Semiparametric Bayesian Partially Identified Models based on Support Function

Yuan Liao [1] & Anna Simoni[2]

[1]University of Maryland

[2]CNRS and THEMA

ESEM, 26-30 August, 2013 - Gothenburg

Partially Identified models I

- We consider a finite dimensional structural parameter $\theta \in \Theta$.
- In many econometric/statistical models this parameter is not point-identified but only *partially identified* to belong to a non-singleton set (*identified set*).
- Partial identification arises when *(i)* limitation of what variables can be observed and/or *(ii)* the plausible constraints coming from economic theory only allow to place the parameter θ into a proper subset of the parameter space Θ.
- The *identified set* is the set of values for θ that are compatible with a particular distribution of observables and economic theory.
- The *identified set* is characterized by ϕ: $\Theta(\phi) \subset \Theta$, where ϕ is a parameter characterizing the distribution of the data, *e.g.* a vector of moments.

Example 1: Interval censored data

- $(Y, Y_1, Y_2) =$ three dimensional random vector such that

$$Y \in [Y_1, Y_2] \qquad w.p.\ 1$$

  $Y =$ unobservable, $Y_1, Y_2 =$ observable.

- $\theta = \mathbf{E}(Y)$;

- $\phi = (\phi_1, \phi_2)'$ with $\phi_1 = \mathbf{E}(Y_1)$and $\phi_2 = \mathbf{E}(Y_2)$.

- Identified set: $\Theta(\phi) = [\phi_1, \phi_2]$.

Partially Identified models: examples I

- *Models with missing observations:* see *e.g.* Imbens and Rubin (1997), Horowitz and Manski (1998, 2000), Gustafson (2011), . . .
- *Interval data:* see *e.g.* Manski and Tamer (2002)
- *Game-theoretic models with multiple equilibria:* entry games (Bresnan and Reiss 1991, Ciliberto and Tamer, . . .), auctions (Ciliberto and Tamer 2009, . . .)
- *Sign restrictions in VAR models:* Canova and De Nicolò (2002), . . .
- *DSGE models:* Lubik and Schorfheide (2004).
- *Finance:* Hansen-Jagannathan bound, see *e.g.* Chernozhukov, Kocatulum and Menzel (2012)

Aim of the paper I

- Inference for θ and $\Theta(\phi)$:
  1. estimating the boundaries of $\Theta(\phi)$;
  2. reporting a confidence set for $\Theta(\phi)$. Distinction btw confidence sets for θ and confidence sets for $\Theta(\phi)$.

- A partial list of recent contributions includes:
  - *frequentist:* Andrews & Guggenberger (2009 ET), Andrews & Soares (2010 ECTA), Beresteanu, Molchanov & Molinari (2010 ECTA), Bontemps, Magnac & Maurin (2012 ECTA), Bugni (2010 ECTA), Canay (2010), Chernozhukov, Hong & Tamer (2007 ECTA), Chiburis (2008), Imbens & Manski (2004 ECTA), Manski (1994), Menzel (2008), Otsu (2006), Romano & Shaikh (2010 ECTA), Rosen (2008 JoE), Stoye (2010 ECTA).
  - *Bayesian:* Gelfand & Sahu (1999), Neath & Samaniego (1997), Norets & Tang (2012), Kitagawa (2012 wp), Florens & Simoni (2011 wp), Moon & Schorfheide (2012 ECTA), Epstein & Seo (2011), Kline (2011), Gustafson (2010, 2012)etc.

Aim of the paper II

- Limit of frequentist approaches: do not naturally tells us anything inside the identified region. So, perfect knowledge of the distribution of the data (*i.e.* $n \to \infty$) will give the (true) identified region.

- Bayesian analysis provides (as $n \to \infty$) a posterior that converges towards a non-degenerate distribution and gives relative weighting of points in the identification region. In a *decision problem setting* we can always take a decision.

- Computational advantage of Bayesian:

  - Bayesian credible sets (BCS) are often easy to construct;
  - BCS are easy to project to a low-dimensional space (if we are interested in just one element of θ).

- Finite sample advantages: when ϕ (which characterizes the support $\Theta(\phi)$ of θ) is integrated out with respect to $p(\phi|Data)$ then the posterior of θ is completely revised by the data. Thus, a Bayesian procedure learns about θ based on the whole posterior distribution of ϕ (more information in finite samples)

Aim of the paper III

- **Our contributions:**

  1. We propose a pure Bayesian procedure without assuming a parametric form of the true likelihood: nonparametric priors on the likelihood, and a prior on (ϕ, θ) (semi-parametric procedure). We use a conditional prior $\pi(\theta|\phi)$ so that we take into account the partial identification. Still, the conditional prior can be very informative inside the identified set.

     ▷ This is different from traditional Likelihood based approaches, see Moon & Schorfheide (2012 ECTA), Gustafson (2011), and also different from moment-inequality-based likelihood (LIL) approaches, see Kim (2002 JoE), Liao & Jiang (2010 AoS).

  2. We construct a (2-sided) BCS for $\Theta(\phi)$ which has (asymptotically) correct frequentist coverage probability. BCS are constructed based on the support function of $\Theta(\phi)$.

  3. We provide a frequentist validation of our procedure:

     - posterior consistency of the posteriors of: θ, $\Theta(\phi)$ and the support function;
     - Bernstein-von Mises (BvM) theorem for the posterior of the support function.

Aim of the paper IV

4. We extend Moon & Schorfheide (2012)'s analysis for θ to a semi-parametric setup (which is relevant in more general moment inequality models). Thus, the (asymptotic) equivalence between BCS and FCS for θ breaks down in partially-identified models.

5. Our procedure is still valid even if there is no prior information on θ available: we do not need to have a prior for θ in order to make Bayesian inference on $\Theta(\phi)$.

6. Projection and subset inference: we show that it is relatively easy for the Bayesian partial identification approach to project onto low-dimensional subspaces for subset inference, and the computation is fast.

7. Our Bayesian inference is valid uniformly over a class of data generating process. In particular, as the identified set shrinks to a singleton so that point identification is (nearly) achieved, the Bayesian inference for the identified set carries over.

8. Applications: we focus on the interval censoring, interval regression and missing data problems. We also study an application problem for the financial asset pricing model.

Semiparametric Bayesian setup I

- Observable random variable X for which one has n i.i.d. observations, denoted by $D_n = \{X_i\}_{i=1}^n$. X takes values in $(\mathcal{X}, \mathfrak{B}_x, F)$.

- Three parameters: $(\theta, \phi, F) \in \Theta \times \Phi \times \mathcal{F}$ where $F \in \mathcal{F}$ is the probability distribution of X.

- With respect to identification:

  - $(\phi, F) =$ identified parameter which characterizes the sampling distribution;
  - $\theta =$ partially identified parameter which is linked to the sampling distribution through ϕ.

- The prior is naturally decomposed in a marginal prior for (ϕ, F) and a conditional prior for θ given ϕ such that

$$\pi(\theta \in \Theta(\phi)|\phi) = 1.$$

Semiparametric Bayesian setup II

- Therefore, partial identification is incorporated into the prior:

$$\pi(\theta|\phi) \propto I_{\theta\in\Theta(\phi)}\, g(\theta).$$

- Two possible schemes for the prior on (ϕ, F):
  - fully nonparametric prior
  - semiparametric prior.

  This prior induces a prior for the identified set $\Theta(\phi)$.

- We develop 2 inferences:
  - Inference for θ: marginal posterior $\pi(\theta|D_n)$
  - Inference for $\Theta(\phi)$: posterior of $\Theta(\phi)$ or posterior of its support function.

  Both these posteriors are justified by good frequentist asymptotic properties.

Prior on (ϕ, F)

Nonparametric Prior:

- The parameter ϕ is a measurable function of F: $\phi = \phi(F)$. *E.g.:* $\phi = \mathbf{E}^F(X) = \int xF(dx)$.

- Nonparametric prior for F and deduce from it the prior for ϕ via $\phi(F)$:

$$X|F \sim F, \qquad F \sim \pi(F), \qquad \theta|\phi = \phi(F) \sim \pi(\theta|\phi(F))$$

Semiparametric Prior:

- Reformulate the model and parameterize F in terms of ϕ and η

$$\mathcal{F} = \{F_{\phi,\eta}; \phi \in \Phi, \; \eta \in \mathcal{P}\}$$

- We assume: $\exists$ a fixed true value for F, denoted by F_0. Then $\exists!\ \phi_0 \in \Phi$ and $\eta_0 \in \mathcal{P}$ such that $F_0 = F_{\phi_0,\eta_0}$ (since both ϕ and F are identified).

