Journal of Econometrics 211 (2019) 338–360

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bayesian inference for partially identified smooth convex models

Yuan Liao[a], Anna Simoni[b,*]

[a] Rutgers University, New Brunswick, NJ, 08901, USA
[b] CREST, CNRS, École Polytechnique, ENSAE, 5, Avenue Henry Le Chatelier, 91120 Palaiseau, France

A R T I C L E I N F O

Article history:
Received 8 June 2016
Received in revised form 28 January 2019
Accepted 4 March 2019
Available online 19 March 2019

JEL classification:
C10
C11
C13

Keywords:
Partial identification
Bayesian credible sets
Bernstein–von Mises theorem
Support function
Moment inequality models

A B S T R A C T

This paper proposes novel Bayesian procedures for partially identified models when the identified set is convex with a smooth boundary, whose support function is locally smooth with respect to the data distribution. Using the posterior of the identified set, we construct Bayesian credible sets for the identified set, the partially identified parameter and their scalar transformations. These constructions, based on the support function, benefit from several computationally attractive algorithms when the identified set is convex, and are proved to have valid large sample frequentist coverages. These results are based on a local linear expansion of the support function of the identified set. We provide primitive conditions to verify such an expansion.

1. Introduction

Bayesian partial identification has emerged as an important area of econometrics. In this paper, we propose a new Bayesian framework for set inferences with a focus on the asymptotic properties of Bayesian credible sets (BCS) for partially identified models. Generally speaking, the BCS is a set in the support of the posterior distribution such that the object of interest lies inside it with a high posterior probability. Usual methods for constructing BCS, such as the highest-posterior-density, would fail, due to the lack of a clear definition of the "posterior density of a set". The problem is even more challenging when the set of interest belongs to a multi-dimensional space.

We focus on convex identified sets with a smooth boundary, which means the identified sets should be characterized by *support functions* that are smooth with respect to the data generating process. Suppose the identified set of θ can be parameterized by a point-identified nuisance parameter ϕ, so it can be denoted by $\Theta(\phi)$. Its support function is defined as

$$S_\phi(\nu) := \sup_{\theta\in\Theta(\phi)} \theta^T\nu, \quad \|\nu\| = 1.$$

In addition to the convexity of the identified set, our main requirement is that the support function is locally smooth with respect to the data distribution. More specifically, we require the following "local linear approximation" (LLA) should

* Corresponding author.
E-mail addresses: yuan.liao@rutgers.edu (Y. Liao), anna.simoni@polytechnique.edu (A. Simoni).

https://doi.org/10.1016/j.jeconom.2019.03.001
0304-4076/

be satisfied: there is a shrinking neighborhood B of the true value ϕ_0, and a continuous vector function $A(\nu)$ such that uniformly in $\phi_1, \phi_2 \in B$,

$$\sup_{\nu \in \mathbb{S}^d} \left| \left(S_{\phi_1}(\nu) - S_{\phi_2}(\nu) \right) - A(\nu)^T (\phi_1 - \phi_2) \right| = o(\|\phi_1 - \phi_2\|), \quad \text{as } \|\phi_1 - \phi_2\| \to 0 \tag{1.1}$$

where $\mathbb{S}^d$ denotes the unit sphere in $\mathbb{R}^d$. To understand the intuition from the frequentist point of view, consider for example models characterized by moment inequalities:

$$E_F m(X, \theta) \le 0$$

where F denotes the underlying distribution of the data X, and E_F denotes the expectation operator with respect to F. Then (1.1) is essentially requiring that the support function of the corresponding identified set should be smooth in F. Models satisfying (1.1) lead to interesting implications. In such models, Kaido and Santos (2014) showed that semiparametric efficient estimations can be achieved for the identified set.

We put a prior on $S_\phi(\cdot)$ (and on $\Theta(\phi)$) via the prior on ϕ, obtain its posterior distribution, and propose new algorithms to make inference about the identified set and partially identified parameters, as well as inference for low-dimensional functionals. Large sample properties of the proposed BCSs are also studied. We show that the Bayesian credible set for the identified set has a correct frequentist coverage asymptotically. In particular, to construct a confidence set for the partially identified parameter θ we use the posterior of the identified set instead of the posterior of θ.

Our approach is built on the LLA condition (1.1). We give primitive conditions for the LLA assumption in three cases:

Case 1: the support function admits a closed-form and is first-order differentiable with respect to ϕ. In this case, LLA can be directly verified by the first-order mean value theorem.

Case 2: θ is one-dimensional, or is a component of a multi-dimensional parameter. In this case the identified set is a closed interval, but the identified set for the multi-dimensional parameter may not be convex. Then the LLA can be verified directly on the parameter of interest θ itself if the end-points of the identified set are differentiable functions of ϕ.

Case 3: the more general multi-dimensional case where the identified set is characterized by (in)equalities:

$$\begin{aligned} \Theta(\phi) = \{\theta \in \Theta :\ & \Psi_i(\theta, \phi) \le 0 \text{ for } i = 1, \ldots, k_1, \\ & \Psi_i(\theta, \phi) = 0 \text{ for } i = k_1 + 1, \ldots, k_1 + k_2\}, \end{aligned} \tag{1.2}$$

where Θ denotes a vector space that contains the unknown set of interest and Ψ_i is a known function of (θ, ϕ). This case is the most challenging. The LLA assumption can be verified using either (i) the implicit function theorem, or (ii) the Lagrangian multiplier method. For the implicit function theorem approach, we rely on the fact that the optimization defining the support function is achieved on the "boundary", that is, the support function is computed based on only equality constraints and binding inequality constraints. This is the case for most applications. For the Lagrange multiplier method, we require that the moment equalities be linear and convex in θ. This approach is similar to Kaido and Santos (2014), and we shall present it in the appendix.

We discuss in detail several specific examples where our assumptions and results apply. These include: two-player entry games, one-dimensional partially identified models (as e.g., Imbens and Manski (2004)), estimating the Hansen–Jagannathan bound in asset pricing (Hansen and Jagannathan, 1991; Chernozhukov et al., 2015), missing data due to nonresponses, and regression with interval outcomes. In particular, in the two-player entry games example, we show that while the identified set for the entire parameter may be non-convex, if we are interested in one of the structural parameters, its marginal identified set can be convex whose support function satisfies the LLA condition. In addition, we refer to Kaido and Santos (2014) for many other examples where they showed that (1.1) is satisfied using a slightly different parameterization in the frequentist context.

1.1. Overview of our results

Computations

We propose several fast algorithms to compute the critical values for inferences on the identified set, partially identified parameters, and their low-dimensional functionals. We develop new algorithms to compute critical values if the linearization of the support function holds. In fact, when we simulate from the posterior of the support function, the LLA (1.1) avoids solving a numerical maximization problem in each step of the MCMC, and instead approximates it using a linear function. The latter is a more tractable problem.

Coverage properties.

Our theoretical findings concerning the coverage properties are as follows.

(i) For the identified set: we show that the constructed BCS not only has a correct Bayesian coverage but also covers the true identified set with correct frequentist probability (asymptotically). The asymptotic frequentist coverage is *exact*.

(ii) For the partially identified parameter: we find that a confidence set constructed based on the posterior distribution for the partially identified parameter θ has valid frequentist coverages asymptotically as long as we use the posterior of the identified set, instead of the posterior of θ. The asymptotic frequentist coverage, however, can be *conservative*. To construct the confidence set we study a Bayesian hypothesis test that tests whether a fixed θ belongs to the random set

$\Theta(\phi)$ (with respect to the posterior distribution of the latter). By inverting the Bayesian test statistic, we construct the confidence set as the collection of all the "accepted" θ's.

(iii) For marginal inferences: we show that it is straightforward to make inference on a scalar function of θ using our procedure. The frequentist coverage of the constructed BCS is asymptotically *exact* for the identified set, but can be *conservative* for the partially identified functions.

The intuition behind the difference on BCS's coverage properties between our results and those of Moon and Schorfheide (2012) is that priors are imposed on different objects. We directly impose the prior on the identified set. Because the identified set is "identified", its posterior will asymptotically concentrate on a "neighborhood" of the true identified set, resulting in a correct frequentist coverage of the BCS. Though this property is not specific to identified sets that are convex, in this paper we show it for such sets. In sharp contrast, as shown by Moon and Schorfheide (2012), imposing the prior directly on the partially identified parameter would result in a posterior supported only within the identified set, leading to under-coverages.

