(\mu + h/\sqrt{T}) - \overline{v}_{k,i,j}(\mu)) - \dot{\overline{v}}_{k,i,j;\mu}(h) \leq \delta.$$

Therefore, for T large enough

$$\inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T}/\sqrt{T}\Big]\Big)$$

is bounded from below by

$$P\Big(\delta + \dot{\underline{v}}_{k,i,j;\mu}(\sqrt{T}(\widehat{\mu}_T - \mu)) \leq z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T} \text{ and}$$

$$-z_{1-\alpha/2}\widehat{\sigma}_{k,i,j;T} \leq \dot{\overline{v}}_{k,i,j;\mu}(\sqrt{T}(\widehat{\mu}_T - \mu)) - \delta, \text{ and } ||\sqrt{T}(\widehat{\mu}_T - \mu)|| \leq M_\epsilon\Big),$$

which, by Assumption 4 (and using the continuity of the directional derivative), converges in distribution to

$$P\Big(\delta + \dot{\underline{v}}_{k,i,j;\mu}(\zeta(P)) \leq z_{1-\alpha/2}\, \sigma \text{ and}$$

$$-z_{1-\alpha/2}\sigma \leq \dot{\overline{v}}_{k,i,j;\mu}((\zeta(P)) - \delta, \text{ and } ||\zeta(P)|| \leq M_\epsilon\Big),$$

where σ is the probability limit of $\widehat{\sigma}_{k,i,j;T}$,

$$\sigma \equiv \max_{r \in R(\mu)} \Big[\dot{v}_{k,i,j}(\mu; r)'\Omega \dot{v}_{k,i,j}(\mu; r)\Big].$$

Consequently, for every $\delta > 0$,

$$\liminf_{T\to\infty} \inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\, \widehat{\sigma}_{k,i,j;T}/\sqrt{T}\Big]\Big)$$

is larger than or equal

$$1 - P\Big(\dot{\underline{v}}_{k,i,j;\mu}(\zeta(P)) > z_{1-\alpha/2}\, \sigma - \delta\Big) - P\Big(\dot{\overline{v}}_{k,i,j;\mu}(\zeta(P)) < -z_{1-\alpha/2}\, \sigma + \delta\Big)$$

$$-P\Big(||\zeta(P)|| > M_\epsilon\Big).$$

Take some $x \in X_*(\mu)$ for which $\underline{\sigma}(x) \equiv \dot{v}_{k,i,j}(\mu; r(\mu; x))'\Omega \dot{v}_{k,i,j}(\mu; r(\mu; x)) > 0$ (one such x must exist by the assumption of this theorem). The fact that $\zeta(P)$ is symmetric and using our formula for the directional derivative of $\underline{v}_{k,i,j}$ we have that

$$\begin{aligned} P\Big(\dot{\underline{v}}_{k,i,j;\mu}(\zeta(P)) > z_{1-\alpha/2}\, \sigma - \delta\Big) &\leq P\Big(\dot{v}_{k,i,j}(\mu; r(\mu; x))'\zeta(P) > z_{1-\alpha/2}\sigma - \delta\Big) \\ &\leq \mathbb{P}\Big(N(0,1) > z_{1-\alpha/2}\frac{\sigma}{\underline{\sigma}(x)} - \frac{\delta}{\underline{\sigma}(x)}\Big), \\ &\leq \mathbb{P}\Big(N(0,1) > z_{1-\alpha/2} - \frac{\delta}{\underline{\sigma}(x)}\Big), \end{aligned}$$

for any $\delta > 0$ (since $\sigma \geq \underline{\sigma}(x)$).