- Bayesian experiment:

$$X|\phi,\eta \sim F_{\phi,\eta}, \qquad (\phi,\eta) \sim \pi(\phi,\eta) = \pi(\phi) \times \pi(\eta),$$
$$\theta|\phi,\eta \sim \pi(\theta|\phi).$$

Outline:

Posterior Consistency for Nonparametric prior I

- Conditional posterior of θ, given ϕ:

$$p(\theta|\phi(F), D_n) = \pi(\theta|\phi(F)).$$

Conditionally on ϕ, the Bayesian experiment is completely uninformative about θ: the prior distribution of θ is revised by the data only through the information brought by the identified parameter $\phi(F)$.

- Marginal posterior of θ:

$$p(\theta|D_n) = \int_{\mathcal{F}} p(\theta|\phi(F), D_n)p(F|D_n)dF = \int_{\mathcal{F}} \pi(\theta|\phi(F))p(F|D_n)dF.$$

The shape of $p(\theta|D_n)$ still relies upon the prior distribution of θ even asymptotically. So, the Bernstein-von Mises theorem does not hold.

Posterior Consistency for Nonparametric prior II

- Posterior consistency: let $d(\theta, \Theta(\phi)) = \inf_{x \in \Theta(\phi)} \|\theta - x\|$. Our goal is to show

$$P(\theta \in \Theta(\phi_0)^{\epsilon} | D_n) \to^p 1 \quad \text{and} \quad P(\theta \in \Theta(\phi_0)^{-\epsilon} | D_n) \to^p (1 - \tau)$$

for any $\epsilon > 0$ and some $\tau > 0$ where

$$\Theta(\phi)^{\epsilon} = \{\theta : d(\theta, \Theta(\phi)) \leq \epsilon\} \quad \text{and} \quad \Theta(\phi)^{-\epsilon} = \{\theta : d(\theta, \Theta \backslash \Theta(\phi)) \geq \epsilon\}$$

are the ϵ-envelope and ϵ-contraction of $\Theta(\phi)$.

Posterior Consistency for Nonparametric prior III

- **Assumption 1:** At least one of the following holds:

  (i). The measurable function $\phi : \mathcal{F} \to \Phi$ is continuous and $\pi(F)$ is such that:

  $$\int_{\mathcal{F}} m(F)p(F|D_n)dF \to^p \int_{\mathcal{F}} m(F)\delta_{F_0}(dF)$$

  for any bounded and continuous function $m(\cdot)$ on $\mathcal{F}$;

  (ii). the prior $\pi(\phi)$ is such that:

  $$\int_{\Phi} m(\phi)p(\phi|D_n)d\phi \to^p \int_{\Phi} m(\phi)\delta_{\phi_0}(d\phi)$$

  for any bounded and continuous function $m(\cdot)$ on Φ.

- **Assumption 2:** For any $\epsilon > 0$ there are measurable sets $A_1, A_2 \subset \Phi$ such that $0 < \pi(\phi \in A_i) \leq 1, i = 1, 2$ and
  (i) for all $\phi \in A_1$, $\Theta(\phi_0)^{\epsilon} \cap \Theta(\phi) \neq \emptyset$; for all $\phi \notin A_1$, $\Theta(\phi_0)^{\epsilon} \cap \Theta(\phi) = \emptyset$,
  (ii) for all $\phi \in A_2$, $\Theta(\phi_0)^{-\epsilon} \cap \Theta(\phi) \neq \emptyset$; for all $\phi \notin A_2$, $\Theta(\phi_0)^{-\epsilon} \cap \Theta(\phi) = \emptyset$.

- **Assumption 3:** For any $\epsilon > 0$, and $\phi \in \Phi$, $\pi(\theta \in \Theta(\phi)^{-\epsilon}|\phi) < 1$.

Posterior Consistency for Nonparametric prior IV

Theorem 1:

Let $\pi(\theta|\phi)$ be a regular conditional distribution. Under assumptions 1-3,for any $\epsilon > 0$, there is $\tau \in (0, 1]$ such that

$$P(\theta \in \Theta(\phi_0)^{\epsilon}|D_n) \to^{p} 1 \quad \text{and} \quad P(\theta \in \Theta(\phi_0)^{-\epsilon}|D_n) \to^{p} (1-\tau).$$

▶ Proof

Outline:

Moment inequality models:

Let us consider a more specific partially identified model which assumes that θ satisfies:

$$\Psi(\theta,\phi) \leq 0, \quad \Psi(\theta,\phi) = \begin{pmatrix} \Psi_1(\theta,\phi) \\ \vdots \\ \Psi_k(\theta,\phi) \end{pmatrix}, \tag{1}$$

where $\Psi : \Theta \times \Phi \to \mathbb{R}^k$ is a known function of (θ,ϕ). The identified set is given by

$$\Theta(\phi) = \{\theta \in \Theta : \Psi(\theta,\phi) \leq 0\}.$$

Many partially identified models can be characterized as moment inequality models, see Andrews, Berry & Jia (2004), Otsu (2006), Pakes, Porter, Ho & Ishii (2006), Chernozhukov, Hong & Tamer (2007), Shaikh (2008, 2010), Rosen (2008), Bugni (2010), and Canay (2010), among others.

Semiparametric prior setup: posterior concentration rate I

- Consistency of the posterior of $\Theta(\phi)$: we aim at deriving the rate r_n such that

$$P\left(d_H(\Theta(\phi), \Theta(\phi_0)) < Cr_n | D_n\right) \to^p 1$$

for some $C > 0$, where d_H denotes the Hausdorff distance. The ***Hausdorff distance*** between two sets A and B is defined as

$$\begin{aligned} d_H(A, B) &= \max\left\{\sup_{a\in A} d(a, B),\ \sup_{b\in B} d(b, A)\right\} \\ &= \max\left\{\sup_{a\in A}\inf_{b\in B} \|a-b\|, \sup_{b\in B}\inf_{a\in A} \|b-a\|\right\}. \end{aligned}$$

Semiparametric prior setup: posterior concentration rate II

- **Assumption PC:** The marginal posterior of ϕ is such that

$$P(\|\phi - \phi_0\| \leq Cn^{-1/2}(\log n)^{1/2}|D_n) \to^p 1 \tag{2}$$

see Rivoirard & Rousseau (2012 - nonparametric) or Bickel & Kleijn (2012 - semiparametric).

▶Conditions

Semiparametric prior setup: posterior concentration rate III

Theorem 2:

Assume that:

H1. $\Theta \times \Phi$ is compact;

H2. Lipschitz equi-continuity on Φ: for some $K > 0$, $\forall \phi_1, \phi_2 \in \Phi$:

$$\sup_{\theta \in \Theta} \|\Psi(\theta, \phi_1) - \Psi(\theta, \phi_2)\| \leq K\|\phi_1 - \phi_2\|;$$

H3. $\exists$ a closed neighborhood $U(\phi_0)$, such that for any $a_n = O(1)$, and any $\phi \in U(\phi_0)$, $\exists\ C_\phi > 0$ that might depend on ϕ such that

$$\inf_{\theta : d(\theta, \Theta(\phi)) \geq C_\phi a_n} \max_{i \leq k} \Psi_i(\theta, \phi) > a_n.$$

Then, under assumption PC:

$$P\left(d_H(\Theta(\phi), \Theta(\phi_0)) > Cn^{-1/2}(\log n)^{1/2} | D_n\right) \to^p 0$$

for some $C > 0$.

▶Proof ▶Example 1

Outline:

1. Introduction
2. Setup and Prior specification
3. Inference for θ: posterior consistency
4. Inference for $\Theta(\phi)$
   - Posterior of $\Theta(\phi)$
   - Posterior of the Support Function
5. Bayesian Credible Region

Bayesian Inference of Support Function I

- Aim: to develop inference for $\Theta(\phi)$ through its support function $S_\phi(p)$. We assume $\Theta(\phi)$ is a convex set for each ϕ.

Definition:

Let $\mathbb{S}^d \subset \mathbb{R}^d$, $d = \dim(\theta)$. $\forall \phi \in \Phi$ the *support function for* $\Theta(\phi)$ is a function $S_\phi(\cdot) : \mathbb{S}^d \to \mathbb{R}$ such that

$$S_\phi(p) = \sup_{\theta \in \Theta(\phi)} \theta^T p.$$

- We consider the *moment inequality model* previously described:

$$\Theta(\phi) := \{\theta \in \Theta;\ \Psi(\theta, \phi) \leq 0\}.$$

- **Assumption S1.** $\Psi(\theta, \phi)$ is continuous in (θ, ϕ) and convex in θ $\forall \phi \in \Phi$.

