Supplementary appendix to the paper

Bayesian inference for partially identified convex models

Yuan Liao — Rutgers University

Anna Simoni — CREST, CNRS

February 27, 2019

Contents

Appendix

We remind some technical notation that will be used throughout the Appendix.

- for some $\delta > 0$, $B(\phi_0, \delta) = \{\phi \in \Phi : \|\phi - \phi_0\| \le \delta\}$;
- for a matrix M, $\|M\|$ denotes the Frobenius norm.

A Coverages of general nonlinear functional transformations

In this section we consider the coverage properties for $g(\theta)$ where $g(\cdot)$ is a one-dimensional continuous nonlinear transformation. Since $g(\cdot)$ is continuous and $\Theta(\phi)$ is convex, we have

$$G(\phi) = \{g(\theta) : \theta \in \Theta(\phi)\} = [\underline{g}(\phi), \bar{g}(\phi)]$$

where $\underline{g}(\phi) = \inf_{\theta\in\Theta(\phi)} g(\theta)$ and $\bar{g}(\phi) = \sup_{\theta\in\Theta(\phi)} g(\theta)$. The support function of $G(\phi)$ is then given by

$$\widetilde{S}_\phi(\nu) = \sup_{\theta\in\Theta(\phi)} \nu g(\theta), \quad \nu \in \{-1, 1\}$$

which is

$$\widetilde{S}_\phi(1) = \bar{g}(\phi), \quad \widetilde{S}_\phi(-1) = -\underline{g}(\phi).$$

Functional identified set: nonlinear case. Let c_τ be the $1-\tau$ quantile of the posterior of

$$L(\phi) := \sqrt{n} \max_{\nu=\pm 1} \left(\widetilde{S}_\phi(\nu) - \widetilde{S}_{\hat{\phi}}(\nu) \right) = \sqrt{n} \max \left(\bar{g}(\phi) - \bar{g}(\hat{\phi}), \underline{g}(\hat{\phi}) - \underline{g}(\phi) \right).$$

Then, Theorem 4.1 implies that the interval $[\underline{g}(\hat{\phi}) - \frac{c_\tau}{\sqrt{n}}, \bar{g}(\hat{\phi}) + \frac{c_\tau}{\sqrt{n}}]$ is a $1-\tau$ BCS for the marginal identified set of $g(\theta)$.

Functional partially identified parameter: nonlinear case. Define

$$\widehat{\Omega}_g := \left\{ t : P\left(t \in [\underline{g}(\phi) - \frac{2c_\tau}{\sqrt{n}}, \bar{g}(\phi) + \frac{2c_\tau}{\sqrt{n}}] \middle| D_n \right) \ge 1 - \tau \right\}.$$

Then $\widehat{\Omega}_g$ is an asymptotically valid confidence set for $g(\theta)$.

The frequentist coverage properties of these confidence sets can be proved very similarly to the linear case. But in the nonlinear case, we also require a high-level LLA assumption: there is $A_g(\nu)$ so that for

$$f_g(\phi_1, \phi_2) \;\; := \;\; \max\{P_1(\phi_1, \phi_2), P_2(\phi_1, \phi_2)\}$$

where

$$P_1(\phi_1, \phi_2) \;\; = \;\; \left| (\bar{g}(\phi_1) - \bar{g}(\phi_2)) - A_g(1)^T(\phi_1 - \phi_2) \right|$$

$$P_2(\phi_1,\phi_2) \quad = \quad \left|\left(\underline{g}(\phi_1)-\underline{g}(\phi_2)\right)-A_g(-1)^T(\phi_1-\phi_2)\right|$$

we have, for $r_n=\sqrt{1/n}$, and any $C>0$, as $n\to\infty$,

$$\sup_{\phi_1,\phi_2\in B(\phi_0,Cr_n)}\frac{f_g(\phi_1,\phi_2)}{\|\phi_1-\phi_2\|}\to 0.$$

This high-level assumption, similar to the one-dimensional case in the main text, can be verified by verifying the differentiability of $\underline{g}(\phi)$ and $\bar{g}(\phi)$ with respect to ϕ. As the main body of the proof should be very similar, we omit the formal presentation of this result for brevity.

Below we present the algorithms to compute these confidence sets.

Algorithm 5 (inference for $G(\phi)$: nonlinear case.)

1. Let $\{\phi_i\}_{i\leq M}$ be the MCMC draws from the posterior of ϕ.

2. For each $\nu=\pm1$, solve the following constrained convex problem

$$\widetilde{S}_{\hat{\phi}}(\nu)=\max_{\theta}\{\nu g(\theta):\Psi(\theta,\hat{\phi})\leq 0\}$$

   and obtain $\theta^*(\nu)=\arg\max_\theta\{\nu g(\theta):\Psi(\theta,\hat{\phi})\leq 0\}$ and the corresponding KT-vector $\lambda_g(\nu,\hat{\phi})$.

3. For each $i\leq M$, let

$$L_i \quad = \quad \sqrt{n}\max_{\nu=\pm1}\left\{A_g(\nu)^T(\phi_i-\hat{\phi})\right\}$$

   and let c_τ be the $(1-\tau)$-th quantile of $\{L_i\}_{i\leq M}$. Then the BCS for $G(\phi)$ is

$$[\underline{g}(\hat{\phi})-\frac{c_\tau}{\sqrt{n}},\bar{g}(\hat{\phi})+\frac{c_\tau}{\sqrt{n}}].$$

Algorithm 5' (inference for $g(\theta)$: nonlinear case)

1. Obtain $A_g(1), A_g(-1)$, and $\widetilde{S}_{\hat{\phi}}(\nu)$ from the above algorithm.

2. Uniformly generate $\{\tilde{\theta}_b\}_{b\leq B}$ from the parameter space Θ.

3. For each $b=1,\ldots,B$, if

$$\frac{1}{M}\sum_{i=1}^{M}1\left\{g(\tilde{\theta}_b)\in[-\underline{S}_{\phi_i}-\frac{2c_\tau}{\sqrt{n}},\bar{S}_{\phi_i}+\frac{2c_\tau}{\sqrt{n}}]\right\}\geq 1-\tau, \tag{A.1}$$

   then set $\theta_b^*:=\tilde{\theta}_b$; otherwise discard $\tilde{\theta}_b$. Here

$$\bar{S}_{\phi_i}=\widetilde{S}_{\hat{\phi}}(1)+A_g(1)^T(\phi_i-\hat{\phi}),\quad \underline{S}_{\phi_i}=\widetilde{S}_{\hat{\phi}}(-1)+A_g(-1)^T(\phi_i-\hat{\phi}).$$

4. Approximate $\widehat{\Omega}_g$ by the following interval:

$$[\min\{g(\theta_b^*)\}, \max\{g(\theta_b^*)\}].$$

B Frontier estimation in finance

In this appendix we consider Example 2.2 and provide the explicit expression of the set $\Xi(\nu, \phi)$ defined in Section 4.2. In particular, in this example the boundary Assumption 4.4 is satisfied. Having an explicit expression for $\Xi(\nu, \phi)$ is computationally convenient because it allows to avoid step 2 in Algorithm 4 to compute $A(\nu)$. The identified set of interest is

$$\Theta(\phi) = \{\theta \in \mathbb{R} \times \mathbb{R}_+; \Psi(\theta, \phi) \leq 0\}, \quad \text{where} \quad \Psi(\theta, \phi) = \phi_1\mu^2 - 2\phi_2\mu + \phi_3 - \sigma^2,$$

$\theta := (\mu, \sigma^2)$ and $\phi = (\phi_1, \phi_2, \phi_3)^T$ are as defined in Example 2.2. Because here there are no equality constraints, then $\Xi(\nu, \phi)$ is equal to the support set of $\Theta(\phi)$.

B.1 The support function

Let $\nu = (\nu_1, \nu_2) \in \mathbb{S}^2$ and denote $D := \phi_2^2 - \phi_1\phi_3$. By the Cauchy-Schwarz inequality, $\phi_2^2 \leq \phi_1\phi_3$ and then $D \leq 0$. The support function is given by $S_\phi(\nu) = \nu^T\Xi(\nu, \phi)$. The explicit expression of the support set $\Xi(\nu, \phi)$ in this example, for different values of ν, is the following:

1. for $\nu_2 > 0$, $\nu_1 > 0$:

$$\Xi(\nu, \phi) = \begin{cases} (\bar{\mu}, \bar{\sigma}^2) & \text{if } \bar{\sigma}^2 \geq \phi_1\bar{\mu}^2 - 2\phi_2\bar{\mu} + \phi_3 \\ \left(\frac{\phi_2 + \sqrt{D + \phi_1\bar{\sigma}^2}}{\phi_1}, \bar{\sigma}^2\right) & \text{if } \bar{\sigma}^2 < \phi_1\bar{\mu}^2 - 2\phi_2\bar{\mu} + \phi_3. \end{cases}$$

2. for $\nu_2 < 0$, $\nu_1 < 0$:

$$\Xi(\nu, \phi) = \begin{cases} \left(\frac{\phi_2}{\phi_1}, \phi_3 - \frac{\phi_2^2}{\phi_1}\right) \text{ if } \phi_2 > 0, \text{ and } \frac{\nu_1}{\nu_2} \leq \phi_2 I(\phi_3 \leq \bar{\sigma}^2) + \sqrt{D + \phi_1\bar{\sigma}^2} I(\phi_3 > \bar{\sigma}^2) \\ \left(\frac{\phi_2 - \sqrt{D + \phi_1\bar{\sigma}^2}}{\phi_1} I(\phi_3 > \bar{\sigma}^2), \phi_3 I(\phi_3 \leq \bar{\sigma}^2) + \bar{\sigma}^2 I(\phi_3 > \bar{\sigma}^2)\right) \quad \text{otherwise.} \end{cases}$$

3. for $\nu_2 < 0$, $\nu_1 > 0$:

   3.1. $\Xi(\nu, \phi) = \left(\frac{\phi_2}{\phi_1} - \frac{\nu_1}{2\nu_2\phi_1}, \frac{\nu_1^2}{4\nu_2\phi_1} - \frac{\phi_2^2}{\phi_1} + \phi_3\right)$, $\forall \nu$ and ϕ such that I and II below are satisfied:

   I. $2\phi_2 - 2\phi_1\bar{\mu} \leq \frac{\nu_1}{\nu_2} < 2\phi_2$ and

   II. for $\mu = \frac{\phi_2}{\phi_1} - \frac{\nu_1}{2\nu 2\phi_1}$ one of the following two conditions is verified:

   II.a $D < 0$, $D + \phi_1\bar{\sigma}^2 \geq 0$ and $\frac{\phi_2 - \sqrt{D + \phi_1\bar{\sigma}^2}}{\phi_1} \leq \mu \leq \frac{\phi_2 + \sqrt{D + \phi_1\bar{\sigma}^2}}{\phi_1}$,

   II.b $D = 0$ and $\frac{\phi_2 - \sqrt{D + \phi_1\bar{\sigma}^2}}{\phi_1} \leq \mu \leq \frac{\phi_2 + \sqrt{D + \phi_1\bar{\sigma}^2}}{\phi_1}$;

3.2. $\Xi(\nu,\phi) = \left(0, \phi_3 I(\phi_3 \leq \bar{\sigma}^2) + \bar{\sigma}^2 I(\phi_3 > \bar{\sigma}^2)\right)$, if $\frac{\nu_1}{\nu_2} \geq 2\phi_2$;

3.3. $\Xi(\nu,\phi) = \left(\bar{\mu}, \phi_1\bar{\mu}^2 - 2\phi_2\bar{\mu} + \phi_3\right)$ if $2\phi_2 - 2\phi_1\bar{\mu} > \frac{\nu_1}{\nu_2}$ and either II.a or II.b above is satisfied for $\mu = \bar{\mu}$;

3.4. (imaginary solution) $\Xi(\nu,\phi) = \left(\frac{\phi_2 + \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}, \bar{\sigma}^2\right)$ if $2\phi_2 - 2\phi_1\bar{\mu} > \frac{\nu_1}{\nu_2}$, $D < 0$ and $D + \phi_1\bar{\sigma}^2 < 0$;

3.6. $\Xi(\nu,\phi) = \left(\bar{\mu}, \bar{\sigma}^2\right)$ if $2\phi_2 - 2\phi_1\bar{\mu} > \frac{\nu_1}{\nu_2}$, $D + \phi_1\bar{\sigma}^2 \geq 0$ and either $\bar{\mu} < \frac{\phi_2 - \sqrt{D+\bar{\sigma}^2\phi_1}}{\phi_1}$ or $\bar{\mu} > \frac{\phi_2 + \sqrt{D+\bar{\sigma}^2\phi_1}}{\phi_1}$;

3.7. (imaginary solution) $\Xi(\nu,\phi) = \left(\frac{\phi_2 + \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}, \bar{\sigma}^2\right)$ if I above is satisfied, $D < 0$ and $D + \phi_1\bar{\sigma}^2 < 0$;

3.8. $\Xi(\nu,\phi) = \left(\frac{\phi_2}{\phi_1} - \frac{\nu_1}{2\nu_2\phi_1}, \bar{\sigma}^2\right)$ if I above is satisfied, $D < 0$, $D + \phi_1\bar{\sigma}^2 \geq 0$ and either $\mu < \frac{\phi_2 - \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}$ or $\mu > \frac{\phi_2 + \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}$ for $\mu = \frac{\phi_2}{\phi_1} - \frac{\nu_1}{2\nu_2\phi_1}$;

3.9. $\Xi(\nu,\phi) = \left(\frac{\phi_2}{\phi_1} - \frac{\nu_1}{2\nu_2\phi_1}, \bar{\sigma}^2\right)$ if I above is satisfied, $D = 0$ and either $\mu < \frac{\phi_2 - \sqrt{\phi_1\bar{\sigma}^2}}{\phi_1}$ or $\mu > \frac{\phi_2 + \sqrt{\phi_1\bar{\sigma}^2}}{\phi_1}$ for $\mu = \frac{\phi_2}{\phi_1} - \frac{\nu_1}{2\nu_2\phi_1}$;

4. $\nu_2 > 0$, $\nu_1 < 0$: $\Xi(\nu,\phi) = \left(\frac{\phi_2 - \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1} I(\phi_3 > \bar{\sigma}^2), \bar{\sigma}^2\right)$.

5. $\nu_2 = 0$, $\nu_1 = 1$:

$$\Xi(\nu,\phi) = \begin{cases} (\bar{\mu}, \sigma^2) & \forall\sigma^2 \in [\phi_1\bar{\mu}^2 - 2\phi_2\bar{\mu} + \phi_3, \bar{\sigma}^2] \text{ if } \bar{\sigma}^2 \geq \phi_1\bar{\mu}^2 - 2\phi_2\bar{\mu} + \phi_3 \\ \left(\frac{\phi_2 + \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}, \bar{\sigma}^2\right) & \text{if } \bar{\sigma}^2 < \phi_1\bar{\mu}^2 - 2\phi_2\bar{\mu} + \phi_3. \end{cases}$$

6. $\nu_2 = 0$, $\nu_1 = -1$:

$$\Xi(\nu,\phi) = \begin{cases} (0, \sigma^2) & \forall\sigma^2 \in [\phi_3, \bar{\sigma}^2] \text{ if } \phi_3 < \bar{\sigma}^2 \\ \left(\frac{\phi_2 - \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}, \bar{\sigma}^2\right) & \text{if } \phi_3 \geq \bar{\sigma}^2. \end{cases}$$

7. $\nu_2 = 1$, $\nu_1 = 0$: $\Xi(\nu,\phi) = (\mu, \bar{\sigma}^2)$, $\forall\mu \in [\max\left(0, \frac{\phi_2 - \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}\right), \min\left(\bar{\mu}, \frac{\phi_2 + \sqrt{D+\phi_1\bar{\sigma}^2}}{\phi_1}\right)]$.

8. $\nu_2 = -1$, $\nu_1 = 0$:

$$\Xi(\nu,\phi) = \begin{cases} \left(\frac{\phi_2}{\phi_1}, \phi_3 - \frac{\phi_2^2}{\phi_1}\right) & \text{if } \phi_2 \geq 0 \text{ and } \phi_1\phi_3 - \phi_2^2 > 0 \\ (0, \phi_3) & \text{if } \phi_2 < 0. \end{cases}$$

B.2 Simulations

The simulated data contain (gross) asset returns $\{R_t\}_{t=1}^N$, where $R_{t+1} := (P_{t+1}^1/P_t^1, \ldots, P_{t+1}^N/P_t^N)^T$ and P_t^i is the observed price of asset i at time t. Let $m := E(R_{t+1})$ and $\Sigma := E(R_{t+1} - m)(R_{t+1} - m)^T$ be the mean vector and covariance matrix of R_{t+1}. Recall the framework where M_{t+1} is the stochastic discount factor, which allows to price financial assets at time t.

Dirichlet process prior. Let F denote a probability distribution. The Bayesian model is $R_t|F \sim F$ and $(m, \Sigma) = (m(F), \Sigma(F))$, where

$$m(F) = \int rF(dr), \quad \Sigma(F) = \int rr^T F(dr) - \int rF(dr) \int rF(dr)^T.$$

Let us impose a Dirichlet process prior for F, with parameter v_0 and base probability measure Q_0 on $\mathbb{R}^N$. By Sethuraman (1994)'s decomposition, the Dirichet process prior induces a prior for (m, Σ) as: $m = \sum_{j=1}^{\infty} \alpha_j \xi_j$, and $\Sigma = \sum_{j=1}^{\infty} \alpha_j \xi_j \xi_j^T - \sum_{i=1}^{\infty} \alpha_i \xi_i \sum_{j=1}^{\infty} \alpha_j \xi_j^T$ where ξ_j are independently sampled from Q_0, $\alpha_j = u_j \prod_{l=1}^{j}(1-u_l)$ with $\{u_i\}_{i=1}^n$ drawn from Beta$(1, v_0)$. These priors then induce a prior for ϕ through the definition of ϕ_1, ϕ_2 and ϕ_3. The posterior distribution for (m, Σ) can be calculated explicitly:

$$\begin{aligned}
\Sigma|D_n \quad \sim \quad & (1-\gamma)\sum_{j=1}^{\infty} \alpha_j \xi_j \xi_j^T + \gamma \sum_{t=1}^{n} \beta_t R_t R_t^n \\
& -\left((1-\gamma)\sum_{j=1}^{\infty} \alpha_j \xi_j + \gamma \sum_{t=1}^{n} \beta_t R_t\right)\left((1-\gamma)\sum_{j=1}^{\infty} \alpha_j \xi_j + \gamma \sum_{t=1}^{n} \beta_t R_t\right)^T,
\end{aligned}$$

$$m|D_n \sim (1-\gamma)\sum_{j=1}^{\infty} \alpha_j \xi_j + \gamma \sum_{t=1}^{n} \beta_t R_t, \quad \gamma \sim \text{Beta}(n, v_0), \quad \{\beta_j\}_{j=1}^n \sim Dir(1, ..., 1).$$

We can set a truncation $K > 0$, so the infinite sums in the posterior representation are replaced with a truncated sum. We can then simulate the posterior for ϕ based on the distributions of $\Sigma|D_n$ and $m|D_n$.

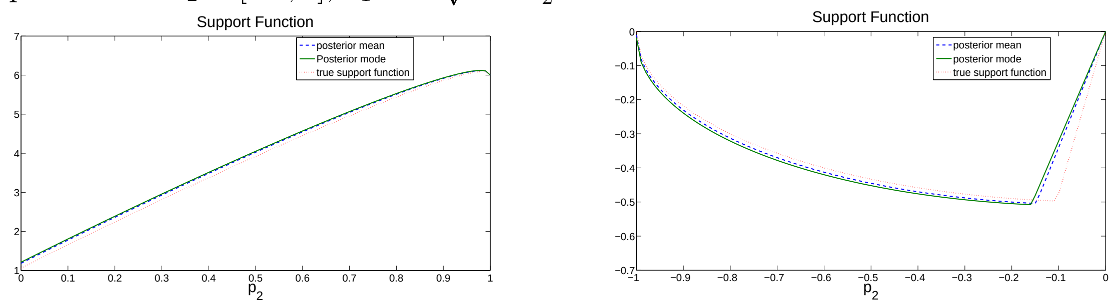
Simulation results. We present a simple simulated example. The returns R_t are generated from a 2-factor model: $R_t = \Lambda f_t + u_t + 2\iota$, where Λ is a $N \times 2$ matrix of factor loadings. The error terms $\{u_{it}\}_{i \le N, t \le n}$ are i.i.d. uniform U$[-2, 2]$. Besides, the components of Λ are standard normal, and the factors are also uniform U$[-2, 2]$. The true $m = ER_t = 2\iota$, $\Sigma = \Lambda\Lambda^T + I_N$.

We set $N = 5$ assets and $n = 200$ observations. We specify a nonparametric Dirichlet Process prior on the distribution of $R_t - m$, with parameter $v_0 = 3$, and base measure $Q_0 = \mathcal{N}(0, 1)$. Then, we obtain the posterior distributions for (m, Σ, ϕ) and sample 1,000 draws from the posterior of ϕ. Each time, we first draw (m, Σ) from their marginal posterior distributions, based on which we obtain the posterior draw of ϕ. The posterior mean $(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$ of ϕ is calculated, based on which we calculate a Bayesian estimate of the boundary of the identified set (we set $\bar{\mu} = 1.4$ and $\bar{\sigma}^2 = 6$):

$$\overline{\partial\Theta(\hat{\phi})} = \{\mu \in [0, \bar{\mu}], \sigma^2 \in [0, \bar{\sigma}^2] : \sigma^2 = \hat{\phi}_1 \mu^2 - 2\hat{\phi}_2 \mu + \hat{\phi}_3\},$$

which is used to compute the BCS for the identified set. In addition, we estimate the support function $S_\phi(\nu)$ using the posterior mean of ϕ. In Figure 1, we plot the Bayesian estimates of the support function for two cases: $\nu_2 \in [0, 1]$, $\nu_1 = \sqrt{1-\nu_2^2}$, and $\nu_2 \in [-1, 0]$, $\nu_1 = -\sqrt{1-\nu_2^2}$.

Figure 1: Posterior estimates of support function. Left panel is for $\nu_2 \in [0,1]$, $\nu_1 = \sqrt{1-\nu_2^2}$; right panel is for $\nu_2 \in [-1,0]$, $\nu_1 = -\sqrt{1-\nu_2^2}$



Using the support function, we calculate the 95% posterior quantile q_τ for $J(\phi)$, based on which we construct the BCS $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ for the identified set. The boundary of $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ is given by

$$\partial\Theta(\hat{\phi})^{q_\tau/\sqrt{n}} = \left\{\mu \in [0,\bar{\mu}], \sigma^2 \in [0,\bar{\sigma}^2] : \inf_z \sqrt{|z-\mu|^2 + |\sigma^2_{\hat{\phi}}(z) - \sigma^2|^2} = q_\tau/\sqrt{n}\right\}.$$

In Figure 2, we plot the posterior estimate $\partial\Theta(\hat{\phi})$, $\partial\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ and the boundary of the true identified set. The estimated boundaries and the BCS show that the true set is well estimated. Note that we truncate at $\mu = 0$ due to the assumed restriction on the parameter space.

C Proofs for Section 4

Theorem 4.1 has been proved in the main text. We now prove Theorem 4.2.

C.1 Proof of Theorem 4.2

Lemma C.1. *For any $x \geq 0$,*

$$P(\sqrt{n}\sup_{\|\nu\|=1} S_\phi(\nu) - S_{\hat{\phi}}(\nu) \leq x|D_n) - P_{D_n}(\sqrt{n}\sup_{\|\nu\|=1} S_{\phi_0}(\nu) - S_{\hat{\phi}}(\nu) \leq x) = o_P(1).$$

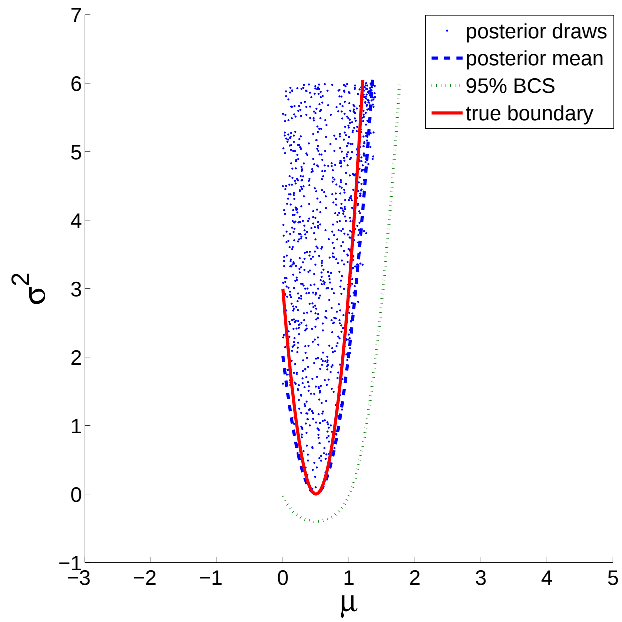
Proof. For $A(\nu)$ defined in Assumption 4.2, define

$$f_\nu^n(\phi_1,\phi_2) = \sqrt{n}A(\nu)^T(\phi_1 - \phi_2).$$

For notational simplicity, we further write $g_n(\phi_1,\phi_2) = \sqrt{n}\sup_{\|\nu\|=1} S_{\phi_1}(\nu) - S_{\phi_2}(\nu)$. Then

$$\begin{aligned}
& |P(g_n(\phi,\hat{\phi}) \leq x|D_n) - P_{D_n}(g_n(\phi_0,\hat{\phi}) \leq x)| \\
\leq & |P(\sup_{\|\nu\|=1} f_\nu^n(\phi,\hat{\phi}) \leq x|D_n) - P_{D_n}(\sup_{\|\nu\|=1} f_\nu^n(\phi_0,\hat{\phi}) \leq x)| \\
& + |P(g_n(\phi,\hat{\phi}) \leq x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi,\hat{\phi}) \leq x|D_n)|
\end{aligned}$$

Figure 2: Solid line is the boundary of the true identified set; dashed line represents the estimated boundary using the posterior mean; dotted line gives the 95% BCS of the identified set. Plots are obtained based on a single set of simulated data.



$$+|P_{D_n}(\sup_{\|\nu\|=1} f_\nu^n(\phi_0,\hat{\phi}) \leq x) - P_{D_n}(g_n(\phi_0,\hat{\phi}) \leq x)| = a_1 + a_2 + a_3.$$

We now show that $a_i = o_P(1)$ for $i = 1, 2, 3$.

$$a_1 = |P(\sqrt{n} \sup_{\|\nu\|=1} A(\nu)^T(\phi - \hat{\phi}) \leq x|D_n) - P_{D_n}(\sqrt{n} \sup_{\|\nu\|=1} A(\nu)^T(\phi_0 - \hat{\phi}) \leq x)|.$$

Assumption 2.1 *(ii)*-*(iii)* imply $\|P_{\sqrt{n}(\phi-\hat{\phi})|D_n} - \mathcal{N}(0, I_0^{-1})\|_{TV} \to^P 0$, hence the posterior distribution of $\sqrt{n}(\phi-\hat{\phi})|D_n$ and the sampling distribution of $\sqrt{n}(\hat{\phi}-\phi_0)$ are asymptotically identically distributed. In addition, the proof similar to that of Lemma D.3 in Section D.2 below shows that both the stochastic processes $\{\sqrt{n}A(\nu)^T(\phi-\hat{\phi}) : \nu \in \mathbb{S}^d\}$ (with ϕ being random, distributed according to its posterior) and $\{\sqrt{n}A(\nu)^T(\phi_0-\hat{\phi}) : \nu \in \mathbb{S}^d\}$ (with $\hat{\phi}$ being random, distributed according to its sampling distribution) are asymptotically tight. In addition, it can be easily verified that the finite dimensional posterior of $\{\sqrt{n}A(\nu_i)^T(\phi-\hat{\phi}) : i = 1, \ldots, k\}$ and the finite dimensional sampling distribution of $\{\sqrt{n}A(\nu_i)^T(\phi_0 - \hat{\phi}) : i = 1, \ldots, k\}$ both converge to the same multivariate Gaussian. Hence the processes $\{\sqrt{n}A(\cdot)^T(\phi - \hat{\phi})\}$ (with ϕ being random, distributed according to its posterior) and $\{\sqrt{n}A(\cdot)^T(\phi_0 - \hat{\phi})\}$ (with $\hat{\phi}$ being random, distributed according to its sampling distribution) are asymptotically identically distributed. Thus by the continuous mapping theorem, $a_1 = o_P(1)$.

As for a_2, define $\Delta_\phi = |g_n(\phi,\hat{\phi}) - \sup_{\|\nu\|=1} f_\nu^n(\phi,\hat{\phi})|$. For any $C > 0$, let $B_{n,C} := B(\phi_0, C/\sqrt{n})$. Then Assumption 4.1 implies that there is $c_n = o(1)$ such that for any

$C > 0$ and any $\phi \in B_{n,C}$, when $\hat{\phi} \in B_{n,C}$, we have $\Delta_\phi \le c_n$, where we used the inequality $|\sup_x h_1(x) - \sup_x h_2(x)| \le 3 \sup_x |h_1(x) - h_2(x)|$ for any two functions $h_1(x)$ and $h_2(x)$.