Study of the posterior of the support function.

In multi-dimensional models, the support function may not have a closed form or may depend on ϕ in a complicated way. In these cases, the LLA (1.1) is a useful tool to characterize $\phi \mapsto S_\phi(\nu)$, which in addition shows the sensitivity of the support function with respect to perturbations of ϕ. By denoting with ϕ_0 the true value of ϕ that generates the data, we show that, for every $\nu \in \mathbb{R}^d$: (i) the posterior distribution of $S_\phi(\nu)$ contracts in a neighborhood of $S_{\phi_0}(\nu)$ at the same rate of contraction of the posterior of ϕ; (ii) the posterior distribution of $S_\phi(\nu)$ converges in total variation towards a normal distribution (strong Bernstein–von Mises theorem); (iii) the posterior distribution of the stochastic process $S_\phi(\cdot)$ weakly converges towards a Gaussian process (weak Bernstein–von Mises theorem).

1.2. Related literature

We are aware of at least three closely related works in the literature that also address the question of frequentist coverages of Bayesian procedures for partially identified models.

Moon and Schorfheide (2012) was one of the first papers that constructed the BCS for the partially identified parameter θ. They impose a prior on θ and show that when such a prior has support equal to the identified set, the BCS for θ can be strictly smaller than the frequentist confidence set, so the BCS does not have a correct frequentist coverage for the partially identified parameter even asymptotically. In addition, in the working paper version of their paper, Moon and Schorfheide (2009) also studied the Bayesian and frequentist coverage of BCS for the identified set.

Kline and Tamer (2016) also provide BCS for the identified set that have the correct frequentist coverage asymptotically without requiring the convexity of the identified set. The correct frequentist coverage relies on the fact that, similar to ours, priors are imposed only on the identified parameter ϕ. They do not develop fast algorithms to compute critical values which might be obtained for convex identified sets. In addition, in our simulation study for missing data we show that our constructed BCS still has the correct coverage even when the identified set shrinks to a singleton. In contrast, their construction is conservative in this case (see Kline and Tamer (2016, Remark 8)).

More recently, Chen et al. (2018) have proposed credible sets for partially identified models that are relatively simple to compute, based on a quasi-Bayesian Monte Carlo approach. They show that their credible sets have asymptotically *exact* frequentist coverages for the *identified set* of the full parameter of interest or its subvectors, and provide fast algorithms for computations. They do not require the convexity of the identified set and so do not rely on the support function. They also develop uniformly valid confidence sets for subvector inference on both the partially identified subvector parameter and its identified set.

The (quasi-) Bayesian literature on partial identification also includes the following contributions whose research questions are substantially different from ours, *e.g.*, Poirier (1998), Liao and Jiang (2010), Florens and Simoni (2011), Gustafson (2012), Kitagawa (2012), Norets and Tang (2014), Wan (2013), etc. There is an extensive literature on partially identified models using frequentist approaches. A partial list includes Manski and Tamer (2002), Chernozhukov et al. (2007), Beresteanu and Molinari (2008), Andrews and Guggenberger (2009), Romano and Shaikh (2010), Andrews and Soares (2010), Canay (2010), Stoye (2009), Rosen (2008), Bugni (2010), among many others. The literature on the support function approach has also grown in recent years. See, *e.g.*, Mammen et al. (2001), Beresteanu et al. (2011), Bontemps et al. (2011), Chandrasekhar et al. (2012), Guntuboyina (2012), Kaido and Santos (2014), Bontemps and Magnac (2017), among others.

Our results complement the literature on the Bernstein–von Mises theorem and the frequentist coverage probabilities of Bayesian credible sets, see *e.g.* Severini (1991), Leahu et al. (2011), Sweeting (2001), Chang et al. (2009), Belloni and Chernozhukov (2009), Rivoirard and Rousseau (2012), Bickel and Kleijn (2012), Castillo and Rousseau (2015), Kato (2013), Bontemps (2011), and Norets (2015). In the framework of partially identified models, Chen et al. (2018) established a version of the Bernstein–von Mises theorem using the likelihood approach.

The paper is organized as follows. Section 2 presents the model, examples, and the prior on ϕ. Section 3 constructs Bayesian credible sets and provides the computational algorithms. Moreover, it provides inference for linear scalar functions of θ. Section 4 shows the asymptotic frequentist validity of our Bayesian procedure. Frequentist asymptotic properties of the posterior of the support function are established in Section 5. Numerical simulations are in Section 6 and Section 7 concludes. All the proofs are in a Supplementary Appendix.

Throughout the paper, the frequentist distribution of the data D_n (based on the true data distribution) will be denoted by P_{D_n}. The prior distribution and its Lebesgue density will be denoted by π while the posterior distribution and its Lebesgue density will be denoted by $P(\cdot|D_n)$. When illustrating asymptotic properties of our Bayesian procedure, we denote by ϕ_0 the true value of ϕ that generates the data. Hence, the true set and its support function will be denoted by $\Theta(\phi_0)$ and $S_{\phi_0}(\cdot)$, respectively. Moreover, '$\to^P$' denotes convergence in probability with respect to P_{D_n}.

2. General setup

2.1. The model

Let $\phi \in \Phi \subset \mathbb{R}^{d_\phi}$ be an identifiable parameter, $\Theta \subset \mathbb{R}^d$, and $\Theta(\phi) \subset \Theta$ be the identified set on which we are interested in making inference. The set $\Theta(\phi)$ is assumed to be closed and convex and contains the parameter of interest θ. In many examples the set $\Theta(\phi)$ is characterized through inequalities as follows. Let $\Psi : \Theta \times \Phi \to \mathbb{R}^k$ be a known and continuous vector-function of (θ, ϕ) that is convex in θ for every $\phi \in \Phi$, then $\Theta(\phi)$ is characterized as

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

Because $\Theta(\phi)$ is closed and convex, it is completely characterized by its support function $S_\phi(\cdot) : \mathbb{S}^d \to \mathbb{R}$, defined as, for every $\phi \in \Phi$ such that $\Theta(\phi)$ is non-empty (see, e.g. Rockafellar (1970)):

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

where $\mathbb{S}^d$ denotes the unit sphere in $\mathbb{R}^d$. The domain of the support function is restricted to the unit sphere $\mathbb{S}^d$ in $\mathbb{R}^d$ since $S_\phi(\nu)$ is positively homogeneous in ν.

When a particular $\tilde{\phi}$ corresponds to an empty $\Theta(\tilde{\phi})$, its support function is defined using the following argument: for a generic ϕ so that $\Theta(\phi)$ is not empty, define $S(\phi, \nu) := S_\phi(\nu)$, which is well defined. Then let $S_{\tilde{\phi}}(\nu) := S(\tilde{\phi}, \nu)$. For instance, suppose the identified set for θ is given by a simple closed interval $[\phi_1, \phi_2]$, with $\phi_j = EY_j$ for some observable variable $Y_j, j = 1, 2$. Let $\phi = (\phi_1, \phi_2)$ be a generic ϕ so that $\phi_1 < \phi_2$, then

$$S_\phi(1) = \phi_2, \quad S_\phi(-1) = -\phi_1.$$

We simply define a function of (ϕ, ν) to be $S(\phi, \nu) := S_\phi(\nu)$ as above where $\nu \in \{-1, 1\}$. Suppose in fact θ is point identified so that $\phi_1 = \phi_2$ at the true value. Then there is $\tilde{\phi} = (\tilde{\phi}_1, \tilde{\phi}_2)$ in a neighborhood of the true value of ϕ, so that $\tilde{\phi}_1 > \tilde{\phi}_2$. Then we let $S_{\tilde{\phi}}(\nu) := S(\tilde{\phi}, \nu)$, that is,

$$S_{\tilde{\phi}}(1) = \tilde{\phi}_2, \quad S_{\tilde{\phi}}(-1) = -\tilde{\phi}_1.$$

This completes the definition of $S_\phi(\nu)$ for all ϕ on its parameter space. As such, we view $S_\phi(\nu)$ as a well-defined function as long as $\Theta(\phi)$ is non-empty at the true value of ϕ.

Below we list two examples of partial identification to illustrate our notation.