Bayesian Inference of Support Function II

Study of frequentist asymptotic properties: we admit the existence of true values of the parameters: ϕ_0, η_0. We show: posterior consistency and BvM theorem.

- We first linearize $S_\phi(p)$ in ϕ and then use it to show consistency and asymptotic Normality of the posterior of $S_\phi(p)$.

- The support function $S_\phi(\cdot) : \mathbb{S}^d \to \mathbb{R}$ of the identified set $\Theta(\phi)$ is the optimal value of an *ordinary convex program*:

$$S_\phi(p) = \sup_{\theta \in \Theta} \{< p, \theta >;\ \Psi(\theta, \phi) \leq 0\}$$

and it also admits a Lagrangian representation (see Rockafellar):

$$S_\phi(p) = \sup_{\theta \in \Theta} \{< p, \theta > - \lambda(p, \phi)^T \Psi(\theta, \phi)\}, \tag{3}$$

where $\lambda(p, \phi) : \mathbb{S}^d \times \mathbb{R}^{d_\phi} \to \mathbb{R}^k_+$ are the Lagrange multipliers.

Bayesian Inference of Support Function III

- Let $B(\phi_0, \delta) = \{\phi \in \Phi;\ \|\phi - \phi_0\| \leq \delta\}$.

Assumption S2. There is $\delta > 0$ such that $\forall \phi \in B(\phi_0, \delta)$, we have:

(i) $\nabla_\phi \Psi(\theta, \phi)$ $\exists$ and is continuous in $(\theta, \phi) \in \Theta \times B(\phi_0, \delta)$;

(ii) the set $\Theta(\phi)$ is non empty;

(iii) $\exists \theta \in \Theta$ such that $\Psi(\theta, \phi) < 0$;

(iv) $\Theta(\phi) \subset int(\Theta)$;

(v) for every $i \in Act(\theta, \phi_0)$, with $\theta \in \Theta(\phi_0)$, $\nabla_\theta \Psi_i(\theta, \phi)$ $\exists$ and is continuous in $(\theta, \phi) \in \Theta \times B(\phi_0, \delta)$.

Assumption S3. The gradient vectors $\{\nabla_\theta \Psi_i(\theta, \phi)\}_{i \in Act(\theta, \phi_0)}$, are linearly independent $\forall \theta \in \Theta(\phi_0)$.

Bayesian Inference of Support Function IV

- Let $\Xi(p,\phi) = \arg\max_{\theta\in\Theta}\{< p,\theta >;\ \Psi(\theta,\phi) \leq 0\}$ be the *support set* of $\Theta(\phi)$.

Assumption S4. At least one of the following holds:

(i) for the ball $B(\phi_0,\delta)$ in S2, $\forall(p,\phi) \in \mathbb{S}^d \times B(\phi_0,\delta)$, $\Xi(p,\phi)$ is a singleton;

(ii) there are linear constraints in $\Psi(\theta,\phi_0)$ which are separable in θ, that is, $\Psi_L(\theta,\phi_0) = A_1\theta + A_2(\phi_0)$ for some function $A_2 : \Phi \to \mathbb{R}^{k_L}$ (not necessarily linear) and some $(k_L \times d)$-matrix A_1.

Bayesian Inference of Support Function V

Theorem 3:

Let $\theta_* : \mathbb{S}^d \to \Theta$ be a Borel measurable mapping satisfying $\theta_*(p) \in \Xi(p, \phi_0)$ for all $p \in \mathbb{S}^d$. Assume that:

- ϕ_0 is in the interior of Φ;
- Θ is convex and compact;

If also assumptions S1 - S4 hold with $\delta = r_n = o(1)$, then $\exists\, N \in \mathbb{N}$ such that for every $n \geq N$ there exist: *(i)* a real function $f(\phi_1, \phi_2)$ defined $\forall \phi_1, \phi_2 \in B(\phi_0, r_n)$ and *(ii)* a function $\lambda(p, \phi_0) : \mathbb{S}^d \times \mathbb{R}^{d_\phi} \to \mathbb{R}^k_+$ such that $\forall \phi_1, \phi_2 \in B(\phi_0, r_n)$:

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

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 $n \to \infty$.

▶Proof

Bayesian Inference of Support Function VI

Theorem 4: Posterior consistency

Under assumption 2 and the assumptions of theorem 3 with $r_n = \sqrt{(\log n)/n}$:

$$P\left(\sup_{p\in\mathbb{S}^d} |S_\phi(p) - S_{\phi_0}(p)| < C_S (\log n)^{1/2} n^{-1/2} \Big| D_n\right) \to^p 1$$

for some constant $C_S > 0$. ▶ Proof

- We now state a Bernstein-von Mises (BvM) theorem for $S_\phi(p)$.
- **Assumption S5.** Let $P_{\sqrt{n}(\phi-\phi_0)|D_n}$ denote the posterior distribution of $\sqrt{n}(\phi-\phi_0)$ and $||\cdot||_{TV}$ denote the total variation distance. We assume

$$||P_{\sqrt{n}(\phi-\phi_0)|D_n} - \mathcal{N}_{d_\phi}(\Delta_{n,\phi_0}, I_{\phi_0}^{-1})||_{TV} \to^p 0$$

where $\Delta_{n,\phi_0} := n^{-1/2}\sum_{i=1}^n I_{\phi_0}^{-1}\dot{l}_{\phi_0}(X_i)$, $\dot{l}_{\phi_0}$ is the semiparametric efficient sore function of the model and I_{ϕ_0} denotes the semiparametric efficient information matrix. See *e.g.* Bickel & Kleijn (2012), Rivoirard & Rousseau (2012).

Bayesian Inference of Support Function VII

Theorem 5: (BvM)

Under assumptions S5 and S6 and the assumptions of Theorem 3 with $r_n = \sqrt{(\log n)/n}$: $\forall p \in \mathbb{S}^d$

$$||P_{\sqrt{n}(S_\phi(p)-S_{\phi_0}(p))|D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} \to^p 0$$

where $\tilde{\Delta}_{n,\phi_0} = \lambda(p,\phi_0)^T \nabla_\phi \Psi(\theta_*(p),\phi_0)\Delta_{n,\phi_0}$ and

$$\tilde{I}_{\phi_0}^{-1} = \lambda(p,\phi_0)^T \nabla_\phi \Psi(\theta_*(p),\phi_0) I_{\phi_0}^{-1} \nabla_\phi \Psi(\theta_*(p),\phi_0)^T \lambda(p,\phi_0).$$

- **Assumption S6.** For some $K_1, K_2, K_3 > 0$ and $\forall \phi_1, \phi_2 \in B(\phi_0, r_n)$:
  - (i) $\sup_{p\in\mathbb{S}^d} ||\lambda(p,\phi_1) - \lambda(p,\phi_2)|| \leq K_1||\phi_1 - \phi_2||$;
  - (ii) $\sup_{\theta\in\Theta} ||\nabla_\phi \Psi(\theta,\phi_1) - \nabla_\phi \Psi(\theta,\phi_2)|| \leq K_2||\phi_1 - \phi_2||$;
  - (iii) $||\nabla_\phi \Psi(\theta_1,\phi_0) - \nabla_\phi \Psi(\theta_2,\phi_0)|| \leq K_3||\theta_1 - \theta_2||$, for every $\theta_1, \theta_2 \in \Theta$;
  - (iv) If $\Xi(p,\phi_0)$ is a singleton $\forall p \in W$ for some compact subset $W \subseteq \mathbb{S}^d$ then there exists a $\varepsilon_n = \mathcal{O}(r_n)$ such that $\Xi(p,\phi_1) \subseteq \Xi^{\varepsilon_n}(p,\phi_0)$.

Bayesian Inference of Support Function VIII

- The support function $S_\phi(\cdot)$ is a stochastic process with realizations in $\mathcal{C}(\mathbb{S}^d)$. The posterior distribution of $\sqrt{n}(S_\phi(\cdot) - S_{\phi_0}(\cdot))$ does not converge to a Gaussian measure on $\mathcal{C}(\mathbb{S}^d)$ in the total variation distance. However, a *weak* Bernstein-von Mises theorem holds with respect to the weak topology.