Also,

$$P\left(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x - \Delta_\phi \middle| D_n \right) \le P(g_n(\phi, \hat{\phi}) \le x | D_n) \le P\left(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi \middle| D_n \right).$$

We then use the inequality: if $c \le a \le b$, then $|d - a| \le |d - b| + |d - c|$:

$$\begin{aligned} a_2 &= |P(g_n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n)| \\ &\le |P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi|D_n)| \\ &\quad +|P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x - \Delta_\phi|D_n)|. \end{aligned} \tag{C.1}$$

Let

$$W_n := |P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi|D_n)|.$$

as the first term in (C.1). It is bounded as:

$$\begin{aligned} &|P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi|D_n)| \le 4 \times 1\{\hat{\phi} \notin B_{n,C}\} \\ &+|P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi|D_n)|1\{\hat{\phi} \in B_{n,C}\} \\ \le\; &|P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi, \phi \in B_{n,C}|D_n)|1\{\hat{\phi} \in B_{n,C}\} \\ &+P(\phi \notin B_{n,C}|D_n) + 4 \times 1\{\hat{\phi} \notin B_{n,C}\}. \end{aligned} \tag{C.2}$$

Write $A_n = P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + \Delta_\phi, \phi \in B_{n,C}|D_n)1\{\hat{\phi} \in B_{n,C}\}$. Then

$$\begin{aligned} A_n &\ge P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x, \phi \in B_{n,C}|D_n)1\{\hat{\phi} \in B_{n,C}\} \\ &= [P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x, \phi \notin B_{n,C}|D_n)]1\{\hat{\phi} \in B_{n,C}\} \\ &\ge P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n)1\{\hat{\phi} \in B_{n,C}\} - P(\phi \notin B_{n,C}|D_n). \end{aligned}$$

On the other hand,

$$A_n \le P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x + c_n|D_n)1\{\hat{\phi} \in B_{n,C}\}.$$

We then use the inequality: if $c \le a \le b$, then $|d - a| \le |d - b| + |d - c|$ and further bound (C.2) by:

$$|P(\sup_{\|\nu\|=1} f_\nu^n(\phi, \hat{\phi}) \le x|D_n) - A_n|1\{\hat{\phi} \in B_{n,C}\} + P(\phi \notin B_{n,C}|D_n) + 4 \times 1\{\hat{\phi} \notin B_{n,C}\}$$

$$
\begin{aligned}
&\leq \ |P(\sup_{\|\nu\|=1} f_\nu^n(\phi,\hat\phi)\leq x|D_n) - P(\sup_{\|\nu\|=1} f_\nu^n(\phi,\hat\phi)\leq x+c_n)|D_n)|1\{\hat\phi\in B_{n,C}\}\\
&\quad +2P(\phi\notin B_{n,C}|D_n)+4\times 1\{\hat\phi\notin B_{n,C}\}\\
&= \ (F_n(x+c_n)-F(x))+2P(\phi\notin B_{n,C}|D_n)+4\times 1\{\hat\phi\notin B_{n,C}\} \qquad \text{(C.3)}
\end{aligned}
$$

where $F_n(x)$ denotes the posterior distribution function of $\sup_{\|\nu\|=1} f_\nu^n(\phi,\hat\phi)$. By Assumption 2.1, for any $\epsilon,\delta\in(0,1)$, there is $C_0>0$, so that $P(\phi\notin B_{n,C_0}|D_n)<\delta$ with probability at least $1-\epsilon$. Also,

$$P_{D_n}(1\{\hat\phi\notin B_{n,C_0}\}>\delta)=P_{D_n}(\hat\phi\notin B_{n,C_0})<\epsilon.$$

Now set $C=C_0$ in (C.3). Note that W_n is independent of C_0. Since $0<F_n(x+c_n)-F(x)=o(1)$, Thus, we have proved, for any $\epsilon,\delta>0$, $P_{D_n}(W_n<\delta)>1-\epsilon$. Because ϵ,δ are arbitrary, $W_n=o_P(1)$.

The same argument implies that the second term on the right hand side of (C.1) is $o_P(1)$. Hence $a_2=o_P(1)$. Similarly, $a_3=o_P(1)$. □

Proof of Theorem 4.2

(i) Note that for any two convex compact sets K_1,K_2, $K_1\subset K_2$ if and only if their support functions satisfy $S_{K_1}(v)\leq S_{K_2}(v)$ for all $v\in\mathbb{S}^d$. In addition, $S_{K^\epsilon}(v)=S_K(v)+\epsilon$ where K^ϵ denotes the ϵ-envelope of K for convex compact K. Hence, for $S_{\Theta(\hat\phi)^{q_\tau/\sqrt{n}}}(\nu)=S_{\hat\phi}(\nu)+q_\tau/\sqrt{n}$ being the support function of $\Theta(\hat\phi)^{q_\tau/\sqrt{n}}$,

$$
\begin{aligned}
&P_{D_n}(\Theta(\phi_0)\subset\Theta(\hat\phi)^{q_\tau/\sqrt{n}}) = P_{D_n}(\sup_{\|\nu\|=1} S_{\phi_0}(\nu)-S_{\Theta(\hat\phi)^{q_\tau/\sqrt{n}}}(\nu)\leq 0)\\
&= \ P_{D_n}(\sqrt{n}\sup_{\|\nu\|=1} S_{\phi_0}(\nu)-S_{\hat\phi}(\nu)\leq q_\tau)\\
&\geq \ P(J(\phi)\leq q_\tau|D_n)+o_P(1)=1-\tau+o_P(1),
\end{aligned}
$$

where the inequality follows from Lemma C.1.

(ii) Note that $\Theta(\phi)^{2q_\tau/\sqrt{n}}=\{\theta:\theta^T\nu\leq S_\phi(\nu)+\frac{2q_\tau}{\sqrt{n}},\forall\|\nu\|=1\}$. For any $\theta\in\Theta(\phi_0)$, $\theta^T\nu\leq S_{\phi_0}(\nu)$ for all $\|\nu\|=1$. Hence, for any $\theta\in\Theta(\phi_0)$

$$
\begin{aligned}
&P(\theta\in\Theta(\phi)^{2q_\tau/\sqrt{n}}|D_n) = P(\theta^T\nu\leq S_\phi(\nu)+\frac{2q_\tau}{\sqrt{n}},\forall\|\nu\|=1|D_n)\\
&\geq \ P(\theta^T\nu\leq S_{\phi_0}(\nu), S_{\phi_0}(\nu)\leq S_{\hat\phi}(\nu)+\frac{q_\tau}{\sqrt{n}}, S_{\hat\phi}(\nu)\leq S_\phi(\nu)+\frac{q_\tau}{\sqrt{n}},\forall\|\nu\|=1|D_n)\\
&= \ P(S_{\phi_0}(\nu)\leq S_{\hat\phi}(\nu)+\frac{q_\tau}{\sqrt{n}}, S_{\hat\phi}(\nu)\leq S_\phi(\nu)+\frac{q_\tau}{\sqrt{n}},\forall\|\nu\|=1|D_n).
\end{aligned}
$$

This then implies $B_n(\theta)\cap A_n\subset C_n$, for the events defined as:

$$A_n=\{S_{\phi_0}(\nu)\leq S_{\hat\phi}(\nu)+\frac{q_\tau}{\sqrt{n}}:\forall\|\nu\|=1\},$$

$$B_n(\theta)=\{P(\theta\in\Theta(\phi)^{2q_\tau/\sqrt{n}}|D_n)<1-\tau\},$$

$$C_n=\{P(S_{\hat\phi}(\nu)\leq S_\phi(\nu)+\frac{q_\tau}{\sqrt{n}},\forall\|\nu\|=1|D_n)<1-\tau\}.$$

Also $C_n \subset \{P(\sqrt{n}\sup_{\|\nu\|=1}[S_{\hat{\phi}}(\nu) - S_\phi(\nu)] \le q_\tau|D_n) < 1-\tau\}$. By the definition of q_τ in Section 3.1, $P(\sqrt{n}\sup_{\|\nu\|=1}[S_{\hat{\phi}}(\nu) - S_\phi(\nu)] \le q_\tau|D_n) \ge 1-\tau$, hence $P_{D_n}(C_n) = 0$.
In addition, by the same proof of Lemma C.1,

$$
\begin{aligned}
& P_{D_n}(A_n) = P_{D_n}(S_{\phi_0}(\nu) \le S_{\hat{\phi}}(\nu) + \frac{q_\tau}{\sqrt{n}} : \forall\|\nu\|=1) \ge P_{D_n}(\sqrt{n}\sup_{\|\nu\|=1}[S_{\phi_0}(\nu) - S_{\hat{\phi}}(\nu)] \le q_\tau) \\
\ge \;& P(\sqrt{n}\sup_{\|\nu\|=1}[S_{\hat{\phi}}(\nu) - S_\phi(\nu)] \le q_\tau|D_n) - o_P(1) \ge 1-\tau-o_P(1).
\end{aligned}
$$

Hence

$$
\begin{aligned}
& \sup_{\theta\in\Theta(\phi_0)} P_{D_n}(\theta\notin\widehat{\Omega}) = \sup_{\theta\in\Theta(\phi_0)} P_{D_n}\left(P(\theta\in\Theta(\phi)^{2q_\tau/\sqrt{n}}|D_n) < 1-\tau\right) = \sup_{\theta\in\Theta(\phi_0)} P_{D_n}(B_n(\theta)) \\
\le \;& \sup_{\theta\in\Theta(\phi_0)} P_{D_n}(B_n(\theta)\cap A_n) + P_{D_n}(A_n^c) \le P_{D_n}(C_n) + P_{D_n}(A_n^c) = P_{D_n}(A_n^c) \le \tau + o_P(1).
\end{aligned}
$$

This completes the proof.

(iii) The proof of the frequentist coverage for the support function follows from the same argument as that of Lemma C.1, hence is omitted.

C.2 Proof of Theorem 4.3

Let S_ϕ be the support function of the full parameter's identified set. We first show that the marginal identified set for $g(\theta) = g^T\theta$ can be written as $G(\phi) = [-S_\phi(-g), S_\phi(g)]$. When $\Theta(\phi)$ is convex, it is also connected. We shall use the following fact: once $\Theta(\phi)$ is convex, then $\theta\in\Theta(\phi)$ if and only if $\theta^T\nu \le S_\phi(\nu)$ for all $\|\nu\|=1$.

On one hand, it is not difficult to verify that $G(\phi) \subset [-S_\phi(-g), S_\phi(g)]$. On the other hand, let $\tilde{S}_\phi$ denote the support function of $G(\phi)$. By definition,

$$
\tilde{S}_\phi(\nu) = \sup_{g(\theta)\in G(\phi)} \nu g(\theta) = \sup_{\theta\in\Theta(\phi)} \nu g^T\theta = S_\phi(\nu g).
$$

Hence for any $g(\theta) \in [-S_\phi(-g), S_\phi(g)]$, for $\nu=\pm 1$, we have $\nu g(\theta) \le S_\phi(\nu g) = \tilde{S}_\phi(\nu)$. This then implies $g(\theta)\in G(\phi)$. Hence $[-S_\phi(-g), S_\phi(g)] \subset G(\phi)$.

We now proceed by proving the asymptotic coverage of the marginal identified set. Then we prove the coverage of $g(\theta)$.

(i) Bayesian coverage of the marginal identified set

$$
\begin{aligned}
& P(G(\phi) \subset [-S_{\hat{\phi}}(-g) - \frac{c_\tau}{\sqrt{n}}, S_{\hat{\phi}}(g) + \frac{c_\tau}{\sqrt{n}}]|D_n) \\
= \;& P(S_\phi(g) \le S_{\hat{\phi}}(g) + \frac{c_\tau}{\sqrt{n}}, -S_\phi(-g) \ge -S_{\hat{\phi}}(-g) - \frac{c_\tau}{\sqrt{n}}|D_n) \\
= \;& P(\max_{\nu=\pm1}[S_\phi(\nu g) - S_{\hat{\phi}}(\nu g)] \le \frac{c_\tau}{\sqrt{n}}|D_n) = 1-\tau.
\end{aligned}
$$

(ii) Frequentist coverage of the marginal identified set

$$
P_{D_n}(G(\phi_0) \subset [-S_{\hat{\phi}}(-g) - \frac{c_\tau}{\sqrt{n}}, S_{\hat{\phi}}(g) + \frac{c_\tau}{\sqrt{n}}])
$$

$$
\begin{aligned}
&= P_{D_n}(S_{\phi_0}(g) \leq S_{\hat{\phi}}(g) + \frac{c_\tau}{\sqrt{n}}, -S_{\phi_0}(-g) \geq -S_{\hat{\phi}}(-g) - \frac{c_\tau}{\sqrt{n}}) \\
&= P_{D_n}(\max_{\nu=\pm 1}[S_{\phi_0}(\nu g) - S_{\hat{\phi}}(\nu g)] \leq \frac{c_\tau}{\sqrt{n}}) \geq P(\max_{\nu=\pm 1}[S_{\phi_0}(\nu g) - S_{\hat{\phi}}(\nu g)] \leq \frac{c_\tau}{\sqrt{n}}|D_n) - o_P(1) \\
&\geq 1 - \tau - o_P(1).
\end{aligned}
$$

Here the first inequality follows from the same argument as in the proof of Lemma C.1,

(iii) Coverage of the marginal parameter

We know that $G(\phi_0) = [-S_{\phi_0}(-g), S_{\phi_0}(g)]$. For any $g(\theta) \in G(\phi_0)$,

$$
\begin{aligned}
& P(g(\theta) \in [-S_\phi(-g) - 2c_\tau/\sqrt{n}, S_\phi(g) + 2c_\tau/\sqrt{n}]|D_n) \\
\geq\ & P\left(\begin{array}{c} S_{\phi_0}(g) \leq S_{\hat{\phi}}(g) + c_\tau/\sqrt{n}, S_{\hat{\phi}}(g) \leq S_\phi(g) + c_\tau/\sqrt{n}, \\ -S_{\phi_0}(-g) \geq -S_{\hat{\phi}}(-g) - c_\tau/\sqrt{n} \\ -S_{\hat{\phi}}(-g) \geq -S_\phi(-g) - c_\tau/\sqrt{n}, \\ g(\theta) \in G(\phi_0) \end{array} \middle| D_n \right).
\end{aligned}
$$

This then implies $B_n(g(\theta)) \cap A_n \subset C_n$, for the events defined as:

$$
\begin{aligned}
A_n &= \{S_{\phi_0}(g) \leq S_{\hat{\phi}}(g) + c_\tau/\sqrt{n}, -S_{\phi_0}(-g) \geq -S_{\hat{\phi}}(-g) - c_\tau/\sqrt{n}\}, \\
B_n(g(\theta)) &= \{P(g(\theta) \in [-S_\phi(-g) - 2c_\tau/\sqrt{n}, S_\phi(g) + 2c_\tau/\sqrt{n}]|D_n) < 1 - \tau\}, \\
C_n &= \{P(S_{\hat{\phi}}(g) \leq S_\phi(g) + c_\tau/\sqrt{n}, -S_{\hat{\phi}}(-g) \geq -S_\phi(-g) - c_\tau/\sqrt{n}|D_n) < 1 - \tau\} \\
&= \{P(\sqrt{n}\max_{\nu=\pm 1}(S_{\hat{\phi}}(\nu g) - S_\phi(\nu g)) < c_\tau|D_n) < 1 - \tau\}.
\end{aligned}
$$

Also, by the definition of c_τ, $1 - \tau = P(\sqrt{n}\max_{\nu=\pm 1}(S_{\hat{\phi}}(\nu g) - S_\phi(\nu g)) < c_\tau|D_n)$. Hence $P_{D_n}(C_n) = 0$.

In addition, using the same argument as in the proof of Lemma C.1,

$$
\begin{aligned}
& P_{D_n}(A_n) \geq P_{D_n}(\sqrt{n}\max_{\nu=\pm 1}[S_{\phi_0}(\nu g) - S_{\hat{\phi}}(\nu g)] \leq c_\tau) \\
\geq\ & P(\sqrt{n}\max_{\nu=\pm 1}[S_{\hat{\phi}}(\nu g) - S_\phi(\nu g)] \leq c_\tau|D_n) - o_P(1) \geq 1 - \tau - o_P(1).
\end{aligned}
$$

Hence for $\widehat{\Omega}_g = \{g(\theta) : P(g(\theta) \in [-S_\phi(-g) - 2c_\tau/\sqrt{n}, S_\phi(g) + 2c_\tau/\sqrt{n}]|D_n) \geq 1 - \tau\}$,

$$
\begin{aligned}
& \sup_{g(\theta)\in G(\phi_0)} P_{D_n}(g(\theta) \notin \widehat{\Omega}_g) \\
=\ & \sup_{g(\theta)\in G(\phi_0)} P_{D_n}\left(P(g(\theta) \in [-S_\phi(-g) - 2c_\tau/\sqrt{n}, S_\phi(g) + 2c_\tau/\sqrt{n}]|D_n) < 1 - \tau\right) \\
=\ & \sup_{g(\theta)\in G(\phi_0)} P_{D_n}(B_n(g(\theta))) \leq \sup_{g(\theta)\in G(\phi_0)} P_{D_n}(B_n(g(\theta)) \cap A_n) + P_{D_n}(A_n^c) \\
\leq\ & P_{D_n}(C_n) + P_{D_n}(A_n^c) = P_{D_n}(A_n^c) \leq \tau + o_P(1).
\end{aligned}
$$

This completes the proof.

C.3 Proof of Proposition 4.1

For any $C > 0$, denote $B_n = B(\phi_0, Cr_n)$. By the assumption, h_1, h_2 are both continuously differentiable. For any $\phi_1, \phi_2 \in B_n$, there is $\tilde{\phi} \in B_n$, so that

$$\begin{aligned} S_{\phi_1}(1) - S_{\phi_2}(1) &= h_2(\phi_1) - h_2(\phi_2) = \nabla h_2(\tilde{\phi})^T(\phi_1 - \phi_2) \\ &= \nabla h_2(\phi_0)^T(\phi_1 - \phi_2) + [\nabla h_2(\tilde{\phi}) - \nabla h_2(\phi_0)]^T(\phi_1 - \phi_2). \end{aligned}$$

Let $A(1) = \nabla h_2(\phi_0)$. Then

$$\sup_{\phi_1,\phi_2 \in B_n} \frac{|S_{\phi_1}(1) - S_{\phi_2}(1) - A(1)^T(\phi_1 - \phi_2)|}{\|\phi_1 - \phi_2\|} \leq \sup_{\phi \in B_n} \|\nabla h_2(\phi) - \nabla h_2(\phi_0)\|.$$

Because $\nabla h_2(.)$ is continuous on ϕ_0, for any $\epsilon > 0$, there is $\delta > 0$ so that $\sup_{\|\phi-\phi_0\|<\delta} \|\nabla h_2(\phi) - \nabla h_2(\phi_0)\| < \epsilon$. For all large n, $B_n \subset \{\phi : \|\phi - \phi_0\| < \delta\}$, hence $\sup_{\phi \in B_n} \|\nabla h_2(\phi) - \nabla h_2(\phi_0)\| \to 0$. Similarly, let $A(-1) = \nabla h_1(\phi_0)$, we have

$$\sup_{\phi_1,\phi_2 \in B_n} \frac{|S_{\phi_1}(-1) - S_{\phi_2}(-1) - A(-1)^T(\phi_1 - \phi_2)|}{\|\phi_1 - \phi_2\|} \to 0.$$

This completes the proof.

C.4 Proof of Lemma 4.1 and Lemma 4.2

C.4.1 inference for β_1 in Lemma 4.1

The model parameters are $(\phi, \beta_j, \gamma_j, \rho)$, with $\gamma_j \geq 0$, $j = 1, 2$. We reparametrize it by setting $d_j := \beta_j - \gamma_j$, and require $\underline{\beta} - \bar{\gamma} \leq d_j \leq \beta_j$. The proof proceeds by the following steps:

Step 1: Represent β_2, γ_2 using β_1, γ_1.

Step 2: Derive an upper bound $\beta_1 \leq U(\phi, \gamma_1, \rho)$ given (ϕ, ρ, γ_1).

Step 3: Derive a lower bound $\beta_1 \geq L(\phi, \gamma_1, \rho)$ given (ϕ, ρ, γ_1).

Step 4: Define

$$h_1(\phi) := \min_{1-|\rho|\geq\epsilon} \min_{\gamma_1} L(\phi, \gamma_1, \rho), \quad h_2(\phi) := \max_{1-|\rho|\geq\epsilon} \max_{\gamma_1} U(\phi, \gamma_1, \rho)$$

and prove that h_1, h_2 are continuously differentiable in ϕ.

Step 1. Define $\varrho := |\rho|$,

$$\begin{aligned} F_1(x, y, \varrho) &:= P(\epsilon_{i1} < x; \epsilon_{i2} < y) \\ F_2(x, y, \varrho) &:= P(\epsilon_{i1} < x; \epsilon_{i2} > y) \\ F_3(x, y, \varrho) &:= P(\epsilon_{i1} > x; \epsilon_{i2} > y) \\ F_4(x, y, z, w, \varrho) &:= P(\epsilon_{i1} < x; \epsilon_{i2} > y) + P(\epsilon_{i1} < z; w < \epsilon_{i2} < y), \quad y \geq w; x \geq z. \end{aligned}$$

These functions are defined using the joint distribution of $(\epsilon_{i1}, \epsilon_{i2})$. By the invertibility of F_1, F_2, F_3 and F_4, the following functions, (f_1, f_2, f_3, f_4), are defined: (the invertibility of F_4

follows from claim 3 below)

$$\begin{aligned}
F_1(x,y,\varrho) = m &\;\Leftrightarrow\; y = f_1(x,m,\varrho)\\
F_1(x,y,\varrho) = m &\;\Leftrightarrow\; x = f_2(y,m,\varrho)\\
F_2(x,y,\varrho) = m &\;\Leftrightarrow\; y = f_1(x,\Phi(x;\varrho)-m,\varrho)\\
F_3(x,y,\varrho) = m &\;\Leftrightarrow\; y = f_3(x,m,\varrho)\\
F_4(x,y,z,f_1(z,k,\varrho),\varrho) = m &\;\Leftrightarrow\; z = f_4(x,y,k,m,\varrho),
\end{aligned}$$

where $\Phi(x;\varrho) = P(\epsilon_{i1} \le x)$. In particular, f_2 is the inverse function of f_1; $F_2(x,y,\varrho) = m \Rightarrow y = f_1(x,\Phi(x;\varrho)-m,\varrho)$ is due to: $F_2(x,y,\varrho) = m$ if and only if $F_1(x,y,\varrho) = \Phi(x;\varrho) - m$.

Then the moment equalities:

$$\phi_{00} = P(\epsilon_{i1} < -\beta_1; \epsilon_{i2} < -\beta_2); \quad \phi_{11} = P(\epsilon_{i1} > -d_1; \epsilon_{i2} > -d_2)$$

are equivalent to

$$-\beta_2 = f_1(-\beta_1, \phi_{00}, \varrho), \quad -\beta_1 = f_2(-\beta_2, \phi_{00}, \varrho), \quad -d_2 = f_3(-d_1, \phi_{11}, \varrho). \tag{C.4}$$

In addition, the following claims will be needed.

Claim 1: $f_2(y,m,\varrho)$ is strictly decreasing in y; $f_3(x,m,\varrho)$ is strictly decreasing in x, with all other arguments fixed.

This is because for any $y_1 < y_2$, let $x_j = f_2(y_j,m,\varrho)$, then $F_1(x_j,y_j,\varrho) = m$, $j = 1,2$. F_1 is strinctly increasing in both x and y. If $x_1 \le x_2$, then $F_1(x_1,y_1,\varrho) < F_1(x_2,y_2,\varrho)$, which contradicts with $F_1(x_j,y_j,\varrho) = m$ for $j = 1,2$. Thus $x_1 > x_2$. This proves the claim for f_2. The claim for f_3 follows from the same argument.

Claim 2: $f_1(x,\Phi(x;\varrho)-m,\varrho)$ is strictly increasing in x, with all other arguments fixed.

This is because for any $x_1 < x_2$, let $y_j = f_1(x_j,\Phi(x_j;\varrho)-m,\varrho)$, then $F_2(x_j,y_j,\varrho) = m$, $j = 1,2$. Now $F_2(x,y,\varrho)$ is strictly increasing in x, and strictly decreasing in y. If $y_1 \ge y_2$, then $F_2(x_1,y_1,\varrho) < F_2(x_2,y_2,\varrho)$, which contradicts with $F_2(x_j,y_j,\varrho) = m$.

Claim 3: $F_4(x,y,z,f_1(z,k,\varrho),\varrho)$ is strictly increasing in z when $f_1(z,k,\varrho) \le y$, with all the other arguments fixed.

First, it is clear that $F_4(x,y,z,w,\varrho)$ is increasing in z (strictly if $y > w$) and strictly decreasing in w when $w \le y$. Also, because $f_1(z,k,\varrho)$ is the inverse function of $f_2(y,k,\varrho)$, it is strictly decreasing in z due to claim 1. As a result, $F_4(x,y,z,f_1(z,k,\varrho),\varrho)$ is strictly increasing in z.

Claim 4: $f_4(x,y,k,m,\varrho)$ is strictly decreasing in x and increasing in y (maybe not strictly), with all other arguments fixed.

First, for any $x_1 < x_2$, let $z_j = f_4(x_j,y,k,m,\varrho)$, then $F_4(x_j,y,z_j,f_1(z_j,k,\varrho),\varrho) = m$, $j = 1,2$. By claim 3, $F_4(x,y,z,f_1(z,k,\varrho),\varrho)$ is strictly increasing in z. It is also clear that it is strictly increasing in x. Hence if $z_1 \le z_2$, then $F_4(x_1,y,z_1,f_1(z_1,k,\varrho),\varrho) < F_4(x_2,y,z_2,f_1(z_2,k,\varrho),\varrho)$, which contradicts with $F_4(x_j,y,z_j,f_1(z_j,k,\varrho),\varrho) = m$. This proves $z_1 > z_2$, and thus $f_4(x,y,k,m,\varrho)$ is strictly decreasing in x.

Next, for any $y_1 < y_2$, let $z_j = f_4(x,y_j,k,m,\varrho)$, then $F_4(x,y_j,z_j,f_1(z_j,k,\varrho),\varrho) = m$, $j = 1,2$. We now prove $F_4(x,y,z,w,\varrho)$ is decreasing in y (strictly if $x > z$), with all the other arguments fixed with $x \geq z$.

$$\begin{aligned} & F_4(x,y,z,w,\varrho) = P(\epsilon_{i1} < x; \epsilon_{i2} > y) + P(\epsilon_{i1} < z; w < \epsilon_{i2} < y) \\ = \; & P(\epsilon_{i1} < x) - P(\epsilon_{i1} < x; \epsilon_{i2} < y) + P(\epsilon_{i1} < z; \epsilon_{i2} < y) - P(\epsilon_{i1} < z; \epsilon_{i2} < w) \\ = \; & -P(z < \epsilon_{i1} < x; \epsilon_{i2} < y) + P(\epsilon_{i1} < x) - P(\epsilon_{i1} < z; \epsilon_{i2} < w). \end{aligned}$$

Hence $F_4(x,y,z,w,\varrho)$ is decreasing in y. Also, $F_4(x,y,z,f_1(z,k,\varrho),\varrho)$ is strictly increasing in z by claim 3. Hence if $z_1 > z_2$, $F_4(x,y_1,z_1,f_1(z_1,k,\varrho),\varrho) > F_4(x,y_2,z_2,f_1(z_2,k,\varrho),\varrho)$, which contradicts with $F_4(x,y_j,z_j,f_1(z_j,k,\varrho),\varrho) = m$. Hence $z_1 \leq z_2$.