Example 2.1 (*Interval IV Regression*)**.** Let (Y, Y_1, Y_2) be a 3-dimensional random vector such that $Y \in [Y_1, Y_2]$ with probability one. The random variables Y_1 and Y_2 are observed while Y is unobservable. For instance, the Bureau of Labor Statistics collects salary data from employers as intervals. Assume that

$$Y = x^T\theta + \epsilon$$

where x is a vector of observable regressors. Denote by Z a vector of nonnegative instrumental variables such that $E(Z\epsilon) = 0$. Then

$$E(ZY_1) \leq E(ZY) = E(Zx^T)\theta \leq E(ZY_2). \tag{2.1}$$

This model has been previously considered in Chernozhukov et al. (2007). We denote $\phi := (\phi_1, \phi_2, \mathrm{vec}(\phi_3))$ where $(\phi_1^T, \phi_2^T) := (E(ZY_1)^T, E(ZY_2)^T)$ and $\phi_3 := E(Zx^T)$. It then follows from (2.1) that θ belongs to the following set

$$\Theta(\phi) = \{\theta \in \Theta : \Psi(\theta, \phi) \leq 0\}, \text{ where } \Psi(\theta, \phi) = (\phi_1 - \phi_3\theta, \phi_3\theta - \phi_2)^T. \quad \square$$

Example 2.2 (*Frontier Estimation in Finance*)**.** Consider the equilibrium price P_t^i of a financial asset i at time t which satisfies the following restriction:

$$P_t^i = E_t[M_{t+1}P_{t+1}^i], \tag{2.2}$$

where M_{t+1} is the stochastic discount factor (SDF), which is unobservable, and E_t is the conditional expectation given information at time t. Determining the SDF M_{t+1} is a crucial research problem in finance. In many cases, Eq. (2.2) admits several solutions M_{t+1}. Let the mean and variance of M_{t+1} be μ and σ^2 respectively, assumed to be time-invariant. Hansen and Jagannathan (1991) show that for every SDF M_{t+1} that satisfies (2.2) it should hold $\sigma^2 \geq \phi_1\mu^2 - 2\phi_2\mu + \phi_3$, where $\phi_1 := m^T\Sigma m$, $\phi_2 := m^T\Sigma\iota$, $\phi_3 := \iota^T\Sigma\iota$, ι is a vector of ones and m and Σ denote, respectively, the mean vector and covariance matrix of (gross) returns of assets $1, \ldots, N$ (which are estimable from the data on returns). Therefore, we say "an SDF M_t prices an asset correctly" if its mean and variance, $\theta := (\mu, \sigma^2)$, belong to the set:

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

and $\phi := (\phi_1, \phi_2, \phi_3)^T$. Hence, $\Theta(\phi)$ becomes the object of interest, whose boundary curve $\{\theta : \Psi(\theta, \phi) = 0\}$ is often known as the "frontier". Statistical inference on $\Theta(\phi)$ can be very helpful if one wants to check whether an SDF prices an asset correctly (see *e.g.* Chernozhukov et al. (2015), Gospodinov et al. (2010) among others). □

We denote by X the observable random variable for which we have n independent and identically distributed (i.i.d.) observations $D_n = \{X_i\}_{i=1}^n$. Our model allows an infinite dimensional nuisance parameter F, which is the distribution of X (DGP, hereafter). We specify the prior distribution for $\Theta(\phi)$ and $S_\phi(\cdot)$ via the prior specification for ϕ. The posteriors of the identified set $\Theta(\phi)$ and of the support function $S_\phi(\cdot)$ are deduced from the posterior $P(\phi|D_n)$. We illustrate below three different scenarios concerning the degree of knowledge of F, the prior on it and the relation between ϕ and F.

Nonparametric prior. The likelihood is completely unrestricted and a nonparametric prior is placed directly on the cumulative distribution function (CDF) F of the data. Since ϕ is identifiable, it can be written as an explicit function of F: $\phi = \phi(F)$. The prior distribution for ϕ is then deduced from the one of F via $\phi(F)$. The Bayesian experiment is

$$X|F \sim F, \qquad F \sim \pi(F),$$

where $\pi(F)$ denotes a nonparametric prior for F. The likelihood and the posterior of F are respectively:

$$l_n(F) := \prod_{i=1}^n F(X_i), \quad P(F|D_n) \propto \pi(F)l_n(F),$$

from which we deduce the posterior of ϕ through $\phi = \phi(F)$. For instance, in Example 2.1, suppose the data $X = (ZY_1^T, ZY_2^T, \text{vec}(Zx^T)^T)^T$ has a multivariate CDF F, then

$$\phi(F) := E(X) = \int xF(x)dx.$$

Examples of $\pi(F)$ include Dirichlet process priors and Polya tree. The case where $\pi(F)$ is a Dirichlet process prior in partially identified models is treated by Florens and Simoni (2011).

Semi-parametric prior. Let η be an infinite dimensional nuisance parameter that is unknown and that, together with ϕ, completely characterizes F. Hence, $F \in \{F_{\phi,\eta}; \phi \in \Phi, \ \eta \in \mathcal{P}\}$, where $\mathcal{P}$ is an infinite dimensional set. Let $\pi(\phi, \eta)$ denote the joint prior on (ϕ, η). The Bayesian experiment is

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

Write the likelihood $l_n(\phi, \eta) := \prod_{i=1}^n F_{\phi,\eta}(X_i)$. The marginal posterior of ϕ becomes:

$$P(\phi|D_n) \propto \int_{\mathcal{P}} \pi(\phi, \eta)l_n(\phi, \eta)d\eta.$$

For instance, in Example 2.1, suppose the data $X = (Z^TY_1, Z^TY_2, \text{vec}(Zx^T)^T)^T$ has a continuous multivariate density function, we can then consider a "location model" as in Ghosal et al. (1999b):

$$X = \phi + u, \quad u \sim \eta,$$

where u is a zero-mean random vector with an unknown Lebesgue density function η. Then, the likelihood is given by $l_n(\phi, \eta) = \prod_{i=1}^n \eta(X_i - \phi)$. Examples of priors on the infinite dimensional density parameter η include, e.g., Dirichlet mixture of normals (Ghosal et al., 1999a) and random Bernstein polynomials (Walker et al., 2007).

Parametric prior. The sampling distribution F is known up to a finite dimensional parameter (ϕ, η), where η is a nuisance parameter. We may write $F = F_{\phi,\eta}$. This is a simple parametric framework. Let $\pi(\phi, \eta)$ be a prior on (ϕ, η) and $l_n(\phi, \eta)$ be the likelihood associated with $F_{\phi,\eta}$. Then

$$P(\phi|D_n) \propto \int \pi(\phi, \eta)l_n(\phi, \eta)d\eta.$$

For instance, in Example 2.1, suppose the data $X = (Z^TY_1, Z^TY_2, \text{vec}(Zx^T)^T)^T$ is normally distributed. Then we can parameterize it as:

$$X = \phi + u, \quad u \sim N(0, \eta),$$

for some covariance matrix η. Then $l_n(\phi, \eta) := \prod_{i=1}^n f(X_i; \phi, \eta)$, where $f(\cdot; \phi, \eta)$ denotes the multivariate normal density function with mean vector ϕ and covariance η.

Regardless of the prior specification, since ϕ is point identified from a frequentist perspective, it is well known that under very mild conditions its posterior asymptotically concentrates around a $\sqrt{n}$-neighborhood of the true value ϕ_0, and is asymptotically normally distributed. We present this well known result in the following assumption without pursuing its proofs.

We denote by $\|\cdot\|_{TV}$ the total variation (TV) norm, that is, for two probability measures P and Q,

$$\|P-Q\|_{TV} := \sup_B |P(B)-Q(B)|$$

where B is an element of the Borel σ-algebra on which P and Q are defined.

Assumption 2.1. *(i)* The marginal posterior of ϕ is such that, for any $\epsilon, \delta > 0$, there is $C > 0$ such that with P_{D_n}-probability at least $1-\epsilon$,

$$P(\|\phi-\phi_0\| > Cn^{-1/2}|D_n) < \delta.$$

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

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

where $\mathcal{N}$ denotes the d_ϕ-dimensional normal distribution, $\Delta_{n,\phi_0} := n^{-1/2}\sum_{i=1}^n I_0^{-1}\ell_{\phi_0}(X_i)$, ℓ_{ϕ_0} is the semiparametric efficient score function of the model and I_0 is the semiparametric efficient Fisher information matrix.