Theorem 5: (weak BvM)

Let $\mathbb{G}$ be a Gaussian measure on $\mathcal{C}(\mathbb{S}^d)$ with mean function $\bar{\Delta}_{n,\phi_0}(\cdot) = \lambda(\cdot,\phi_0)^T \nabla_\phi \Psi(\theta_*(\cdot),\phi_0)\Delta_{n,\phi_0}$ and covariance operator with kernel

$$\bar{I}_0^{-1}(p_1,p_2) = \lambda(p_1,\phi_0)^T \nabla_\phi \Psi(\theta_*(p_1),\phi_0) I_0^{-1} \nabla_\phi \Psi(\theta_*(p_2),\phi_0)^T \lambda(p_2,\phi_0), \qquad \forall p_1,p_2 \in \mathbb{S}^d$$

Let '$\Rightarrow$' denote weak convergence on the class of probability measures on $\mathcal{C}(\mathbb{S}^d)$. Then

$$P_{\sqrt{n}(S_\phi(\cdot)-S_{\phi_0}(\cdot))|D_n} \Rightarrow \mathbb{G}(\cdot). \tag{4}$$

Bayesian Inference of Support Function IX

Credible region for θ I

- Finite-sample Bayesian credible sets (BCS) is a set $BCS(\tau)$ such that

$$P(\theta \in \mathrm{BCS}(\tau)|D_n) = 1 - \tau, \qquad \tau \in (0,1) \tag{5}$$

- A frequentist confidence set (FCS) for θ_0 satisfies

$$\lim_{n\to\infty} \inf_{\phi\in\Phi} \inf_{\theta\in\Theta(\phi)} P_0(\theta \in \mathrm{FCS}(\tau)) \geq 1 - \tau, \qquad \tau \in (0,1).$$

- **Assumption 3.**
  (i) The $\mathrm{FCS}(\tau)$ is such that, there is $\hat{\phi}$ with $\|\hat{\phi} - \phi_0\| = o_p(1)$ satisfying $\Theta(\hat{\phi}) \subset \mathrm{FCS}(\tau)$.
  (ii) $\sup_{\theta\in\Theta} g(\theta) < \infty$ (where $\pi(\theta|\phi) \propto g(\theta) I_{\theta\in\Theta(\phi)}$).

Credible region for θ II

Theorem 6:

Under Assumptions 2 and 3, for any $\epsilon > 0$, and any $\tau > 0$,

(i) $P(\theta \in \mathrm{FCS}(\tau) | D_n) \to^p 1$;

(ii) $P(\theta \in \mathrm{FCS}(\tau), \theta \notin \mathrm{BCS}(\tau) | D_n) \to^p \tau$.

▶Proof

Two-sided credible region for $\Theta(\phi)$ I

We now construct the BCS for $\Theta(\phi)$. Aim: constructing two-sided credible sets A_1 and A_2 such that

$$P(A_1 \subset \Theta(\phi) \subset A_2 | D_n) \geq 1 - \tau \qquad w.p. \to 1.$$

The one-sided set A_2 is easy to obtain. We construct A_1 and A_2 with the help of the support function.

- Why support function can help?
- Let $\hat{\phi}_M$ be the posterior mode. Then for any $c_n \geq 0$:

$$P\left(\Theta(\hat{\phi}_M)^{-c_n} \subset \Theta(\phi) \subset \Theta(\hat{\phi}_M)^{c_n} \middle| D_n\right) = P\left(\sup_{\|p\|=1} |S_\phi(p) - S_{\hat{\phi}_M}(p)| \leq c_n \middle| D_n\right).$$

Two-sided credible region for $\Theta(\phi)$ II

- Let q_τ be the $1-\tau$ quantile of the posterior of

$$J(\phi) = \sqrt{n} \sup_{\|p\|=1} |S_\phi(p) - S_{\hat{\phi}_M}(p)|$$

so that

$$P\left(J(\phi) \leq q_\tau \middle| D_n\right) = 1 - \tau. \tag{6}$$

Theorem 6:

Suppose for any $\tau \in (0,1)$, q_τ is defined as in (6), then for every sampling sequence D_n,

$$P(\Theta(\hat{\phi}_M)^{-q_\tau/\sqrt{n}} \subset \Theta(\phi) \subset \Theta(\hat{\phi}_M)^{q_\tau/\sqrt{n}} | D_n) = 1 - \tau.$$

▶ Proof

- The BCS for θ does not have a correct frequentist coverage when θ is partially identified, since the BCS tends to be a subset of the interior of FCS.

Two-sided credible region for $\Theta(\phi)$ III

- In contrast, our two-sided BCS for the identified set has desired frequentist properties.
- **Assumption 4.** The posterior mode $\hat{\phi}_M$ is such that

$$\sqrt{n}(\hat{\phi}_M - \phi_0) \to^d N(0, I_{\phi_0}^{-1})$$

where I_{ϕ_0} denotes the semi-parametric efficient information matrix.

Theorem 7:

The constructed two-sided Bayesian credible set has asymptotically correct frequentist coverage probability, *i.e.*

$$P_0(\Theta(\hat{\phi}_M)^{-q_\tau/\sqrt{n}} \subset \Theta(\phi_0) \subset \Theta(\hat{\phi}_M)^{q_\tau/\sqrt{n}}) \geq 1 - \tau + o_p(1).$$

▶ Proof

Conclusions

- We have shown how to develop a nonparametric/ semiparametric Bayesian inference on the partially identified parameter.
- We establish frequentist asymptotic properties of our procedure (posterior consistency, BvM and BCS that are (asymptotically) FCS).
- Bayesian credible sets for θ and for $\Theta(\phi_0)$.

Semiparametric Bayesian Partially Identified Models based on Support Function

Yuan Liao [1] & Anna Simoni[2]

[1]University of Maryland

[2]CNRS and THEMA

ESEM, 26-30 August, 2013 - Gothenburg

Example 1: Interval censored data

- (Y, Y_1, Y_2) = three dimensional random vector such that

$$Y \in [Y_1, Y_2] \qquad w.p.\ 1$$

  Y = unobservable, Y_1, Y_2 = observable.

- $\theta = \mathbf{E}(Y)$;

- $\phi = (\phi_1, \phi_2)'$ with $\phi_1 = \mathbf{E}(Y_1)$and $\phi_2 = \mathbf{E}(Y_2)$.

- Identified set: $\Theta(\phi) = [\phi_1, \phi_2]$.

Example 2: Hansen-Jagannathan bounds for SDF I

- The equilibrium price P_t^i of a financial asset i is equal to

$$P_t^i = \mathbf{E}[M_{t+1} P_{t+1}^i | \mathcal{I}_t], \qquad i = 1, \dots, N$$

where M_{t+1}= stochastic discount factor (SDF) and $\mathcal{I}_t$= information set at time t. In vectorial form:

$$\iota = \mathbf{E}[M_{t+1} R_{t+1} | \mathcal{I}_t]$$

where $R_{t+1} = (r_{1,t+1}, \dots, r_{N,t+1})'$ with $r_{i,t+1} = P_{t+1}^i / P_t^i$ the gross asset return at time $(t+1)$.

- This model can be reinterpreted as a model of the SDF and may be used to detect the SDFs (*i.e.* the asset-pricing models) that are compatible with asset return data.

Example 2: Hansen-Jagannathan bounds for SDF II

- Hansen and Jagannathan (1991) show that the minimum variance $\sigma_*^2(\mu)$ achievable by a SDF with mean μ and compatible with the observed (m, Σ) is given by

$$\begin{aligned} \sigma_*^2(\mu) &= (1-\mu m)'\Sigma^{-1}(1-\mu m) =: \phi_1\mu^2 - 2\phi_2\mu + \phi_3 \\ \text{with } \phi_1 &= m'\Sigma^{-1}m, \qquad \phi_2 = m'\Sigma^{-1}\iota, \qquad \phi_3 = \iota'\Sigma^{-1}\iota \end{aligned}$$

  and $m = \mathbf{E}(R_{t+1})'$, $\Sigma = \mathbf{E}(R_{t+1}-m)(R_{t+1}-m)'$, $\mu = \mathbf{E}(M_{t+1})$ and $\sigma^2 = Var(M_{t+1})$ (constant over time).

- **Identified set**:

$$\Theta(\phi) = \left\{(\mu, \sigma^2) \in \Theta;\ \sigma_*^2(\mu) - \sigma^2 \leq 0\right\} \tag{7}$$

  where $\phi = (\phi_1, \phi_2, \phi_3)'$.

- If there is a non-risky asset then μ is fixed and the lower bound is just a point.

Example 1: Interval censored data

- Consider the simpler setting: $Y_2 = Y_1 + 1$. Therefore,

$$EY_1 \leq EY \leq EY_1 + 1,$$

where only Y_1 is observable, *i.e.* $Y_1 \equiv X \sim F$.