Step 2. Note that $\phi_{01} \leq P(\epsilon_{i1} < -d_1; \epsilon_{i2} > -\beta_2)$ is equivalent to

$$\phi_{01} \leq F_2(-d_1, -\beta_2, \varrho). \tag{C.5}$$

Also note that when $y = f_1(-d_1, \Phi(-d_1;\varrho) - \phi_{01}, \varrho)$, then $F_2(-d_1, y, \varrho) = \phi_{01}$. Because $F_2(x,y,\varrho)$ is decreasing in y, it is equivalent to $-\beta_2 \leq y$, which means $-\beta_2 \leq f_1(-d_1, \Phi_N(-d_1) - \phi_{01}, \varrho)$. By claim 1 and (C.4), $f_2(y,m,\varrho)$ is strictly decreasing in y, equivalent to

$$-\beta_1 = f_2(-\beta_2, \phi_{00}, \varrho) \geq f_2(f_1(-d_1, \Phi(-d_1;\varrho) - \phi_{01}, \varrho), \phi_{00}, \varrho),$$

and equivalent to $\beta_1 \leq U(\phi, d_1, \varrho)$, where

$$U(\phi, d_1, \varrho) := -f_2(f_1(-d_1, \Phi(-d_1;\varrho) - \phi_{01}, \varrho), \phi_{00}, \varrho). \tag{C.6}$$

Step 3. The constraint $\phi_{01} \geq P(\epsilon_{i1} < -(\beta_1 - \gamma_1); \epsilon_{i2} > -(\beta_2 - \gamma_2)) + P(\epsilon_{i1} < -\beta_1; -\beta_2 < \epsilon_{i2} < -(\beta_2 - \gamma_2))$ is equivalent to $\phi_{01} \geq F_4(-d_1, -d_2, -\beta_1, -\beta_2, \varrho)$. By (C.4),

$$\phi_{01} \geq F_4(-d_1, -d_2, -\beta_1, f_1(-\beta_1, \phi_{00}, \varrho), \varrho).$$

Note that when $z = f_4(-d_1, -d_2, \phi_{00}, \phi_{01}, \varrho)$, then $\phi_{01} = F_4(-d_1, -d_2, z, f_1(z, \phi_{00}, \varrho), \varrho)$. By claim 3, $z \geq -\beta_1$, which is equivalent to $-f_4(-d_1, -d_2, \phi_{00}, \phi_{01}, \varrho) \leq \beta_1$. By (C.4), $\beta_1 \geq L_1(\phi, d_1, \varrho)$, where

$$L_1(\phi, d_1, \varrho) := -f_4(-d_1, f_3(-d_1, \phi_{11}, \varrho), \phi_{00}, \phi_{01}, \varrho). \tag{C.7}$$

Recall that $d_1 \leq \beta_1$, hence $\beta_1 \geq L(\phi, d_1, \varrho)$, where

$$L(\phi, d_1, \varrho) := \max\{L_1(\phi, d_1, \varrho), d_1\}.$$

Step 4. As for the upper bound,

$$h_2(\phi) := \max_{\varrho \in \Theta_\varrho} \max_{\underline{\beta} - \bar{\gamma} \leq d_1 \leq \bar{\beta}} U(\phi, d_1, \rho),$$

by claims 1, 2, $f_2(y,m,\varrho)$ is decreasing in y and $f_1(x, \Phi(x;\varrho) - m, \varrho)$ is increasing in x, so $U(\phi, d_1, \varrho)$ is decreasing in d_1. Thus

$$\max_{\underline{\beta} - \bar{\gamma} \leq d_1 \leq \bar{\beta}} U(\phi, d_1, \varrho) = U(\phi, \underline{\beta} - \bar{\gamma}, \varrho).$$

As for the lower bound,

$$h_1(\phi) := \min_{\varrho\in\Theta_\varrho} \min_{\underline{\beta}-\bar{\gamma}\leq d_1\leq\bar{\beta}} \max\{L_1(\phi, d_1, \varrho), d_1\},$$

by claims 1, 4, $f_4(x, y, k, m, \varrho)$ is strictly decreasing in x and increasing in y and $f_3(x, m, \varrho)$ is strictly decreasing in x. Hence $L_1(\phi, d_1, \varrho)$ is strictly decreasing in d_1. To calculate $\min_{\underline{\beta}-\bar{\gamma}\leq d_1\leq\bar{\beta}} \max\{L_1(\phi, d_1, \varrho), d_1\}$, pick up a fixed $u_0 \in (\underline{\beta} - \bar{\gamma}, \bar{\beta})$. Because the parameter space Φ for ϕ is compact and $L_1(\phi, d_1, \varrho)$ is strictly decreasing in d_1, we have

$$L_1(\phi, \underline{\beta} - \bar{\gamma}, \varrho) \geq L_1(\phi, u_0, \varrho) \geq \inf_{\phi\in\Phi;\varrho\in[0,1-\epsilon]} L_1(\phi, u_0, \varrho) \geq \underline{\beta} - \bar{\gamma}$$

where the last inequality holds for sufficiently small $\underline{\beta} - \bar{\gamma}$. In addition,

$$L_1(\phi, \bar{\beta}, \varrho) \leq L_1(\phi, u_0, \varrho) \leq \sup_{\phi\in\Phi;\varrho\in[0,1-\epsilon]} L_1(\phi, u_0, \varrho) \leq \bar{\beta}$$

where the last inequality holds for sufficiently large $\bar{\beta}$. Then because $L_1(\phi, x, \varrho) - x$ is continuous in x, by the intermediate value theorem, there is $d^* \in [\underline{\beta} - \bar{\gamma}, \bar{\beta}]$, so that

$$L_1(\phi, d^*, \varrho) = d^*.$$

Because $L_1(\phi, d_1, \varrho) := -f_4(-d_1, f_3(-d_1, \phi_{11}, \varrho), \phi_{00}, \phi_{01}, \varrho)$, which is continuously differentiable in (ϕ, d_1, ϱ), we apply the implicit function theorem to $L_1(\phi, d^*, \varrho) = d^*$, leading to a continuously differentiable function $d^*(\phi, \varrho)$, so that

$$L_1(\phi, d^*, \varrho) = d^* \Leftrightarrow d^* = d^*(\phi, \varrho). \tag{C.8}$$

Moreover, because $L_1(\phi, d_1, \varrho)$ is strictly decreasing in d_1, we have

$$\min_{\underline{\beta}-\bar{\gamma}\leq d_1\leq\bar{\beta}} \max\{L_1(\phi, d_1, \varrho), d_1\} = L_1(\phi, d^*, \varrho) = d^*(\phi, \varrho),$$

which is continuously differentiable.

Therefore, $\beta \in [d^*(\phi, \varrho), U(\phi, \underline{\beta} - \bar{\gamma}, \varrho)]$, where both d^* and U are twice continuously differentiable in (ϕ, ϱ), and

$$h_1(\phi) = \inf_{\varrho\in\Theta_\varrho} d^*(\phi, \varrho), \quad h_2(\phi) = \sup_{\varrho\in\Theta_\varrho} U(\phi, \underline{\beta} - \bar{\gamma}, \varrho).$$

Now we show $\inf_{\varrho\in\Theta_\varrho} d^*(\phi, \varrho)$ is continuously differentiable in ϕ. There are two cases: (1) $\inf_{\varrho\in\Theta_\varrho} d^*(\phi, \varrho)$ is achieved for some ϱ^* at the boundary, in which case it is $d^*(\phi, \varrho^*)$, and is continuously differentiable in ϕ; (2) $\inf_{\varrho\in\Theta_\varrho} d^*(\phi, \varrho)$ is achieved by some ϱ^* that satisfies:

$$\frac{\partial d^*(\phi, \varrho^*)}{\partial \varrho} = 0$$

then by the implicit function theorem, it deduces a continuously differentiable function

$\varrho^* = \varrho^*(\phi)$, whose differentiability follows from the twice differentiability of d^*. Thus $\inf_{\varrho\in\Theta_\varrho} d^*(\phi,\varrho) = d^*(\phi,\varrho^*(\phi))$, which is continuously differentiable. The differentiability of $h_2(\phi)$ follows from a similar argument.

Putting together, $\beta \in [\inf_{\varrho\in\Theta_\varrho} d^*(\phi,\varrho), \sup_{\varrho\in\Theta_\varrho} U(\phi,\underline{\beta}-\bar{\gamma},\varrho)]$ where $d^*(\phi,\varrho)$ is defined by (C.7) and (C.8) and $U(\phi,\underline{\beta}-\bar{\gamma},\varrho)$ is defined by (C.6).

C.4.2 inference for γ_1 in Lemma 4.1

We now define two new functions f_5 and f_6 as follows. As $F_2(x,y,\varrho)$ is strictly increasing in x, we have

$$F_2(x,y,\varrho) = m \Leftrightarrow x = f_5(y,m,\varrho).$$

In addition, in step 4 above, we proved $L_1(\phi,y,\varphi)$ is strictly decreasing in y. So

$$L_1(\phi,y,\varrho) = x \Leftrightarrow y = f_6(x,\phi,\varrho).$$

Also, let

$$\begin{aligned} L_\gamma(x,\phi,\varrho) &= x + f_5(f_1(-x,\phi_{00},\varrho),\phi_{01},\varrho) \\ U_\gamma(x,\phi,\varrho) &= x - f_6(x,\phi,\varrho). \end{aligned}$$

By (C.5), $\phi_{01} \leq F_2(-d_1,-\beta_2,\varrho)$. So $-d_1 \geq f_5(-\beta_2,\phi_{01},\varrho)$. With $d_1 = \beta_1 - \gamma_1$, and the expression for β_2 in (C.4), this inequality is equivalent to

$$\gamma_1 \geq L_\gamma(\beta_1,\phi,\varrho).$$

Using the same argument as in step 1, it can be shown that $f_5(y,m,\varrho)$ is strictly increasing in y. Also note that $f_1(x,y,\varrho)$ is strictly decreasing in x, so $L_\gamma(x,\phi,\varrho)$ is strictly increasing in x. Therefore,

$$\gamma_1 \geq \inf_{\varrho\in\Theta_\varrho} \inf_{\beta_1\in[\underline{\beta},\bar{\beta}]} L_\gamma(\beta_1,\phi,\varrho) = \inf_{\varrho\in\Theta_\varrho} L_\gamma(\underline{\beta},\phi,\varrho)$$

which gives the lower bound for γ_1.

In addition, by (C.7), $\beta_1 \geq L_1(\phi,d_1,\varrho)$. With $d_1 = \beta_1 - \gamma_1$, this inequality is equivalent to

$$\gamma_1 \leq U_\gamma(\beta_1,\phi,\varrho).$$

Using the same argument as before, it can be shown that $f_6(x,\phi,\varrho)$ is strictly decreasing in x, so $U_\gamma(x,\phi,\varrho)$ is strictly increasing in x. Hence

$$\gamma_1 \leq \sup_{\varrho\in\Theta_\varrho} \sup_{\beta_1\in[\underline{\beta},\bar{\beta}]} U_\gamma(\beta_1,\phi,\varrho) = \sup_{\varrho\in\Theta_\varrho} U_\gamma(\bar{\beta},\phi,\varrho),$$

which gives the upper bound for γ_1.

The fact that $\inf_{\varrho\in\Theta_\varrho} L_\gamma(\underline{\beta},\phi,\varrho)$ and $\sup_{\varrho\in\Theta_\varrho} U_\gamma(\bar{\beta},\phi,\varrho)$ are both differentiable with respect to ϕ follows from the same argument for the differentiability of $h_1(\phi) = \inf_{\varrho\in\Theta_\varrho} d^*(\phi,\varrho)$, so we omit it for brevity. In addition, due to the continuity, as long as the true value ϕ_0 for

ϕ is such that

$$\inf_{\varrho\in\Theta_\varrho} L_\gamma(\underline{\beta},\phi_0,\varrho)>0$$

then the lower bound for γ_1, $\inf_{\varrho\in\Theta_\varrho} L_\gamma(\underline{\beta},\phi,\varrho)$, is positive for all ϕ in a neighborhood of ϕ_0.

Final step. As a more concrete example, we now calculate $d^*(\phi,\varrho)$, $U(\phi,\underline{\beta}-\bar{\gamma},\varrho)$ for β_1, and $L_\gamma(\underline{\beta},\phi,\varrho)$, $U_\gamma(\bar{\beta},\phi,\varrho)$ for γ_1 when (ϵ_1,ϵ_2) are independent bivariate normal. In this case ϱ denotes the correlation between them and equals zero. In the notation below, Φ_N denotes the CDF of standard normal, $\Phi_N^{-1}(.)$ denotes its inverse function while $\Phi_N(x)^{-1}$ denotes the reciprocal of $\Phi_N(x)$. In this case

$$\begin{array}{rcl}
F_1(x,y,\varrho)=\Phi_N(x)\Phi_N(y) & \Leftrightarrow & f_1(x,m,\varrho)=\Phi_N^{-1}(m\Phi_N(x)^{-1})\\
F_1(x,y,\varrho)=\Phi_N(x)\Phi_N(y) & \Leftrightarrow & f_2(y,m,\varrho)=\Phi_N^{-1}(m\Phi_N(y)^{-1})\\
F_3(x,y,\varrho)=\Phi_N(-x)\Phi_N(-y) & \Leftrightarrow & f_3(x,m,\varrho)=-\Phi_N^{-1}(m\Phi_N(-x)^{-1})\\
F_4(x,y,z,f_1(z,k,\varrho),\varrho) & = & \Phi_N(x)\Phi_N(-y)+\Phi_N(z)\Phi_N(y)-k=m\\
 & \Leftrightarrow & f_4(x,y,k,m,\varrho)=\Phi_N^{-1}\{[m+k-\Phi_N(x)\Phi_N(-y)]\Phi_N(y)^{-1}\}\\
f_5(y,m,\varrho) & = & \Phi_N^{-1}(m\Phi_N(-y)^{-1})\\
f_6(x,\phi,\varrho) & = & \Phi_N^{-1}(-\phi_{11}\Phi_N(x)(\phi_{10}-\Phi_N(x))^{-1}).
\end{array}$$

Then straightforward calculations give

$$\begin{array}{rcl}
U(\phi,d_1,\varrho) & = & -\Phi_N^{-1}\left(\dfrac{\phi_{00}\Phi_N(-d_1)}{\Phi_N(-d_1)-\phi_{01}}\right)\\
L_1(\phi,d_1,\varrho) & = & -\Phi_N^{-1}\left(\dfrac{(\phi_{11}+\phi_{01}+\phi_{00})\Phi_N(d_1)-\phi_{11}}{\Phi_N(d_1)-\phi_{11}}\right)\\
L_\gamma(\beta_1,\phi,\varrho) & = & \beta_1+\Phi_N^{-1}\left(\dfrac{\phi_{01}\Phi_N(-\beta_1)}{\Phi_N(-\beta_1)-\phi_{00}}\right)\\
U_\gamma(\beta_1,\phi,\varrho) & = & \beta_1-\Phi_N^{-1}\left(\dfrac{-\phi_{11}\Phi_N(\beta_1)}{\phi_{10}-\Phi_N(\beta_1)}\right).
\end{array}$$

Solving for $L_1(\phi,d_1,\varrho)=d_1$ gives (note that $1-(\phi_{11}+\phi_{01}+\phi_{00})=\phi_{10}$)

$$d^*(\phi,\varrho)=\Phi_N^{-1}(\phi_{10}+\phi_{11}).$$

So $h_1(\phi)=\Phi_N^{-1}(\phi_{10}+\phi_{11})$ and $h_2(\phi)=-\Phi_N^{-1}\left(\frac{\phi_{00}\Phi_N(-(\underline{\beta}-\bar{\gamma}))}{\Phi_N(-(\underline{\beta}-\bar{\gamma}))-\phi_{01}}\right)$.

C.4.3 Proof of Lemma 4.2

First we consider the support function for the entire vector's identified set. Let $m(\nu)=\phi_3\nu$. Let $1\{m(\nu)<0\}=(1\{m_1(\nu)<0\},\dots,1\{m_d(\nu)<0\})^T$, and $1\{m(\nu)>0\}$ is a vector defined similarly. Let $\circ$ be the component-wise product of two vectors. So

$$S_\phi(\nu)=\sup_{\phi_1\le\phi_3^{-1}\theta\le\phi_2}\nu^T\theta=\sup_{\phi_1\le\phi_3^{-1}\theta\le\phi_2}m(\nu)^T\phi_3^{-1}\theta=m(\nu)^T\xi^*_{\nu,\phi,\alpha}$$

for any $\alpha \in \mathbb{R}^{\dim(\theta)}$, where

$$\xi^*_{\nu,\phi,\alpha} = \phi_2 \circ 1\{m(\nu) > 0\} + \phi_1 \circ 1\{m(\nu) < 0\} + \alpha \circ 1\{m(\nu) = 0\}.$$

Note that $m(\nu)^T \xi^*_{\nu,\phi,\alpha}$ does not depend on α, and

$$\begin{aligned}
m(\nu)^T \xi^*_{\nu,\phi,\alpha} &= m(\nu)^T \phi_2 - m(\nu)^T [(\phi_2 - \phi_1) \circ 1\{m(\nu) < 0\}] \\
&= \frac{1}{2} m(\nu)^T (\phi_1 + \phi_2) \\
&\quad + \frac{1}{2} m(\nu)^T [(\phi_2 - \phi_1) \circ 1\{m(\nu) > 0\}] - \frac{1}{2} m(\nu)^T [(\phi_2 - \phi_1) \circ 1\{m(\nu) < 0\}] \\
&= \frac{1}{2} m(\nu)^T (\phi_1 + \phi_2) \\
&\quad + \frac{1}{2} [m(\nu) \circ 1\{m(\nu) > 0\}]^T (\phi_2 - \phi_1) - \frac{1}{2} [m(\nu) \circ 1\{m(\nu) < 0\}]^T (\phi_2 - \phi_1) \\
&= \frac{1}{2} m(\nu)^T (\phi_1 + \phi_2) + \frac{1}{2} \left| m(\nu)^T \right| (\phi_2 - \phi_1)
\end{aligned}$$

where the second to last equality is due to $a^T(b \circ c) = (a \circ c)^T b = \sum_i a_i b_i c_i$; the last equality follows from $a1\{a \geq 0\} - a1\{a < 0\} = |a|$. The absolute value is taken coordinately. Therefore, the support function for the entire vector θ's identified set is

$$S_\phi(\nu) = \frac{1}{2} \nu^T \phi_3 (\phi_1 + \phi_2) + \frac{1}{2} \left| \nu^T \phi_3 \right| (\phi_2 - \phi_1).$$

Now we study the marginal identified set for θ_k, the k th component of θ. We know that its marginal identified set is given by $[h_1(\phi), h_2(\phi)]$ where

$$h_1(\phi) = -S_\phi(-e_k), \quad h_2(\phi) = S_\phi(e_k).$$

This shows

$$\begin{aligned}
h_1(\phi) &= \phi_{3,k}^T \left(\frac{\phi_1 + \phi_2}{2} \right) - M(\phi), \\
h_2(\phi) &= \phi_{3,k}^T \left(\frac{\phi_1 + \phi_2}{2} \right) + M(\phi), \\
M(\phi) &= |\phi_{3,k}^T| \left(\frac{\phi_1 + \phi_2}{2} \right).
\end{aligned}$$

Next, because all the components of $\phi^0_{3,k}$ are nonzero, there is a neighborhood of $\phi^0_{3,k}$ so that all the components of $\phi_{3,k}$ are nonzero in this neighborhood, and the sign of $\phi_{3,k,j}$ is the same as the sign of $\phi^0_{3,k,j}$ for each j. As a result, it is straightforward to calculate that

$$M(\phi) = \frac{1}{2} \sum_{\phi^0_{3,k,j} > 0} \phi_{3,k,j} (\phi_{2,j} - \phi_{1,j}) - \frac{1}{2} \sum_{\phi^0_{3,k,j} < 0} \phi_{3,k,j} (\phi_{2,j} - \phi_{1,j}),$$

which is continuously differentiable. Q.E.D.

C.5 Proof of Theorem 4.4

For $\nu \in \mathbb{S}^d$, let θ^* be an element of $\Xi(\nu, \phi_0)$, where we leave implicit the dependence of θ^* on ν. Therefore, because $S_{\phi_0}(\nu) = \sup\{\nu^T\theta;\ \theta \in \overline{\partial\Theta(\phi_0)}\}$, then $\theta^* \in \overline{\partial\Theta(\phi_0)}$. Under Assumption 4.5 the Implicit Function Theorem holds (see *e.g.* the version of the theorem in (Dieudonné, 1960, Theorem 10.2.3)). Therefore by applying it to the system of equations $\{\Psi_i(\theta, \phi_0) = 0, \forall i \in \mathcal{I}_{\nu,\phi_0}\}$, whose solution is $(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$, there exist d_{ν,ϕ_0} unique functions

$$h_1(\cdot,\cdot), \ldots, h_{d_{\nu,\phi_0}}(\cdot,\cdot) \tag{C.9}$$

defined on a $(d - d_{\nu,\phi_0} + d_\phi)$-closed neighborhood $\mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0) := \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}) \times \mathcal{V}(\phi_0)$ of $(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$ such that:

(a) $\Psi_i(h_1(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi), \ldots, h_{d_{\nu,\phi_0}}(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi), \theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) = 0$, $\forall i \in \mathcal{I}_{\nu,\phi_0}$ and for every $(\theta_{-\mathcal{I}_{\nu,\phi}}, \phi) \in \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$;

(b) $\theta^*_j = h_j(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$, for every $j = 1, \ldots, d_{\nu,\phi_0}$;

(c) $h_j \in \mathcal{C}^2(\mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0))$.

For these functions $h_1, \ldots, h_{d_{\nu,\phi_0}}$ define, $\forall \theta_{-\mathcal{I}_{\nu,\phi_0}} \in \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}})$ and $\forall \phi \in \mathcal{V}(\phi_0)$:

$$\xi_\phi(\theta_{-\mathcal{I}_{\nu,\phi_0}}) := \nu_1 h_1(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) + \ldots + \nu_{d_{\nu,\phi_0}} h_{d_{\nu,\phi_0}}(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) + \nu^T_{-\mathcal{I}_{\nu,\phi_0}} \theta_{-\mathcal{I}_{\nu,\phi_0}}$$

and remark that

$$S_{\phi_0}(\nu) = \sup_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}})} \xi_{\phi_0}(\theta_{-\mathcal{I}_{\nu,\phi_0}}) \tag{C.10}$$

because $\theta^* \in \Xi(\nu, \phi_0)$. Moreover, for every $(\theta_{-\mathcal{I}_{\nu,\phi}}, \phi) \in \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$ we use the notation $h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) := (h_1(\theta_{-\mathcal{I}_\nu,\phi_0}, \phi), \ldots, h_{d_{\nu,\phi_0}}(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi))^T$ for the vector of functions defined above and $\nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) := \partial h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi)/\partial\phi^T$ for their partial derivative with respect to ϕ.

Lemma C.2. *Under Assumption 4.6 (ii)-(iii) there exists a $N > 0$ such that for every $n \geq N$ and every $\phi \in B(\phi_0, r_n)$ with non-empty $\Theta(\phi)$, the correspondence $\phi \mapsto \Theta(\phi)$ is compact and convex-valued and continuous at all , that is, it is upper and lower hemicontinuous.*

Proof. First, under Assumptions 4.6 *(ii)-(iii)*, $\forall \phi \in B(\phi_0, r_n)$ so that $\Theta(\phi)$ is nonempty, the set is a closed, compact convex set. Then, we have to show that the correspondence $\phi \mapsto \Theta(\phi)$ is continuous (for a definition of continuity of a correspondence see for instance Definition 17.2 in Aliprantis and Border (2006)). First, we show that $\phi \mapsto \Theta(\phi)$ is lower hemicontinuous. We show this by using Theorem 17.19 (ii) in Aliprantis and Border (2006), that is, we have to show that for any $\theta^* \in \Theta(\phi_*)$ and net $\{\phi_j\}_{j\in\mathfrak{U}_1}$ with $\phi_j \to \phi_* \in B(\phi_0, r_n)$, there exists a subnet $\{\phi_{j_\beta}\}_{\beta\in\mathfrak{U}_2}$ and a net $\{\theta_\beta\}_{\beta\in\mathfrak{U}_2}$ such that $\theta_\beta \in \Theta(\phi_{j_\beta})$ for every $\beta \in \mathfrak{U}_2$, and $\theta_\beta \to \theta^*$. Define $\Theta(\phi_*)_{in} := \{\theta \in \Theta;\ \Psi_i(\theta, \phi_*) \leq 0, i = 1, \ldots, k_1\}$ so that $\Theta(\phi_*) \subset \Theta(\phi_*)_{in}$ and consider $\theta^* \in \Theta(\phi_*)$. Since $\Theta(\phi_*)_{in}$ is convex with nonempty interior, there exists a net $\theta_j \in \Theta(\phi_*)_{in}$ such that $\theta_j \to \theta^*$. Let $\{\phi_j\}$ be a convergent net inside $B(\phi_*, r_n)$

such that $\phi_j \to \phi_* \in B(\phi_*, r_n)$ and, by compactness of $B(\phi_*, r_n)$, $\{\phi_j\}$ admits a convergent subnet. Denote by $\{\phi_{j_\beta}\}_{\beta \in \mathfrak{U}_2}$ such a subnet where $\mathfrak{U}_2 := \{\beta; \phi_{j_\beta} \in B(\phi_*, r_n),\ \theta_\beta \in \Theta(\phi_{j_\beta})$ and $\{\theta_\beta\}$ is a convergent subnet of $\{\theta_j\}\}$. Remark that $\{\theta_j\}$ admits a subnet $\{\theta_\beta\}$ since $\Theta(\phi_*)_{in}$ is compact (because it is a closed subset of a compact set). Therefore, $\theta_\beta \to \theta^*$ and $\theta_\beta \in \Theta(\phi_{j_\beta})$ by construction. This establishes lower hemicontinuity.

Next, we show that the correspondence $\phi \mapsto \Theta(\phi)$ is upper hemicontinuous. By Theorem 17.16 in Aliprantis and Border (2006) it is sufficient to show that for every net $\{\phi_j, \theta_j\}$ such that $\theta_j \in \Theta(\phi_j)$ for each j (*i.e.* (ϕ_j, θ_j) is in the graph of $\Theta(\cdot)$), if $\phi_j \to \phi$ then $\theta_j \to \theta^* \in \Theta(\phi)$.

To show this, let $\{\phi_j\}$ be a convergent net in $B(\phi_0, r_n)$ such that $\{\phi_j\} \to \phi \in B(\phi_0, r_n)$. Consider a net $\{\theta_j\} \in \Theta(\phi_j)$ so that $\{(\phi_j, \theta_j)\}$ is in the graph of $\Theta(\cdot)$ and $\theta_j \to \theta^*$. Under Assumption 4.6 *(iv)* the function $\Psi(\cdot, \cdot)$ is continuous in θ and ϕ and so: $0 \geq \Psi_i(\theta_j, \phi_j) \to \Psi_i(\theta^*, \phi)$, for $i = 1, \ldots, k_1$ and $0 = \Psi_i(\theta_j, \phi_j) \to \Psi_i(\theta^*, \phi)$, for $i = k_1 + 1, \ldots, k_1 + k_2$. This shows that $\theta^* \in \Theta(\phi)$ and upper hemicontinuity is established. Since the correspondence $\phi \mapsto \Theta(\phi)$ is lower- and upper- hemicontinuous, then it is continuous and the result of the lemma follows. $\square$

Lemma C.3. *Under Assumptions 4.4 and 4.6 (ii)-(v) there exists a $N > 0$ such that for every $n \geq N$ the correspondence $(\nu, \phi) \mapsto \Xi(\nu, \phi)$ is: (i) compact-valued and upper hemicontinuous for all $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$ such that $\Theta(\phi)$ is nonempty. Moreover, (ii) if $k_2 = 0$, is non-empty for all $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$.*

Proof. We start by considering the case $k_2 = 0$ and prove that the correspondence $(\nu, \phi) \mapsto \Xi(\nu, \phi)$ is non-empty, compact-valued and upper hemicontinuous for all $(\nu, \phi) \in \mathbb{S}^s \times B(\phi_0, r_n)$ and $n \geq N$. This is the case where there are no equality constraints that characterize the identified set. We intend to apply the Berge Maximum Theorem (*e.g.* (Aliprantis and Border, 2006, Theorem 17.31)). Because by Lemma C.2, the correspondence $\phi \mapsto \Theta(\phi)$ is non-empty, convex-valued and continuous at all $\phi \in B(\phi_0, r_n)$ such that $\Theta(\phi)$ is non-empty, then the Berge Maximum Theorem implies that $\forall \nu \in \mathbb{S}^d$ and $\forall \phi \in B(\phi_0, r_n)$, the correspondence $(\nu, \phi) \mapsto \Xi(\nu, \phi)$ has nonempty compact values and is upper hemicontinuous for $n \geq N$.
Next, we consider the case $k_2 > 0$. First, remark that by Assumption 4.4 when $k_2 = 0$ we have: $\text{argmax}_{\theta \in \Theta(\phi)} \nu^T \theta = \text{argmax}_{\theta \in \overline{\partial \Theta(\phi)}} \nu^T \theta$. Denote this set by $\widetilde{\Xi}(\nu, \phi)$, which in the case $k_2 = 0$ coincides with $\Xi(\nu, \phi)$. By the above argument we known that $\widetilde{\Xi}(\nu, \phi)$ is compact-valued and upper hemicontinuous. Moreover, for $k_2 > 0$, $\Xi(\nu, \phi) := \text{argmax}_{\theta \in \overline{\partial \Theta(\phi)}} \nu^T \theta$ is closed since it is a finite union of singleton which are closed under the usual topology of $\mathbb{R}^d$ (see *e.g.* (Rudin, 2006, page 7)). Hence, because $\Xi(\nu, \phi) = \Xi(\nu, \phi) \cap \widetilde{\Xi}(\nu, \phi)$, Theorem 17.25 in Aliprantis and Border (2006) implies that the correspondence $(\nu, \phi) \mapsto \Xi(\nu, \phi)$ is upper hemicontinuous for $n \geq N$. Moreover, because $\Xi(\nu, \phi)$ is the intersection of a closed set and a compact set, then it is compact. $\square$

Lemma C.4. *Under the Assumptions of Lemma C.3 there exists a $N > 0$ such that for every $n \geq N$ the correspondence $\nu \mapsto \Xi(\nu, \phi)$ defined on $\mathbb{S}^d$ and with values in Θ is lower hemicontinuous for every $\phi \in B(\phi_0, r_n)$ such that $\Theta(\phi)$ is non-empty.*

Proof. For every $\phi \in B(\phi_0, r_n)$, let us consider the correspondence $\nu \mapsto \Xi(\nu, \phi)$ defined on $\mathbb{S}^d$ with values in Θ which is compact under Assumption 4.6 *(ii)*. We show lower hemicontinuity of this correspondence by using (Mas-Colell et al., 1995, Definition M.H.4), that is, we have to show that for every sequence $\nu_\alpha \to \nu \in \mathbb{S}^d$ with $\nu_\alpha \in \mathbb{S}^d$ for all α and every $\theta \in \Xi(\nu, \phi)$, we can find a sequence $\theta_\alpha \to \theta$ and an integer α_0 such that $\theta_\alpha \in \Xi(\nu_\alpha, \phi)$ for $\alpha > \alpha_0$.