Now, take some $x \in X^*(\mu)$ for which $\overline{\sigma}(x) \equiv \dot{v}_{k,i,j}(\mu; r(\mu; x))'\Omega \dot{v}_{k,i,j}(\mu; r(\mu; x)) > 0$. Note that

$$\begin{aligned} P\Big(\dot{\overline{v}}_{k,i,j;\mu}(\zeta(P)) < -z_{1-\alpha/2} + \delta\, \sigma\Big) &\leq P\Big(\dot{v}_{k,i,j}(\mu; r(\mu, x))'\zeta(P) < -z_{1-\alpha/2}\sigma + \delta\Big) \\ &\leq \mathbb{P}\Big(N(0,1) < -z_{1-\alpha/2}\frac{\sigma}{\overline{\sigma}(x)} + \frac{\delta}{\overline{\sigma}(x)}\Big), \\ &\leq \mathbb{P}\Big(N(0,1) < -z_{1-\alpha/2} + \frac{\delta}{\overline{\sigma}(x)}\Big), \end{aligned}$$

for any $\delta > 0$ (since $\sigma > \overline{\sigma}(x)$). We conclude that for any $\epsilon > 0$ and $\delta > 0$

$$\liminf_{T\to\infty} \inf_{\lambda \in \mathcal{I}^{\mathcal{R}}_{k,i,j}(\mu(P))} P\Big(\lambda \in \Big[\underline{v}_{k,i,j}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T},\ \overline{v}_{k,i,j}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}_{k,i,j;T}/\sqrt{T}\Big]\Big)$$

is bounded from below by

$$\Phi\left(z_{1-\alpha/2} - \frac{\delta}{\underline{\sigma}(x)}\right) - \Phi\left(-z_{1-\alpha/2} + \frac{\delta}{\overline{\sigma}(x)}\right) - \epsilon,$$

where $\Phi(\cdot)$ is the standard normal c.d.f. Since $\epsilon > 0$, $\delta > 0$ are arbitrary and $\Phi(\cdot)$ is continuous, the desired result follows.

A.6. *Proof of Theorem 3 Part b)*

PROOF: We would like to show that for every $\epsilon > 0, \eta > 0$ there is $T^*(\epsilon,\eta)$ such that for $T \geq T^*(\epsilon,\eta)$ we have that

$$P(RBC(Y_1,\ldots,Y_T) < 1-\alpha-\epsilon) < \eta.$$

We divide the proof into 5 steps.

Step 1 (Definitions of $M_{\epsilon,\eta}$, δ_ϵ): Let ζ be a $\mathcal{N}_d(\mathbf{0},\Omega(P))$ random vector. For given $\epsilon > 0, \eta > 0$ define $M_{\epsilon,\eta} \in \mathbb{R}$ as the scalar such that

$$\mathbb{P}(||\zeta|| > M_{\epsilon,\eta}) < \min\{\epsilon/3, \eta/4\}.$$

Let $\Phi(\cdot)$ denote the standard normal c.d.f. Define $\delta_\epsilon > 0$ to be any scalar such that

$$|\Phi(z_{1-\alpha/2} - \delta_\epsilon/\underline{\sigma}(\mu)) - \Phi(-z_{1-\alpha/2} + \delta_\epsilon/\overline{\sigma}(\mu)) - (1-\alpha)| < \epsilon/3.$$

Such a scalar exists by the continuity of $\Phi(\cdot)$ and the fact that $\underline{\sigma}(\mu)$ and $\overline{\sigma}(\mu)$ are positive.

Step 2 (Definitions of $A_T(\epsilon), B_T(\epsilon), C_T(\epsilon)$). Let

$$Y^T \equiv (Y_1,\ldots,Y_T)$$

denote the data. In a slight abuse of notation, let $\widehat{\sigma}_T$ abbreviate $\widehat{\sigma}_{k,i,j}$ and let σ denote the probability limit of $\widehat{\sigma}_T$. Define the events:

$$\begin{aligned}
A_T(\epsilon,\eta) &\equiv \Big\{Y^T \;\Big|\; ||\sqrt{T}(\widehat{\mu}_T - \mu)|| > M_{\epsilon,\eta}\Big\},\\
B_T(\epsilon) &\equiv \Big\{Y^T \;\Big|\; \sup_{B\in\mathcal{B}(\mathbb{R}^d)} |P^*_\mu\left(\sqrt{T}(\mu^* - \widehat{\mu}_T) \in B \mid Y^T\right) - \mathbb{P}\left(\zeta \in B\right)| > \frac{\epsilon}{3}\Big\},\\
C_T(\epsilon) &\equiv \Big\{Y^T \;\Big|\; |\widehat{\sigma} - \sigma| > \frac{\delta_\epsilon}{2z_{1-\alpha/2}}\Big\}.
\end{aligned}$$