(iii) There exists a regular estimator $\hat{\phi}$ of ϕ that satisfies

$$\sqrt{n}(\hat{\phi}-\phi_0) \to^d \mathcal{N}(0, I_0^{-1}).$$

More precise definition of ℓ_{ϕ_0} and I_0 for parametric and semiparametric models can be found in van der Vaart (2002, Definition 2.15) and Bickel and Kleijn (2012, Section 4), respectively. Assumption 2.1 *(i), (ii)* are standard results known as the posterior concentration and Bernstein–von Mises theorem in (semi-)parametric Bayesian literature and are in general satisfied, under mild restrictions, for both nonparametric and semiparametric prior on (ϕ, F). Primitive conditions for this assumption are given in Shen (2002, e.g.). Other references that provide primitive conditions include Belloni and Chernozhukov (2009), Bickel and Kleijn (2012), and Rivoirard and Rousseau (2012). Assumption 2.1 *(iii)* requires the existence of a best regular estimator. Primitive conditions for semi-parametric models are given in Van der Vaart (e.g. 2000, Lemma 25.23) and more specifically for semi-parametric maximum likelihood estimators are given in Shen (e.g. 2002, Theorems 1 and 2).

3. Algorithms to compute Bayesian credible sets

For a generic set $\mathcal{C}$, and any $\epsilon > 0$, define the "ϵ-expansion" of $\mathcal{C}$ to be $\mathcal{C}^\epsilon := \{\theta : d(\theta, \mathcal{C}) \leq \epsilon\}$, where $d(\theta,\mathcal{C}) := \inf_{c\in\mathcal{C}}\|\theta - c\|$ and $\|\cdot\|$ denotes the Euclidean norm.

3.1. Algorithms for computing critical values for the identified set

We start with constructing the Bayesian Credible Set (BCS) for the identified set $\Theta(\phi)$. Its support function plays a central role in our construction. For a credible level $1-\tau$, $\tau \in (0,1)$, we shall find appropriate critical values ϵ_τ and $\tilde{\epsilon}_\tau$, and construct the BCS as $\Theta(\hat{\phi})^{\epsilon_\tau}$ for the identified set, and the Uniform Bayesian Credible Band (UBCB) $\{S_\phi(\cdot) : S_\phi(\nu) \in [S_{\hat{\phi}}(\nu) \pm \tilde{\epsilon}_\tau], \forall \nu \in \mathbb{S}^d\}$ for the support function.

To obtain the critical values, note that for convex sets, the support function has the following nice property: for any $\epsilon > 0$,

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

In fact, we shall show below that tractable algorithms can be developed based on simulating the posterior quantiles of the support function. For $\tau \in (0,1)$, let q_τ and $\tilde{q}_\tau$ be the $1-\tau$ quantiles of the posterior of

$$J(\phi) := \sqrt{n}\sup_{\|\nu\|=1}\left(S_\phi(\nu) - S_{\hat{\phi}}(\nu)\right) \quad \text{and} \quad \tilde{J}(\phi) := \sqrt{n}\sup_{\|\nu\|=1}\left|S_\phi(\nu) - S_{\hat{\phi}}(\nu)\right|,$$

respectively, so that

$$P(J(\phi) \leq q_\tau | D_n) = 1-\tau, \quad \text{and} \quad P\left(\tilde{J}(\phi) \leq \tilde{q}_\tau \middle| D_n\right) = 1-\tau.$$

Theorem 4.1 shows that

$$P\left(\Theta(\phi) \subset \Theta(\hat{\phi})^{q_\tau/\sqrt{n}} \middle| D_n\right) = 1-\tau, \quad \text{and} \quad P\left(\sup_{\|\nu\|=1}\left|S_\phi(\nu) - S_{\hat{\phi}}(\nu)\right| \leq \frac{\tilde{q}_\tau}{\sqrt{n}} \middle| D_n\right) = 1-\tau.$$

Thus we shall use $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ as the BCS for the identified set, and $\{S_\phi(\cdot) : S_\phi(\nu) \in [S_{\hat{\phi}}(\nu) \pm \tilde{q}_\tau/\sqrt{n}], \forall \nu \in \mathbb{S}^d\}$ as the UBCB for the support function.

In general, calculating the critical values based on Monte Carlo methods relies on evaluating the support function. In complex models where $S_\phi(\cdot)$ does not have a closed form, this would in principle require a Monte Carlo procedure as follows. Uniformly generate $\{\nu_j\}_{j\le G}$ such that $\|\nu_j\|=1$. In addition, sample $\{\phi_i\}_{i\le M}$ from the posterior distribution $P(\phi|D_n)$ of ϕ. For each ν_j, and for a given estimator $\hat\phi$ of ϕ,

- (outer-loop) Solve an optimization problem to calculate $S_{\hat\phi}(\nu_j)$.
- (inner-loop) Solve M optimization problems to calculate $S_{\phi_i}(\nu_j)$, for every $i=1,\dots,M$.

Then q_τ and $\tilde q_\tau$ are respectively calculated as the $1-\tau$ quantiles of

$$\left\{\sqrt{n}\max_{j\le G}(S_{\phi_i}(\nu_j)-S_{\hat\phi}(\nu_j)) : i\le M\right\},\quad \text{and } \left\{\sqrt{n}\max_{j\le G}\left|S_{\phi_i}(\nu_j)-S_{\hat\phi}(\nu_j)\right| : i\le M\right\}.$$

Generating several $\{\nu_j\}$ allows to approximately solve the "outside" optimization problem "$\sup_{\|\nu\|=1}$" using the outer-loop. But for each ν_j, the above procedure requires solving M optimization problems in the inner-loop, which makes the overall computational task very intensive.

Instead, below we propose several algorithms to carry out the inner-loop, which are based on the following local linear approximation (LLA) assumption: uniformly over ϕ in a neighborhood of $\hat\phi$, there is a vector $A(\nu)$ that depends on ν but is independent of ϕ, so that (the formal statement is given by Assumption 4.1 in Section 4)

$$S_\phi(\nu)-S_{\hat\phi}(\nu)\approx A(\nu)^T(\phi-\hat\phi). \tag{3.1}$$

The approximation error in (3.1) is smaller than the first-order statistical error $n^{-1/2}$. As a result, $S_{\phi_i}(\nu_j)$ can be approximated by:

$$S_{\phi_i}(\nu_j)\approx S_{\hat\phi}(\nu_j)+A(\nu_j)^T(\phi_i-\hat\phi),$$

which avoids solving M optimization problems in the inner-loop.

We shall verify the LLA under various primitive conditions. From a computational point of view, $A(\nu)$ can be derived using either the implicit function theorem (as described in Section 4.2) or the Lagrange multiplier approach (as described in Appendix E in the supplement). With the Lagrange multiplier approach, $A(\nu)$ is computed as:

$$A(\nu)^T=\lambda(\nu,\hat\phi)^T\nabla_\phi\Psi(\theta^*(\nu),\hat\phi),$$

where $\theta^*(\nu)$ is the optimizer that is obtained when calculating $S_{\hat\phi}(\nu)$, $\lambda(\nu,\hat\phi)$ is the Kuhn–Tucker (KT) vector arising from the calculation of $S_{\hat\phi}(\nu)$ as described in Theorem E.1, and $\nabla_\phi\Psi(\theta^*(\nu),\hat\phi)$ is the partial gradient of Ψ with respect to ϕ. We now state our algorithm for calculating the critical values q_τ and $\tilde q_\tau$ as follows:

Algorithm 1 (Identified Set)

1. Fix a prior $\pi(\phi)$, and construct the posterior of ϕ. Let $\{\phi_i\}_{i\le M}$ be the MCMC draws from the posterior of ϕ. Let $\hat\phi=\frac{1}{M}\sum_{i=1}^M\phi_i$. In addition, uniformly generate $\{\nu_j\}_{j\le G}$ such that $\|\nu_j\|=1$ for each j.
2. (outer-loop): For each $j\le G$, solve the following constrained convex problem for $S_{\hat\phi}(\nu_j)$:

   $$\max_\theta \nu_j^T\theta \qquad \text{subject to} \quad \Psi(\theta,\hat\phi)\le 0$$

   and obtain $\theta_j^*=\arg\max_\theta\{\nu_j^T\theta:\Psi(\theta,\hat\phi)\le 0\}$ and the corresponding KT-vector $\lambda(\nu_j,\hat\phi)$. If the set of maximizers of this problem is not a singleton, then arbitrarily pick one of the elements of the set.
   Calculate $A(\nu_j)$ either by

   $$A(\nu_j)^T=\lambda(\nu_j,\hat\phi)^T\nabla_\phi\Psi(\theta_j^*,\hat\phi)$$

   or from Algorithm 4 below.
3. (inner-loop): For each $i\le M$, let

   $$J_i=\sqrt{n}\max_{j\le G}\left\{A(\nu_j)^T[\phi_i-\hat\phi]\right\},\quad \tilde J_i=\sqrt{n}\max_{j\le G}\left\{|A(\nu_j)^T[\phi_i-\hat\phi]|\right\}.$$

4. Let q_τ and $\tilde q_\tau$ be the $(1-\tau)$ th quantile of $\{J_i\}_{i\le M}$ and $\{\tilde J_i\}_{i\le M}$, respectively. Use $\Theta(\hat\phi)^{q_\tau/\sqrt{n}}$ as the BCS for the identified set, and $\{S_\phi(\cdot):S_\phi(\nu)\in[S_{\hat\phi}(\nu)\pm\tilde q_\tau/\sqrt{n}],\forall\nu\in\mathbb{S}^d\}$ as the UBCB for the support function.