To show this we distinguish two cases: (i) the case where the set $\Theta(\phi)$ has no faces, so that $\Xi(\nu, \phi)$ is a singleton for every ν, and (ii) the case where the set $\Theta(\phi)$ has faces, that is, there is some $\nu \in \mathbb{S}^d$ for which $\Xi(\nu, \phi)$ is not a singleton. We consider these two cases separately and start by considering the first one. Because in this case the correspondence $\nu \mapsto \Xi(\nu, \phi)$ is singleton-valued and because it is upper hemicontinuous by Lemma C.3 for every $\phi \in B(\phi_0, r_n)$ and $n \geq N$, then we can apply (Aliprantis and Border, 2006, Lemma 17.6) which implies that $\nu \mapsto \Xi(\nu, \phi)$ is lower hemicontinuous.

Let us now consider case (ii). First, consider the case where we take a sequence $\nu_\alpha \to \nu \in \mathbb{S}^d$ such that $\Xi(\nu, \phi)$ is not a singleton and it is a face, and take a point $\theta \in \Xi(\nu, \phi)$. Because $\Xi(\nu, \phi)$ is a face (and so it is convex) then this point θ can be characterized as a convex combination of two specific points on the boundary of $\Xi(\nu, \phi)$, denoted by $\partial\Xi(\nu, \phi)$ and defined as the difference between the closure and the interior of $\Xi(\nu, \phi)$ in $\mathbb{R}^d$. For instance, one can take $\theta_1(\nu) := \mathrm{argmin}_{\theta \in \partial\Xi(\nu,\phi)}\|\theta\|$, $\theta_2(\nu) := \mathrm{argmax}_{\theta \in \partial\Xi(\nu,\phi)}\|\theta\|$ (plus some other characteristics if these norms are the same) and so, by convexity, there exists $\lambda \in [0, 1]$ such that $\theta = \lambda\theta_1(\nu) + (1 - \lambda)\theta_2(\nu)$. Remark that if $\Xi(\nu_\alpha, \phi)$ is a singleton, then $\theta = \theta_1(\nu_\alpha) = \theta_2(\nu_\alpha)$. Then, we can construct an $\alpha_0 \in \mathbb{N}$ and a sequence θ_α as follows (we suppose here that $\Theta(\phi)$ has more than one face; if not the construction can be easily adapted). Define

$$\alpha_0 = \min_\alpha \Big\{\alpha \in \mathbb{N}; \Xi(\nu_\alpha, \phi) \text{ is a singleton, } \Xi(\nu_{\alpha-1}, \phi) \text{ is not a singleton, and} \tag{C.11}$$

$$\forall\alpha' > \alpha : \text{ either } \Xi(\nu_{\alpha'}, \phi) = \Xi(\nu_\alpha, \phi) \text{ or } \Xi(\nu_{\alpha'}, \phi) = \Xi(\nu, \phi)\Big\}. \tag{C.12}$$

Let $\tilde{\theta}_\beta$ be a sequence in $\Theta(\phi)$ such that $\tilde{\theta}_\beta \to \theta(\nu_{\alpha_0})$, where $\theta(\nu_{\alpha_0}) = \Xi(\nu_{\alpha_0}, \phi)$. Remark that because $\Xi(\nu_\alpha, \phi) \subset \Theta(\phi)$ for every $\nu_\alpha \in \mathbb{S}^d$, such a sequence exists since every element of $\Xi(\nu_\alpha, \phi)$ is a limit point of $\Theta(\phi)$. Moreover, because $\Theta(\phi)$ is compact then $\tilde{\theta}_\beta$ admits a convergent subsequence $\tilde{\theta}_{\beta_\alpha}$. This subsequence can be constructed by finding a set of indices β_α such that $\tilde{\theta}_{\beta_\alpha} \to \theta(\nu_{\alpha_0})$. We can therefore construct the sequence θ_α as: $\forall\alpha \leq \alpha_0$, $\theta_\alpha = \tilde{\theta}_{\beta_\alpha}$, and $\forall\alpha > \alpha_0$, $\theta_\alpha = \lambda\theta_1(\nu_\alpha) + (1 - \lambda)\theta_2(\nu_\alpha)$, where λ is as defined above. Remark that $\theta_\alpha \to \theta$ and $\theta_\alpha \in \Xi(\nu_\alpha, \phi)$ for all $\alpha > \alpha_0$ so that lower hemicontinuity is established.

Second, consider the case where we take a sequence $\nu_\alpha \to \nu \in \mathbb{S}^d$ such that $\Xi(\nu, \phi)$ is not a singleton and contains a finite number of elements, and take a point $\theta \in \Xi(\nu, \phi)$. This case proceeds exactly as the previous case with the only difference that $\Xi(\nu, \phi)$ is not convex and so θ cannot be characterized as a convex combination of other elements of $\Xi(\nu, \phi)$. Instead, one can order the elements of $\Xi(\nu, \phi)$ based on their norm. For instance, suppose $\Xi(\nu, \phi) = \{\theta_1, \theta_2, \theta_3\}$ and $\|\theta_2\| < \|\theta_1\| < \|\theta_3\|$, so we denote $\theta_{(1)} := \theta_2, \theta_{(2)} := \theta_1, \theta_{(3)} := \theta_3$ and the $\theta \in \Xi(\nu, \phi)$ chosen has to be characterized either as $\theta_{(1)}, \theta_{(2)}$ or $\theta_{(3)}$.

Third, consider the case where we take a sequence $\nu_\alpha \to \nu \in \mathbb{S}^d$ such that $\Xi(\nu, \phi)$ is a singleton. In this case $\theta \equiv \Xi(\nu, \phi)$ and we can construct a sequence θ_α as follows: $\theta_\alpha = \mathrm{argmin}_{\theta \in \Xi(\nu_\alpha,\phi)}\|\theta\|$. It is clear that $\theta_\alpha = \Xi(\nu_\alpha, \phi)$ if $\Xi(\nu_\alpha, \phi)$ is a singleton, $\theta_\alpha \in \Xi(\nu_\alpha, \phi)$ and $\theta_\alpha \to \nu$. So, lower hemicontinuity is proved. □

Lemma C.5. *Let* $\nu \mapsto I_{\nu,\phi_0} := \{i \in \{1, 2, \ldots, k\};\ \Psi_i(\theta, \phi_0) = 0, \forall\theta \in \Xi(\nu, \phi_0)\}$ *be the corre-*

spondence defined on $\mathbb{S}^d$ with values in the set $\{1,\dots,k\}$ endowed with the discrete topology. This correspondence is continuous at every $\nu \in \mathbb{S}^d$.

Proof. Let us consider the correspondence $\nu \mapsto I_{\nu,\phi_0}$ defined on $\mathbb{S}^d$ and with values in the set $\{1,\dots,k\}$ endowed with the discrete topology. We show continuity of this correspondence by showing that it is both upper and lower hemicontinuous. We start by showing upper hemicontinuity. For that, define the constant correspondence $\widetilde{I}_{\cdot,\phi_0} : \mathbb{S}^d \to \{1,\dots,k\}$, that associates to every $\nu \in \mathbb{S}^d$ the fixed set $\{1,\dots,k\}$. This correspondence is continuous, see *e.g.* (Aliprantis and Border, 2006, page 559), and compact-valued since finite sets are compact for the discrete topology. Because for every $\nu \in \mathbb{S}^d$, $I_{\nu,\phi_0} \subset \widetilde{I}_{\nu,\phi_0}$ then I_{ν,ϕ_0} is a subcorrespondence of $\widetilde{I}_{\nu,\phi_0}$. Moreover, the graph of I_{ν,ϕ_0} is closed. Hence, by (Aliprantis and Border, 2006, Corollary 17.18), the correspondence $\nu \mapsto I_{\nu,\phi_0}$ is upper hemicontinuous at every $\nu \in \mathbb{S}^d$.
We now show lower hemicontinuity by using (Aliprantis and Border, 2006, Theorem 17.19). By compactness of $\mathbb{S}^d$, every convergent net $\nu_\alpha \to \nu$ admits a convergent subnet $\{\nu_{\alpha_j}\}_{j\in J}$ where J is a set to be defined. Let us distinguish three cases: the case where ν is such that I_{ν,ϕ_0} contains only one element and the corresponding $\Xi(\nu,\phi_0)$ is a face, the case where ν is such that I_{ν,ϕ_0} contains only one element but $\Xi(\nu,\phi_0)$ is not a face of the set, and the case where ν is such that I_{ν,ϕ_0} contains more than one element. Let us start from the first case. Because I_{ν,ϕ_0} contains only one element, say $\tilde{k} \equiv I_{\nu,\phi_0}$, then the set J can be constructed such that for every ν_{α_j}, $j \in J$, $I_{\nu_{\alpha_j},\phi_0}$ contains $\tilde{k}$ and other elements (*i.e.* $I_{\nu_{\alpha_j},\phi_0}$ is a vertex of the set). Therefore, we can construct a constant net $\{k_j\}_{j\in J} = \{\tilde{k}\}$. By construction, $k_j \to \tilde{k}$ and $k_j \in I_{\nu_{\alpha_j},\phi_0}$ for each $j \in J$. Next, we consider the second case. In this case, for every subnet $\{\nu_{\alpha_j}\}_{j\in J}$ of ν_α we can construct a net $\{k_j\}_{j\in J} = \{I_{\nu_{\alpha_j},\phi_0}\}_{j\in J}$ because $I_{\nu_{\alpha_j},\phi_0}$ contains only one element. Hence, $k_j \to I_{\nu,\phi_0}$ and $k_j \in I_{\nu_{\alpha_j},\phi_0}$ for each $j \in J$. Finally, to deal with the third case, let $\tilde{k} \in I_{\nu,\phi_0}$ and consider every subnet $\{\nu_{\alpha_j}\}_{j\in J}$ of $\nu_\alpha \to \nu$ such that $I_{\nu_{\alpha_j},\phi_0}$ contains $\tilde{k}$ for every $j \in J$. Therefore, by constructing a constant net $\{k_j\}_{j\in J} = \{\tilde{k}\}$ we have that $k_j \to \tilde{k}$ and $k_j \in I_{\nu_{\alpha_j},\phi_0}$ for each $j \in J$. □

Proof of Theorem 4.4 Under Assumptions 4.5 and 4.6 and because $\mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}})$ is compact (since it is a closed subset of a compact space) we can apply (Shapiro, 1991, Theorem 3.1) which states that, for every $\phi \in B(\phi_0, r_n)$ and n sufficiently large, the value function $\sup_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}})} \xi_\phi(\theta_{-\mathcal{I}_{\nu,\phi_0}})$ is Hadamard directionally differentiable at $\xi_{\phi_0}(\theta_{-\mathcal{I}_{\nu,\phi_0}})$ and with Hadamard directional derivative given by

$$\max_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi^*(\nu,\phi_0)} \xi_\phi(\theta_{-\mathcal{I}_{\nu,\phi_0}}) \tag{C.13}$$

where $\Xi^*(\nu,\phi_0) = \operatorname{argmax}_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \mathcal{V}(\theta^*_{-\mathcal{I}_{\nu,\phi_0}})} \xi_{\phi_0}(\theta_{-\mathcal{I}_{\nu,\phi_0}}) \subseteq \Xi(\nu,\phi_0)$. Remark that (C.13) is equal to $\max_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi(\nu,\phi_0)} \xi_\phi(\theta_{-\mathcal{I}_{\nu,\phi_0}})$ by definition of $\Xi(\nu,\phi_0)$, where $\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi(\nu,\phi_0)$ means that $\theta_{-\mathcal{I}_{\nu,\phi_0}}$ is the vector made of the last $d - d_{\nu,\phi_0}$ components of an element of $\Xi(\nu,\phi_0)$. If $\Xi^*(\nu,\phi_0)$ is a singleton, then (C.13) is replaced by $\xi_\phi(\theta^*_{-\mathcal{I}_{\nu,\phi_0}}(\nu,\phi_0))$ where $\theta^*_{-\mathcal{I}_{\nu,\phi_0}}(\nu,\phi_0)$ denotes the last $d - d_{\nu,\phi_0}$ components of $\Xi^*(\nu,\phi_0)$, and every "max" in the following of the proof has to be eliminated.

Therefore, by using Assumption 4.6 *(v)*, the definition of the Hadamard directional differential of $S_\phi(\nu)$ (see *e.g.* Shapiro (1991)) and its expression given in (C.13), then there exists a $N > 0$ such that $\forall n \geq N$, $\forall \phi \in B(\phi_0, r_n)$ and $\forall \nu \in \mathbb{S}^d$,

$$\sqrt{n}\left(S_\phi(\nu) - S_{\phi_0}(\nu)\right) = \max_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi(\nu,\phi_0)} \sqrt{n}\nu^T_{1:d_{\nu,\phi_0}} \Big(h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) - h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)\Big) + o_{a.e.}(1) \tag{C.14}$$

where $o_{a.e.}(1)$ means that the convergence to zero is almost everywhere. It follows from the proof of (Shapiro, 1991, Theorem 2.1) that, because the Hadamard directional derivative is continuous (although possibly nonlinear) and ν takes values in a compact set $\mathbb{S}^d$, then the $o_{a.e.}(1)$ is uniform in ν, so that the Hadamard directional derivative of $S_\phi(\nu)$ is bounded uniformly in ν. Remark also that the correspondence $\nu \mapsto \Xi(\nu, \phi_0)$ is continuous because by Lemmas C.3 and C.4 it is both upper and lower hemicontinuous, see *e.g.* (Aliprantis and Border, 2006, Definition 17.2). Therefore (Shapiro, 1991, Theorem 2.1) also implies that the previous linear approximation is uniform in $\nu \in \mathbb{S}^d$:

$$\sup_{\nu \in \mathbb{S}^d} \sqrt{n}\left|\left(S_\phi(\nu) - S_{\phi_0}(\nu)\right) - \max_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi(\nu,\phi_0)} \nu^T_{1:d_{\nu,\phi_0}} \Big(h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) - h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)\Big)\right| = o_{a.e.}(1). \tag{C.15}$$

In order to make the expansion (C.15) linear in $(\phi - \phi_0)$, we now apply a second order Mean Value expansion to $h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi)$ around ϕ_0. Because $h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi)$ is two times continuously differentiable with respect to ϕ, we get the following result: $\forall \phi \in B(\phi_0, r_n)$ there exists a $\bar{\phi}$ lying between ϕ and ϕ_0 such that

$$\begin{aligned} h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi) - h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0) &= \nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)(\phi - \phi_0) + \sum_{k,j} (\phi_k - \phi_{0,k}) \frac{\partial^2 h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \bar{\phi})}{\partial \phi_k \partial \phi_j} (\phi_j - \phi_{0,j}) \\ &= \nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)(\phi - \phi_0) + o\left(\|\phi - \phi_0\|\right) \end{aligned} \tag{C.16}$$

where $o\left(\|\phi - \phi_0\|\right)$ is uniform in $(\nu, \theta_{-\mathcal{I}_{\nu,\phi_0}}) \in \mathbb{S}^d \times \Xi(\nu, \phi_0)$. This is because for every j-th component of the vector h we have by the Cauchy Schwartz inequality, continuity of the second derivative of h and compactness of Θ and $B(\phi_0, r_n)$:

$$\begin{aligned} \sum_{\ell,i} (\phi_\ell - \phi_{0,\ell}) \frac{\partial^2 h_j(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \bar{\phi})}{\partial \phi_\ell \partial \phi_i} (\phi_j - \phi_{0,j}) &= (\phi - \phi_0)^T \frac{\partial^2 h_j(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \bar{\phi})}{\partial \phi \partial \phi^T} (\phi - \phi_0) \\ &\leq \|\phi - \phi_0\|^2 \left\| \frac{\partial^2 h_j(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \bar{\phi})}{\partial \phi \partial \phi^T} \right\| \\ &\leq \|\phi - \phi_0\|^2 \sup_{\theta \in \Theta} \sup_{\phi \in B(\phi_0, r_n)} \left\| \frac{\partial^2 h_j(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi)}{\partial \phi \partial \phi^T} \right\| \\ &\leq C_{\nu,\phi_0} \|\phi - \phi_0\|^2 \leq \sup_{\nu \in \mathbb{S}^d} C_{\nu,\phi_0} \|\phi - \phi_0\|^2 = O(\|\phi - \phi_0\|^2) \end{aligned}$$

where $C_{\nu,\phi_0} < \infty$ and to get the last term we have used the fact that $\sup_{\nu\in\mathbb{S}^d} C_{\nu,\phi_0} < \infty$ because, since $k < \infty$, we have at most a finite number of values of $\nu \in \mathbb{S}^d$ for which the functions h differ in the form and in the number. Therefore, we have at most a finite number of different constants C_{ν,ϕ_0}, so that the supremum becomes a maximum over a finite set which is bounded. By putting together (C.15) and (C.16) we obtain that there exists a $N > 0$ such that for every $n \geq N$ and for every $\phi \in B(\phi_0, \sqrt{1/n})$:

$$\sup_{\nu\in\mathbb{S}^d} \sqrt{n} \left| (S_\phi(\nu) - S_{\phi_0}(\nu)) - \max_{\theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi(\nu,\phi_0)} \nu^T_{1:d_{\nu,\phi_0}} \nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)(\phi - \phi_0) \right| = o(1) \qquad \text{(C.17)}$$

where the $o(1)$ term is uniform in $\phi \in B(\phi_0, \sqrt{1/n})$. Finally, by using once again the Implicit Function Theorem we have an explicit expression for $\nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$:

$$\forall \theta_{-\mathcal{I}_{\nu,\phi_0}} \in \Xi(\nu, \phi_0), \qquad \nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0) := \left.\frac{\partial h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi)}{\partial \phi^T}\right|_{\phi=\phi_0}$$

$$= \left(- \left(\frac{\partial \Psi_i(h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0), \theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)}{\partial \theta_{\mathcal{I}_{\nu,\phi_0}, i'}} \right)^{-1}_{i,i'\in\mathcal{I}_{\nu,\phi_0}} \left(\frac{\partial \Psi_i(h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0), \theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)}{\partial \phi_j} \right)_{i\in\mathcal{I}_{\nu,\phi_0}} \right)_{j=1,\dots,d_\phi}$$

$$= \left(- \left(\frac{\partial \Psi_i(\theta^*, \phi_0)}{\partial \theta_{\mathcal{I}_{\nu,\phi_0}, i'}} \right)^{-1}_{i,i'\in\mathcal{I}_{\nu,\phi_0}} \left(\frac{\partial \Psi_i(\theta^*, \phi_0)}{\partial \phi_j} \right)_{i\in\mathcal{I}_{\nu,\phi_0}} \right)_{j=1,\dots,d_\phi}, \qquad \forall \theta^* \in \Xi(\nu, \phi_0).$$

To show continuity of $A(\nu)$, we notice that the correspondence $\nu \mapsto \Xi(\nu, \phi_0)$ is continuous because it is upper and lower hemicontinuous by Lemmas C.3 and C.4. Moreover, let Gr denote the graph of a correspondence, then the function $(\nu, \theta) \mapsto \nu^T_{1:d_{\nu,\phi_0}} \nabla_\phi h(\theta_{-\mathcal{I}_{\nu,\phi_0}}, \phi_0)$ is continuous on $Gr\Xi$ by Assumption 4.5 *(i)*, Lemma C.5 and because the projection functions $\pi_1 : 2^{\{1,\dots,k\}} \times \mathbb{S}^d \to \mathbb{S}^d$ and $\pi_2 : 2^{\{1,\dots,k\}} \times \Theta \to \Theta$ are continuous in ν, θ and in their first component (the latter is due to the fact that any function on a space with discrete topology is continuous). Therefore, the Berge Maximum Theorem (see *e.g.* (Aliprantis and Border, 2006, Theorem 17.31)) implies that the value function $A(\nu)$ is continuous. This concludes the proof of the theorem.

D Proofs for Section 5

D.1 Proof of Theorem 5.1

For a given $\nu \in \mathbb{S}^d$, denote $h_n := \sqrt{n}(S_\phi(\nu) - S_{\phi_0}(\nu))$ and, for $\nu \mapsto A(\nu)$ as in Assumption 4.1 denote $g_n := \sqrt{n}A(\nu)^T[\phi - \phi_1]$ and $\zeta_n := (h_n - g_n)$. Denote by $P_{h_n|D_n}$ and $P_{g_n|D_n}$ the posterior distributions of h_n and g_n, respectively, and by Φ_n the $\mathcal{N}(\bar{\Delta}_{n,\phi_0}(\nu), \bar{I}_0^{-1}(\nu))$ distribution. Let $r_n = n^{-1/2}$ and for any bounded $C > 0$, let $B_{n,C} := \{h_n; |h_n| \leq C\}$. For a given set B and a distribution P denote by P^B the distribution obtained by restricting P to B and next renormalizing. The proof consists of two steps. In the first step we show that the difference between $P^{B_{n,C}}_{h_n|D_n}$ and $\Phi_n^{B_{n,C}}$ converges to zero in probability. Second, we show that the difference between $P^{B_{n,C}}_{h_n|D_n}$, $\Phi^{B_{n,C}}_n$ and their respective unrestricted versions converges to zero in probability.

Denoting by $d\Phi^{B_{n,C}}_n(u)$, $dP^{B_{n,C}}_{h_n|D_n}(u|D_n)$ and $dP^{B_{n,C}}_{g_n|D_n}(u|D_n)$ the Lebesgue densities of $\Phi^{B_{n,C}}_n$, $P^{B_{n,C}}_{h_n|D_n}$ and $dP^{B_{n,C}}_{g_n|D_n}$, respectively, evaluated at u, we can write the total variation distance as:

$$\begin{aligned}\frac{1}{2}\|P^{B_{n,C}}_{h_n|D_n} - \Phi^{B_{n,C}}_n\|_{TV} &= \int \left(1 - \frac{d\Phi^{B_{n,C}}_n(u)}{dP^{B_{n,C}}_{h_n|D_n}(u|D_n)}\right)_+ dP^{B_{n,C}}_{h_n|D_n}(u|D_n) \\ &\to^p \int \left(1 - \frac{d\Phi^{B_{n,C}}_n(u)}{dP^{B_{n,C}}_{g_n|D_n}(u|D_n)}\right)_+ dP^{B_{n,C}}_{g_n|D_n}(u|D_n) \quad \text{(D.1)}\end{aligned}$$

where we have used the fact that by Assumption 4.1 $P_{h_n|D_n}$ is the convolution of the posterior distributions of g_n and ζ_n and the distribution of ζ_n converges towards a Dirac measure in zero since $\zeta_n = o(\sqrt{n}\|\phi - \phi_0\|) = o(1)$ for every $\phi \in B_{n,C}$. Moreover, by Assumption 2.1 *(ii)* the right hand side of (D.1) converges to zero in probability. Therefore, for any $\epsilon, \delta > 0$ there exists $C > 0$ such that

$$\begin{aligned}&E\left[\frac{1}{2}\|P^{B_{n,C}}_{h_n|D_n} - \Phi^{B_{n,C}}_n\|_{TV}\right] \\ &= E\left[\frac{1}{2}\|P^{B_{n,C}}_{h_n|D_n} - \Phi^{B_{n,C}}_n\|_{TV}1\{P(B_{n,C}|D_n) > 1-\delta\} + \frac{1}{2}\|P^{B_{n,C}}_{h_n|D_n} - \Phi^{B_{n,C}}_n\|_{TV}1\{P(B^c_{n,C}|D_n) > \delta\}\right] \\ &\le E\left[\frac{1}{2}\|P^{B_{n,C}}_{h_n|D_n} - \Phi^{B_{n,C}}_n\|_{TV}1\{P(B_{n,C}|D_n) > 1-\delta\}\right] + P_{D_n}(P(B^c_{n,C}|D_n) > \delta) \\ &\le o(1) + \epsilon\end{aligned}$$

where we have used Assumption 2.1 *(i)* to get $P_{D_n}(P(B^c_{n,C}|D_n) > \delta) = 1 - P_{D_n}(P(B^c_{n,C}|D_n) < \delta) \le \epsilon$.

The second step of the proof is easily developed by noticing that: $\|\Phi_n - \Phi^{B_{n,C}}_n\|_{TV} = \|\Phi^{B^c_{n,C}}_n\|_{TV} = o_p(1)$ and $\|P_{h_n|D_n} - P^{B_{n,C}}_{h_n|D_n}\|_{TV} \le 2P(B^c_{n,C}|D_n)$ which is less than δ with P_{D_n}-probability at least $1-\epsilon$ for any $\delta, \epsilon > 0$ by Lemma D.1. Since this is true for any $\epsilon, \delta > 0$ then we have shown that $E\left[\frac{1}{2}\|P_{h_n|D_n} - \Phi_n\|_{TV}\right] \to 0$.

Lemma D.1. *Under Assumptions 2.1 (i) and 4.1, for any $\epsilon, \delta > 0$, there exists $C > 0$, such that with P_{D_n}-probability $1-\epsilon$,*

$$P\left(\sup_{\nu \in \mathbb{S}^d} |S_\phi(\nu) - S_{\phi_0}(\nu)| > Cn^{-1/2} \,\middle|\, D_n\right) < \delta.$$

Proof. Denote $r_n = n^{-1/2}$, $B_{n,C_0} = B(\phi_0, C_0 r_n)$ and $\Omega = \{\phi \in B_{n,C_0}\}$. Under Assumption 2.1 *(i)*: for any $\epsilon, \delta > 0$, there exists $C > 0$ such that with probability at least $1-\epsilon$, $P(\Omega^c|D_n) < \delta$. Also define $\mathfrak{A} := \{\sup_{\nu \in \mathbb{S}^d} |S_\phi(\nu) - S_{\phi_0}(\nu)| > Cr_n\}$ for some constant $C > 0$.

Then, by using the expansion of the support function given in Assumption 4.1 we have

$$P_{D_n}\left(P\Big(\mathfrak{A}\Big|D_n\Big) < \delta\right) = P_{D_n}\left(P\Big(\mathfrak{A}\cap\Omega\Big|D_n\Big) + P\Big(\mathfrak{A}\cap\Omega^c\Big|D_n\Big) < \delta\right)$$
$$\geq P_{D_n}\left(P\Big(\mathfrak{A}\cap\Omega\Big|D_n\Big) + P(\Omega^c|D_n) < \delta\right)$$
$$\geq P_{D_n}\left(P\Big(\{f(\phi,\phi_0) + \sup_{\nu\in\mathbb{S}^d}|A(\nu)^T[\phi-\phi_0]| > Cr_n\}\cap\Omega\Big|D_n\Big) + P(\Omega^c|D_n) < \delta\right)$$
$$\geq P_{D_n}\left(P\Big(\{\Big(\sup_{\phi\in B(\phi_0,r_n)}\frac{f(\phi,\phi_0)}{\|\phi-\phi_0\|} + \sup_{\nu\in\mathbb{S}^d}\|A(\nu)\|\Big)\|\phi-\phi_0\| > Cr_n\}\cap\Omega\Big|D_n\Big) + P(\Omega^c|D_n) < \delta\right).$$

Remark that $\sup_{\nu\in\mathbb{S}^d}\|A(\nu)\| < \infty$ by continuity of $\nu\mapsto A(\nu)$ and compactness of $\mathbb{S}^d$. Therefore, by choosing $C = C_0\left(\sup_{\phi\in B(\phi_0,r_n)}\frac{f(\phi,\phi_0)}{\|\phi-\phi_0\|} + \sup_{\nu\in\mathbb{S}^d}\|A(\nu)\|\right)$ we conclude that

$$P_{D_n}\left(P\Big(\mathfrak{A}\Big|D_n\Big) < \delta\right) \geq P_{D_n}\left(P(\|\phi-\phi_0\| > C_0 r_n\cap\Omega|D_n) + P(\Omega^c|D_n) < \delta\right)$$
$$= P_{D_n}\left(P(\Omega^c|D_n) < \delta\right) \geq 1-\epsilon.$$

since on Ω: $\{\|\phi-\phi_0\| > C_0 r_n\}$ has posterior probability zero. This concludes the proof. □

D.2 Proof of Theorem 5.2

We start by stating and proving two lemmas that will be used to prove the theorem.