We will show that if the Robust Bayes Credibility of our delta-method interval falls below $1-\alpha-\epsilon$ then one of the events above occurs *a fortiori*. We will then argue that our assumptions imply that the probability of each of these events becomes arbitrarily small for large T (implying the event in which the Robust Bayes Credibility is below $1-\alpha-\epsilon$ happens with an arbitrarily small probability).

Note that the CLT for $\widehat{\mu}_T$ (Assumption 4) implies that for any $\epsilon > 0$ and any $\eta > 0$

$$P(A_T(\epsilon,\eta)) \to \mathbb{P}(||\zeta|| > M_{\epsilon,\eta}). \tag{A.11}$$

The Bernstein von-Mises Theorem for μ^* (Assumption 5) implies that for any $\epsilon > 0$

$$P(B_T(\epsilon)) \to 0. \tag{A.12}$$

Finally, the definition of probability limit implies that

$$P(C_T(\epsilon)) \to 0. \tag{A.13}$$

Therefore, for any $\epsilon > 0, \eta > 0$ there exists $T_1(\epsilon, \eta)$ such that for any $T \geq T_1(\epsilon, \eta)$

$$|P(A_T(\epsilon, \eta)) - P(||\zeta|| > M_{\epsilon,\eta})| < \eta/4, \quad |P(B_T(\epsilon))| < \eta/4, \quad P(C_T(\epsilon)) < \eta/4. \tag{A.14}$$

Step 3 (First order approximations of the bounds of the identified set). Let μ denote the true parameter and define $Z_T^* \equiv \sqrt{T}(\mu^* - \widehat{\mu}_T)$ and $Z_T \equiv \sqrt{T}(\widehat{\mu}_T - \mu)$. Let $\underline{v}(\cdot)$ abbreviate $\underline{v}_{k,i,j}(\cdot)$ and, likewise, let $\overline{v}(\cdot)$ abbreviate $\overline{v}_{k,i,j}(\cdot)$. Note that

$$\sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)) = \sqrt{T}(\underline{v}(\mu + Z_T^*/\sqrt{T} + Z_T/\sqrt{T}) - \underline{v}(\mu)) - \sqrt{T}(\underline{v}(\mu + Z_T/\sqrt{T}) - \underline{v}(\mu)).$$

The differentiability of $\underline{v}(\cdot)$ at μ (which follows from Theorem 2 and the fact that $X_*(\mu)$ is a singleton) implies that whenever $||Z_T^*|| \leq M_\epsilon$ and $||Z_T|| \leq M_\epsilon$ there is $T_2(\epsilon, \eta)$ such that for $T \geq T_2(\epsilon, \eta)$,

$$|\sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)) - \dot{\underline{v}}_\mu(Z_T^* + Z_T) - \dot{\underline{v}}_\mu(Z_T)| = |\sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)) - \dot{\underline{v}}_\mu(Z_T^*)| < \delta_\epsilon/2.$$

Analogously, we can find $T_3(\epsilon, \eta)$ such that for $T \geq T_3(\epsilon, \eta)$ we have

$$|\sqrt{T}(\overline{v}(\mu^*) - \overline{v}(\widehat{\mu}_T)) - \dot{\overline{v}}_\mu(Z_T^*)| < \delta_\epsilon/2.$$