For each generated ν_j, we only need to solve the constraint optimization once in the outer-loop for $S_{\hat\phi}(\nu_j)$, and both the minimizer θ^* and the KT-vector $\lambda(\nu_j,\hat\phi)$ are automatically obtained in the outer-loop. This step takes advantage of convex optimizations, and both θ_j^* and $\lambda(\nu_j,\hat\phi)$ can be computed fast using the Matlab function `fmincon`. The maximizations in the inner-loop are also quite easy as they involve optimizations on a finite set $j\in\{1,\dots,G\}$. The computational time for implementing Algorithm 1 depends on the difficulty in computing the support function in the outer loop, which is also related to the dimension of θ. When the dimension of θ is large, the computational burden can be partially avoided in at least two situations: when we are interested only in some components of θ and when the support function has a closed form.

3.2. Algorithm for covering the partially identified parameter using the posterior distribution

In this subsection we show that we can construct a confidence set for the partially identified parameter $\theta \in \Theta(\phi)$ using the posterior distribution of $\Theta(\phi)$ that has a desired frequentist coverage.

Consider a Bayesian testing problem for the null hypothesis

$$H_0(\theta) : \theta \in \Theta(\phi), \tag{3.2}$$

for a fixed and known $\theta \in \Theta$, where $\Theta(\phi)$ is drawn from the posterior distribution of ϕ. We shall derive an "acceptance" criterion from the posterior distribution of $\Theta(\phi)$ so that

$$\inf_{\theta\in\Theta(\phi_0)} P_{D_n}(\theta : H_0(\theta) \text{ is "accepted"}) \geq 1-\tau. \tag{3.3}$$

Inverting test statistics has been one of the most popular methods for constructing confidence intervals for partially identified parameters in the frequentist literature, e.g., Beresteanu and Molinari (2008), Andrews and Soares (2010), Canay (2010), Rosen (2008), etc. Here we apply a similar idea to construct a confidence set for θ. The main novelty of our method, to be presented below, is that here we investigate the testing problem $H_0(\theta)$ from a Bayesian point of view. Here $\{H_0(\theta)$ is "accepted"$\}$ is an event in which a given θ is accepted if the event $H_0(\theta)$ holds with a large posterior probability. This standard Bayesian procedure has some optimality properties under certain circumstances from the perspective of Bayesian decision theory (Casella and Berger, 2002). From the practical point of view, as we shall illustrate in Algorithm 2 below, the Bayesian testing procedure takes advantage of the merit of the MCMC algorithm, leading to a very convenient algorithm to simulate the confidence set.

Recall that for a fixed θ, $H_0(\theta)$ is true if and only if $\theta^T\nu \leq S_\phi(\nu)$ for all $\|\nu\| = 1$. Therefore, for a to-be-determined critical value δ_τ, we can define

$$\Omega_\tau(\phi) := \left\{\theta : \theta^T\nu \leq S_\phi(\nu) + \frac{\delta_\tau}{\sqrt{n}}, \forall \|\nu\| = 1\right\}. \tag{3.4}$$

Intuitively, $\theta \in \Omega_\tau(\phi)$ means θ is "close" to $\Theta(\phi)$. In fact, it can be shown that $\Omega_\tau(\phi) = \Theta(\phi)^{\delta_\tau/\sqrt{n}}$. Therefore, we construct a Bayesian test for (3.2) by investigating whether θ is "posteriorly covered" with a high probability:

$$\text{accept } H_0(\theta) \Leftrightarrow P(\theta \in \Theta(\phi)^{\delta_\tau/\sqrt{n}}|D_n) \geq 1-\tau, \tag{3.5}$$

for the confidence level $1-\tau$. Here $P(\theta \in \Theta(\phi)^{\delta_\tau/\sqrt{n}}|D_n)$ is the posterior probability with respect to the posterior distribution of ϕ, treating θ as fixed. Combining (3.3)–(3.5), we construct the frequentist confidence set for θ as:

$$\widehat{\Omega} := \{\theta \in \Theta : P(\theta \in \Theta(\phi)^{\delta_\tau/\sqrt{n}}|D_n) \geq 1-\tau\},$$

which depends on the critical value δ_τ. We will choose $\delta_\tau = 2q_\tau$, where q_τ is the critical value of $J(\phi)$ defined in Section 3.1.

Using the results of the Bayesian credible band for the support function, we shall show in Section 4 that

$$\inf_{\theta\in\Theta(\phi_0)} P_{D_n}\left(\underbrace{P\left(\theta \in \Theta(\phi)^{2q_\tau/\sqrt{n}}\big|D_n\right) \geq 1-\tau}_{\theta\in\widehat{\Omega}}\right) \geq 1-\tau-o_p(1).$$

To explain this in words: if we define $\widehat{\Omega}$ as the set of all the "accepted" θ's (covered by $\Theta(\phi)^{2q_\tau/\sqrt{n}}$ with posterior probability at least $1-\tau$), then $\widehat{\Omega}$ covers the partially identified parameter with a sampling (frequentist) probability of at least $1-\tau$.

The set $\widehat{\Omega}$ can be computed using the following MCMC-based algorithm.

Algorithm 2 (Partially Identified Parameter)

1. Let $\{\phi_i\}_{i\leq M}$ be the MCMC draws from the posterior of ϕ. Compute $\hat{\phi} = \frac{1}{M}\sum_i \phi_i$. In addition, uniformly generate $\{\tilde{\theta}_b\}_{b\leq B}$ from the parameter space Θ, and uniformly generate $\{\nu_j\}_{j\leq G}$ such that $\|\nu_j\| = 1$ for each j.
2. Solve for $S_{\hat{\phi}}(\nu_j)$ and $A(\nu_j)$ as in Algorithm 1 for $j = 1, \ldots, G$.
3. For each $b = 1, \ldots, B$ and for a $\tau \in (0, 1)$, if $\tilde{\theta}_b$ satisfies:
$$\frac{1}{M}\sum_{i=1}^{M} 1\left\{\max_{j\leq G}[\tilde{\theta}_b^T\nu_j - S_{\hat{\phi}}(\nu_j) - A(\nu_j)^T(\phi_i - \hat{\phi})] \leq \frac{2q_\tau}{\sqrt{n}}\right\} \geq 1-\tau \tag{3.6}$$
then accept $\tilde{\theta}_b$; otherwise discard $\tilde{\theta}_b$. The critical value q_τ is obtained in Algorithm 1.
4. Collect all the accepted $\tilde{\theta}_b$'s as a set $\widehat{\Omega}^*$, which is an approximation of $\widehat{\Omega}$.