Lemma D.2. *Let $\mathcal{M}_\kappa$ denote any finite set of κ points in $\mathbb{S}^d$: $\mathcal{M}_\kappa = (\nu_1,\ldots,\nu_\kappa)$, and $H(\phi,\mathcal{M}_\kappa)$ denote the κ-tuple $H(\phi,\mathcal{M}_\kappa) = (S_\phi(\nu_i) - S_{\phi_0}(\nu_i))_{i=1}^{\kappa}$. If Assumption 4.1 holds, then under Assumption 2.1 (ii):*

$$\|P_{\sqrt{n}H(\phi,\mathcal{M}_\kappa)|D_n} - \mathcal{N}(\bar{\Delta}_{n,\nu_0},\bar{I}_0^{-1})\|_{TV} \to^p 0$$

for any finite subset $\mathcal{M}_\kappa = (\nu_1,\ldots,\nu_\kappa)$ of $\mathbb{S}^d$ where $\bar{\Delta}_{n,\phi_0} = \Lambda_\kappa^T\Delta_{n,\phi_0}$ and $\bar{I}_0^{-1} = \Lambda_\kappa^T I_0^{-1}\Lambda_\kappa$ with $\Lambda_\kappa = (A(\nu_1),\ldots,A(\nu_\kappa))$.

Proof. The proof proceeds exactly as the proof of Theorem 5.1 by replacing $h_n(\nu) = \sqrt{n}(S_\phi(\nu) - S_{\phi_0}(\nu))$ with the vector $h_n(\mathcal{M}_\kappa) := \sqrt{n}H(\phi,\mathcal{M}_\kappa)$ and by remarking that Assumption 4.1 implies that $h_n(\mathcal{M}_\kappa)$ is asymptotically equal to $\sqrt{n}\Lambda_\kappa^T[\phi-\phi_0]$ uniformly on $\mathbb{S}^d$. □

Lemma D.3. *Let Assumptions 2.1 (i) and 4.1 hold. Then, the process $h_n(\cdot) = \sqrt{n}\left(S_\phi(\cdot) - S_{\phi_0}(\cdot)\right)$ is asymptotically tight in probability, that is, for all $\varepsilon > 0$*

$$\lim_{\rho\to 0}\limsup_{n\to\infty} P\left\{\sup_{||\nu_1-\nu_2||\leq\rho}|h_n(\nu_1) - h_n(\nu_2)| > \varepsilon\,\middle|\, D_n\right\} = 0. \tag{D.2}$$

Proof. Let $r_n = n^{-1/2}$ and for any bounded $C > 0$, let $B_{n,C} := B(\phi_0, Cr_n)$. For every $\nu\in\mathbb{S}^d$ and for $A(\nu)$ as in Assumption 4.1 define $g_n(\nu) := \sqrt{n}A(\nu)^T[\phi-\phi_0]$ and $\zeta_n(\nu) := h_n(\nu) - g_n(\nu)$. By Assumption 4.1, for every $\phi\in B_{n,C}$, $\zeta_n(\nu) \leq \sup_{\nu\in\mathbb{S}^d}|\zeta_n(\nu)| = \sqrt{n}f(\phi,\phi_0) =$

$o(\sqrt{n}\|\phi - \phi_0\|) = o(1)$ where $f(\phi, \phi_0)$ is as defined in the statement of Assumption 4.1. Therefore, on $B_{n,C}$

$$\sup_{\|\nu_1-\nu_2\|\leq\rho} |h_n(\nu_1) - h_n(\nu_2)| \leq \sup_{\|\nu_1-\nu_2\|\leq\rho} |g_n(\nu_1) - g_n(\nu_2)| + 2\sup_{\nu\in\mathbb{S}^d} |\zeta_n(\nu)|$$
$$= \sup_{\|\nu_1-\nu_2\|\leq\rho} |g_n(\nu_1) - g_n(\nu_2)| + o(1). \quad \text{(D.3)}$$

Next, let us consider the probability in (D.2)

$$P\left\{\sup_{\|\nu_1-\nu_2\|\leq\rho} |h_n(\nu_1) - h_n(\nu_2)| > \varepsilon \middle| D_n\right\} = P\left\{\sup_{\|\nu_1-\nu_2\|\leq\rho} |h_n(\nu_1) - h_n(\nu_2)| > \varepsilon \middle| D_n, B_{n,C}\right\} P(B_{n,C}|D_n)$$
$$+ P\left\{\sup_{\|\nu_1-\nu_2\|\leq\rho} |h_n(\nu_1) - h_n(\nu_2)| > \varepsilon \middle| D_n, B_{n,C}^c\right\} P(B_{n,C}^c|D_n)$$
$$\leq P\left\{\sup_{\|\nu_1-\nu_2\|\leq\rho} |g_n(\nu_1) - g_n(\nu_2)| + o(1) > \varepsilon \middle| D_n, B_{n,C}\right\} P(B_{n,C}|D_n) + P(B_{n,C}^c|D_n). \quad \text{(D.4)}$$

We start by considering the first probability in (D.4). For this, remark that

$$\sup_{\|\nu_1-\nu_2\|\leq\rho} |g_n(\nu_1) - g_n(\nu_2)| \leq \sup_{\|\nu_1-\nu_2\|\leq\rho} \|A(\nu_1) - A(\nu_2)\|\sqrt{n}\|\phi - \phi_0\|.$$

Since the function $\nu \mapsto A(\nu)$ is uniformly continuous in $\nu \in \mathbb{S}^d$ by Assumption 4.1 and compactness of $\mathbb{S}^d$ it follows that $\forall\delta > 0$ there exists a $\rho > 0$ such that

$$\sup_{||\nu_1-\nu_2||\leq\rho} \|A(\nu_1) - A(\nu_2)\| < \delta.$$

Since this is valid for an arbitrary δ and $\rho \to 0$, then $\sup_{||\nu_1-\nu_2||\leq\rho} \|A(\nu_1) - A(\nu_2)\| \to 0$ as $\rho \to 0$. Therefore, for every $\epsilon > 0$ and $\delta > 0$ there exists a ρ such that

$$P\left\{\sup_{\|\nu_1-\nu_2\|\leq\rho} |g_n(\nu_1) - g_n(\nu_2)| + o(1) > \varepsilon \middle| D_n, B_{n,C}\right\} \leq P\left\{\delta + o(1) > \epsilon \middle| D_n, B_{n,C}\right\} \quad \text{(D.5)}$$

that converges to zero as $\rho \to 0$ and $n \to \infty$. Let us analyze the second term in (D.4). By Assumption 2.1 *(i)*, for every $\epsilon, \delta > 0$ there exists a $C_0 > 0$ such that $P(\phi \notin B_{n,C_0}|D_n) < \delta$ with P_{D_n}-probability at least $1 - \epsilon$. Take $C = C_0$ in (D.4). Because this is valid for all $\epsilon, \delta > 0$ and $P\left\{\sup_{||\nu_1-\nu_2||\leq\rho} |h_n(\nu_1) - h_n(\nu_2)| > \varepsilon \middle| D_n\right\}$ in (D.4) does not depend on C, then we conclude that $P\left\{\sup_{||\nu_1-\nu_2||\leq\rho} |h_n(\nu_1) - h_n(\nu_2)| > \varepsilon \middle| D_n\right\} = o_p(1)$ as $n \to \infty$ and $\rho \to 0$. □

We are now ready to prove the result of Theorem 5.2. Let the finite dimensional posterior distributions of $S_\phi(\cdot)$ be the posterior distributions of $\{S_\phi(\nu_i) - S_{\phi_0}(\nu_i) : i = 1, \ldots, \kappa\}$ for each κ-tuple $(\nu_1, \ldots, \nu_\kappa)$ of points in $\mathbb{S}^d$ and any finite κ. By Lemma D.2 the finite dimensional posterior distributions of the stochastic process $S_\phi(\cdot)$ converge, in total variation distance, to the marginal distributions of the Gaussian measure $\mathbb{G}$. Moreover, Lemma D.3 shows that $S_\phi(\cdot)$ is asymptotically tight in probability. Therefore, by Theorem 1.5.4 in van der Vaart and Wellner (1996), the stochastic process $S_\phi(\cdot)$ converges weakly to the Gaussian measure $\mathbb{G}$. Finally, by using one of the characterization of weak convergence given by the Portmanteau theorem (*e.g.* van der Vaart and Wellner (1996, Theorem 1.3.4)), we can write

$$P_{\sqrt{n}(S_\phi(\cdot)-S_{\phi_0}(\cdot))|D_n}(B) \to^p \mathbb{G}(B)$$

for every Borel set B with $\mathbb{G}(\partial B) = 0$.

E LLA in the Multi-Dimensional Case based on the Lagrangian representation

In this section we provide an alternative way to verify the LLA condition of Assumption 4.1 in the more complex multi-dimensional case when the support function does not necessarily have a closed-form. The approach and the set of assumptions used here are an alternative to the ones used in Section 4.2. We still allow the set of interest $\Theta(\phi)$ to be characterized by both equalities and inequalities but put s little bit more structure on it. More precisely, we consider the structure of the identified set given in the following assumption. For some $\delta > 0$, recall that $B(\phi_0, \delta) := \{\phi \in \Phi : \|\phi - \phi_0\| \leq \delta\}$.

Assumption E.1. *The identified set $\Theta(\phi)$ in (1.2) is defined as*

$$\Theta(\phi) := \left\{ \begin{array}{lc} \theta \in \Theta : & a_i^T\theta + \Psi_{s,i}(\theta, \phi) \leq 0,\ i = 1, \dots, k_1 \\ & \text{and} \quad a_i^T\theta + b_i(\phi) = 0,\ i = k_1 + 1, \dots, k_1 + k_2 \end{array} \right\}$$

where $k_1 + k_2 = k$, $\{a_i\}_{i=1}^k$ are known d-vectors, $\{b_i(\cdot)\}_{i=k_1+1}^k$ are known functions that depend only on ϕ and $\{\Psi_{s,i}(\cdot,\cdot)\}_{i=1}^{k_1}$ are known functions that depend on both θ and ϕ. Moreover, (i) there is a $\delta > 0$ such that for all $\phi \in B(\phi_0, \delta)$ and all $i = 1, \dots, k_1$, the function $\theta \mapsto \Psi_{s,i}(\theta, \phi)$ may depend only on a subvector of θ and is strictly convex in this subvector; (ii) for $i = 1, \dots, k_1$, $\Psi_{s,i}(\theta, \phi)$ is continuous in (θ, ϕ) and for $i = k_1 + 1, \dots, k_1 + k_2$, $b_i(\cdot)$ is a continuous real-valued function of ϕ.

Note that we allow the cases $k_1 = 0$ (equality constraints only) or $k_2 = 0$ (inequality constraints only). Under this assumption, the first k_1 constraints (inequality constraints on θ) that define the set allow: linear (in θ) constraints-only, strictly convex (in θ) constraints-only and, the sum of these two types of constraints. On the other hand, we restrict to the linear (in θ) equality constraints, and admit this as a potential drawback in applications when the support function does not have a closed form. In these cases, we intend to use a Lagrange representation of the support function and this is often possible if the equality constraints are affine functions of θ (see *e.g.* Rockafellar (1970)).

In the following, we denote by $\Psi_s(\theta, \phi) := \{\Psi_{s,i}(\theta, \phi)\}_{i=1}^{k_1}$ the k_1-vector that collects the k_1 functions $\Psi_{s,i}(\theta, \phi)$. Moreover, we denote $\Psi(\theta, \phi)$ as the k-vector that contains all the moment functions, that is, $\Psi(\theta, \phi) := (\{a_i^T\theta + \Psi_{s,i}(\theta, \phi)\}_{i=1}^{k_1}, \{a_i^T\theta + b_i(\phi)\}_{i=k_1+1}^{k})$. For each

(θ, ϕ), define

$$Act(\theta, \phi) := \{i \leq k;\ \Psi_i(\theta, \phi) = 0\}$$

as the set of the inequality active constraint indices and equality constraint indices. By definition, for every $\theta \in \Theta(\phi)$ the number of elements in $Act(\theta, \phi)$ is at least k_2.

We make the following further assumptions to derive the LLA condition in Assumption 4.1 for the identified set of Assumption E.1. Denote by $\nabla_\phi \Psi(\theta, \phi)$ the $k \times d_\phi$ matrix of partial derivatives of Ψ with respect to ϕ, and by $\nabla_\theta \Psi_i(\theta, \phi)$ the d-vector of partial derivatives of Ψ_i with respect to θ for each $i \leq k$. Their existence and continuity is assumed in the Assumption E.4 below.

Assumption E.2. *(i) The true parameter value for ϕ_0 is in the interior of Φ;*
(ii) The parameter space $\Theta \subset \mathbb{R}^d$ is convex, compact and has nonempty interior (relative to $\mathbb{R}^d$).

Assumption E.3. *For any $\theta \in \Theta(\phi_0)$, the gradient vectors $\{\nabla_\theta \Psi_i(\theta, \phi_0)\}_{i \in Act(\theta, \phi_0)}$ are linearly independent.*

Assumption E.4. *There is $\delta > 0$ such that for all $\phi \in B(\phi_0, \delta)$, we have:*
(i) the matrix $\nabla_\phi \Psi(\theta, \phi)$ exists and is continuous in $(\theta, \phi) \in \Theta(\phi) \times B(\phi_0, \delta)$;
(ii) $\Theta(\phi) \neq \emptyset$ and $\Theta(\phi)$ is contained in the interior of Θ (relative to $\mathbb{R}^d$);
(iii) the vector $\nabla_\theta \Psi_i(\theta, \phi)$ exists and is continuous in $(\theta, \phi) \in \Theta(\phi) \times B(\phi_0, \delta)$ for every $i \leq k$.

Consider the ordinary convex problem that defines the support function: $S_\phi(\nu) := \sup_\theta\{\nu^T\theta;\ \theta \in \Theta(\phi)\}$ where $\Theta(\phi)$ is characterized as in Assumption E.1, and assume that this optimal value is finite. Assumption E.3 guarantees: existence of a unique Kuhn-Tucker (KT) vector for this problem and that the strong duality holds, so that the KT-conditions are necessary and sufficient optimality conditions. Therefore, under the previous assumptions and if $S_\phi(\cdot) < \infty$, the support function admits a Lagrangian representation, see (Rockafellar, 1970, Theorem 28.2): $\forall \phi \in B(\phi_0, \delta)$ with δ as in Assumptions E.1, and $\forall \nu \in \mathbb{S}^d$

$$S_\phi(\nu) = \sup_{\theta \in \Theta} \left\{\nu^T\theta - \lambda(\nu, \phi)^T \Psi(\theta, \phi)\right\}, \tag{E.1}$$

where $\lambda(\nu, \phi) : \mathbb{S}^d \times B(\phi_0, \delta) \to \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$ is a k-vector of KT multipliers. Moreover, define

$$\Xi(\nu, \phi) := \arg\max_{\theta \in \Theta}\{\nu^T\theta : \Psi(\theta, \phi) \leq 0\} \tag{E.2}$$

as the *support set* of $\Theta(\phi)$. Then, by definition,

$$\nu^T\theta = S_\phi(\nu), \quad \forall \theta \in \Xi(\nu, \phi)$$

and the maximizers in (E.2) consist of the boundary points of $\Theta(\phi)$ at which the set $\Theta(\phi)$ is tangent to the hyperplane $\{\theta \in \Theta; \nu^T\theta = S_\phi(\nu)\}$.

Discussion of the assumptions. Our imposed assumptions are very similar to those of Kaido and Santos (2014)'s. Assumption E.1 is more general in the sense that we allow (linear) equality constraints while they do not. We also place the same type of restrictions on the convexity of the maps $\theta \mapsto \Psi_i(\theta, \phi)$, $i = 1, \ldots, k_1$. Assumption E.1 also requires the slopes a_i of the moment functions to be known. When the slope of a linear constraint is unknown (depending on the unknown data distribution F), there might be no asymptotically linear regular estimator for the identified set. This could be the case for instance when the constraint is linear in θ. In fact, in this case Kaido and Santos (2014) show through an example that the identified set may be nondifferentiable in ϕ. Lack of differentiability implies that there is no asymptotically linear regular estimator.

Assumption E.3 requires that the active inequality and equality gradients $\nabla_\theta \Psi_i(\theta, \phi_0)$ be linearly independent, which guarantees that the strong duality holds. Even though the strong duality can be guaranteed under weaker assumptions (e.g., like the Slater's condition, see *e.g.* (Rockafellar, 1970, Theorem 28.3)), Assumption E.3 also ensures the uniqueness of the KT-vector which we need in our proofs. Moreover, as remarked by Kaido and Santos (2014), it is possible to construct testing procedures to detect cases where Assumption E.3 does not hold.

Assumption E.4 *(i)* and *(iii)* are used to prove directional differentiability of the function $\phi \mapsto S_\phi(\nu)$. Assumption E.4 *(ii)* means that the boundary of $\Theta(\phi)$ is determined by the inequalities/equalities and not by the parameter space Θ. Assumptions very similar to Assumption E.4 are made also in Kaido and Santos (2014).

E.1 LLA for the support function

Assumptions E.1-E.4 imply that the support function of the closed and convex set $\Theta(\phi)$ admits directional derivatives in ϕ and that it is differentiable at ϕ. The next theorem exploits this fact and states that the support function can be locally approximated by a linear function of ϕ establishing, in this way, Assumption 4.1.

Theorem E.1 (LLA). *If Assumptions E.2, E.3 hold and Assumptions E.1 and E.4 hold with $\delta = r_n$ for some $r_n = o(1)$, then for all large n, there exist: (i) a real function $f(\phi_1, \phi_2)$ defined for every $\phi_1, \phi_2 \in B(\phi_0, r_n)$, (ii) a vector function of KT multipliers $\lambda(\cdot, \cdot) : \mathbb{S}^d \times \mathbb{R}^{d_\phi} \to \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$, and (iii) a Borel measurable mapping $\theta_*(\cdot) : \mathbb{S}^d \to \Theta$ satisfying $\theta_*(\nu) \in \Xi(\nu, \phi_0)$ for all $\nu \in \mathbb{S}^d$, such that :*

$$\sup_{\nu \in \mathbb{S}^d} \left| (S_{\phi_1}(\nu) - S_{\phi_2}(\nu)) - \lambda(\nu, \phi_0)^T \nabla_\phi \Psi(\theta_*(\nu), \phi_0)[\phi_1 - \phi_2] \right| = f(\phi_1, \phi_2)$$

and

$$\frac{f(\phi_1, \phi_2)}{\|\phi_1 - \phi_2\|} \to 0$$

uniformly in $\phi_1, \phi_2 \in B(\phi_0, r_n)$ as $n \to \infty$.

We remark that the functions λ and θ_* do not depend on the specific choice of ϕ_1 and ϕ_2 inside $B(\phi_0, r_n)$, but only on ν and the true value ϕ_0. The linear expansion can also be viewed as stochastic when ϕ_1, ϕ_2 are interpreted as random variables associated with the posterior distribution $P(\phi | D_n)$.

The linear approximation given in Theorem E.1 is particularly helpful to implement our procedure. In fact, as shown in Section 3 in numerical simulations and implementations, one can compute the support function by using a simple linear transformation, as stated in Theorem E.1, instead of solving an optimization problem which might be challenging in some cases.

The LLA of the support function provided in Theorem E.1 is similar to the one proved in Kaido and Santos (2014) to characterize the influence function of the support function, and our proof of Theorem E.1 rests in many places on the proof of Kaido and Santos (2014). However, there are some differences in the two results: (1) we state this result for every ϕ in a shrinking ball centred on the true ϕ_0 while their result is stated at $\hat{\phi}$ (by using our notation); (2) our LLA is valid on a ball of radius r_n, for any $r_n = o(1)$, while their result is established for a rate $r_n = n^{-1/2}$; (3) we establish the LLA of the support function when the set is more generally characterized by moment inequalities and equalities, while moment equalities are not allowed in Kaido and Santos (2014).

Remark E.1. A local linear approximation similar to the one given in Theorem E.1 can be obtained in the more general case where ϕ is the data distribution, $\phi := F$, if Assumption 2.1 *(i)-(ii)* is replaced by Assumption 2.1$''$ below. In this case, the identified set $\Theta(\phi)$, as defined in Assumption E.1, depends on the unknown distribution ϕ of X through an index function that depends on θ, contrarily to the previous case where the index function did not depend on θ. More precisely, for two known real-valued functions $\Psi^1_{s,i}$ and $\Psi^2_{s,i}$, the identified set writes (by omitting the equality constraints for simplicity)

$$\Theta(\phi) := \left\{\theta \in \Theta; a_i^T\theta + \Psi^1_{s,i}\left(\int \Psi^2_{s,i}(x,\theta)d\phi(x)\right) \leq 0, i = 1,\dots,k_1\right\}$$

where $\{a_i\}_{i=1}^{k_1}$ are as defined in Assumption E.1 and $\theta \mapsto \Psi^1_{s,i}\left(\int \Psi^2_{s,i}(x,\theta)d\phi(x)\right)$ is strictly convex in θ, for every $i = 1,\dots,k_1$ (or strictly convex in a subvector of θ if it depends only on this subvector). This case is the general case considered in Kaido and Santos (2014). With this characterization of $\Theta(\phi)$ the prior on ϕ has to be nonparametric. Let us denote the vectors of indices $g(\phi,\theta) := \int \Psi^2_s(x,\theta)d\phi(x)$ and $g_0(\theta) := g(\phi_0,\theta)$ where $\phi_0 = F_0$. Assumption 2.1 *(i)* and *(ii)* has to be replaced by the following one:

Assumption 2.1$''$**.** *(i) The posterior of ϕ is such that, for any $\epsilon,\delta > 0$, there is $C > 0$ such that, for every $\nu \in \mathbb{S}^d$ and $\theta_*(\nu) \in \Xi(\nu,\phi_0)$, with P_{D_n}-probability at least $1-\epsilon$,*

$$P(\|g(\phi,\theta_*(\nu)) - g_0(\theta_*(\nu))\| > Cn^{-1/2}|D_n) < \delta.$$

(ii) For a given $\nu \in \mathbb{S}^d$, let $P_{\sqrt{n}(g_*-g_0)|D_n}$ denote the posterior distribution of $\sqrt{n}(g(\phi,\theta_*(\nu)) - g_0(\theta_*(\nu)))$ with $\theta_*(\nu) \in \Xi(\nu,\phi_0)$. We assume that, for every $\nu \in \mathbb{S}^d$ and $\theta_*(\nu) \in \Xi(\nu,\phi_0)$,

$$\|P_{\sqrt{n}(g_*-g_0)|D_n} - \mathcal{N}(\Delta_{n,g_0}(\nu), I_0^{-1}(\nu))\|_{TV} \to^P 0$$

where $\mathcal{N}$ denotes the d_ϕ-dimensional normal distribution, $\Delta_{n,g_0}(\nu) := n^{-1/2}\sum_{i=1}^n I_0^{-1}(\nu)\ell_{g_0}(X_i)$, ℓ_{g_0} is the semiparametric efficient score function of the model and $I_0^{-1}(\nu) := E[\tilde{\psi}(\nu)\tilde{\psi}(\nu)^T]$, where $\tilde{\psi}(\nu)$ is the semiparametric efficient influence function for estimating $g_0(\theta_*(\nu))$.

We remark that $\Delta_{n,g_0}(\nu)$ and $I_0^{-1}(\nu)$ in Assumption 2.1$''$ depend on ν. This is because the index g depends on θ which in turn is taken to be an element of $\Xi(\nu,\phi_0)$. Moreover, by slightly modifying Assumptions E.1-E.4 the linear expansion in Theorem E.1 holds with $\nabla_\phi\Psi(\theta_*(\nu),\phi_0)[\phi_1-\phi_2]$ replaced by

$$\nabla\Psi_s^1\left(\int\Psi_s^2(x,\theta_*(\nu))d\phi_0(x)\right)\left[\int\Psi_s^2(x,\theta_*(\nu))d\phi_2(x)-\int\Psi_s^2(x,\theta_*(\nu))d\phi_1(x)\right] \quad \text{(E.3)}$$

where $\theta_*(\nu)\in\Xi(\nu,\phi_0)$, and for every probability distribution ϕ_1,ϕ_2 such that $g(\phi_i,\theta_*(\nu))\in B(g_0(\theta_*(\nu)),r_n)$, for $i\in\{1,2\}$ and $\forall\nu\in\mathbb{S}^d$. The notation $\nabla\Psi_s^1$ means the gradient of the function $\Psi_s^1(\cdot)$. This result is obtained by computing the Gâteaux differential of the functional $\phi\mapsto\Psi_s^1\left(\int\Psi_s^2(x,\theta_*(\nu))d\phi(x)\right)$ at ϕ_1 in the direction ϕ_2.

E.2 Proof of Theorem E.1

We use the following notation. Under Assumption E.1, we denote by A the $k_1\times d$ matrix obtained by stacking the vectors $\{a_i^T\}_{i=1}^{k_1}$ row-wise. For every $1\le i\le k_1$, define $\mathcal{S}_i\subseteq\{1,\dots,d\}$ as the set of indices of the arguments of $\theta\mapsto\Psi_{s,i}(\theta,\phi_0)$. So, if $\mathcal{S}_i=\emptyset$, then the i-th constraint is linear in θ. If $\mathcal{S}_i=\{1,\dots,d\}$, then the i-th constraint is strictly convex in all the components of θ.

Moreover, we recall the notation:

- $\Xi(\nu,\phi)=\arg\max_{\theta\in\Theta}\{\nu^T\theta;\ \Psi(\theta,\phi)\le 0\}$ is the *support set* of $\Theta(\phi)$;
- $\nabla_\phi\Psi(\theta,\phi)$ is the $k\times d_\phi$ matrix of partial derivatives of Ψ with respect to ϕ;
- $\nabla_\theta\Psi(\theta,\phi)$ is the $d\times k$ matrix of partial derivatives of Ψ with respect to θ;
- for some $r_n=o(1)$, for every $\phi\in B(\phi_0,r_n)$ and $\theta\in\Theta(\phi)$, we denote by $Act(\theta,\phi)\equiv\{i\le k;\ \Psi_i(\theta,\phi)=0\}$ the set of the inequality active constraint indices and equality constraint indices at θ and we denote by $d_A(\theta,\phi)$ the number of its elements;
- $\forall i\in Act(\theta,\phi)$, $\nabla_\theta\Psi_i(\theta,\phi)$ denotes the d-vector of partial derivatives of Ψ_i with respect to θ;
- for a set A, $int(A)$ denotes its interior and ∂A its boundary.

The proof of Theorem E.1 is composed of three steps. First, we show the differentiability of $S_{(\cdot)}(\nu)$, with respect to ϕ, for every $\nu\in\mathbb{S}^d$. Then, we apply the mean value theorem. Finally, we show that for any $\phi_1,\phi_2\in B(\phi_0,r_n)$ the first derivative of $S_{(\cdot)}(\nu)$ is equal to a linear function of $(\phi_1-\phi_2)$ which depends on the true ϕ_0 plus a negligible term. These three steps require several technical lemmas to be proved, and for ease of presentation we state and prove these lemmas in Appendix F.

Step 1: Differentiability. For any $\tau\in[0,1]$ and any $\phi_1,\phi_2\in B(\phi_0,r_n)$, define $\phi_\tau=\tau\phi_1+(1-\tau)\phi_2$ with $\phi_2=\phi_\tau|_{\tau=0}$ and $\phi_1=\phi_\tau|_{\tau=1}$. For every $\nu\in\mathbb{S}^d$, the support function, as a function of ϕ and restricted to $B(\phi_0,r_n)$, may be rewritten as a function of τ: $S_{\phi_{(\cdot)}}(\nu):[0,1]\to\mathbb{R}$ which is continuous under Assumption E.1 *(ii)*. Lemma E.1 shows that the function $S_{\phi_{(\cdot)}}(\nu)$ is differentiable at $\tau=\tau_0\in(0,1)$.