Step 4 (Lower bound on the Robust Bayesian Credibility of a set). Define the posterior probability that the bounds of the identified set are contained in our delta-method interval as

$$c(Y^T) \equiv P_\mu^*\left([\underline{v}(\mu^*), \overline{v}(\mu^*)] \subseteq \left[\underline{v}(\widehat{\mu}_T) - z_{1-\alpha/2}\,\widehat{\sigma}/\sqrt{T},\ \overline{v}(\widehat{\mu}_T) + z_{1-\alpha/2}\,\widehat{\sigma}/\sqrt{T}\right] |Y^T\right).$$

Note that for every data realization

$$c(Y^T) \leq RBC(Y^T),$$

which follows from the fact that for any (A, B) we have that $\lambda(A, B) \in [\underline{v}(\mu), \overline{v}(\mu)]$. Therefore for any $\epsilon > 0$

$$P(RBC(Y^T) < 1 - \alpha - \epsilon) \leq P(c(Y^T) < 1 - \alpha - \epsilon) \tag{A.15}$$

Thus, to establish Theorem 4 it suffices to show that for any $\epsilon > 0$

$$\lim_{T\to\infty} P(c(Y^T) < 1 - \alpha - \epsilon) = 0.$$

We establish such a result in the following step.

Step 5: We now show that for any $\epsilon > 0$, $\eta > 0$ there is T large enough such that

$$P(c(Y^T) < 1 - \alpha - \epsilon) \leq P(A_T(\epsilon, \eta) \cup B_T(\epsilon) \cup C_T(\epsilon)),$$

or equivalently, that

$$P(A_T^c(\epsilon) \cap B_T^c(\epsilon) \cap C_T^c(\epsilon)) \leq P(c(Y^T) \geq 1 - \alpha - \epsilon)$$

for T large enough. We start by re-writing $c(Y^T)$ as

$$P_\mu^*\Big(- z_{1-\alpha/2}\widehat{\sigma} \leq \sqrt{T}(\underline{v}(\mu^*) - \underline{v}(\widehat{\mu}_T)), \text{ and } \sqrt{T}(\overline{v}(\mu^*) - \overline{v}(\widehat{\mu}_T)) \leq z_{1-\alpha/2}\widehat{\sigma} |Y^T\Big),$$

and noting that

$$c(Y^T) \geq \tag{A.16}$$

$$P_\mu^*\Big(-z_{1-\alpha/2}\widehat{\sigma} \leq \sqrt{T}(\underline{v}(\mu^*)-\underline{v}(\widehat{\mu}_T)), \text{ and } \sqrt{T}(\overline{v}(\mu^*)-\overline{v}(\widehat{\mu}_T)) \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||\sqrt{T}(\mu^*-\widehat{\mu}_T)|| \leq M_{\epsilon,\eta}|Y^T\Big).$$

Take $T^*(\epsilon,\eta) = \max\{T_1(\epsilon,\eta), T_2(\epsilon,\eta), T_3(\epsilon,\eta)\}$. From Equation (A.16) and Step 2 it follows that

$$Y^T \in A^c_T(\epsilon,\eta) \implies$$

$$\text{(A.17)} \quad c(Y^T) \geq P^*_\mu\Big(-z_{1-\alpha/2}\widehat{\sigma} \leq \underline{\dot{v}}_\mu(Z^*_T) - \delta_\epsilon/2, \text{ and } \dot{\overline{v}}_\mu(Z^*_T) + \delta_\epsilon/2 \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||Z^*_T|| \leq M_{\epsilon,\eta}|Y^T\Big)$$

for $T \geq T^*$. In addition,

$$Y^T \in C^c_T(\epsilon)$$

implies that the right-hand side of equation (A.17) is larger than or equal

$$P^*_\mu\Big(-z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(Z^*_T) - \delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(Z^*_T) + \delta_\epsilon \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||Z^*_T|| \leq M_{\epsilon,\eta}|Y^T\Big).$$