To explain that $\widehat{\Omega}^*$ is an approximation of $\widehat{\Omega}$, we now explain that (3.6) is an MCMC approximation of the event $\tilde{\theta}_b \in \widehat{\Omega}$. First note that for any $\epsilon > 0$, we have $d(\tilde{\theta}_b, \Theta(\phi_i)) \leq \epsilon$ if and only if $\tilde{\theta}_b^T\nu \leq S_{\phi_i}(\nu) + \epsilon$ for all $\|\nu\| = 1$. In addition, the LLA entails $S_{\phi_i}(\nu_j) \approx S_{\hat{\phi}}(\nu_j) + A(\nu_j)^T(\phi_i - \hat{\phi})$. Therefore, for large G,

$$
\begin{aligned}
& 1\left\{\max_{j\le G}[\tilde{\theta}_b^T \nu_j - S_{\hat{\phi}}(\nu_j) - A(\nu_j)^T(\phi_i - \hat{\phi})] \le \frac{2q_\tau}{\sqrt{n}}\right\} \\
\approx\ & 1\left\{\max_{j\le G}[\tilde{\theta}_b^T \nu_j - S_{\phi_i}(\nu_j)] \le \frac{2q_\tau}{\sqrt{n}}\right\} \\
\approx\ & 1\left\{d(\tilde{\theta}_b, \Theta(\phi_i)) \le \frac{2q_\tau}{\sqrt{n}}\right\}.
\end{aligned}
$$

Thus, since $\{\phi_i\}_{i\le M}$ are the MCMC draws from the posterior of ϕ, the left hand side of (3.6) is approximately equal to,

$$
\frac{1}{M}\sum_{i=1}^{M} 1\left\{d(\tilde{\theta}_b, \Theta(\phi_i)) \le \frac{2q_\tau}{\sqrt{n}}\right\} = \frac{1}{M}\sum_{i=1}^{M} 1\left\{\tilde{\theta}_b \in \Theta(\phi_i)^{2q_\tau/\sqrt{n}}\right\} \approx P\left(\tilde{\theta}_b \in \Theta(\phi)^{2q_\tau/\sqrt{n}}\middle|D_n\right).
$$

Therefore, (3.6) is approximately the same as requiring $P\left(\tilde{\theta}_b \in \Theta(\phi)^{2q_\tau/\sqrt{n}}\middle|D_n\right) \ge 1-\tau$, which means, by definition, $\tilde{\theta}_b \in \widehat{\Omega}$. Therefore, $\widehat{\Omega}^*$ is an approximation of $\widehat{\Omega}$.

Remark 3.1. To explain the factor "2" in the choice of the critical value, we briefly explain the key step of the technical proof here. Note that for any $\theta \in \Theta(\phi_0)$, we decompose the event $\{\theta^T\nu \le S_\phi(\nu) + \frac{2q_\tau}{\sqrt{n}}\}$ into the intersection of the following two events:

$$
\underbrace{\{S_{\phi_0}(\nu) \le S_{\hat{\phi}}(\nu) + \frac{q_\tau}{\sqrt{n}}\}}_{\text{frequentist event}}, \quad \underbrace{\{S_{\hat{\phi}}(\nu) \le S_{\phi}(\nu) + \frac{q_\tau}{\sqrt{n}}\}}_{\text{Bayesian event}}
$$

so that the following lower bound holds for the posterior $P(\theta \in \Theta(\phi)^{2q_\tau/\sqrt{n}}|D_n)$:

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

Hence, we see that the factor "2" appears from the above application of a "triangular-type" inequality. This allows us to link the event $\{\theta \in \Theta(\phi)^{2q_\tau/\sqrt{n}}\}$ with two more concrete events, whose large sample probabilities are easier to characterize. By the definition of q_τ, both sub events hold with high probabilities.

3.3. Algorithms for set inference for scalar functions

Suppose we are interested in a one-dimensional continuous transformation of the partially identified parameter $g(\theta)$. Our method provides a simple procedure to construct a BCS for $g(\theta)$ and for its identified set

$$
G(\phi) := \{g(\theta) : \theta \in \Theta(\phi)\}.
$$

We refer to Kaido et al. (2019), Bugni et al. (2017) for many interesting examples. When $g(\cdot)$ is a linear transformation, $G(\phi)$ can be simply derived. Suppose $g(\theta) = g^T\theta$ for some vector g. We consider the normalized case $\|g\| = 1$. Then its support function is

$$
\widetilde{S}_\phi(\nu) = \sup\{\theta^T g\nu : \theta \in \Theta(\phi)\} = S_\phi(g\nu) : \quad \nu \in \{-1, 1\}
$$

and the identified set is given by $G(\phi) = [-S_\phi(-g), S_\phi(g)]$. This case is interesting, for instance, when $g(\theta)$ denotes the k th component of θ. Then we take $g = e_k$ as the vector in $\mathbb{R}^d$ with one in the kth coordinate and zeros elsewhere. In the main text we focus on the linear transformation, and discuss the more general nonlinear case in Appendix A.

Identified set of scalar functions: linear case. Let c_τ be the $1-\tau$ quantile of the posterior of

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

that is, $P(L(\phi) \le c_\tau|D_n) = 1-\tau$. Then, Theorem 4.1 immediately implies:

$$
P\left(G(\phi) \subset \left[-S_{\hat{\phi}}(-g) - \frac{c_\tau}{\sqrt{n}}, S_{\hat{\phi}}(g) + \frac{c_\tau}{\sqrt{n}}\right]\middle| D_n\right) = 1-\tau.
$$

Therefore, the interval $[-S_{\hat{\phi}}(-g) - \frac{c_\tau}{\sqrt{n}}, S_{\hat{\phi}}(g) + \frac{c_\tau}{\sqrt{n}}]$ is a $1-\tau$ BCS for the marginal identified set of $g(\theta) = g^T\theta$.

Partially identified scalar functions: linear case. Similarly to the construction in Section 3.2, the critical value for the marginal confidence interval of $g(\theta)$ is also simple to compute. Define

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

We shall show in Section 4 that $\widehat{\Omega}_g$ is an asymptotically valid confidence set for $g(\theta)$.

Importantly, we directly construct confidence sets for the scalar functions instead of projecting from the set for the full vector of the partially identified parameter as in the projection method which often leads to very conservative coverages (see Kaido et al. (2019) for discussions). This is an appealing feature even computationally. Indeed, $L(\phi)$ can be approximated using the local linear approximation based on either the implicit function approach as described in Section 4.2 or on the Lagrange multipliers approach. In the latter case:

$$L(\phi) \approx \sqrt{n} \max_{\nu=\pm1} \left\{ \lambda(\nu g, \hat{\phi})^T \nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})[\phi - \hat{\phi}] \right\}.$$

As already stressed in Algorithm 1, thanks to this approximation we do not need to solve an optimization problem for each value of ϕ drawn from the posterior of ϕ. The algorithm below shows that computing these critical values c_τ is straightforward using the MCMC sampler.

Algorithm 3 (Inference for $G(\phi)$ (The Identified Set for $g(\theta)$): linear case)

1. Let $\{\phi_i\}_{i\leq M}$ be the MCMC draws from the posterior of ϕ.
2. For $\nu = \pm 1$, solve the following constrained convex problem

$$\max_\theta \nu g^T \theta \qquad \text{subject to} \quad \Psi(\theta, \hat{\phi}) \leq 0$$

   and obtain $\theta^*(\nu) = \arg\max_\theta \{\nu g^T\theta : \Psi(\theta, \hat{\phi}) \leq 0\}$ and the corresponding Kuhn–Tucker vector $\lambda(\nu g, \hat{\phi})$ (respectively for $\nu = \pm 1$). If the set of maximizers of this problem is not a singleton, then arbitrarily pick one element of the set.
3. For each $i \leq M$, let either $A_g(\nu)^T = \lambda(\nu g, \hat{\phi})^T \nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})$ or $A_g(\nu)$ be as described in Algorithm 4 below, then

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

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

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

Algorithm 3' (Inference for $g(\theta) = g^T\theta$: Linear Case)

1. Obtain $A_g(1)$, $A_g(-1)$ and c_τ 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

$$\frac{1}{M} \sum_{i=1}^{M} 1\left\{ g(\tilde{\theta}_b) \in [-\underline{S}_{\phi_i} - \frac{2c_\tau}{\sqrt{n}}, \bar{S}_{\phi_i} + \frac{2c_\tau}{\sqrt{n}}] \right\} \geq 1 - \tau, \tag{3.7}$$

   then set $\theta_b^* := \tilde{\theta}_b$ ("accepted draws"); otherwise discard $\tilde{\theta}_b$. Here

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

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

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

   where the end points are respectively the minimum and maximum of the "accepted draws".