Step 2: Applying the mean value theorem. The mean value theorem applied to the map $\tau \to S_{\phi_\tau}(\nu)$ implies that there exists a point $\tau_0 \in (0,1)$ such that:

$$S_{\phi_1}(\nu) - S_{\phi_2}(\nu) = \frac{\partial}{\partial \tau} S_{\phi_\tau}(\nu)\bigg|_{\tau=\tau_0\in(0,1)} \tag{E.4}$$

where $\phi_\tau = \tau\phi_1 + (1-\tau)\phi_2$, $\phi_1, \phi_2 \in B(\phi_0, r_n)$. By Lemma E.1, there exists a N such that $\forall n \geq N$ and for all $\nu \in \mathbb{S}^d$, $\phi_1, \phi_2 \in B(\phi_0, r_n)$, the function $S_{\phi_{(\cdot)}}(\nu)$ is differentiable at $\tau = \tau_0 \in (0,1)$ with derivative given by (E.8). By plugging (E.8) in (E.4), and developing further, we obtain

$$\begin{aligned} S_{\phi_1}(\nu) - S_{\phi_2}(\nu) &= \lambda(\nu, \phi_{\tau_0})^T \nabla_\phi \Psi(\tilde{\theta}(\nu), \phi_{\tau_0})[\phi_1 - \phi_2] \\ &= \lambda(\nu, \phi_0)^T \nabla_\phi \Psi(\theta_*(\nu), \phi_0)[\phi_1 - \phi_2] \\ &+ \Big(\lambda(\nu, \phi_{\tau_0})^T \nabla_\phi \Psi(\tilde{\theta}(\nu), \phi_{\tau_0}) - \lambda(\nu, \phi_0)^T \nabla_\phi \Psi(\theta_*(\nu), \phi_0)\Big)[\phi_1 - \phi_2] \end{aligned} \tag{E.5}$$

where $\theta_*(\cdot) : \mathbb{S}^d \to \Theta$ is a Borel measurable mapping satisfying $\theta_*(\nu) \in \Xi(\nu, \phi_0)$ and $\tilde{\theta}(\nu) \in \Xi(\nu, \phi_{\tau_0})$.

Step 3: Linearization. To carry the third step out we first analyze terms in (E.5). By Lemma F.5 and Lemma F.7 in Appendix F, the function $\nu \mapsto \lambda(\nu, \phi_0)$ is unique and continuous in $\nu \in \mathbb{S}^d$ and therefore it attains its supremum. Moreover, $\sup_{\nu\in\mathbb{S}^d} \nabla_\phi \Psi(\theta_*(\nu), \phi_0) \leq \sup_{\theta\in\Theta} \nabla_\phi \Psi(\theta, \phi_0)$ and the supremum is attained since, under Assumptions E.2 *(ii)* and E.4 *(i)*, $\theta \mapsto \nabla_\phi \Psi(\theta, \phi_0)$ is uniformly continuous on Θ. Thus, we can write:

$$\sup_{\nu\in\mathbb{S}^d} \left|(S_{\phi_1}(\nu) - S_{\phi_2}(\nu)) - \lambda(\nu, \phi_0)^T \nabla_\phi \Psi(\theta_*(\nu), \phi_0)[\phi_1 - \phi_2]\right| \tag{E.6}$$

$$= \sup_{\nu\in\mathbb{S}^d} \left|\Big(\lambda(\nu, \phi_{\tau_0})^T \nabla_\phi \Psi(\tilde{\theta}(\nu), \phi_{\tau_0}) - \lambda(\nu, \phi_0)^T \nabla_\phi \Psi(\theta_*(\nu), \phi_0)\Big)[\phi_1 - \phi_2]\right|$$

$$=: \quad f(\phi_1, \phi_2). \tag{E.7}$$

It remains to analyze the local behavior of $f(\phi_1, \phi_2)$, which is given in Lemma E.2 below. The lemma implies that $\frac{f(\phi_1,\phi_2)}{\|\phi_1-\phi_2\|}$ converges to 0 uniformly in $(\phi_1, \phi_2) \in B(\phi_0, r_n)$ as $r_n \to 0$. This establishes the claim of the theorem.

Lemma E.1. *For any $\phi_1, \phi_2 \in B(\phi_0, r_n)$ and $\tau \in [0,1]$ define $\phi_\tau = \tau\phi_1 + (1-\tau)\phi_2$ with $\phi_2 = \phi_\tau|_{\tau=0}$ and $\phi_1 = \phi_\tau|_{\tau=1}$. Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 hold with $\delta = r_n$. Then, there exists a N such that $\forall n \geq N$ and for all $\nu \in \mathbb{S}^d$, $\phi_1, \phi_2 \in B(\phi_0, r_n)$, the function $\tau \mapsto S_{\phi_\tau}(\nu)$ is differentiable at every $\tau \in (0,1)$ and, for $\tau_0 \in (0,1)$, we have*

$$\frac{\partial}{\partial \tau} S_{\phi_\tau}(\nu)\bigg|_{\tau=\tau_0} = \lambda(\nu, \phi_{\tau_0})^T \nabla_\phi \Psi(\tilde{\theta}, \phi_{\tau_0})[\phi_1 - \phi_2] \tag{E.8}$$

for any $\tilde{\theta} \in \Xi(\nu, \phi_{\tau_0})$ and where $\nabla_\phi \Psi(\tilde{\theta}(\nu), \phi_{\tau_0})$ denotes the $(k \times d_\phi)$-matrix of partial derivatives of $\Psi(\theta, \phi)$, with respect to ϕ, evaluated at $(\theta, \phi) = (\tilde{\theta}(\nu), \phi_{\tau_0})$.

Proof. Define $L(\theta, \lambda, \tau; \nu) = \nu^T\theta - \lambda(\nu, \phi_\tau)^T\Psi(\theta, \phi_\tau)$ for $\lambda(\nu, \phi) : \mathbb{S}^d \times B(\phi_0, r_n) \to \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$, and denote by $\frac{d}{d\tau+}$ and $\frac{d}{d\tau-}$ the right and left derivatives, respectively. Under Assumptions E.1 - E.3 and E.4 *(i)-(ii)* we can apply Corollary 5 in Milgrom and Segal (2002) which guarantees that the function $\tau \mapsto S_{\phi_\tau}(\nu)$ (for a fixed ν) is directionally differentiable in τ and its directional derivatives (in direction τ) are given by: $\frac{d}{d\tau+}S_{\phi_\tau}(\nu) = \max_{\theta\in\Xi(\nu,\phi_\tau)} \frac{d}{d\tau}L(\theta, \lambda, \tau; \nu)$ and $\frac{d}{d\tau-}S_{\phi_\tau}(\nu) = \min_{\theta\in\Xi(\nu,\phi_\tau)} \frac{d}{d\tau}L(\theta, \lambda, \tau; \nu)$ (where, in addition, we have used Lemma F.5 in Appendix F which demonstrates that the Lagrange multipliers are unique for n sufficiently large). Remark that we have extended Corollary 5 in Milgrom and Segal (2002) to the case with equality constraints by replacing each equality constraint $\Psi_i(\theta, \phi) = 0$, $i = k_1+1, \ldots, k$, by two constraints: $\Psi_i(\theta, \phi) \leq 0$ and $\Psi_i(\theta, \phi) \geq 0$, for $i = k_1+1, \ldots, k$, as suggested in Rockafellar (1970). Since under Assumption E.4 (i) the right and left derivatives at τ_0 are

$$\begin{aligned}
\left.\frac{dS_{\phi_\tau}(\nu)}{d\tau+}\right|_{\tau=\tau_0} &= \max_{\theta\in\Xi(\nu,\phi_{\tau_0})} \Big[-(\phi_1-\phi_2)^T \left.\left(\frac{\partial}{\partial\phi_\tau}\lambda(\nu,\phi_\tau)\right)^T\right|_{\tau=\tau_0} \Psi(\theta,\phi_\tau)|_{\tau=\tau_0} \\
&\qquad -\lambda(\nu,\phi_{\tau_0})^T\,\nabla_\phi\Psi(\theta,\phi_\tau)|_{\tau=\tau_0}(\phi_1-\phi_2)\Big] \\
\left.\frac{dS_{\phi_\tau}(\nu)}{d\tau-}\right|_{\tau=\tau_0} &= \min_{\theta\in\Xi(\nu,\phi_{\tau_0})} \Big[-(\phi_1-\phi_2)^T \left.\left(\frac{\partial}{\partial\phi_\tau}\lambda(\nu,\phi_\tau)\right)^T\right|_{\tau=\tau_0} \Psi(\theta,\phi_\tau)|_{\tau=\tau_0} \\
&\qquad -\lambda(\nu,\phi_{\tau_0})^T\,\nabla_\phi\Psi(\theta,\phi_\tau)|_{\tau=\tau_0}(\phi_1-\phi_2)\Big].
\end{aligned}$$

The first term on the right hand side of both these two equations is equal to zero because $\Psi(\theta, \phi_\tau)|_{\tau=\tau_0} = 0$ for $\theta \in \Xi(\nu, \phi_{\tau_0})$ since this is the first order condition of the optimization problem in $\Xi(\nu, \phi_{\tau_0})$ evaluated at the optimum value θ. Thus,

$$\left.\frac{dS_{\phi_\tau}(\nu)}{d\tau+}\right|_{\tau=\tau_0} = \max_{\theta\in\Xi(\nu,\phi_{\tau_0})} \Big[-\lambda(\nu,\phi_{\tau_0})^T\,\nabla_\phi\Psi(\theta,\phi_\tau)|_{\tau=\tau_0}(\phi_1-\phi_2)\Big] \tag{E.9}$$

$$\left.\frac{dS_{\phi_\tau}(\nu)}{d\tau-}\right|_{\tau=\tau_0} = \min_{\theta\in\Xi(\nu,\phi_{\tau_0})} \Big[-\lambda(\nu,\phi_{\tau_0})^T\,\nabla_\phi\Psi(\theta,\phi_\tau)|_{\tau=\tau_0}(\phi_1-\phi_2)\Big]. \tag{E.10}$$

By Lemma F.9, there exists a N such that for every $n \geq N$, every $\phi_1, \phi_2 \in B(\phi_0, r_n)$, every $1 \leq i \leq k_1$ such that $\lambda_i(\nu, \phi_{\tau_0}) \neq 0$, we must have $\theta_{1,j} = \theta_{2,j}$ for all $j \in \mathcal{S}_i$ and all $\theta_1, \theta_2 \in \Xi(\nu, \phi_{\tau_0})$. Therefore, by Assumption E.1 *(i)*, and since $\nabla_\phi\Psi_i(\theta, \phi_{\tau_0})$ does not depend on θ, $\forall i = k_1 + 1, \ldots, k$ and $A\theta$ does not depend on ϕ, we can conclude:

$$\begin{aligned}
\left.\frac{dS_{\phi_\tau}(\nu)}{d\tau+}\right|_{\tau=\tau_0} &= \max_{\theta\in\Xi(\nu,\phi_{\tau_0})} \Big[-\sum_{i;\lambda_i(\nu,\phi_{\tau_0})\neq 0} \lambda_i(\nu,\phi_{\tau_0})^T\,\nabla_\phi\Psi_{s,i}(\theta,\phi_\tau)|_{\tau=\tau_0}(\phi_1-\phi_2)\Big] \\
&= \min_{\theta\in\Xi(\nu,\phi_{\tau_0})} \Big[-\sum_{i;\lambda_i(\nu,\phi_{\tau_0})\neq 0} \lambda_i(\nu,\phi_{\tau_0})^T\,\nabla_\phi\Psi_{s,i}(\theta,\phi_\tau)|_{\tau=\tau_0}(\phi_1-\phi_2)\Big] \\
&= \left.\frac{dS_{\phi_\tau}(\nu)}{d\tau-}\right|_{\tau=\tau_0}.
\end{aligned}$$

Because the left and right derivatives at τ_0 exist and are equal at every $\tau_0 \in [0, 1]$, we conclude that $S_{\phi_\tau}(\nu)$ is differentiable at every $\tau_0 \in (0,1)$ and its derivative takes the form given in (E.8). $\square$

Lemma E.2. *Let $\theta_*(\nu)$ be a Borel measurable mapping satisfying $\theta_*(\nu) \in \Xi(\nu,\phi_0)$ for $\nu \in \mathbb{S}^d$. For every $\phi_1, \phi_2 \in B(\phi_0, r_n)$ and $\tau_0 \in [0,1]$ define $\phi_{\tau_0} = \tau_0\phi_1 + (1-\tau_0)\phi_2$ and:*

$$f(\phi_1,\phi_2) := \sup_{\nu\in\mathbb{S}^d} \left| \left(\lambda(\nu,\phi_{\tau_0})^T \nabla_\phi \Psi(\tilde{\theta}(\nu),\phi_{\tau_0}) - \lambda(\nu,\phi_0)^T \nabla_\phi \Psi(\theta_*(\nu),\phi_0)\right)[\phi_1-\phi_2]\right|$$

where $\tilde{\theta}(\nu) \in \Xi(\nu,\phi_{\tau_0})$. Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 hold with $\delta = r_n$. Then, for every $\epsilon > 0$ there exists a N such that for every $n \geq N$

$$\sup_{\phi_1,\phi_2\in B(\phi_0,r_n)} \frac{f(\phi_1,\phi_2)}{\|\phi_1-\phi_2\|} < C_{\phi_0}\epsilon.$$

where C_{ϕ_0} denotes a generic constant that depends on ϕ_0.

Proof. Because $\tilde{\theta}(\nu) \in \Xi(\nu,\phi_{\tau_0})$, $\tilde{\theta}(\nu)$ depends also on ϕ_1 and ϕ_2. By the Cauchy-Schwartz inequality:

$$\begin{aligned}
&\left| \left(\lambda(\nu,\phi_{\tau_0})^T \ \nabla_\phi \Psi(\tilde{\theta},\phi_{\tau_0}) - \lambda(\nu,\phi_0)^T \nabla_\phi \Psi(\theta_*(\nu),\phi_0)\right)[\phi_1-\phi_2]\right| \\
&\qquad\qquad \leq \ \|\nabla_\phi \Psi(\tilde{\theta}(\nu),\phi_{\tau_0})^T \lambda(\nu,\phi_{\tau_0}) - \nabla_\phi \Psi(\theta_*(\nu),\phi_0)^T \lambda(\nu,\phi_0)\| \ \|\phi_1-\phi_2\|
\end{aligned}$$

so that

$$\begin{aligned}
\frac{f(\phi_1,\phi_2)}{\|\phi_1-\phi_2\|} \leq \ & \sup_{\nu\in\mathbb{S}^d} \|\nabla_\phi \Psi(\tilde{\theta}(\nu),\phi_{\tau_0})^T \left(\lambda(\nu,\phi_{\tau_0}) - \lambda(\nu,\phi_0)\right)\| + \\
& \sup_{\nu\in\mathbb{S}^d} \| \left(\nabla_\phi \Psi(\tilde{\theta}(\nu),\phi_{\tau_0}) - \nabla_\phi \Psi(\theta_*(\nu),\phi_0)\right)^T \lambda(\nu,\phi_0)\| \ =: \mathcal{A}_1 + \mathcal{A}_2.
\end{aligned}$$

We start by analyzing term $\mathcal{A}_1$. Since $B(\phi_0,r_n)$ is convex then $\phi_1,\phi_2 \in B(\phi_0,r_n)$ implies $\phi_\tau \in B(\phi_0,r_n)$, $\forall \tau \in [0,1]$. Thus,

$$\begin{aligned}
\sup_{\phi_1,\phi_2\in B(\phi_0,r_n)} \mathcal{A}_1 = & \sup_{\phi_1,\phi_2\in B(\phi_0,r_n)} \sup_{\nu\in\mathbb{S}^d} \|\nabla_\phi \Psi(\tilde{\theta}(\nu),\phi_{\tau_0})^T \left(\lambda(\nu,\phi_{\tau_0}) - \lambda(\nu,\phi_0)\right)\| \\
\leq & \sup_{\phi_1,\phi_2\in B(\phi_0,r_n)} \sup_{\nu\in\mathbb{S}^d} \sup_{\tau\in[0,1]} \sup_{\theta\in\Xi(\nu,\phi_\tau)} \|\nabla_\phi \Psi(\theta,\phi_\tau)^T \left(\lambda(\nu,\phi_\tau) - \lambda(\nu,\phi_0)\right)\| \\
\leq & \sup_{\phi\in B(\phi_0,r_n)} \sup_{\nu\in\mathbb{S}^d} \sup_{\theta\in\Xi(\nu,\phi)} \|\nabla_\phi \Psi(\theta,\phi)\| \sup_{\phi\in B(\phi_0,r_n)} \sup_{\nu\in\mathbb{S}^d} \|\lambda(\nu,\phi) - \lambda(\nu,\phi_0)\| \\
& \leq \sup_{\phi\in B(\phi_0,r_n)} \sup_{\theta\in\Theta} \|\nabla_\phi \Psi(\theta,\phi)\| \sup_{\phi\in B(\phi_0,r_n)} \sup_{\nu\in\mathbb{S}^d} \|\lambda(\nu,\phi) - \lambda(\nu,\phi_0)\|.
\end{aligned}$$

By Assumptions E.4 *(i)*, $\nabla_\phi \Psi(\theta,\phi)$ exists and is continuous in $(\theta,\phi) \in \Theta \times B(\phi_0,r_n)$. Since Θ and $B(\phi_0,r_n)$ are compact, by Assumption E.2, it follows that $\|\nabla_\phi \Psi(\theta,\phi)\|$ is uniformly bounded on $\Theta \times B(\phi_0,r_n)$, that is, there exists a constant $0 < C_{\phi_0} < \infty$ such that

$$\sup_{\phi\in B(\phi_0,r_n)} \sup_{\theta\in\Theta} \|\nabla_\phi \Psi(\theta,\phi_\tau)\| < C_{\phi_0}.$$

By Lemma F.8 in Appendix F, for every $\epsilon > 0$, there exists a N such that for every $n \geq N$ we have

$$\sup_{\phi \in B(\phi_0, r_n)} \sup_{\nu \in \mathbb{S}^d} \|\lambda(\nu, \phi) - \lambda(\nu, \phi_0)\| < \epsilon.$$

Thus, we conclude that $\sup_{\phi_1, \phi_2 \in B(\phi_0, r_n)} \mathcal{A}_1 < C_{\phi_0} \epsilon$.

Next, let us consider term $\mathcal{A}_2$. By Lemma F.10 in Appendix F, for any $\epsilon > 0$ there exists a N such that for every $n \geq N$ the second inequality in the following display holds

$$\sup_{\phi_1, \phi_2 \in B(\phi_0, r_n)} \mathcal{A}_2 \leq$$

$$\sup_{\phi_1, \phi_2 \in B(\phi_0, r_n)} \sup_{\nu \in \mathbb{S}^d} \sup_{\theta \in \Xi(\nu, \phi_{\tau_0})} \left\| \Big(\nabla_\phi \Psi(\theta, \phi_{\tau_0}) - \nabla_\phi \Psi(\theta_*(\nu), \phi_0) \Big)^T \lambda(\nu, \phi_0) \right\| < C_{\phi_0} \epsilon.$$

By collecting the two upper bounds we get the result. □

F Auxiliary lemmas

Lemma F.1. *Let Assumptions E.1 (ii) and E.2 be satisfied with $\delta = r_n$. For every $\epsilon > 0$ there exists a N such that for every $n \geq N$ and $\phi \in B(\phi_0, r_n)$*

$$\sup_{\theta \in \Theta} \|\Psi(\theta, \phi) - \Psi(\theta, \phi_0)\| < \epsilon. \tag{F.1}$$

Proof. Under Assumption E.1 *(ii)*, the function $\phi \mapsto \Psi(\theta, \phi)$ is continuous on Φ, for every $\theta \in \Theta$, and uniformly continuous on $B(\phi_0, r_n)$, due to the compactness of $B(\phi_0, r_n)$. Therefore, for every $\theta \in \Theta$ and every $\epsilon > 0$ there exists a $\delta_\theta > 0$ such that $\forall \phi \in B(\phi_0, \delta_\theta)$: $\|\Psi(\theta, \phi) - \Psi(\theta, \phi_0)\| < \epsilon$. Now, denote $f_\phi(\theta) := \|\Psi(\theta, \phi) - \Psi(\theta, \phi_0)\|$ and

$$A_{\delta_\theta} := \left\{ \tilde{\theta} \in \Theta;\ f_\phi(\tilde{\theta}) < \epsilon,\ \forall \phi \in B(\phi_0, \delta_\theta) \right\}$$

for every $\theta \in \Theta$, $\epsilon > 0$ and $\delta_\theta > 0$. This means that $\forall \theta \in \Theta$ there is a δ_θ such that $\theta \in A_{\delta_\theta}$. Under Assumption E.1 *(ii)* $f_\phi(\theta)$ is continuous in θ for every $\phi \in B(\phi_0, \delta_\theta)$, hence A_{δ_θ} is an open set and $\bigcup_{\theta \in \Theta} A_{\delta_\theta}$ is an open cover of Θ: $\Theta \subset \bigcup_{\theta \in \Theta} A_{\delta_\theta}$. Due to compactness of Θ (by Assumption E.2 *(ii)*) there exists a finite set $\{\delta_1, \ldots, \delta_K\}$, $K < \infty$ such that $\{A_{\delta_i}\}_{i=1}^K$ is a subcover of Θ, that is, $\Theta \subset \bigcup_{i=1}^K A_{\delta_i}$. Let $\delta^* = \min\{\delta_1, \ldots, \delta_K\}$ so that $A_{\delta_i} \subseteq A_{\delta^*}$ for every $i = 1, \ldots, K$, and $\theta \in A_{\delta^*}$ for every $\theta \in \Theta$. Remark that this δ^* does not depend on θ. This then implies that for any $\phi \in B(\phi_0, r_n)$ and $r_n < \delta^*$: $\sup_{\theta \in \Theta} \|\Psi(\theta, \phi) - \Psi(\theta, \phi_0)\| < \epsilon$. □

Lemma F.2. *Under Assumption E.2 and Assumptions E.1 and E.4 (ii) with $\delta = r_n$, there exists a N such that for every $n \geq N$ the correspondence $\phi \mapsto \Theta(\phi)$ is well defined, convex-valued and continuous at all $\phi \in B(\phi_0, r_n)$, that is, it is upper and lower hemicontinuous.*

Proof. First, under Assumptions E.1 and E.2 *(ii)*, $\forall \phi \in B(\phi_0, r_n)$, the set $\Theta(\phi)$ is a closed convex set, with empty interior if $k_2 > 0$, since it is the intersection of a closed convex set with closed hyperplanes, see (Rockafellar, 1970, Corollary 2.1.1) and (Aliprantis and Border, 2006, Lemma 5.55).

Then, we have to show that the correspondence $\phi \mapsto \Theta(\phi)$ is continuous (for a definition of continuity of a correspondence see for instance Definition 17.2 in Aliprantis and Border (2006)). First, we show that $\phi \mapsto \Theta(\phi)$ is lower hemicontinuous at any $\phi \in B(\phi_0, r_n)$. We show this by showing Theorem 17.19 (ii) in Aliprantis and Border (2006), that is, we have to show that for any $\theta^* \in \Theta(\phi_*)$ and net $\{\phi_j\}_{j\in\mathfrak{U}_1}$ with $\phi_j \to \phi_* \in B(\phi_0, r_n)$, there exists a subnet $\{\phi_{j_\beta}\}_{\beta\in\mathfrak{U}_2}$ and a net $\{\theta_\beta\}_{\beta\in\mathfrak{U}_2}$ such that $\theta_\beta \in \Theta(\phi_{j_\beta})$ for every $\beta \in \mathfrak{U}_2$, and $\theta_\beta \to \theta^*$. Define $\Theta(\phi_*)_{in} := \{\theta \in \Theta;\ \Psi_i(\theta, \phi_*) \leq 0, i = 1, \ldots, k\}$ so that $\Theta(\phi_*) \subset \Theta(\phi_*)_{in}$ and consider $\theta^* \in \Theta(\phi_*)$. Since $\Theta(\phi_*)_{in}$ is convex with nonempty interior, there exists a net $\theta_j \in \Theta(\phi_*)_{in}$ such that $\theta_j \to \theta^*$. Let $\{\phi_j\}$ be a convergent net inside $B(\phi_*, r_n)$ such that $\phi_j \to \phi_* \in B(\phi_*, r_n)$ and, by compactness of $B(\phi_*, r_n)$, $\{\phi_j\}$ admits a convergent subnet. Denote by $\{\phi_{j_\beta}\}_{\beta\in\mathfrak{U}_2}$ such a subnet where $\mathfrak{U}_2 := \{\beta; \phi_{j_\beta} \in B(\phi_*, r_n),\ \theta_\beta \in \Theta(\phi_{j_\beta})$ and $\{\theta_\beta\}$ is a convergent subnet of $\{\theta_j\}\}$. Remark that $\{\theta_j\}$ admits a subnet $\{\theta_\beta\}$ since $\Theta(\phi_*)_{in}$ is compact (because it is a closed subset of a compact set). Therefore, $\theta_\beta \to \theta^*$ and $\theta_\beta \in \Theta(\phi_{j_\beta})$ by construction. This establishes lower hemicontinuity.

Next, we show that the correspondence $\phi \mapsto \Theta(\phi)$ is upper hemicontinuous at any $\phi \in B(\phi_0, r_n)$. By Theorem 17.16 in Aliprantis and Border (2006) it is sufficient to show that for every net $\{\phi_j, \theta_j\}$ such that $\theta_j \in \Theta(\phi_j)$ for each j (*i.e.* (ϕ_j, θ_j) is in the graph of $\Theta(\cdot)$), if $\phi_j \to \phi$ then $\theta_j \to \theta^* \in \Theta(\phi)$.

To show this, let $\{\phi_j\}$ be a convergent net in $B(\phi_0, r_n)$ such that $\{\phi_j\} \to \phi \in B(\phi_0, r_n)$. Consider a net $\{\theta_j\} \in \Theta(\phi_j)$ so that $\{(\phi_j, \theta_j)\}$ is in the graph of $\Theta(\cdot)$ and $\theta_j \to \theta^*$. Under Assumption E.1 *(ii)* the function $\Psi(\cdot, \cdot)$ is continuous in θ and ϕ and so:

$$0 \geq \Psi_i(\theta_j, \phi_j) \to \Psi_i(\theta^*, \phi), \qquad i = 1, \ldots k_1 \tag{F.2}$$

$$0 = a_i^T \theta_j + b_i(\phi_j) \to a_i^T \theta^* + b_i(\phi), \qquad i = k_1 + 1, \ldots k. \tag{F.3}$$

This shows that $\theta^* \in \Theta(\phi)$ and upper hemicontinuity is established. Since the correspondence $\phi \mapsto \Theta(\phi)$ is lower- and upper- hemicontinuous, then it is continuous and the result of the lemma follows. □

Lemma F.3. *Let Assumption E.2 holds and Assumptions E.1 and E.4 (ii) hold with $\delta = r_n$. Then, there exists a N such that for every $n \geq N$ the correspondence*

$$(\nu, \phi) \mapsto \Xi(\nu, \phi) = \arg\max_{\theta\in\Theta}\{\nu^T\theta;\ \Psi(\theta, \phi) \leq 0\}$$

has non-empty compact values and it is upper hemicontinuous on $\mathbb{S}^d \times B(\phi_0, r_n)$.

Proof. Under Assumptions E.1 and E.2 *(ii)*, for every $\phi \in \Phi$, the set $\Theta(\phi)$ is a closed convex set because it is a finite intersection of a closed convex set with closed hyperplanes. Because $\forall\phi \in \Phi$, $\Theta(\phi) \subseteq \Theta$ and Θ is compact (under Assumption E.2 *(ii)*) then the set $\Theta(\phi)$ is also compact for every $\phi \in \Phi$. Moreover, $\Theta(\phi)$ is non-empty for every $\phi \in B(\phi_0, r_n)$ under Assumption E.4 *(ii)*. Hence, $\Xi(\nu, \phi)$ is well defined as the maximum is attained for every $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$.

Now, under assumption E.4 *(ii)* and by the continuity result of Lemma F.2, we can apply the "Berge Maximum Theorem", see *e.g.* Theorem 17.31 in Aliprantis and Border (2006), which guarantees that the correspondence

$$(\nu, \phi) \mapsto \Xi(\nu, \phi) = \arg \max_{\theta \in \Theta(\phi)} \nu^T \theta$$

has nonempty compact values and it is upper hemicontinuous, for every $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$. $\square$

Lemma F.4. *Let Assumptions E.2 and E.3 and Assumptions E.1 and E.4 (ii)-(iii) be satisfied with $\delta = r_n$. Then, there exists an N such that for every $n \geq N$, $\phi \in B(\phi_0, r_n)$ and $\theta \in \Theta(\phi)$ the vectors*

$$\{\nabla_\theta \Psi_i(\theta, \phi)\}_{i \in Act(\theta, \phi)}$$

are linearly independent.