This means that for $T \geq T^*(\epsilon,\eta)$

$$Y^T \in A^c_T(\epsilon) \cap C^c_T(\epsilon) \implies$$

$$\text{(A.18)} \quad c(Y^T) \geq P^*_\mu\Big(-z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(Z^*_T) - \delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(Z^*_T) + \delta_\epsilon \leq z_{1-\alpha/2}\widehat{\sigma}, \text{ and } ||Z^*_T|| \leq M_{\epsilon,\eta}|Y^T\Big).$$

Define the set

$$B = \Big\{z \in \mathbb{R}^d \;\Big|\; -z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(z) - \delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(z) + \delta_\epsilon \leq z_{1-\alpha/2}\sigma, \text{ and } ||z|| \leq M_{\epsilon,\eta}\Big\}.$$

By definition, $\underline{\dot{v}}_\mu(\cdot)$ and $\dot{\overline{v}}_\mu(\cdot)$ are linear and thus measurable functions. This means that B is a Borel Set (as it is the inverse image of a Borel subset on the real line under a measurable function). Consequently,

$$Y^T \in A^c_T(\epsilon,\eta) \cap B^c_T(\epsilon) \cap C^c_T(\epsilon)$$

implies that

$$\begin{aligned} C(Y^T) &\geq \mathbb{P}\Big(-z_{1-\alpha/2}\sigma \leq \underline{\dot{v}}_\mu(\zeta) - \delta_\epsilon, \text{ and } \dot{\overline{v}}_\mu(\zeta) + \delta_\epsilon \leq z_{1-\alpha/2}\sigma, \text{ and } ||\zeta|| \leq M_{\epsilon,\eta}\Big) - \epsilon/3 \\ &= \mathbb{P}\Big(-\underline{\dot{v}}_\mu(\zeta) \leq z_{1-\alpha/2}\sigma - \delta_\epsilon, \text{ and } -z_{1-\alpha/2}\sigma + \delta_\epsilon \leq -\dot{\overline{v}}_\mu(\zeta), \text{ and } ||\zeta|| \leq M_{\epsilon,\eta}\Big) - \epsilon/3. \end{aligned}$$

Note further that because the distribution of ζ is the same as that of $-\zeta$ and because $\underline{\dot{v}}_\mu(\cdot)$, $\dot{\overline{v}}_\mu(\cdot)$ are linear functions (by definition of derivative) we have that

$$\begin{aligned} C(Y^T) &\geq \mathbb{P}\Big(\underline{\dot{v}}_\mu(\zeta) \leq z_{1-\alpha/2}\sigma - \delta_\epsilon, \text{ and } -z_{1-\alpha/2}\sigma + \delta_\epsilon \leq \dot{\overline{v}}_\mu(\zeta), \text{ and } ||\zeta|| \leq M_\epsilon\Big) - \epsilon/3 \\ &\geq 1 - \mathbb{P}\Big(\underline{\dot{v}}_\mu(\zeta) > z_{1-\alpha/2}\sigma - \delta_\epsilon\Big) - \mathbb{P}\Big(-z_{1-\alpha/2}\sigma + \delta_\epsilon > \dot{\overline{v}}_\mu(\zeta)\Big) - 2\epsilon/3 \\ &= 1 - \mathbb{P}\Big(N(0,1) > z_{1-\alpha/2}\frac{\sigma}{\underline{\sigma}(\mu)} - \frac{\delta_\epsilon}{\underline{\sigma}(\mu)}\Big) - \mathbb{P}\Big(-z_{1-\alpha/2}\frac{\sigma}{\overline{\sigma}(\mu)} + \frac{\delta_\epsilon}{\overline{\sigma}(\mu)} > N(0,1)\Big) - 2\epsilon/3 \\ &\geq 1 - \mathbb{P}\Big(N(0,1) > z_{1-\alpha/2} - \frac{\delta_\epsilon}{\underline{\sigma}(\mu)}\Big) - \mathbb{P}\Big(-z_{1-\alpha/2} + \frac{\delta_\epsilon}{\overline{\sigma}(\mu)} > N(0,1)\Big) - 2\epsilon/3 \\ &\geq \Phi\Big(z_{1-\alpha/2} - \frac{\delta_\epsilon}{\underline{\sigma}(\mu)}\Big) - \Phi\Big(-z_{1-\alpha/2} + \frac{\delta_\epsilon}{\overline{\sigma}(\mu)}\Big) - 2\epsilon/3 \\ &\geq 1 - \alpha - \epsilon. \end{aligned}$$