- Let $\phi = EY_1$ and $\theta = EY$, then

$$\Theta(\phi) = [\phi, \phi + 1].$$

- Dirichlet process prior for F:

$$\pi(F) = \mathcal{D}ir(\nu_0, Q_0),$$

where $\nu_0 \in \mathbb{R}_+$ and Q_0 is a base probability on $(\mathcal{X}, \mathfrak{B}_x)$ such that $Q_0(x) = 0, \forall x \in (\mathcal{X}, \mathfrak{B}_x)$.

- Induced prior on $\phi = \phi(F)$:

$$\pi(\phi \in A) = P\left(\sum_{j=1}^{\infty} \alpha_j \xi_j \in A\right), \quad \forall A \subset \Phi$$

where (see Sethuraman 1994 SS):

- $\xi_j \sim i.i.d. Q_0$, for $j \geq 1$,
- $\alpha_j = v_j \prod_{l=1}^{j}(1 - v_l)$ with $v_l \sim i.i.d. \mathcal{B}e(1, \nu_0)$, for $l \geq 1$
- $\{v_l\}_{l \geq 1}$ are independent of $\{\xi_j\}_{j \geq 1}$.

◄

Example 1: Interval censored data. We reformulate the model as

$$\begin{aligned} Y_1 = \phi_1 + u, \qquad & Y_2 = \phi_2 + v \\ u \sim f_1, \qquad & v \sim f_2, \qquad \mathbf{E}^{f_1}(u) = 0,\ \mathbf{E}^{f_2}(v) = 0 \\ & u \perp\!\!\!\perp v | f_1, f_2 \text{ and disjoint supports.} \end{aligned}$$

- Therefore, $\eta = (f_1, f_2)$, $X = (Y_1, Y_2)|\phi, \eta \sim F_{\phi,\eta}$ and the likelihood function is

$$l_n(\phi, \eta) = \prod_{i=1}^{n} f_1(Y_{1i} - \phi_1) f_2(Y_{2i} - \phi_2).$$

- We put priors on (ϕ, f_1, f_2):

$$\pi(\theta, \phi, \eta) = \pi(\theta|\phi) \times \pi(\phi, \eta) = \pi(\theta|\phi) \times \pi(\phi) \times \pi(\eta). \qquad (8)$$

This is the location model, see Ghosal et al. (1999) and Amewou-Atisso et al. (2003).

Examples of priors on a density function η:

- mixture of Dirichlet process priors;
- Gaussian process priors (Lenk, 1991; Van der Vaart & Van Zanten, 2008);
- Finite mixture of Normals:

$$\eta(x) = \sum_{i=1}^{k} w_i \phi(x - \mu_i; \Sigma_i);$$

  such that $\sum_{i=1}^{k} w_i \mu_i = 0$.
- Dirichlet mixture of Normals: if $H \sim \mathcal{D}ir(\nu_0, Q_0)$

$$\eta(x) = \int \phi(x - z; \Sigma) dH(z)$$

  where ϕ is a standard Normal density. H must be such that $\int zH(z)dz = 0$. We may also place a prior on Σ independent of H.

- Random Bernstein polynomial

$$\eta(x) = \sum_{j=1}^{k} [H(j/k) - H((j-1)/k)]\mathcal{B}e(x; j, k-j+1).$$

where $\mathcal{B}e(x; a, b)$ is the beta density, $H \sim \mathcal{D}ir(\nu_0, Q_0)$ and $H \perp\!\!\!\perp k$ is independent of the prior on k. Then

$$p(\phi|D_n) \propto \int \pi(\phi) \prod_{i=1}^{n} \eta(X_i - \phi)\pi(H)\pi(k)dHdk.$$

- Other examples: wavelet expansions, Polya tree priors (Lavine (1992), etc.

Example 1: Interval censored data.

- $\Psi(\theta, \phi) = (\theta - \phi_2, \phi_1 - \theta)^T$;
- for any $\phi = (\phi_1, \phi_2)$ and $\tilde{\phi} = (\tilde{\phi}_1, \tilde{\phi}_2)$,

$$\|\Psi(\theta, \phi) - \Psi(\theta, \tilde{\phi})\| = \|\phi - \tilde{\phi}\|.$$

  This verifies assumption H2.

- Moreover, for any θ such that $d(\theta, \Theta(\phi)) \geq a_n$, either

$$\theta \leq \phi_1 - a_n$$

  or

$$\theta \geq \phi_2 + a_n.$$

  If $\theta \leq \phi_1 - a_n$, then $\Psi_2(\theta, \phi) = \phi_1 - \theta \geq a_n$; if $\theta \geq \phi_2 + a_n$, then $\Psi_1(\theta, \phi) = \theta - \phi_2 \geq a_n$. This verifies assumption H3.

Proof of Theorem 1 (sketch)

- $\pi(\theta|\phi)$ is a regular conditional distribution $\rightarrow \exists$ a transition probability from $(\Phi, \mathfrak{B}_\phi)$ to $(\Theta, \mathfrak{B}_\theta)$ that characterizes it $\rightarrow \pi(\Theta_0(\phi_0)^\epsilon|\phi)$ is a measurable function of ϕ.

- $\pi(\Theta_0(\phi_0)^\epsilon|\phi) = 0, \forall \phi \notin A$.

- since $\forall \phi \in \Phi$, $|\pi(\Theta_0(\phi_0)^\epsilon|\phi)| \leq 1$, by the the Lusin's theorem $\exists\, h_m \in \mathcal{C}(\Phi)$, $|h_m| \leq 1$ such that

$$\pi(\Theta_0(\phi_0)^\epsilon|\phi) = \lim_{m\to\infty} h_m(\phi), \qquad \pi_\phi - a.s.$$

- Now,

$$\begin{aligned}
P(\theta \in \Theta(\phi_0)^\epsilon|D_n) &= \int_\Phi \pi(\theta \in \Theta(\phi_0)^\epsilon|\phi)\pi(\phi|D_n)d\phi \\
&= \int_\Phi \lim_{m\to\infty} h_m(\phi)\pi(\phi|D_n)d\phi \\
&= \lim_{m\to\infty} \int_\Phi h_m(\phi)\pi(\phi|D_n)d\phi, \quad \text{by D.C.T.}
\end{aligned}$$

$$= \lim_{m\to\infty} \int_{\mathcal{F}} h_m(\phi(F))\pi(F|D_n)dF.$$

- Since ϕ is a continuous function of F, $h_m \circ \phi$ is a continuous and bounded function of F and under assumption 1:

$$\begin{aligned}
\lim_{n\to\infty} P(\theta \in \Theta(\phi_0)^{\epsilon}|D_n) &= \lim_{n\to\infty}\lim_{m\to\infty} \int_{\mathcal{F}} h_m(\phi(F))\pi(F|D_n)dF \\
&= \lim_{m\to\infty} \int_{\mathcal{F}} h_m(\phi(F))\delta_{F_0}(dF) \\
&= \lim_{m\to\infty} h_m(\phi(F_0)) = \pi(\theta \in \Theta(\phi_0)^{\epsilon}|\phi(F_0)) \\
&= 1, \quad F_0 - a.s.
\end{aligned}$$

Q.E.D.

Posterior concentration for ϕ: lower level assumptions

- Let $E = E_{\eta_0,\phi_0}$ and $U = \{\phi; ||\phi - \phi_0|| \leq Mr_n\}$ for $M > 0$ and $r_n = n^{-1/2}(\log n)^{1/2}$.
- Then it suffices to show that for some $M > 0$, $EP(\phi \in U^c | D_n) = o(1)$.
- Existence of test functions:

  Assumption A1: for all n large enough, $\mathcal{N}(n^{-1/2}(\log n)^{1/2}, G, \|.\|_G) \leq n$ where
  $$G = \{l(\cdot; \phi, \eta) : \phi \in \Phi, \eta \in \mathcal{P}\}$$
  .