Proof. By Lemma F.2, the correspondence $\phi \mapsto \Theta(\phi)$ is upper hemicontinuous and $\Theta(\phi)$ is compact $\forall \phi \in B(\phi_0, r_n)$, then for every net $\{\phi_\alpha, \theta_\alpha\}$ in the graph of $\Theta(\cdot)$ (*i.e.* such that $\theta_\alpha \in \Theta(\phi_\alpha)$, $\forall \alpha$ such that $\phi_\alpha \in B(\phi_0, r_n)$) we have that if $\phi_\alpha \to \phi_0$ then $\theta_\alpha \to \theta^*$ for some θ^* in $\Theta(\phi_0)$ (see *e.g.* Theorem 17.16 in Aliprantis and Border (2006)). Moreover, by continuity of $(\theta, \phi) \mapsto \Psi(\theta, \phi)$ (under Assumption E.1 *(ii)*), we have that $\forall i \in Act^c(\theta^*, \phi_0)$

$$\Psi_i(\theta_\alpha, \phi_\alpha) \to \Psi_i(\theta^*, \phi_0) < 0.$$

Therefore, there exists N such that for every $n \geq N$ and every $\phi \in B(\phi_0, r_n)$, $\theta \in \Theta(\phi)$, the constraints that are inactive under (θ^*, ϕ_0) are also inactive under (θ, ϕ), where θ^* is the limit of $\theta \in \Theta(\phi)$ as $\phi \to \phi_0$, that is, $Act(\theta, \phi) \subseteq Act(\theta^*, \phi_0)$, $\forall (\theta, \phi) \in \Theta(\phi) \times B(\phi_0, r_n)$.

By Assumption E.4 *(iii)* the vectors $\nabla_\theta \Psi_i(\theta, \phi)$ are continuous in $(\theta, \phi) \in \Theta(\phi) \times B(\phi_0, r_n)$ and hence

$$\nabla_\theta \Psi_i(\theta_\alpha, \phi_\alpha) \to \nabla_\theta \Psi_i(\theta^*, \phi_0), \quad \forall i \in Act(\theta^*, \phi_0). \tag{F.4}$$

Denote by $\nabla_\theta \Psi^A(\theta, \phi)$ the $(d \times d_A(\theta, \phi))$-matrix obtained by stacking columnwise the vectors $\{\nabla_\theta \Psi_i(\theta, \phi)\}_{i \in Act(\theta, \phi)}$ and by $\{\rho_i(\theta, \phi)\}_{i \in Act(\theta, \phi)}$ its singular values for some $\theta \in \Theta(\phi)$. By Assumption E.3 the matrix $\nabla_\theta \Psi^A(\theta, \phi_0)$ is full-column rank $\forall \theta \in \Theta(\phi_0)$ and then there exists a $\epsilon > 0$ such that $\inf_{i \in Act(\theta, \phi_0)} \rho_i(\theta, \phi_0) > \epsilon$, $\forall \theta \in \Theta(\phi_0)$. Continuity (in (θ, ϕ)) of the singular values (which follows from continuity of $\nabla_\theta \Psi^A(\theta, \phi)$, see *e.g.* Theorem II.5.1 in Kato (1995)) and (F.4) imply

$$\rho_i(\theta_\alpha, \phi_\alpha) \to \rho_i(\theta^*, \phi_0), \qquad \forall i \in Act(\theta^*, \phi_0)$$

which is larger than $\epsilon > 0$ because $\theta^* \in \Theta(\phi_0)$. We conclude that there exists a N such that for every $n \geq N$, $\phi \in B(\phi_0, r_n)$ and $\theta \in \Theta(\phi)$ the eigenvalues $\{\rho_i(\theta, \phi)\}_{i \in Act(\theta, \phi_0)}$ are strictly positive which implies that $\nabla_\theta \Psi^A(\theta, \phi)$ is non-singular. Henceforth, $\{\nabla_\theta \Psi_i(\theta, \phi)\}_{i \in Act(\theta, \phi)}$ are linearly independent and this prove the lemma. $\square$

Lemma F.5. *Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 (ii)-(iii) hold with $\delta = r_n$. Then, there exists a N such that for every $n \geq N$, $\phi \in B(\phi_0, r_n)$ and $\nu \in \mathbb{S}^d$ there exists a unique $\lambda(\nu, \phi) \in \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$ satisfying*

$$\sup_{\theta \in \Theta(\phi)} \nu^T \theta = \sup_{\theta \in \Theta} \left\{ \nu^T \theta - \lambda(\nu, \phi)^T \Psi(\theta, \phi) \right\} \tag{F.5}$$

provided $\sup_{\theta \in \Theta(\phi)} \nu^T \theta < \infty$.

Proof. Under Assumptions E.1, E.2 *(ii)*, and E.3, Lemma F.4 and Theorem 28.2 in Rockafellar (1970) guarantee that there exists an N such that for every $n \geq N$ and $\phi \in B(\phi_0, r_n)$, KT multipliers exists for the problem $\sup_{\theta \in \Theta(\phi)} \nu^T \theta$ provided $\sup_{\theta \in \Theta(\phi)} \nu^T \theta < \infty$. Therefore, the equality in (F.5) holds for some $\lambda(\nu, \phi) \in \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$. In the case where we have only inequality constraints, or only linear (in θ) constraints, existence of KT multipliers is guaranteed by Corollary 28.2.1 and Corollary 28.2.2 in Rockafellar (1970), respectively, instead of Theorem 28.2 in Rockafellar (1970).

Next, we have to show that there exists a N such that for all $n \geq N$ and all $\phi \in B(\phi_0, r_n)$, $\lambda(\nu, \phi)$ is unique $\forall \nu \in \mathbb{S}^d$. By Lemma F.3 and Assumption E.4 (ii), there exists a N such that for every $n \geq N$, $\emptyset \neq \Xi(\nu, \phi) \subseteq \Theta(\phi) \subset int\Theta$, $\forall \nu \in \mathbb{S}^d$, where $int\Theta$ denotes the interior of Θ. Consider a N such that the result of Lemma F.4 holds and let $n \geq N$. Then, by strong duality, any $\tilde{\theta} \in \Xi(\nu, \phi)$ with $\phi \in B(\phi_0, r_n)$ satisfies the KT-conditions and then

$$\nu - \nabla_\theta \Psi(\tilde{\theta}, \phi) \lambda(\nu, \phi) = 0, \qquad \forall \nu \in \mathbb{S}^d, \ \forall \phi \in B(\phi_0, r_n) \tag{F.6}$$

under Assumption E.4 (iii). Suppose that for $(\phi, \nu) \in B(\phi_0, r_n) \times \mathbb{S}^d$ there exist two different vectors $\lambda_1(\nu, \phi)$ and $\lambda_2(\nu, \phi)$ that satisfy equation (F.5). By the complementary slackness condition, the Lagrange multipliers of the non-binding constraints are equal to 0. Therefore, equation (F.6) implies

$$\nu - \sum_{i=1}^{d_A(\tilde{\theta}, \phi)} \lambda_{1,i}(\nu, \phi) \nabla_\theta \Psi_i(\tilde{\theta}, \phi) = \nu - \sum_{i=1}^{d_A(\tilde{\theta}, \phi)} \lambda_{2,i}(\nu, \phi) \nabla_\theta \Psi_i(\tilde{\theta}, \phi) = 0 \tag{F.7}$$

which, after simplifications, gives

$$\sum_{i=1}^{d_A(\tilde{\theta}, \phi)} \left(\lambda_{1,i}(\nu, \phi) - \lambda_{2,i}(\nu, \phi)\right) \nabla_\theta \Psi_i(\tilde{\theta}, \phi) = 0. \tag{F.8}$$

By Lemma F.4, there exists a N such that for all $n \geq N$ and all $\phi \in B(\phi_0, r_n)$ the vectors $\{\nabla_\theta \Psi_i(\tilde{\theta}, \phi)\}_{i \in Act(\tilde{\theta}, \phi)}$ are linearly independent. This and (F.8) contradict $\lambda_1(\nu, \phi) \neq \lambda_2(\nu, \phi)$ and show uniqueness of the KT multipliers. $\square$

Lemma F.6. *Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 (ii)-(iii) be satisfied with* $\delta = r_n$. *Then, there exists an* N *such that for every* $n \geq N$, $\|\lambda(\nu, \phi)\|$ *is uniformly bounded in* $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$.

Proof. For every $\nu \in \mathbb{S}^d$, $\phi \in B(\phi_0, r_n)$ and $\theta \in \Theta(\phi)$, denote by $\lambda^A(\nu, \phi, \theta)$ the $d_A(\theta, \phi)$-vector with components $\{\lambda^i(\nu, \phi)\}_{i \in Act(\theta, \phi)}$ and by $\nabla_\theta \Psi^A(\theta, \phi)$ the $(d \times d_A(\theta, \phi))$ matrix obtained by stacking columnwise the vectors $\{\nabla_\theta \Psi_i(\theta, \phi)\}_{i \in Act(\phi)}$. By Lemma F.4, there exists a N such that for every $n \geq N$, $\phi \in B(\phi_0, r_n)$ and $\theta \in \Theta(\phi)$, the matrix $[\nabla_\theta \Psi^A(\theta, \phi)]^T \nabla_\theta \Psi^A(\theta, \phi)$ is invertible and let $n \geq N$ in the rest of the proof. It follows from (F.7), which is valid under Assumption E.4 *(iii)*, that for n sufficiently large we can write

$$\lambda^A(\nu, \phi, \theta) = \left([\nabla_\theta \Psi^A(\theta, \phi)]^T \nabla_\theta \Psi^A(\theta, \phi)\right)^{-1} [\nabla_\theta \Psi^A(\theta, \phi)]^T \nu,$$

for $\theta \in \Xi(\nu, \phi)$ and every $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$. Since $\lambda_i(\nu, \phi) = 0$ for every $i \notin Act(\theta, \phi)$ then $\|\lambda(\nu, \phi)\| = \|\lambda^A(\nu, \phi, \theta)\|$ and for any $\theta(\nu, \phi) \in \Xi(\nu, \phi)$

$$\begin{aligned}
\sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \|\lambda(\nu, \phi)\| &= \sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \|\lambda^A(\nu, \phi, \theta)\| \\
&\leq \sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \left|\left|[\nabla_\theta \Psi^A(\theta(\nu, \phi), \phi)]^T \nabla_\theta \Psi^A(\theta(\nu, \phi), \phi)\right|\right|^{-1} \|\nabla_\theta \Psi^A(\theta(\nu, \phi), \phi)\| \|\nu\| \\
&\leq \sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \left(\underline{\rho}(\theta(\nu, \phi), \phi)\right)^{-2} d_A(\theta, \phi)^{-1/2} \sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \|\nabla_\theta \Psi^A(\theta(\nu, \phi), \phi)\|
\end{aligned}$$

where $\underline{\rho}(\theta, \phi)$ denotes the smallest singular values of $\nabla_\theta \Psi^A(\theta, \phi)$. By Lemma F.4, $\underline{\rho}(\theta, \phi)$ is strictly positive. Under Assumption E.4 *(iii)*, $(\theta, \phi) \mapsto \nabla_\theta \Psi_i(\theta, \phi)$ is continuous for n sufficiently large, which implies continuity of the matrix $\nabla_\theta \Psi^A(\theta, \phi)$. Furthermore, because $\Xi(\nu, \phi) \subset \Theta(\phi)$ and $\mathbb{S}^d \times B(\phi_0, r_n) \times \Xi(\nu, \phi)$ is compact (compactness of $\Xi(\nu, \phi)$ follows from Lemma F.3), it follows from the extreme value theorem that $\nabla_\theta \Psi^A(\theta, \phi)$ attains its maximum value on $\mathbb{S}^d \times B(\phi_0, r_n) \times \Xi(\nu, \phi)$ so that there exists a constant $0 < C_1 < \infty$ such that

$$\sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \|\nabla_\theta \Psi^A(\theta(\nu, \phi), \phi)\| \leq \sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \sup_{\theta \in \Xi(\nu, \phi)} \|\nabla_\theta \Psi^A(\theta, \phi)\| < C_1.$$

Continuity of the singular values (which follows from the continuity of $\nabla_\theta \Psi^A(\theta, \phi)$, see *e.g.* Theorem II.5.1 in Kato (1995)) and compactness of $\mathbb{S}^d \times B(\phi_0, r_n) \times \Xi(\nu, \phi)$ implies that there exists a constant $0 < C_2 < \infty$ such that

$$\sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \left(\underline{\rho}(\tilde{\theta}(\nu, \phi), \phi)\right)^{-2} \leq \sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \sup_{\theta \in \Xi(\nu, \phi)} \left(\underline{\rho}(\theta, \phi)\right)^{-2} < C_2$$

since the maximum is attained. This proves the statement of the lemma that $\lambda(\nu, \phi)$ is uniformly bounded in $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$. □

Lemma F.7. *Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 (ii)-(iii) hold with $\delta = r_n$. Then, there exists an N such that for every $n \geq N$ the vector $\lambda(\nu, \phi)$ is continuous in $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$.*

Proof. The proof follows the line of the proof of Lemma A.12 in Kaido and Santos (2014) and uses the results of Lemmas F.5, F.6 and Assumption E.1 *(i)*. □

Lemma F.8. *Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 (ii)-(iii) hold with $\delta = r_n$. Then, for any $\epsilon > 0$ there exists a N such that for every $n \geq N$ we have*

$$\sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \|\lambda(\nu, \phi) - \lambda(\nu, \phi_0)\| < \epsilon.$$

Proof. By Lemmas F.5 and F.7 there exists a N such that for every $n \geq N$ the function $\lambda : \mathbb{S}^d \times B(\phi_0, r_n) \to \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$ is singleton valued and continuous in $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$. Therefore, by compactness of $B(\phi_0, r_n)$ the function $\phi \mapsto \lambda(\nu, \phi)$ is uniformly continuous on $B(\phi_0, r_n)$ for every $\nu \in \mathbb{S}^d$. This means that for every ν and any $\epsilon > 0$ there exists a natural number N_ν that depends on ν such that for all $n \geq N_\nu$,

$$\sup_{\phi \in B(\phi_0, r_n)} \|\lambda(\nu, \phi) - \lambda(\nu, \phi_0)\| < \epsilon. \tag{F.9}$$

Define $f_n(\nu) := \sup_{\phi \in B(\phi_0, r_n)} \|\lambda(\nu, \phi) - \lambda(\nu, \phi_0)\|$ and, for a given ϵ,

$$A_{N_\nu} := \left\{\tilde{\nu} \in \mathbb{S}^d;\ f_n(\tilde{\nu}) < \epsilon,\ \forall n \geq N_\nu\right\}$$

for every $\nu \in \mathbb{S}^d$. This means that for any $\nu \in \mathbb{S}^d$ there is a N_ν so that $\nu \in A_{N_\nu}$. Since $f_n(\nu)$ is continuous in ν, A_{N_ν} is an open set and $\bigcup_{\nu \in \mathbb{S}^d} A_{N_\nu}$ is an open cover of $\mathbb{S}^d$: $\mathbb{S}^d \subset \bigcup_{\nu \in \mathbb{S}^d} A_{N_\nu}$. Due to compactness of $\mathbb{S}^d$ there exists a finite set $\{N_1, \ldots, N_K\}$ with $K < \infty$ such that $\{A_{N_i}\}_{i=1}^K$ is a subcover of $\mathbb{S}^d$: $\mathbb{S}^d \subset \bigcup_{i=1}^K A_{N_i}$. Let $N^* = \max\{N_1, \ldots, N_K\}$ so that $A_{N_i} \subseteq A_{N^*}$ for every $i = 1, \ldots, K$ and, for any $\nu \in \mathbb{S}^d$ we have $\nu \in A_{N^*}$. Remark that this N^* does not depend on ν. This then implies that for any $n > N^*$

$$\sup_{\nu \in \mathbb{S}^d} \sup_{\phi \in B(\phi_0, r_n)} \|\lambda(\nu, \phi) - \lambda(\nu, \phi_0)\| < \epsilon.$$

□

Lemma F.9. *Let Assumption E.2 holds and Assumptions E.1 and E.4 (ii) hold with $\delta = r_n$. Then there exists an N such that for every $n \geq N$, every $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$ and all $1 \leq i \leq k_1$, one of the following must hold: (i) $\lambda_i(\nu, \phi) = 0$, or (ii) $\theta_{1,j} = \theta_{2,j}$ for all $j \in \mathcal{S}_i$ and all $\theta_1, \theta_2 \in \Xi(\nu, \phi)$.*

Proof. This proof is a slight modification of the proof of Lemma A.10 in Kaido and Santos (2014). By Lemma F.3, $\Xi(\nu, \phi) \neq \emptyset$ is compact for every $(\nu, \phi) \in \mathbb{S}^d \times B(\phi_0, r_n)$ for n sufficiently large, then $\Xi(\nu, \phi)$ is closed and bounded (by the Heine-Borel Theorem) and convex (since it is a closed subset of the convex set $\Theta(\phi)$). We aim at show that condition (i) must hold whenever (ii) fails. To establish this, suppose there exists a $1 \leq i \leq k_1$ such that $\theta_{1,j} \neq \theta_{2,j}$ for some $\theta_1, \theta_2 \in \Xi(\nu, \phi)$ and $j \in \mathcal{S}_i$. Moreover, let $\kappa \in [0, 1]$ and $\tilde{\theta} := \kappa\tilde{\theta}_1 + (1 - \kappa)\tilde{\theta}_2$. It follows that $\tilde{\theta} \in \Xi(\nu, \phi)$ and, $\forall \phi \in B(\phi_0, r_n)$,

$$\Psi_{s,i}(\tilde{\theta}, \phi) < \kappa \Psi_{s,i}(\theta_1, \phi) + (1 - \kappa)\Psi_{s,i}(\theta_2, \phi) \leq 0, \tag{F.10}$$

where the last inequality is due to the fact that $\theta_1, \theta_2 \in \Xi(\nu, \phi) \subset \Theta(\phi)$. Therefore, the strict inequality in (F.10) and the complementary slackness condition imply that $\lambda_i(\nu, \phi) = 0$. This establishes the lemma. □

Lemma F.10. *Let $\tau_0 \in (0, 1)$ and $\theta_*(\nu)$ be a Borel measurable mapping satisfying $\theta_*(\nu) \in \Xi(\nu, \phi_0)$ for $\nu \in \mathbb{S}^d$. For every $\phi_1, \phi_2 \in B(\phi_0, r_n)$ define $\phi_{\tau_0} = \tau_0\phi_1 + (1 - \tau_0)\phi_2$. Let Assumptions E.2 and E.3 hold and Assumptions E.1 and E.4 hold with $\delta = r_n$. Then, for any $\varepsilon > 0$ there exists a N such that $\forall n \geq N$*

$$\sup_{\phi_1, \phi_2 \in B(\phi_0, r_n)} \sup_{\nu \in \mathbb{S}^d} \sup_{\theta \in \Xi(\nu, \phi_{\tau_0})} \| \left[\nabla_\phi \Psi(\theta, \phi_{\tau_0}) - \nabla_\phi \Psi(\theta_*(\nu), \phi_0)\right]^T \lambda(\nu, \phi_0)\| < C_{\phi_0}\varepsilon$$

for a generic constant C_{ϕ_0} that depends on ϕ_0.

Proof. First, remark that

$$\|\left[\nabla_\phi \Psi(\theta, \phi_{\tau_0}) - \nabla_\phi \Psi(\theta_*(\nu), \phi_0)\right]^T \lambda(\nu, \phi_0)\| \le \|\left[\nabla_\phi \Psi(\theta, \phi_{\tau_0}) - \nabla_\phi \Psi(\theta, \phi_0)\right]^T \lambda(\nu, \phi_0)\| + \|\left[\nabla_\phi \Psi(\theta, \phi_0) - \nabla_\phi \Psi(\theta_*(\nu), \phi_0)\right]^T \lambda(\nu, \phi_0)\| =: \mathcal{A}_1 + \mathcal{A}_2 \tag{F.11}$$

and the function $\nu \mapsto \lambda(\nu, \phi_0)$ is unique and continuous in $\nu \in \mathbb{S}^d$ by Lemmas F.5 and F.7. Therefore, by compactness of $\mathbb{S}^d$ there exists a constant C_{ϕ_0} that depends on ϕ_0 such that $\sup_{\nu \in \mathbb{S}^d} \|\lambda(\nu, \phi_0)\| < C_{\phi_0}$.

Next, we analyze term $\mathcal{A}_1$:

$$\sup_{\phi_1, \phi_2 \in B(\phi_0, r_n)} \sup_{\nu \in \mathbb{S}^d} \sup_{\theta \in \Xi(\nu, \phi_{\tau_0})} \|\left[\nabla_\phi \Psi(\theta, \phi_{\tau_0}) - \nabla_\phi \Psi(\theta, \phi_0)\right]^T \lambda(\nu, \phi_0)\| \le \sup_{\phi \in B(\phi_0, r_n)} \sup_{\theta \in \Theta} \|\nabla_\phi \Psi(\theta, \phi) - \nabla_\phi \Psi(\theta, \phi_0)\| \sup_{\nu \in \mathbb{S}^d} \|\lambda(\nu, \phi_0)\|$$

since by convexity of $B(\phi_0, r_n)$ we have that $\phi_1, \phi_2 \in B(\phi_0, r_n)$ implies $\phi_{\tau_0} \in B(\phi_0, r_n)$, $\forall \nu \in \mathbb{S}^d$. By compactness of $B(\phi_0, r_n)$ and under Assumption E.4 *(i)*, the function $\phi \mapsto \nabla_\phi \Psi(\theta, \phi)$ is uniformly continuous on $B(\phi_0, r_n)$ for n sufficiently large and every $\theta \in \Theta$. Hence, for every θ and any $\varepsilon > 0$ there exists a natural number N_θ that depends on θ such that for all $n \ge N_\theta$ we have

$$f_n(\theta) := \sup_{\phi \in B(\phi_0, r_n)} \|\nabla_\phi \Psi(\theta, \phi) - \nabla_\phi \Psi(\theta, \phi_0)\| < \varepsilon.$$

Define $A_{N_\theta} \equiv \left\{\tilde{\theta} \in \Theta;\ f_n(\tilde{\theta}) < \varepsilon,\ \forall n > N_\theta\right\}$ for every $\theta \in \Theta$. Since $f_n(\theta)$ is continuous in θ, hence A_{N_θ} is an open set and $\bigcup_{\theta \in \Theta} A_{N_\theta}$ is an open cover of Θ. Due to compactness of Θ there exists a finite set $\{N_1, \ldots, N_K\}$ with $K < \infty$ such that $\{A_{N_i}\}_{i=1}^K$ is a subcover of Θ: $\Theta \subset \bigcup_{i=1}^K A_{N_i}$. Let $N^* = \max\{N_1, \ldots, N_K\}$ so that $A_{N_i} \subseteq A_{N^*}$ for every $i = 1, \ldots, K$ and for any $\theta \in \Theta$ we have $\theta \in A_{N^*}$. Remark that this N^* does not depend on θ so, for any $n > N^*$,

$$\sup_{\theta \in \Theta} \sup_{\phi \in B(\phi_0, r_n)} \|\nabla_\phi \Psi(\theta, \phi) - \nabla_\phi \Psi(\theta, \phi_0)\| < \varepsilon \tag{F.12}$$

and this establishes uniform convergence to zero of $\mathcal{A}_1$.

Finally, we analyze term $\mathcal{A}_2$ of (F.11). For every $\nu \in \mathbb{S}^d$, define

$$\mathcal{I}(\nu) := \bigcup_{\{i : \lambda_{I,i}(\nu, \phi_0) \neq 0\}} \mathcal{S}_i$$

where $\lambda_I(\nu, \phi_0)$ denotes the k_1-subvector of $\lambda(\nu, \phi_0)$ relative to the first k_1 moment functions $\Psi_I(\theta, \phi_0) := \{\Psi_i(\theta, \phi_0)\}_{i=1}^k$. Consider the map $\Pi_\nu : \Theta \to \mathbb{R}^d$ constructed in (Kaido and Santos, 2014, proof of Lemma B.7) which satisfies: $(\Pi_\nu \theta)_j = \theta_{*,j}(\nu)$ if $j \notin \mathcal{I}(\nu)$ and $(\Pi_\nu \theta)_j = \theta_j$ if $j \in \mathcal{I}(\nu)$. Lemma B.7 in Kaido and Santos (2014) establishes that for every $\nu \in \mathbb{S}^d$ and $\theta \in \Theta$,

$$\|\theta_*(\nu) - \Pi_\nu \theta\| \le \inf_{\tilde{\theta} \in \Xi(\nu, \phi_0)} \sqrt{d} \|\tilde{\theta} - \theta\| \tag{F.13}$$

and because $\Xi(\nu, \phi_{\tau_0})$ is well defined for all $(\nu, \tau_0) \in \mathbb{S}^d \times [0,1]$ by Lemma F.3, we have

$$\sup_{\phi_1,\phi_2 \in B(\phi_0,r_n)} \sup_{\nu \in \mathbb{S}^d} \sup_{\theta \in \Xi(\nu,\phi_{\tau_0})} \|\Pi_\nu \theta - \theta_*(\nu)\| \leq \sup_{\phi_1,\phi_2 \in B(\phi_0,r_n)} \sup_{\nu \in \mathbb{S}^d} \sup_{\theta \in \Xi(\nu,\phi_{\tau_0})} \inf_{\tilde{\theta} \in \Xi(\nu,\phi_0)} \sqrt{d}\|\theta - \tilde{\theta}\|. \tag{F.14}$$

Because $\nabla_\phi \Psi_i(\theta, \phi) = \nabla_\phi b_i(\phi)$, $\forall i = k_1 + 1, \ldots, k$, and does not depend on θ, term $\mathcal{A}_2$ simplifies as

$$\begin{aligned}
\mathcal{A}_2 &= \| \left[\nabla_\phi \Psi_s(\theta, \phi_0) - \nabla_\phi \Psi_s(\theta_*(\nu), \phi_0)\right]^T \lambda_I(\nu, \phi_0)\| \\
&= \left\| \sum_{\{i;\lambda_{I,i}(\nu,\phi_0)\neq 0\}} \left[\nabla_\phi \Psi_{s,i}(\theta, \phi_0) - \nabla_\phi \Psi_{s,i}(\theta_*(\nu), \phi_0)\right] \lambda_{I,i}(\nu, \phi_0) \right\| \\
&\leq \left\| \sum_{\{i;\lambda_{I,i}(\nu,\phi_0)\neq 0\}} \left[\nabla_\phi \Psi_{s,i}(\Pi_\nu\theta, \phi_0) - \nabla_\phi \Psi_{s,i}(\theta_*(\nu), \phi_0)\right] \lambda_{I,i}(\nu, \phi_0) \right\|.
\end{aligned} \tag{F.15}$$

The inequality in the third line is due to the triangular inequality, the definition of the map Π_ν and Assumption E.1 *(i)*. By upper hemicontinuity of the correspondence $\phi \mapsto \Xi(\nu, \phi)$ at the point ϕ_0 (as established in Lemma F.3), for every $\epsilon > 0$ and every $\nu \in \mathbb{S}^d$, there exists a N_ν such that for every $n \geq N_\nu$ and every $\phi \in B(\phi_0, r_n)$, $\Xi(\nu, \phi) \subseteq \Xi(\nu, \phi_0)^\epsilon$ where $\Xi(\nu, \phi_0)^\epsilon := \{\theta \in \Theta; \inf_{\tilde{\theta} \in \Xi(\nu,\phi_0)} \|\theta - \tilde{\theta}\| < \epsilon\}$. Moreover, by Lemma F.9, if $\lambda_{I,i}(\nu, \phi_0) \neq 0$ then $\theta_{1,j} = \theta_{2,j}$, for all $j \in \mathcal{S}_i$ and all $\theta_1, \theta_2 \in \Xi(\nu, \phi_0)$. Therefore, by Assumption E.1 *(i)* and the definition of $\mathcal{S}_i$, $\nabla_\phi \Psi_{s,i}(\theta_1, \phi_0) = \nabla_\phi \Psi_{s,i}(\theta_2, \phi_0)$ for all $\theta_1, \theta_2 \in \Xi(\nu, \phi_0)$ and $i \in \{i; \lambda_{I,i}(\nu, \phi_0) \neq 0\}$. Hence, by also using (F.14), we obtain that for some deterministic sequence $\delta_n \to 0$ there exists an $N = \sup_{\nu \in \mathbb{S}^d} N_\nu$ such that for every $n \geq N$:

$$\begin{aligned}
&\sup_{\phi_1,\phi_2 \in B(\phi_0,r_n)} \sup_{\nu \in \mathbb{S}^d} \sup_{\theta \in \Xi(\nu,\phi_{\tau_0})} \mathcal{A}_2 \\
&\qquad \leq \sup_{\|\theta_1 - \theta_2\| < \delta_n} \left\| \sum_{\{i;\lambda_{I,i}(\nu,\phi_0)\neq 0\}} \left[\nabla_\phi \Psi_{s,i}(g(\theta), \phi_0) - \nabla_\phi \Psi_{s,i}(\theta_2, \phi_0)\right] \lambda_{I,i}(\nu, \phi_0) \right\|
\end{aligned} \tag{F.16}$$

which converges to zero as $\delta_n \to 0$ by continuity of the function $\theta \mapsto \nabla_\phi \Psi_{s,i}(\theta, \phi_0)$ under Assumption E.4 *(i)*. This result, together with the Cauchy-Schwartz inequality, (F.11), (F.12) and $\sup_{\nu \in \mathbb{S}^d} \|\lambda(\nu, \phi_0)\| < C_{\phi_0}$, proves the claim of the lemma.