Thus, we have shown that if $T \geq T^*(\epsilon,\eta)$, then

$$Y^T \in A^c_T(\epsilon,\eta) \cap B^c_T(\epsilon) \cap C^c_T(\epsilon) \implies c(Y^T) \geq 1 - \alpha - \epsilon.$$

This means that if $T \geq T^*(\epsilon,\eta)$, then

$$\begin{aligned} P\Big(c(Y^T) < 1-\alpha-\epsilon\Big) &\leq P(A_T(\epsilon,\eta)) + P(B_T(\epsilon)) + P(C_T(\epsilon)) \\ &\leq |P(A_T(\epsilon)) - \mathbb{P}(||\zeta|| > M_{\epsilon,\eta})| + \mathbb{P}(||\zeta|| > M_{\epsilon,\eta}) \\ &+ P(B_T(\epsilon)) + P(C_T(\epsilon)) \\ &\leq 4(\eta/4) \quad \text{(by equation (A.14))} . \end{aligned}$$

Therefore, for any $\epsilon > 0, \eta > 0$ there is $T^*(\epsilon,\eta)$ such that

$$P(RBC(Y^T) < 1-\alpha-\epsilon) \leq P(c(Y^T) < 1-\alpha-\epsilon) < \eta.$$

A.7. *Implementation details*

A.7.1. *Bonferroni confidence set*

This section describes the implementation of the Bonferroni-type method proposed by Granziera et al. (2017). The following algorithm is a variation of the algorithm outlined on p.17 in Granziera et al. (2017). There is one minor difference between the algorithms. To avoid dealing with degenerate $\hat{D}$, we add Step 3.c instead of implicitly adjusting the criterion function as proposed in Section 4.2 of Granziera et al. (2017). The rate of the sequence $\underline{\sigma}_T$ guaranties that the additional noise $\underline{\sigma}_T\epsilon_b$ does not affect the asymptotic distribution of $\mathcal{G}(\xi_g)$.

1. Generate N_B draws $\left\{\mu_b^*\right\}_{b=1}^{N_B} \sim N\left(\widehat{\mu}_T, \widehat{\Omega}_T\right)$.
2. Generate N_G grid points $\{x_g\}_{g=1}^{N_g}$ on a unit $d-$sphere distributed uniformly using the algorithm from Uhlig (2005).
3. For every grid point x_g, we implement the following statistical test (of size $1-\alpha/2$) of whether $B_{1g} = \widehat{\Sigma}_T^{1/2} x_g$ satisfies all identification restrictions. This is done by following steps a) to g) below.
   (a) Compute estimated residuals[21],
   $$\xi_g = \left(S'\left(\hat{\mu}_T\right), Z'\left(\hat{\mu}_T\right)\right)' B_{1g}.$$
   (b) Compute re-centered bootstrap residuals $\left\{\xi_{g;b}^*\right\}_{b=1}^{N_B}$,
   $$\tilde{\xi}_{g;b}^\star = \left(S'\left(\mu_b^*\right), Z'\left(\mu_b^*\right)\right)' \Sigma_b^{*\,1/2} x_g - \xi_g.$$
   (c) Add independent normally distributed noise with $\epsilon_b \sim N(0, I)$ and $\underline{\sigma}_T = 10^{-6}/\sqrt{T \ln(\ln T)}$,
   $$\xi_{g;b}^* = \tilde{\xi}_{g;b}^\star + \underline{\sigma}_T \epsilon_b.$$
   (d) Compute standard errors for $\left\{\xi_{g;b}^*\right\}_{b=1}^{N_B}$. The diagonal matrix $\hat{D}^{1/2}$ has the corresponding standard errors on the diagonal.
   (e) Select binding inequities as inequalities corresponding to the components ℓ of ξ_g such that
   $$e_\ell' \hat{D}^{-1/2} \xi_g \leq \kappa_T = 1.96 \ln(\ln T).$$
   (f) Compute the criterion function $\mathcal{G}(\xi_g)$ and $\left\{\mathcal{G}\left(\xi_{g;b}^*\right)\right\}_{b=1}^{N_B}$ which includes only the equalities