Lemma A.1:

Under Assumption A1, there exists a test T and a constant $L > 4$ and $L \geq M + 2$ (for M defined in Assumption A2) such that
(i)
$$ET = o(1)$$
(ii) for $r_n = \sqrt{(\log n)/n}$,
$$\sup_{\eta \in \mathcal{P}, \|\phi - \phi_0\| > Lr_n} E_{\phi,\eta}(1 - T) \leq \exp\left(-\frac{9}{16}L^2 nr_n^2\right).$$

- Therefore,

$$\begin{aligned} EP(\phi \in U^c|D_n) &= E[P(\phi \in U^c|D_n)T] + E[P(\phi \in U^c|D_n)(1-T)] \\ &\leq ET + EP(\phi \in U^c|D_n)(1-T) = o(1) + EP(\phi \in U^c|D_n)(1-T) \\ &= o(1) + E[P(\phi \in U^c|D_n)(1-T)I_A] + E[P(\phi \in U^c|D_n)(1-T)I_{A^c}] \\ &\leq EP(\phi \in U^c|D_n)(1-T)I_A + o(1) \end{aligned}$$

where

- $A := \left\{ \int\int \frac{l_n(\phi,\eta)}{l_n(\phi_0,\eta_0)} \pi(\phi,\eta) d\phi d\eta \geq \beta_n \right\}$,
- $\beta_n := \frac{1}{2n^2} \pi(K_{\phi,\eta} \leq \log n/n, V_{\phi,\eta} \leq \log n/n)$ and
- it can be shown that $P(A) \to 1$ if the prior $\pi(\phi,\eta)$ satisfies (**Assumption A2**):

$$\pi\left(K_{\phi,\eta} \leq \frac{\log n}{n}, \quad V_{\phi,\eta} \leq \frac{\log n}{n} \right) n^M \to \infty$$

for some $M > 2$.

- Finally, we need to lower bound the denominator of the posterior probability, and upper bound the numerator as well.
- Then

$$E[P(\phi \in U^c|D_n)(1-T)I_A] \leq \frac{1}{\beta_n} E\{\iint_{U^c \times \mathcal{P}} \prod_{i=1}^{n} \frac{l(X_i;\phi,\eta)}{l(X_i;\phi_0,\eta_0)} \pi(d\eta, d\phi)(1-T)\}$$

$$= \beta_n^{-1} \iiint_{X \times U^c \times \mathcal{P}} \prod_{i=1}^{n} l(X_i;\phi,\eta)(1-T)\pi(d\eta,d\phi)dX_1...dX_n$$

$$= \beta_n^{-1} \iint_{U^c \times \mathcal{P}} E_{\phi,\eta}(1-T)\pi(d\eta,d\phi)$$

$$\leq \beta_n^{-1}\pi(\phi \in U^c) \sup_{\phi \in U^c, \eta \in \mathcal{P}} E_{\phi,\eta}(1-T) \leq \exp(-Lnr_n^2)\beta_n^{-1} = o(1).$$

Proof of Theorem 2 (sketch)

Let us define $r_n = \sqrt{\frac{\log n}{n}}$, $A = \{\Theta(\phi_0) \subset \Theta(\phi)^{Lr_n} \text{ and } \Theta(\phi) \subset \Theta(\phi_0)^{Kr_n}\}$ and $C \geq \max\{L, K\}$ for some $K, L > 0$ which satisfy Lemmas C.4 and C.5. Then,

$$
\begin{aligned}
P\Big(d_H(\Theta(\phi),\Theta(\phi_0)) \leq Cr_n|D_n\Big) &= P\Big(\{d_H(\Theta(\phi),\Theta(\phi_0)) \leq Cr_n\} \cap A|D_n\Big) \\
&\quad + P\Big(\{d_H(\Theta(\phi),\Theta(\phi_0)) \leq Cr_n\} \cap A^c|D_n\Big) \\
&\leq P\Big(d_H(\Theta(\phi),\Theta(\phi_0)) \leq Cr_n|A, D_n\Big)P(A|D_n) \\
&\quad + P\Big(\{d_H(\Theta(\phi),\Theta(\phi_0)) \leq Cr_n\}|A^c, D_n\Big)P(A^c|D_n) \\
&\leq P\Big(\max\{L,K\}r_n \leq Cr_n|D_n\Big)P(A|D_n) + o_p(1) \to^p 1
\end{aligned}
$$

if

- $P(\{\Theta(\phi_0) \subset \Theta(\phi)^{Lr_n} \text{ and } \Theta(\phi) \subset \Theta(\phi_0)^{Kr_n}\}|D_n) \to^p 1$ (see Lemmas C.4 and C.5) and
- $d_H(\Theta(\phi),\Theta(\phi_0)) \leq \max\{L,K\}r_n$ on A (see Lemma C.6).

Define $Q(\theta,\phi) = \|\max(\Psi(\theta,\phi),0)\| = \left[\sum_{i=1}^{k}(\max(\Psi_i(\theta,\phi),0))^2\right]^{1/2}$.

Lemma C.4:

$\exists$ a constant $K > 0$ so that

$$P(\Theta(\phi) \subset \Theta(\phi_0)^{Kr_n}|D_n) \to^p 1,$$

where $\Theta(\phi_0)^{Kr_n} = \{\theta \in \Theta : d(\theta, \Theta(\phi_0)) \leq Kr_n\}$, and $P(.|D_n)$ denotes the marginal posterior probability of ϕ.

Proof. For any $K > 0$, let $\Theta \backslash \Theta(\phi_0)^{Kr_n} = \{\theta \in \Theta : d(\theta, \Theta(\phi_0)) > Kr_n\}$.

- It suffices to show that $\exists$ $K > 0$ such that

$$P\left(\inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Cr_n}} Q(\theta,\phi) > \sup_{\theta\in\Theta(\phi)} Q(\theta,\phi)\middle|D_n\right) \to^p 1. \tag{9}$$

- Note that $\sup_{\theta\in\Theta(\phi)} Q(\theta,\phi) = 0$, since $\forall\theta \in \Theta(\phi)$, $\Psi(\theta,\phi) \leq 0$, which is equivalent to $Q(\theta,\phi) = 0$.

- We have $P(\|\phi-\phi_0\|<r_n|D_n)\to^p 1$. Therefore, it remains to show

$$P\left(\inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Kr_n}} Q(\theta,\phi)>0\middle|D_n\right)\to^p 1. \tag{10}$$

- In fact, for any ϕ so that $\|\phi-\phi_0\|\leq r_n$, by Lemma C.1, $\exists\,\tilde{K}>0$ such that for any $K>0$,

$$\begin{aligned}\inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Kr_n}} Q(\theta,\phi) &\geq \inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Kr_n}} Q(\theta,\phi_0)-\sup_{\theta\in\Theta}|Q(\theta,\phi)-Q(\theta,\phi_0)|\\ &\geq \inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Kr_n}} Q(\theta,\phi_0)-\tilde{K}r_n. \end{aligned} \tag{11}$$

- By Lemma C.3, $\exists\,\tilde{K}>0$ such that

$$\inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Kr_n}} Q(\theta,\phi_0)=\inf_{d(\theta,\Theta(\phi_0))\geq Kr_n} Q(\theta,\phi_0)\geq 3\tilde{K}r_n.$$

- Hence we have shown that whenever $\|\phi-\phi_0\|\leq r_n$,

$$\inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Kr_n}} Q(\theta,\phi)\geq 2\tilde{K}r_n>0.$$

- Therefore, by the posterior concentration for ϕ, (10) holds from

$$P\left(\inf_{\theta\in\Theta\backslash\Theta(\phi_0)^{Cr_n}} Q(\theta,\phi)>0\middle|D_n\right)\geq P(\|\phi-\phi_0\|\leq r_n|D_n)\to^p 1.$$

Lemma C.5:

There exists $L > 0$ so that

$$P(\Theta(\phi_0) \subset \Theta(\phi)^{Lr_n} | D_n) \to^p 1.$$

Proof. By Lemma C.1, $\exists \tilde{L} > 0$ such that whenever $\|\phi - \phi_0\| \leq r_n$,

$$\sup_{\theta \in \Theta} |Q(\theta, \phi) - Q(\theta, \phi_0)| \leq \tilde{L} r_n.$$

- Now, fix such a ϕ, then for all large enough n, $\phi \in U(\phi_0)$ where $U(\phi_0)$ is the neighborhood satisfying Lemma C.3. For such a $\tilde{L}$, by Lemma C.3, $\exists L > 0$ that does not depend on ϕ such that

  $$\inf_{d(\theta, \Theta(\phi)) \geq Lr_n} Q(\theta, \phi) > \tilde{L} r_n,$$

  which then implies that $\{\theta : Q(\theta, \phi) \leq \tilde{L} r_n\} \subset \{\theta : d(\theta, \Theta(\phi)) \leq Lr_n\}$.
- On the other hand, for any $\theta \in \Theta(\phi_0)$, $Q(\theta, \phi_0) = 0$, which implies

  $$Q(\theta, \phi) \leq 0 + |Q(\theta, \phi) - Q(\theta, \phi_0)| \leq \tilde{L} r_n.$$

  Therefore, $\Theta(\phi_0) \subset \{\theta : Q(\theta, \phi) \leq \tilde{L} r_n\} \subset \{\theta : d(\theta, \Theta(\phi)) \leq Lr_n\}$.
- Hence we have in fact shown that, the event $\|\phi - \phi_0\| \leq r_n$ implies the event $\Theta(\phi_0) \subset \Theta(\phi)^{Lr_n}$.