□

G Illustrating the IFT approach (Theorem 4.4) for fixed design interval regression

In this subsection we provide more detailed derivations of the model described in Section 6.2, and further explain the introduced notations used in the Implicit Function Theorem approach.

Lemma G.1. *Write $d = \dim(\theta)$. Then, for $m(\nu) := \Sigma^{-1}\nu$,*

$$\begin{aligned} S_\phi(\nu) &:= \sup_{\Psi(\theta,\phi)\leq 0} \nu^T\theta \\ &= m(\nu)^T(\phi_2 \circ 1\{m(\nu) > 0\} + \phi_1 \circ 1\{m(\nu) < 0\}) \qquad \text{(G.1)} \\ &= \frac{1}{2}m(\nu)^T(\phi_1+\phi_2) + \frac{1}{2}\left|m(\nu)^T\right|(\phi_2-\phi_1), \qquad \text{(G.2)} \end{aligned}$$

where $1\{m(\nu) < 0\} = (1\{m_1(\nu) < 0\}, \dots, 1\{m_d(\nu) < 0\})^T$, $1\{m(\nu) > 0\}$ is a vector defined similarly, and $\circ$ is the component-wise product of two vectors.

Proof. Write $\xi = \Sigma\theta$. Recall that $\Psi(\theta,\phi) \leq 0$ is equivalent to

$$\begin{pmatrix} \Psi_I(\theta,\phi) \\ \Psi_{II}(\theta,\phi) \end{pmatrix} := \begin{pmatrix} \Sigma\theta - \phi_2 \\ \phi_1 - \Sigma\theta \end{pmatrix} = \begin{pmatrix} \xi - \phi_2 \\ \phi_1 - \xi \end{pmatrix} \leq 0$$

So

$$S_\phi(\nu) = \sup_{\Psi(\theta,\phi)\leq 0} \nu^T\theta = \sup_{\phi_1\leq\xi\leq\phi_2} \nu^T\theta = \sup_{\phi_1\leq\xi\leq\phi_2} m(\nu)^T\xi = m(\nu)^T\xi^*_{\nu,\phi,\alpha}$$

for any $\alpha \in \mathbb{R}^{\dim(\theta)}$, where

$$\xi^*_{\nu,\phi,\alpha} = \phi_2 \circ 1\{m(\nu) > 0\} + \phi_1 \circ 1\{m(\nu) < 0\} + \alpha \circ 1\{m(\nu) = 0\}.$$

(G.1) then follows. To show (G.2), note that $m(\nu)^T\xi^*_{\nu,\phi,\alpha}$ does not depend on α, and

$$\begin{aligned} m(\nu)^T\xi^*_{\nu,\phi,\alpha} &= m(\nu)^T\phi_2 - m(\nu)^T[(\phi_2-\phi_1)\circ 1\{m(\nu) < 0\}] \\ &= \frac{1}{2}m(\nu)^T(\phi_1+\phi_2) \\ &\quad +\frac{1}{2}m(\nu)^T[(\phi_2-\phi_1)\circ 1\{m(\nu) > 0\}] - \frac{1}{2}m(\nu)^T[(\phi_2-\phi_1)\circ 1\{m(\nu) < 0\}] \\ &= \frac{1}{2}m(\nu)^T(\phi_1+\phi_2) \\ &\quad +\frac{1}{2}[m(\nu)\circ 1\{m(\nu) > 0\}]^T(\phi_2-\phi_1) - \frac{1}{2}[m(\nu)\circ 1\{m(\nu) < 0\}]^T(\phi_2-\phi_1) \\ &= \frac{1}{2}m(\nu)^T(\phi_1+\phi_2) + \frac{1}{2}\left|m(\nu)^T\right|(\phi_2-\phi_1) \end{aligned}$$

where the second to last equality is due to $a^T(b\circ c) = (a\circ c)^T b = \sum_i a_i b_i c_i$; the last equality follows from $a1\{a \geq 0\} - a1\{a < 0\} = |a|$. $\square$

We now provide a step-by-step illustration of Theorem 4.4 using this model.

Lemma G.2. $\sup\{\nu^T\theta;\, \theta \in \Theta(\phi)\} = \sup\{\nu^T\theta;\, \theta \in \overline{\partial\Theta(\phi)}\}$.

Proof. For any α, set

$$\widetilde{\theta}(\nu,\phi,\alpha) = \Sigma^{-1}\xi^*_{\nu,\phi,\alpha}.$$

Then $S_\phi(\nu) = m(\nu)^T\xi^*_{\nu,\phi,\alpha} = \nu^T\widetilde{\theta}(\nu,\phi,\alpha)$. It then remains to show that $\widetilde{\theta}(\nu,\phi,\alpha) \in \partial\Theta(\phi)$. It then suffices to show $\Psi_i(\widetilde{\theta}(\nu,\phi,\alpha),\phi) = 0$ for some i, which is equivalent to: there exists

j, so that either the j th component of $\Sigma\theta - \phi_2$ is zero, or the j th component of $\phi_1 - \Sigma\theta$ is zero.

In fact, due to the invertibility of Σ and that $\nu \neq 0$, there must be j, so that $m_j(\nu) \neq 0$. Let e_j denote the j-th unit vector $e_j = (0, \ldots, 1, 0, \ldots, 0)$. Then,

$$e_j' \Sigma \widetilde{\theta}(\nu, \phi, \alpha) = e_j' \xi^*_{\nu,\phi,\alpha} = (\xi^*_{\nu,\phi,\alpha})_j = \begin{cases} \phi_{2,j} & \text{if } m_j(\nu) > 0 \\ \phi_{1,j} & \text{if } m_j(\nu) < 0. \end{cases}$$

This means the j-th component of $\Sigma\widetilde{\theta}(\nu, \phi, \alpha)$ equals either ϕ_{2j} or ϕ_{1j}. This proves $\Psi_i(\widetilde{\theta}(\nu, \phi, \alpha), \phi) = 0$ for some i.

□

The above proof in fact has shown that $\Sigma^{-1}\xi^*_{\nu,\phi,\alpha} \in \overline{\partial\Theta(\phi)}$, and that $S_\phi(\nu) = \nu^T \Sigma^{-1} \xi^*_{\nu,\phi,\alpha}$ for any $\alpha \in \mathbb{R}^{\dim(\theta)}$, where $\xi^*_{\nu,\phi,\alpha}$ is as defined in the proof of Lemma G.1. By definition,

$$\Xi(\nu, \phi) := \arg \max_{\theta \in \overline{\partial\Theta(\phi)}} \nu^T \theta = \{\Sigma^{-1}\xi^*_{\nu,\phi,\alpha} : \alpha \in \mathbb{R}^{\dim(\theta)}\}.$$

We now characterize $\mathcal{I}_{\nu,\phi}$. Write $d = \dim(\theta)$. Let $I = \{1, \ldots, d\}$, $II = \{d+1, \ldots, 2d\}$, corresponding to the indices of

$$\Psi(\theta, \phi) := \begin{pmatrix} \Psi_I(\theta, \phi) \\ \Psi_{II}(\theta, \phi) \end{pmatrix} := \begin{pmatrix} \Sigma\theta - \phi_2 \\ \phi_1 - \Sigma\theta \end{pmatrix}.$$

Let $(m(\nu) \neq 0) = \{j \leq d : m_j(\nu) \neq 0\}$.

Lemma G.3.

$$\begin{aligned} \mathcal{I}_{\nu,\phi} &:= \{i \in \{1, 2, \ldots, 2d\};\ \Psi_i(\theta, \phi) = 0, \forall \theta \in \Xi(\nu, \phi)\} \\ &= \{i \in I : m_i(\nu) > 0\} \cup \{i \in II : m_{i-d}(\nu) < 0\}. \end{aligned}$$

Therefore, $d_{\nu,\phi} := card(\mathcal{I}_{\nu,\phi}) = |(m(\nu) \neq 0)|_0$.

Proof. Recall $\widetilde{\theta}(\nu, \phi, \alpha) = \Sigma^{-1}\xi^*_{\nu,\phi,\alpha}$.

$$\begin{aligned} \mathcal{I}_{\nu,\phi} &:= \left\{i \in \{1, 2, \ldots, 2d\};\ \Psi_i(\widetilde{\theta}(\nu, \phi, \alpha), \phi) = 0, \forall \alpha \in \mathbb{R}^d\right\} \\ &= \{i \leq d : [\Sigma\widetilde{\theta}(\nu, \phi, \alpha)]_i = \phi_{2,i}, \forall \alpha \in \mathbb{R}^d\} \cup \{i = d + j : [\Sigma\widetilde{\theta}(\nu, \phi, \alpha)]_j = \phi_{1,j}, \forall \alpha \in \mathbb{R}^d\}. \end{aligned}$$

The j-th component of $\Sigma\widetilde{\theta}(\nu, \phi, \alpha)$ equals either ϕ_{2j} or ϕ_{1j}, or α_j.

$$[\Sigma\widetilde{\theta}(\nu, \phi, \alpha)]_j = e_j' \Sigma \widetilde{\theta}(\nu, \phi, \alpha) = (\xi^*_{\nu,\phi,\alpha})_j = \begin{cases} \phi_{2,j} & \text{if } m_j(\nu) \geq 0 \\ \phi_{1,j} & \text{if } m_j(\nu) < 0. \\ \alpha_j & \text{if } m_j(\nu) = 0. \end{cases}$$

This is always either $\phi_{2,j}$ or $\phi_{1,j}$ regardless of the choice of α if and only if $m_j(\nu) \neq 0$. So $\mathcal{I}_{\nu,\phi} = \{i \leq d : m_i(\nu) > 0\} \cup \{i = d + j : m_j(\nu) < 0\}$. □

Now we are ready to describe $A(\nu)$.

Lemma G.4.

$$\begin{aligned} A(\nu) &= \max_{\theta\in\Xi(\nu,\phi_0)} -\nu^T_{1:d_{\nu,\phi_0}} M_1(\theta,\nu)^{-1}M_2(\theta,\nu) \\ &= \frac{1}{2}\left(m(\nu)^T - |m(\nu)^T|, m(\nu)^T + |m(\nu)^T|\right). \end{aligned}$$

Proof. The first equality is rewriting Theorem 4.4. The second equality can be easily proved by using a first order Taylor expansion of S_ϕ around ϕ_0 and noting that it is in fact $\partial S_\phi(\nu)/\partial\phi|_{\phi=\phi_0} = A(\nu)$. We now verify that the right hand side on the second equality is indeed $\max_{\theta\in\Xi(\nu,\phi_0)} -\nu^T_{1:d_{\nu,\phi_0}} M_1(\theta,\nu)^{-1}M_2(\theta,\nu)$. To do so, we characterize $M_1(\theta,\nu)$ and $M_2(\theta,\nu)$. We show that $M_1(\theta,\nu)$ is a row-permutation of Σ, and $M_2(\theta,\nu)$ is a permutation matrix.

Let $\Sigma_{(m>0)}$ and $\Sigma_{(m<0)}$ respectively denote rows of $\Sigma_{(m\neq0)}$ indexed by $(m(\nu)>0)$ and $(m(\nu)<0)$. Then

$$M_1(\theta,\nu) = \frac{\partial\Psi_{\mathcal{I}_{\nu,\phi}}(\theta,\phi_0)}{\partial\theta_{\mathcal{I}_{\nu,\phi_0}}} = \begin{pmatrix}\Sigma_{(m>0)}\\ -\Sigma_{(m<0)}\end{pmatrix}, \quad d_{\nu,\phi}\times d_{\nu,\phi}, \quad d_{\nu,\phi} = |(m(\nu)\neq 0)|_0.$$

Note that this is simply a row-permutation of $\Sigma_{(m\neq0)}$, up to a sign change. Let $D(\nu)$ be a $d_{\nu,\phi}\times d_{\nu,\phi}$ matrix, whose first $|(m(\nu)>0)|_0 := card((m(\nu)>0))$ rows are like $(0,\ldots,0,1,0,\ldots,0)$. Here, the locations of 1 in the rows are indexed by $(m(\nu)>0)$. Let the remaining $|(m(\nu)<0)|_0$ rows of $D(\nu)$ be like $(0,\ldots,0,-1,0,\ldots,0)$, where the locations of -1 in the rows are indexed by $(m(\nu)<0)$. Then, $D(\nu)$ is a permutation matrix, and we can write

$$M_1(\theta,\nu) = D(\nu)\Sigma_{(m\neq0)}.$$

Also, $D(\nu)^{-1} = D(\nu)^T$.

As for $M_2(\theta,\nu)$, by definition,

$$M_2(\theta,\nu) := \frac{\partial\Psi_{\mathcal{I}_{\nu,\phi}}(\theta,\phi_0)}{\partial\phi} = \begin{pmatrix}0 & \tilde{D}_{-1}(\nu)\\ \tilde{D}_1(\nu) & 0\end{pmatrix}.$$

where $\tilde{D}_1(\nu)$ is a $|(m(\nu)<0)|_0\times d$ matrix whose rows are like $(0,\ldots,0,1,0,\ldots,0)$. Here the location of 1 in the j-th row is given by the j-th element of the set $(m(\nu)<0)$; $\tilde{D}_{-1}(\nu)$ is a $|(m(\nu)>0)|_0\times d$ matrix whose rows are like $(0,\ldots,0,-1,0,\ldots,0)$. Here, the location of -1 in the j-th row of $\tilde{D}_{-1}(\nu)$ is given by the j-th element of the set $(m(\nu)>0)$. Then,

$$\max_{\theta\in\Xi(\nu,\phi_0)} -\nu^T_{1:d_{\nu,\phi_0}} M_1(\theta,\nu)^{-1}M_2(\theta,\nu) = -\nu^T_{1:d_{\nu,\phi_0}}\Sigma^{-1}_{(m\neq0)}D(\nu)^TM_2(\theta,\nu).$$

We now respectively describe $\nu^T_{1:d_{\nu,\phi_0}}\Sigma^{-1}_{(m\neq0)}$ and $D(\nu)^TM_2(\theta,\nu)$.

As for $\nu^T_{1:d_{\nu,\phi_0}}\Sigma^{-1}_{(m\neq0)}$, we show it equals the vector of nonzero elements of $m(\nu)^T$, denoted by $m_{\neq 0}(\nu)^T$. Consider the permutation version of $\Sigma^{-1}\nu = m(\nu)$:

$$\Bigg[\underbrace{\begin{pmatrix}\Sigma_{(m\neq 0)} & G_1\\ G_1^T & G_2\end{pmatrix}}_{\text{permutation of }\Sigma}\Bigg]^{-1} \underbrace{\begin{pmatrix}\nu_{1:d_{\nu,\phi_0}}\\ \tilde{\nu}\end{pmatrix}}_{\text{permutation of }\nu} = \underbrace{\begin{pmatrix}m_{\neq 0}(\nu)\\ 0\end{pmatrix}}_{\text{permutation of }m(\nu)}.$$

Using the blockwise matrix inversion formula, we can write

$$\begin{pmatrix}\Sigma^{-1}_{(m\neq 0)} + R_1 & R_2^T\\ R_2 & H\end{pmatrix}\begin{pmatrix}\nu_{1:d_{\nu,\phi_0}}\\ \tilde{\nu}\end{pmatrix} = \begin{pmatrix}m_{\neq 0}(\nu)\\ 0\end{pmatrix},$$

where $H := (G_2 - G_1^T\Sigma^{-1}_{(m\neq 0)}G_1)^{-1}$, $R_1 = \Sigma^{-1}_{(m\neq 0)}G_1HG_1^T\Sigma^{-1}_{(m\neq 0)}$, and $R_2 = -HG_1^T\Sigma^{-1}_{(m\neq 0)}$. Here the second block of equalities $R_2\nu_{1:d_{\nu,\phi_0}} + H\tilde{\nu} = 0$ implies $\tilde{\nu} = G_1^T\Sigma^{-1}_{(m\neq 0)}\nu_{1:d_{\nu,\phi_0}}$. Substitute to the first block of equalities,

$$\begin{aligned}
m_{\neq 0}(\nu) &= \Sigma^{-1}_{(m\neq 0)}\nu_{1:d_{\nu,\phi_0}} + R_1\nu_{1:d_{\nu,\phi_0}} + R_2^T\tilde{\nu}\\
&= \Sigma^{-1}_{(m\neq 0)}\nu_{1:d_{\nu,\phi_0}} + \Sigma^{-1}_{(m\neq 0)}G_1H[G_1^T\Sigma^{-1}_{(m\neq 0)}\nu_{1:d_{\nu,\phi_0}} - \tilde{\nu}]\\
&= \Sigma^{-1}_{(m\neq 0)}\nu_{1:d_{\nu,\phi_0}}.
\end{aligned}$$

This shows the permutation version of $m(\nu)^T$ is $(\nu^T_{1:d_{\nu,\phi_0}}\Sigma^{-1}_{(m\neq 0)}, 0)$.

As for $D(\nu)^TM_2(\theta,\nu)$, note that it is a $|(m(\nu)\neq 0)|_0 \times 2d$ matrix, whose rows are like $(0,\dots,0,-1,0,\dots,0)$. We can partition it into $D(\nu)^TM_2(\theta,\nu) = [A;B]$, both A, B are $|(m(\nu)\neq 0)|_0 \times d$ matrices. To describe A, B, we order all elements of $(m(\nu) > 0)$ and $(m(\nu) < 0)$ as $i_1 < i_2 < ...$, and $\{i_1, i_2, ...\} = (m(\nu) > 0) \cup (m(\nu) < 0)$. For any j, if $i_j \in (m(\nu) < 0)$, then set the (j, i_j) th element of A to -1; if $i_j \in (m(\nu) < 0)$, then set the (j, i_j) th element of B to -1. In either case, set all other components of $[A;B]$ on the same row to zero. For instance, suppose $d = 4$, $(m(\nu) > 0) = \{1, 3\}$ and $(m(\nu) < 0) = 4$. Then

$$D(\nu)^TM_2(\theta,\nu) = [A;B],\quad A = \begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&-1\end{pmatrix},\quad B = \begin{pmatrix}-1&0&0&0\\0&0&-1&0\\0&0&0&0\end{pmatrix}.$$

Then

$$\begin{aligned}
-\nu^T_{1:d_{\nu,\phi}}D(\nu)^TM_2(\theta,\nu) &= -m_{\neq 0}(\nu)^T[A;B]\\
&= (m(\nu)^T \circ 1\{m(\nu) < 0\}, m(\nu)^T \circ 1\{m(\nu) > 0\})\\
&= \frac{1}{2}\left(m(\nu)^T - |m(\nu)^T|, m(\nu)^T + |m(\nu)^T|\right).
\end{aligned}$$

□

Example G.1. Suppose $d = 3$, and $m_1(\nu) < 0, m_3(\nu) > 0$ and $m_2(\nu) = 0$, then $(m(\nu) > 0) = \{3\}$, and $(m(\nu) < 0) = \{1\}$. Then

$$\Sigma_{(m\neq 0)} = \begin{pmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{pmatrix}, \quad M_1(\theta,\nu) = \begin{pmatrix} \sigma_{31} & \sigma_{33} \\ -\sigma_{11} & -\sigma_{13} \end{pmatrix}, \quad \xi^*_{\nu,\phi} = \begin{pmatrix} \phi_{11} \\ \phi_{23} \end{pmatrix}$$

$$D(\nu) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad M_2(\theta,\nu) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$D(\nu)^T M_2(\theta,\nu) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Then,

$$\begin{aligned} -\nu^T_{1:d_{\nu,\phi}} \Sigma^T_{(m\neq 0)} D(\nu)^T M_2(\theta,\nu) &= (m_1(\nu), 0, 0, 0, 0, m_3(\nu)) \\ &= \frac{1}{2}\left(m(\nu)^T - |m(\nu)^T|, m(\nu)^T + |m(\nu)^T|\right). \end{aligned}$$

H Uniformity

It is possible to achieve the coverage results uniformly over a set of DGP's, allowing sequences of DGP that converge to the point identification. This is true as long as Assumptions 2.1 and 4.1 hold uniformly on a class of DGP's. Let $\mathcal{F}$ denote a collection of DGP's F, and let the data $D_n = D_n(F)$ be generated from a given F. Given F, the parameter of interest becomes $\phi(F)$, whose true value is $\phi_0 = \phi_0(F)$. We also denote $\Psi(\theta, \phi(F)) \leq 0$ as the moment constraints, which depends on the given DGP F. The deduced support function is denoted by:

$$\forall \nu \in \mathbb{S}^d, \qquad S_\phi(\nu, F) := \sup_{\theta\in\Theta} \left\{\nu^T\theta; \Psi(\theta, \phi(F)) \leq 0\right\}.$$

To prove the uniform coverages, we need to strengthen Assumptions 2.1 and 4.1 to hold uniformly.

Assumption H.1. *Let $\mathcal{F}$ be a class of distribution functions so that:*
(i) The marginal posterior of $\phi(F)$ is such that, for any $\epsilon, \delta > 0$, there is $C > 0$ such that

$$\inf_{F\in\mathcal{F}} P_{D_n(F)} \left(P\left(\|\phi(F) - \phi_0(F)\| > Cn^{-1/2} | D_n(F)\right) < \delta\right) > 1 - \epsilon.$$

(ii) Let $P_{\sqrt{n}(\phi(F)-\phi_0(F))|D_n}$ denote the posterior distribution of $\sqrt{n}(\phi(F) - \phi_0(F))$. Then,

$$\sup_{F\in\mathcal{F}} \|P_{\sqrt{n}(\phi(F)-\phi_0(F))|D_n} - \mathcal{N}(\Delta_{n,\phi_0}(F), I_0(F)^{-1})\|_{TV} \to^P 0$$

where $\Delta_{n,\phi_0}(F) := n^{-1/2}\sum_{i=1}^n I_0(F)^{-1}\ell_{\phi_0}(X_i)$.
(iii) There exists a regular estimator $\hat{\phi}(F)$ of $\phi_0(F)$ that satisfies

$$\sup_{F\in\mathcal{F}} \|P_{\sqrt{n}(\hat{\phi}(F)-\phi_0(F))} - \mathcal{N}(0, I_0(F)^{-1})\|_{TV} \to^P 0$$

where $P_{\sqrt{n}(\phi(F)-\phi_0(F))}$ denotes the sampling distribution of $\sqrt{n}(\hat{\phi}(F) - \phi_0(F))$, $I_0(F)^{-1}$ is the semiparametric efficient Fisher information matrix for the DGP F.

(iv) There is a continuous vector function (with respect to v for each fixed F) $A(\nu, F)$ such that for

$$f_F(\phi_1, \phi_2) := \sup_{\nu \in \mathbb{S}^d} \left|(S_{\phi_1}(\nu, F) - S_{\phi_2}(\nu, F)) - A(\nu, F)^T(\phi_1(F) - \phi_2(F))\right|,$$

we have, for $r_n = n^{-1/2}$, and any $C > 0$, as $n \to \infty$,

$$\sup_{F \in \mathcal{F}} \sup_{\phi_1, \phi_2 \in B(\phi_0, Cr_n)} \frac{f_F(\phi_1, \phi_2)}{\|\phi_1 - \phi_2\|} \to 0.$$

Theorem H.1 (Uniformity of frequentist coverages)**.** *Let $\mathcal{F}$ be a collection of DGP's so that Assumption H.1 holds. Suppose $\Theta(\phi)$ is convex and closed for every ϕ in its parameter space. In addition, let $g(\theta) = g^T\theta$ for some $\|g\| = 1$. Then, as $n \to \infty$, for any $\tau \in (0, 1)$,*

(i) Identified set: $\inf_{F \in \mathcal{F}} P_{D_n(F)}\left(\Theta(\phi_0(F)) \subset \Theta(\hat{\phi}(F))^{q_\tau/\sqrt{n}}\right) = 1 - \tau + o_P(1)$*;*

(ii) Partially identified parameter: $\inf_{F \in \mathcal{F}} \inf_{\theta \in \Theta(\phi_0(F))} P_{D_n(F)}\left(\theta \in \widehat{\Omega}(F)\right) \geq 1 - \tau - o_P(1)$*;*

(iii) Identified set (for scalar functions):

$$\inf_{F \in \mathcal{F}} P_{D_n(F)}\left(G(\phi_0(F)) \subset [-S_{\hat{\phi}}(-g, F) - \frac{c_\tau}{\sqrt{n}}, S_{\hat{\phi}}(g, F) + \frac{c_\tau}{\sqrt{n}}]\right) = 1 - \tau + o_P(1),$$

where $G(\phi_0(F)) = \{g(\theta) : \theta \in \Theta(\phi_0(F))\}$.

(iv) Scalar functions: $\inf_{F \in \mathcal{F}} \inf_{g(\theta) \in G(\phi_0)} P_{D_n(F)}\left(g(\theta) \in \widehat{\Omega}_g(F)\right) \geq 1 - \tau - o_P(1)$*, where*
$$\widehat{\Omega}_g(F) := \left\{g(\theta) : P\left(g(\theta) \in [-S_\phi(-g, F) - \tfrac{2c_\tau}{\sqrt{n}}, S_\phi(g, F) + \tfrac{2c_\tau}{\sqrt{n}}]\middle| D_n(F)\right) \geq 1 - \tau\right\}.$$

Proof. (i) Let $R_n(F) := P_{D_n(F)}\left(\Theta(\phi_0(F)) \subset \Theta(\hat{\phi}(F))^{q_\tau/\sqrt{n}}\right)$. Let $\{F_n : n \geq 1\} \subset \mathcal{F}$ be a sequence such that $\liminf_n \inf_{F \in \mathcal{F}} R_n(F) = \liminf_n R_n(F_n)$. Such a sequence always exists. Let $\{F_{n_k} : k \geq 1\}$ be a subsequence of $\{F_n : n \geq 1\}$ so that $\liminf_n R_n(F_n) = \lim_n R_n(F_{n_k})$. Thus for any $\epsilon > 0$, for all large k,

$$\inf_{F \in \mathcal{F}} R_n(F) \geq \liminf_n \inf_{F \in \mathcal{F}} R_n(F) - \epsilon = \lim_n R_n(F_{n_k}) - \epsilon \geq R_n(F_{n_k}) - 2\epsilon.$$

This implies $\inf_{F \in \mathcal{F}} R_n(F) \geq R_n(F_{n_k}) - o(1)$. Now by Assumption H.1, we have

(a) $P_{F_{n_k}}(\|\phi(F_{n_k}) - \phi_0(F_{n_k})\| \leq Cn^{-1/2}|D_n) > 1 - \delta$ with probability at least $1 - \epsilon$.

(b) $\|P_{\sqrt{n}(\phi(F_{n_k}) - \phi_0(F_{n_k}))|D_n} - \mathcal{N}(\Delta_{n,\phi_0}(F_{n_k}), I_0(F_{n_k})^{-1})\|_{TV} \to^p 0$.

(c) $\|P_{\sqrt{n}(\hat{\phi}(F_{n_k}) - \phi_0(F_{n_k}))} - \mathcal{N}(0, I_0(F_{n_k})^{-1})\|_{TV} \to^P 0$.

(d) There is a continuous vector function $A(\nu, F_{n_k})$ such that for

$$f_{F_{n_k}}(\phi_1, \phi_2) := \sup_{\nu \in \mathbb{S}^d} \left|(S_{\phi_1}(\nu, F_{n_k}) - S_{\phi_2}(\nu, F_{n_k})) - A(\nu, F_{n_k})^T(\phi_1 - \phi_2)\right|,$$

we have, $\sup_{\phi_1,\phi_2 \in B(\phi_0,r_n)} \frac{f_{F_{n_k}}(\phi_1,\phi_2)}{\|\phi_1-\phi_2\|} \to 0$. Then the conditions of Theorem 4.2 hold. Hence Theorem 4.2 implies $R_n(F_{n_k}) \geq 1-\tau+o_P(1)$. Thus $\inf_{F\in\mathcal{F}} R_n(F) \geq 1-\tau+o_P(1)$. This proves part (i). Parts (ii)-(iv) can be proved using the same argument, and hence the proof is omitted for brevity.

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite dimensional analysis - A Hitchhiker's Guide.* Springer-Verlag, Berlin.

Dieudonné, J. (1960). *Foundations of Modern Analysis.* Academic Press Inc.

Kaido, H. and Santos, A. (2014). Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica* **82** 387–413.

Kato, T. (1995). *Perturbation theory for linear operators.* Springer-Verlag Berlin Heidelberg.

Mas-Colell, A., Whinston, A. and Green, J. (1995). *Microeconomic Theory.* Oxford Univeristy Press.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica* **70** 583–601.

Rockafellar, R. (1970). *Convex Analysis.* The first edition ed. Princeton University Press, Princeton.

Rudin, W. (2006). *Functional Analysis.* McGraw-Hill.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research* **30** 169–186.

van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes.* The first ed. Springer.