[21] We only compute matrices $\left(S'\left(\hat{\mu}_T\right), Z'\left(\hat{\mu}_T\right)\right)' \sqrt{\widehat{\Sigma}_T}$ and $\left(S'\left(\mu_b^*\right), Z'\left(\mu_b^*\right)\right)' \sqrt{\Sigma_b^*}$ once to speed up the costly matrix multiplication.

and the binding inequalities, where

$$\mathcal{G}\left(\xi_{g;b}^{*}\right)=\sum_{\ell=1}^{m_z}\left(e_{\ell}'\hat{D}^{-1/2}\xi_{g;b}^{*}\right)^2+\sum_{\ell=m_z+1}^{m_s+m_s}\left(e_{\ell}'\hat{D}^{-1/2}\xi_{g;b}^{*}\right)^2\mathbf{1}\left\{e_{\ell}'\hat{D}^{-1/2}\xi_g\leq\kappa_T\right\} \tag{A.19}$$

(g) Grid point x_g passes the test if $\mathcal{G}(\xi_g)$ is less that $1-\alpha/2$ sample quantile of $\left\{\mathcal{G}\left(\xi_{g;b}^{*}\right)\right\}_{b=1}^{N_B}$.

4. If x_g passes the test in Step 3, compute $\underline{\lambda}_{k,i,j}^{(g)}$ and $\bar{\lambda}_{k,i,j}^{(g)}$ as $\alpha/4$ and $1-\alpha/4$ sample quantiles of $\left\{e_i'C_k\left(A_b^{*}\right)\sqrt{\Sigma_b^{*}}x_g\right\}_{b=1}^{N_B}$ correspondingly. Otherwise set $\underline{\lambda}_{k,i,j}^{(g)}=+\infty$ and $\bar{\lambda}_{k,i,j}^{(g)}=-\infty$.
5. Report

$$\mathrm{CS}_T^{GSM}(1-\alpha)=\left[\min_{g=1,N_G}\underline{\lambda}_{k,i,j}^{(g)},\max_{g=1,N_G}\bar{\lambda}_{k,i,j}^{(g)}\right].$$

Our implementation corresponds to a generalized version of the criterion function considered in Section 6 of Granziera et al. (2017). This generalized criterion function can potentially be applied to a combination of zero and sign restrictions. In our baseline empirical application, however, the acceptance rate of Step 3 is so low that we could not find a single point out of 10000 grid points that would pass the test. For this reason, we report the results for the alternative identification scheme with the zero restriction on the FFR being replaced by a negative sign restriction.

The number of grid points that pass Step 4 of the algorithm depends crucially on the number of the identifying restrictions imposed. In our experiment, every additional sign restriction reduces the acceptance rate almost by half and, correspondingly, requires twice more grid points and computational time to achieve the same level of accuracy. For the UMP example with 4 sign restrictions the acceptance rate is 9.1%.

A.7.2. *Joint Confidence Sets*

To implement Inoue and Kilian (2013)'s algorithm, we first sample 10,000 joint draws from the posterior of reduced-form parameters and structural coefficients that satisfy all identification restriction. We use those draws to compute 10,000 structural impulse response function. Second, we sample 20,000 draws of reduced-form parameters to compute the marginal posterior density for each structural response. Third, we compute a joint 68% credible set by keeping all of structural responses which have marginal density higher than the lowest 32%. The second step is computationally costly. In our implementation it takes 2.5 hours when using 50 parallel workers in Matlab.