Remark 3.2. In step 3 of Algorithm 3', we respectively use $\bar{S}_{\phi_i}$ and $\underline{S}_{\phi_i}$ to approximate $S_{\phi_i}(g)$ and $S_{\phi_i}(-g)$ based on the LLA, avoiding in this way to solve M optimization problems (one for each ϕ_i, $i \in \{1, \ldots, M\}$) to calculate the support function for each ϕ_i. In addition, since we only need to find the minimum and the maximum $g(\tilde{\theta}_b)$ such that (3.7) holds, step 3 of Algorithm 3' can be simplified and replaced by the following step. Thus, we do not need to evaluate (3.7) for all the $b \leq B$.

Step 3': Re-arrange $g(\tilde{\theta}_{(1)}) \leq \cdots \leq g(\tilde{\theta}_{(B)})$, where the notation $g(\tilde{\theta}_{(j)})$ means that $g(\cdot)$ is evaluated at the $\tilde{\theta}_b$ that gives the jth smallest value of g. Starting from $g(\tilde{\theta}_{(1)})$ to gradually increase, find the smallest $g(\tilde{\theta}_{(b)})$ such that (3.7) holds, and set it to be $\min\{g(\theta_b^*)\}$. Starting from $g(\tilde{\theta}_{(B)})$ to gradually decrease, find the largest $g(\tilde{\theta}_{(b)})$ such that (3.7) holds, and set it to be $\max\{g(\theta_b^*)\}$.

3.4. Connections in computations with frequentist confidence sets

The support function has been used in the frequentist literature as well to construct FCS for the set. For instance, Beresteanu and Molinari (2008) constructed FCS for the set by:

$$P(\Theta(\phi_0) \subset \widehat{\Theta}^{\bar{c}_\tau}) \geq 1-\tau,$$

where $\widehat{\Theta}$ is the sample analogue of the true identified set. They proposed a Bootstrap procedure to simulate the critical value $\bar{c}_\tau$: Let $\{\widehat{\Theta}_i\}_{i\leq M}$ be the sample analogues of the identified set based on the M bootstrap samples. Then, $\bar{c}_\tau$ is determined as the $(1-\tau)$th quantile of: (by letting $S_{\mathcal{C}}(\cdot)$ denote the support function for a generic set $\mathcal{C}$)

$$\left\{\sqrt{n}\sup_{\|\nu\|=1}|S_{\widehat{\Theta}_i}(\nu)-S_{\widehat{\Theta}}(\nu)|: \quad i\leq M\right\}. \tag{3.8}$$

Computing (3.8) may be difficult without a linear approximation. Kaido and Santos (2014) derived a linear approximation of $S_{\widehat{\Theta}_i}(\nu)-S_{\widehat{\Theta}}(\nu)$ for moment inequality models, which allows to avoid solving the optimization in $S_{\widehat{\Theta}_i}$ for each of the bootstrap sample. The Lagrange multiplier approach that we propose to linearize the support function is similar to theirs when $\widehat{\Theta}_i$ can be parameterized by ϕ.

The major computational difference between our proposed BCS and Kaido and Santos (2014)'s FCS for the identified set is that in our procedure, our "sampled" identified sets $\{\Theta(\phi_i)\}_{i\leq M}$ are the MCMC draws from the posterior distribution of ϕ, while the FCS's "sampled" sets $\{\widehat{\Theta}_i\}_{i\leq M}$ are the bootstrap samples from the empirical distribution. The differences and connections are therefore essentially those between the Bayesian's MCMC and the frequentist's Bootstrap. While both algorithms share many similarities in the current context, the Bayesian approach makes good use of the prior information of ϕ, which may be informative in practice. In particular, if the prior is informative it is likely to have areas of the support where the posterior concentrates a lot of mass. Therefore, we need a smaller number of MCMC draws (after the burn-in period) to have a good description of the posterior of ϕ. This reduces the computational costs in particular if ϕ is multidimensional.

On the other hand, the proposed confidence set $\widehat{\Omega}$ for the partially identified parameter is computationally different from all the existing frequentist confidence sets to the best of our knowledge. While it is also computationally simple, we do not claim that it is advantageous over existing methods in the literature. Instead, we use it to clearly show that with the help of our Bayesian analysis based on the support function, a FCS for the partially identified parameter can be also constructed using the posterior distribution of $\Theta(\phi)$. This property is both computationally and theoretically attractive, and complements the Bayesian literature for partially identified models.

4. Large sample coverages

In this section we study the coverage properties of the constructed BCS's. First of all, we note that the constructed BCS for the identified set has a correct *Bayesian coverage*, which is stated in the following theorem.

Theorem 4.1 (*Bayesian Coverage*). *Suppose $\Theta(\phi)$ is convex and closed for every ϕ in its parameter space. For every sampling sequence D_n, and any $\tau \in (0,1)$,*

$$P\left(\Theta(\phi)\subset\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}\Big|D_n\right)=1-\tau, \quad \text{and} \quad P\left(\sup_{\|\nu\|=1}\left|S_\phi(\nu)-S_{\hat{\phi}}(\nu)\right|\leq\frac{\tilde{q}_\tau}{\sqrt{n}}\Big|D_n\right)=1-\tau.$$

Below we shall focus on the large sample frequentist coverage properties. We aim to show:

$$P_{D_n}\left(\Theta(\phi_0)\subset\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}\right)=1-\tau+o_P(1), \tag{4.1}$$

and

$$P_{D_n}\left(\sup_{\|\nu\|=1}|S_{\phi_0}(\nu)-S_{\hat{\phi}}(\nu)|\leq\frac{\tilde{q}_\tau}{\sqrt{n}}\right)=1-\tau+o_P(1). \tag{4.2}$$

Hence, both $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ and the set $\{S_\phi(\cdot): S_\phi(\nu)\in[S_{\hat{\phi}}(\nu)\pm\tilde{q}_\tau/\sqrt{n}], \forall\nu\in\mathbb{S}^d\}$ also have (asymptotically) correct frequentist coverage probabilities. Consequently, our BCS and UBCB are useful for both Bayesian and frequentist inference. Note that in the Bayesian coverage of Theorem 4.1, $\Theta(\phi)$ is the random set (with respect to the posterior of ϕ), while $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ is treated as fixed. On the contrary, in the frequentist coverage (4.1), $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ is the random set (with respect to the sampling distribution of $\hat{\phi}$), while $\Theta(\phi_0)$ is the true and fixed set object.

The general Bayesian–frequentist duality relies on the asymptotic equivalence between the posterior distribution of $S_\phi(\cdot)-S_{\phi_0}(\cdot)$ and the sampling (frequentist) distribution of $S_{\hat{\phi}}(\cdot)-S_{\phi_0}(\cdot)$. To establish this equivalence, we rely on a local linear expansion of the support function, which we first present as a high-level condition and then we provide primitive conditions for it. We denote by $B(\phi_0,\delta)$ the closed ball centered on ϕ_0 with radius $\delta>0$. Recall that $\Theta(\phi_0)$ denotes the true identified set and that P_{D_n} denotes the probability measure in the frequentist sense.

Assumption 4.1 (*Local Linear Approximation (LLA)*)**.** There is a continuous vector function $A(\nu)$ such that for

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

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

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

Theorem 4.2 (*Frequentist Coverage*)**.** *Suppose $\Theta(\phi)$ is convex and closed for every ϕ in its parameter space such that $\Theta(\phi)$ is nonempty. Suppose Assumptions 2.1 and 4.1 hold. Then, the frequentist coverage probabilities of the BCS and UBCB constructed in Section 3 satisfy: for any $\tau\in(0,1)$ and for q_τ and $\tilde{q}_\tau$ as defined in Section 3.1, as $n\to\infty$,*

(i) $P_{D_n}\left(\Theta(\phi_0)\subset\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}\right) = 1-\tau+o_P(1);$[1]

(ii) $\inf_{\theta\in\Theta(\phi_0)} P_{D_n}\left(\theta\in\widehat{\Omega}\right)\geq 1-\tau-o_P(1)$, *where*

$$\widehat{\Omega} := \{\theta\in\Theta : P(\theta\in\Theta(\phi)^{2q_\tau/\sqrt{n}}|D_n)\geq 1-\tau\};$$

(iii) $P_{D_n}\left(\sup_{\|\nu\|=1}|S_{\phi_0}(\nu)-S_{\hat{\phi}}(\nu)|\leq\frac{\tilde{q}_\tau}{\sqrt{n}}\right) = 1-\tau+o_P(1).$

Concerning the inference about the transformed parameter $g(\theta)$, let $G(\phi)$ be the identified set for $g(\theta)$. The following theorem presents the case of linear transformations. We shall present the more general case of nonlinear transformations in the appendix.