- Moreover, the event $\|\phi - \phi_0\| \leq r_n$ occurs with probability approaching one under the posterior distribution of ϕ, which then implies the result.

Q.E.D.

Lemma C.6:

For two sets A, B, if $A \subset B^{r_1}$ and $B \subset A^{r_2}$ for some r_1, r_2, then

$$d_H(A, B) \leq \max\{r_1, r_2\}.$$

Proof. $d_H(A, B) = \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\}$. Then $\forall a \in A$, since $A \subset B^{r_1}$, $a \in B^{r_1}$, that is $d(a, B) \leq r_1$. This implies $\sup_{a \in A} d(a, B) \leq r_1$. Similarly we can show $\sup_{b \in B} d(b, A) \leq r_2$. Q.E.D.

Lemma C.1:

There exists $C > 0$ such that for any $\phi_1, \phi_2 \in \Phi$,

$$\sup_{\theta \in \Theta} |Q(\theta, \phi_1) - Q(\theta, \phi_2)| \leq C\|\phi_1 - \phi_2\|.$$

Proof. For any $\phi_1, \phi_2 \in \Phi$,

$$\begin{aligned}
& |Q(\theta, \phi_1) - Q(\theta, \phi_2)| = \Big| \|\max(\Psi(\theta, \phi_1), 0)\| - \|\max(\Psi(\theta, \phi_2), 0)\| \Big| \qquad (12) \\
\leq \quad & \|\max(\Psi(\theta, \phi_1), 0) - \max(\Psi(\theta, \phi_2), 0)\| \\
= \quad & \left(\sum_{i=1}^{d} [\max(\Psi_i(\theta, \phi_1), 0) - \max(\Psi_i(\theta, \phi_2), 0)]^2 \right)^{1/2} \\
= \quad & \left(\sum_{i=1}^{d} [f(\Psi_i(\theta, \phi_1)) - f(\Psi_i(\theta, \phi_2))]^2 \right)^{1/2} \leq \left(\sum_{i=1}^{d} [\Psi_i(\theta, \phi_1) - \Psi_i(\theta, \phi_2)]^2 \right)^{1/2} \\
= \quad & \|\Psi\theta, \phi_1) - \Psi(\theta, \phi_2)\| \leq C\|\phi_1 - \phi_2\| \qquad \text{by H2.} \qquad (13)
\end{aligned}$$

Q.E.D.

Lemma C.2:

$\exists$ a closed neighborhood $U(\phi_0)$, for any $a_n = O(1)$, there exists $K > 0$ that does not depend on ϕ, so that

$$\inf_{\phi \in U(\phi_0)} \inf_{d(\theta,\Theta(\phi)) \geq Ka_n} \max_{i \leq k} \Psi_i(\theta, \phi) > a_n.$$

Proof. For any $C > 0$, define

$$A_C = \{\phi \in U(\phi_0) : \inf_{\theta : d(\theta,\Theta(\phi)) \geq Ca_n} \max_{i \leq k} \Psi_i(\theta, \phi) > a_n\}.$$

- By assumption H3, $\forall \phi \in U(\phi_0)$, there exists $C_\phi > 0$ so that $\phi \in A_{C_\phi}$. Thus, $U(\phi_0) \subset \cup_{\phi \in U(\phi_0)} A_{C_\phi}$.
- Since $U(\phi_0)$ is compact $\exists$ constants $C_1, ..., C_N$ for some finite $N > 0$ to form a finite cover so that

  $$U(\phi_0) \subset \bigcup_{i=1}^{N} A_{C_i}.$$

  Then $\forall \phi \in U(\phi_0)$, there exists $j \leq N$ so that $\phi \in A_{C_j}$, that is

  $$\inf_{\theta : d(\theta,\Theta(\phi)) \geq C_j a_n} \max_{i \leq d} \Psi_i(\theta, \phi) > a_n.$$

- Let $K = \max\{C_i : i \leq N\}$, then

$$\inf_{\theta: d(\theta, \Theta(\phi)) \geq K a_n} \max_{i \leq k} \Psi_i(\theta, \phi) \geq \inf_{\theta: d(\theta, \Theta(\phi)) \geq C_j a_n} \max_{i \leq k} \Psi_i(\theta, \phi) > a_n.$$

  This is true for any $\phi \in U(\phi_0)$.

Q.E.D.

Lemma C.3:

For any $M > 0$, $\exists\, \delta > 0$, and a neighborhood $U(\phi_0)$ so that

$$\inf_{\phi \in U(\phi_0)} \inf_{d(\theta,\Theta(\phi)) \geq \delta\sqrt{(\log n)/n}} Q(\theta, \phi) > M\sqrt{\frac{\log n}{n}}.$$

Proof. For any $M > 0$, by Lemma C.2, $\exists\, U(\phi_0)$ and $\delta > 0$ so that

$$\inf_{\phi \in U(\phi_0)} \inf_{d(\theta,\Theta(\phi)) \geq \delta\sqrt{\log n/n}} \max_{i \leq d} \Psi_i(\theta, \phi) > M\sqrt{\frac{\log n}{n}}. \tag{14}$$

- Now, for any $(\theta, \phi) \in \left\{(\theta, \phi) \in \Theta \times U(\phi_0) : d(\theta, \Theta(\phi)) \geq \delta\sqrt{\log n/n}\right\}$, we have

  $$\max_{i \leq k} \Psi_i(\theta, \phi) > 0$$

  since $\theta \notin \Theta(\phi)$, which then implies that

  $$\max_{i \leq k} \Psi_i(\theta, \phi) = \max_{i \leq k} \Psi_i(\theta, \phi) I(\Psi_i(\theta, \phi) > 0).$$

- Let $\Psi_i = \Psi_i(\theta, \phi)$, and $\Psi = (\Psi_1, ..., \Psi_k)^T$. Then using the fact that $\max_i A_i^2 = (\max_i A_i)^2$ if $A_i \geq 0$, we have,

$$
\begin{aligned}
Q(\theta, \phi) &= \|\max(\Psi, 0)\| = \left(\sum_{i=1}^{k}[\max(\Psi_i, 0)]^2\right)^{1/2} &(15)\\
&\geq \left(\max_{i\leq k}[\max(\Psi_i, 0)]^2\right)^{1/2} = \left([\max_{i\leq k}\max(\Psi_i, 0)]^2\right)^{1/2} &(16)\\
&= \max_{i\leq k}\max(\Psi_i, 0) = \max_{i\leq k}\Psi_i I(\Psi_i \geq 0) = \max_{i\leq k}\Psi_i(\theta, \phi). &(17)
\end{aligned}
$$

The result follows immediately from (14).

Q.E.D.

Proof of Lemma 1 (sketch)

- $\forall \tau \in [0,1]$ define $\phi_\tau := \tau\phi_1 + (1-\tau)\phi_2$ with $\phi_2 = \phi_\tau|_{\tau=0}$ and $\phi_1 = \phi_\tau|_{\tau=1}$.
- Under Assumption S3 $S_{\phi_\tau}(p)$ is differentiable at $\tau = \tau_0 \in (0,1)$. By the mean value theorem:

$$S_{\phi_1}(p) - S_{\phi_2}(p) = \frac{\partial}{\partial\tau} S_{\phi_\tau}(p)\Big|_{\tau=\tau_0\in(0,1)}. \tag{18}$$

- Define $\tau_0 : \mathbb{S}^d \to (0,1)$ a measurable and differentiable function of p. We prove that

$$\begin{aligned}\frac{\partial}{\partial\tau} S_{\phi_\tau}(p)\Big|_{\tau=\tau_0(p)} &= \frac{dS_{\phi_\tau}(p)}{d\tau+}\Big|_{\tau=\tau_0(p)} = \frac{dS_{\phi_\tau}(p)}{d\tau-}\Big|_{\tau=\tau_0(p)} \\ &= \lambda(p, \phi_{\tau_0(p)})^T \nabla_\phi \Psi(\tilde{\theta}(p), \phi_{\tau_0(p)})[\phi_1 - \phi_2]\end{aligned}$$

for some $\tilde{\theta}(p) \in \Xi(p, \phi_{\tau_0(p)})$, $p \in \mathbb{S}^d$.

Proof of Lemma 1 (sketch)

- Therefore,

$$
\begin{aligned}
S_{\phi_1}(p) - S_{\phi_2}(p) = &\lambda(p, \phi_{\tau_0(p)})^T \nabla_\phi \Psi(\tilde{\theta}(p), \phi_{\tau_0(p)})[\phi_1 - \phi_2] \\
= &\lambda(p, \phi_0)^T \nabla_\phi \Psi(\theta_*(p), \phi_0)[\phi_1 - \phi_2] \\
&+ \left(\lambda(p, \phi_{\tau_0(p)})^T \nabla_\phi \Psi(\tilde{\theta}(p), \phi_{\tau_0(p)}) - \lambda(p, \phi_0)^T \nabla_\phi \Psi(\theta_*(p), \phi_0) \right) [\phi_1 - \phi_2]
\end{aligned}
$$

where $\theta_* : \mathbb{S}^d \to \Theta$ is a Borel measurable mapping s.t. $\theta_*(p) \in \Xi(p, \phi_0)$.

- We show:

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

- The functions $\lambda(p, \phi_0)$ and $\nabla_\phi \Psi(\theta_*(p), \phi_0)$ are uniformly bounded in p, then:

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

Proof of Theorem 3 (sketch)

- Denote $r_n = (\log n)^{1/2} n^{-1/2}$ and $\Omega = \{\phi \in B(\phi_0, r_n)\}$.
- Under assumption 2:

$$
\begin{aligned}
P\Big(\sup_{p\in\mathbb{S}^d} |S_\phi(p) - S_{\phi_0}(p)| \geq Cr_n \Big| D_n\Big) &= P\Big(\{\sup_{p\in\mathbb{S}^d} |S_\phi(p) - S_{\phi_0}(p)| \geq Cr_n\} \cap \Omega \Big| D_n\Big) \\
&\quad + P\Big(\sup_{p\in\mathbb{S}^d} |S_\phi(p) - S_{\phi_0}(p)| \geq Cr_n \cap \Omega^c \Big| D_n\Big) \\
&\leq P\Big(\sup_{p\in\mathbb{S}^d} |S_\phi(p) - S_{\phi_0}(p)| \geq Cr_n \cap \Omega \Big| D_n\Big) + P(\Omega^c|D_n) \\
&\leq P\Big(o(||\phi-\phi_0||) + \sup_{p\in\mathbb{S}^d} |\lambda(p,\phi_0)'\nabla_\phi \Psi(\theta_*(p),\phi_0)[\phi-\phi_0]| \geq Cr_n \cap \Omega \Big| D_n\Big) + o_p(
\\
&\leq P\Big(o(||\phi-\phi_0||) + \sup_{p\in\mathbb{S}^d} |\lambda(p,\phi_0)'\nabla_\phi \Psi(\theta_*(p),\phi_0)|||\phi-\phi_0|| \geq Cr_n \Big| D_n\Big) P(\Omega|D_n) +
\end{aligned}
$$

which converges to 0 in probability.

Proof of Theorem 4 (sketch)

- Denote $r_n = (\log n)^{1/2} n^{-1/2}$, $\Omega := \{\phi \in B(\phi_0, r_n)\}$ and $h_n := \sqrt{n}\sup_{p\in\mathbb{S}^d}(S_\phi(p) - S_{\phi_0}(p))$.
- Since the $||\cdot||_{TV} \leq 2$:

$$
\begin{aligned}
\mathbf{E}||P_{h_n|D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} &= \mathbf{E}||P_{h_n|D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} I_\Omega + \\
&\quad \mathbf{E}||P_{h_n|D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} I_{\Omega^c} \\
&\leq \mathbf{E}||P_{h_n|D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} I_\Omega + 2P(\Omega^c).
\end{aligned}
$$

- Under assumption 2, $P(\Omega^c) = o(1)$ so that, under assumption S6:

$$
\begin{aligned}
&\mathbf{E}||P_{h_n|D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} = \\
&\qquad \mathbf{E}||P_{\sqrt{n}\sup_{p\in\mathbb{S}^d}|\lambda(p,\phi_0)'\nabla_\phi\Psi(\theta_*(p),\phi_0)[\phi-\phi_0]||D_n} - \mathcal{N}(\tilde{\Delta}_{n,\phi_0}, \tilde{I}_{\phi_0}^{-1})||_{TV} I_\Omega + o(1)
\end{aligned}
$$

which converges to 0 under assumption S4.

Q.E.D.

Proof of Theorem 5 (sketch)

Since $\Theta(\hat{\phi}) \subset \text{FCS}(\tau)$, we have:

- *Part (i).* $P(\theta \notin \text{FCS}(\tau)|D_n) \leq P(\theta \notin \Theta(\hat{\phi})|D_n)$.
- *Part (ii).* $P(\text{FCS}(\tau)\backslash\text{BCS}(\tau)|D_n)$ is lower bounded by

$$\geq P(\Theta(\hat{\phi})\backslash\text{BCS}(\tau)|D_n) \geq P(\theta \in \Theta(\hat{\phi})|D_n) - P(\theta \in \text{BCS}(\tau)|D_n) \to^p \tau.$$

We then have to show that $P(\theta \notin \Theta(\hat{\phi})|D_n) = o_p(1)$.

$$\begin{aligned} P(\theta \notin \Theta(\hat{\phi})|D_n) &= \int \pi(\theta \notin \Theta(\hat{\phi})|\phi)p(\phi|D_n)d\phi \leq \int \pi(\theta \notin \Theta(\phi_0)|\phi)p(\phi|D_n)d\phi \\ &+ \int \Big|\pi(\theta \notin \Theta(\hat{\phi})|\phi) - \pi(\theta \notin \Theta(\phi_0)|\phi)\Big| p(\phi|D_n)d\phi. \end{aligned}$$

The result follows by the posterior concentration of ϕ (at the rate r_n) and by the asymptotic expansion of $S_\phi(\cdot)$ since $||\phi - \phi_0|| \leq r_n$ implies

$$P\left(d_H(\Theta(\phi), \Theta(\phi_0)) \leq Cr_n|D_n\right) \to 1, \qquad \text{for some } C > 0.$$

Hence, $P\left(\Theta(\phi) \subset \Theta(\phi_0)^{Cr_n}\right) \to 1$.

Proof of Theorem 6 (sketch)

By definition of q_τ:

$$P\left(\Theta(\hat{\phi}_M)^{-q_\tau/\sqrt{n}} \subset \Theta(\phi) \subset \Theta(\hat{\phi}_M)^{q_\tau/\sqrt{n}} \middle| D_n\right)$$
$$= P\left(\sup_{||p||=1} |S_\phi(p) - S_{\hat{\phi}_M}(p)| \leq \frac{q_\tau}{\sqrt{n}} \middle| D_n\right) = 1 - \tau.$$

Q.E.D. ◄

Proof of Theorem 7 (sketch)

- We first show that for any $q \geq 0$,

$$P(\sqrt{n} \sup_{\|p\|=1} |S_\phi(p) - S_{\hat{\phi}_M}(p)| \leq x | D_n) - P_0(\sqrt{n} \sup_{\|p\|=1} |S_{\phi_0}(p) - S_{\hat{\phi}_M}(p)| \leq x) = o_p(1).$$

- This implies

$$P_0(\Theta(\hat{\phi}_M)^{-q_\tau/\sqrt{n}} \subset \Theta(\phi_0) \subset \Theta(\hat{\phi}_M)^{q_\tau/\sqrt{n}}) = P_0(\sqrt{n} \sup_{\|p\|=1} |S_{\phi_0}(p) - S_{\hat{\phi}_M}(p)| \leq q_\tau)$$

$$\geq P\left(\sqrt{n} \sup_{\|p\|=1} |S_{\phi_0}(p) - S_{\hat{\phi}_M}(p)| \leq q_\tau \middle| D_n\right)$$

$$-\left|P_0(\sqrt{n} \sup_{\|p\|=1} |S_{\phi_0}(p) - S_{\hat{\phi}_M}(p)| \leq q_\tau) - P\left(\sqrt{n} \sup_{\|p\|=1} |S_{\phi_0}(p) - S_{\hat{\phi}_M}(p)| \leq q_\tau \middle| D_n\right)\right|$$

$$= P(J(\phi) \leq q_\tau | D_n) + o_p(1) = 1 - \tau + o_p(1),$$