Theorem 4.3 (*Scalar Functions*)**.** *Let $g(\theta) = g^T\theta$ for some $\|g\| = 1$. Suppose $\Theta(\phi)$ is closed and convex for every $\phi\in\Phi$ such that $\Theta(\phi)$ is nonempty and that the assumptions of Theorem 4.2 hold. Then, for any $\tau\in(0,1)$ and for c_τ as defined in Section 3.3,*

(i) Bayesian coverage of the identified set for scalar functions: for almost all data D_n,

$$P\left(G(\phi)\subset\left[-S_{\hat{\phi}}(-g)-\frac{c_\tau}{\sqrt{n}}, S_{\hat{\phi}}(g)+\frac{c_\tau}{\sqrt{n}}\right]\middle|D_n\right) = 1-\tau;$$

(ii) Frequentist coverage of the identified set for scalar functions: as $n\to\infty$

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

(iii) Frequentist coverage of scalar functions: as $n\to\infty$

$$\inf_{g(\theta)\in G(\phi_0)} P_{D_n}\left(g(\theta)\in\widehat{\Omega}_g\right)\geq 1-\tau-o_P(1),$$

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

Assumption 4.1 essentially requires that the support function has to be smooth respect to the DGP of observed data. Models of this type lead to interesting implications. In such models, Kaido and Santos (2014) showed that semiparametric efficient estimations can be achieved for the identified set. We refer to Kaido and Santos (2014) for more practical examples where Assumption 4.1 is satisfied. In the current context, it remains to verify this high-level Assumption. We shall verify it in two setups: (1) in Section 4.1, we verify it in the one-dimensional case where the identified set is a closed interval; (2) in Section 4.2 , we verify it in the multi-dimensional case, where the support function does not necessarily have an analytical form.

4.1. Verifying the LLA in the one-dimensional case

Consider the case

$$\Theta(\phi) = [h_1(\phi), h_2(\phi)]\subset\Theta\subset\mathbb{R},$$

where h_1, h_2 are known functions taking values in $\mathbb{R}$. In particular, we allow $\sup_\phi|h_1(\phi)-h_2(\phi)| = o(1)$ as a drifting parameter sequence that converges to zero. This drifting sequence depends on n and h_1, h_2 can be indexed by n, with

[1] The result presented here is understood as: There is a random sequence $\Delta(D_n)$ that depends on D_n such that $\Delta(D_n) = o_P(1)$, and for any sampling sequence D_n, we have $P_{D_n}(\Theta(\phi_0)\subset\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}) = 1-\tau+\Delta(D_n)$.

$h_1(\cdot) = h_2(\cdot)$ as a special case. Hence, the identified set can shrink to a singleton. For notational simplicity, we suppress the subscript n.

It is easy to verify that the support function is given by: $S_\phi(1) = h_2(\phi)$, and $S_\phi(-1) = -h_1(\phi)$. Hence, the critical values q_τ and $\tilde{q}_\tau$ are obtained from the posteriors of

$$J(\phi) := \sqrt{n} \sup_{\|\nu\|=1} \left(S_\phi(\nu) - S_{\hat{\phi}}(\nu)\right) = \sqrt{n} \max\{h_2(\phi) - h_2(\hat{\phi}), h_1(\hat{\phi}) - h_1(\phi)\}$$

and

$$\tilde{J}(\phi) := \sqrt{n} \sup_{\|\nu\|=1} |S_\phi(\nu) - S_{\hat{\phi}}(\nu)| = \sqrt{n} \max\{|h_2(\phi) - h_2(\hat{\phi})|, |h_1(\hat{\phi}) - h_1(\phi)|\}.$$

We now provide a primitive condition to verify Assumption 4.1 in this case.

Assumption 4.2. $h_1(\phi)$ and $h_2(\phi)$ are continuously differentiable on a closed neighborhood of ϕ_0.

We have the following proposition.

Proposition 4.1. *In the one-dimensional setup, Assumption 4.2 implies Assumption 4.1.*

Remark 4.1. Sometimes the functions h_1 and h_2 may be only partially known, up to an additional infinite-dimensional parameter η, representing the unknown (but identifiable) DGP. Then we can write them as $h_1(\phi, \eta)$, $h_2(\phi, \eta)$, or $h_1(F)$, $h_2(F)$, where F denotes the data distribution. Our method can be adapted to cover this case as well by defining $\tilde{\phi} = (\phi, \eta)$, and impose a semi-(non) parametric prior on it. When $\tilde{\phi}$ is infinite-dimensional, an LLA similar to that of Assumption 4.1 can still be verified.

4.1.1. A two-player entry game

We consider the entry game in Tamer (2003), Ciliberto and Tamer (2009). In this example, we show that while the identified set for the entire parameter may be non-convex, if we are interested in one of the components, its marginal identified set is a closed interval that satisfies Assumption 4.2, and thus its support function satisfies the high-level Assumption 4.1 due to Proposition 4.1.

Suppose there are two players: firm 1 and firm 2. Firm j $(= 1, 2)$ makes an entry decision and either does not enter market i, operates as a monopolist, or operates as a duopolist, depending on the entry decision of the competing firm. We use the notation of Moon and Schorfheide (2012) to model the potential monopoly (M) and duopoly (D) profits as:

$$\pi_{ij}^M = \beta_j + \epsilon_{ij}, \quad \pi_{ij}^D = \beta_j - \gamma_j + \epsilon_{ij}, \quad j = 1, 2, \quad i = 1, \ldots, n,$$

where we assume the parameter space to be: for some known $(\bar{\gamma}, \underline{\beta}, \bar{\beta}) \in \mathbb{R}^3$,

$$0 \le \gamma_j \le \bar{\gamma}; \quad \underline{\beta} \le \beta_j \le \bar{\beta}, \qquad j = 1, 2,$$

and both players play a pure strategy Nash equilibrium.

We assume that $(\epsilon_{i1}, \epsilon_{i2})$ are independent identically distributed, sampled from a joint distribution

$$F(x, y; \varrho) := P(\epsilon_{i1} \le x; \epsilon_{i2} \le y)$$

where the joint distribution function $F(x, y; \varrho)$ is known up to a finite dimensional parameter $\varrho \in \Theta_\varrho$; here Θ_ϱ denotes a compact parameter space for ϱ. For instance, in Moon and Schorfheide (2012), Chen et al. (2018), who also studied the two-player entry game model, it is assumed that $(\epsilon_{i1}, \epsilon_{i2})$ are jointly normally distributed with variance 1 and correlation $\varrho \in [-1 + \epsilon, 1 - \epsilon]$ for some $\epsilon > 0$. This example generalizes their studies by allowing for dependent and possibly non-Gaussian shocks.

We observe which firm enters each of the n markets, and use n_{11}, n_{00}, n_{10}, and n_{01} to denote the frequency across the n markets of: duopoly, no firm enters, monopoly of firm 1 and monopoly of firm 2, respectively. In addition, we use $\phi := [\phi_{11}, \phi_{00}, \phi_{10}]$, with $\sum_{ij} \phi_{ij} = 1$, to denote the probabilities of observing a duopoly, no entry, or the monopoly of firm 1. Then ϕ is point identified, whose maximum likelihood estimator is given by $\hat{\phi}_{lm} = \frac{n_{lm}}{n}$, $l = 0, 1, m = 0, 1$. Then, for $\hat{\phi} = [\hat{\phi}_{10}, \hat{\phi}_{11}, \hat{\phi}_{00}]$, under suitable conditions we have $\sqrt{n}(\hat{\phi} - \phi_0) \to^d \mathcal{N}(0, V)$, whose covariance matrix V is easy to obtain. It is well known that in this case the pure strategy Nash equilibrium implies (e.g., Tamer (2003)):

$$\begin{aligned}
&\phi_{00} = P(\epsilon_{i1} < -\beta_1; \epsilon_{i2} < -\beta_2), \qquad \phi_{11} = P(\epsilon_{i1} > -(\beta_1 - \gamma_1); \epsilon_{i2} > -(\beta_2 - \gamma_2))\\
&\phi_{01} \ge P(\epsilon_{i1} < -(\beta_1 - \gamma_1); \epsilon_{i2} > -(\beta_2 - \gamma_2)) + P(\epsilon_{i1} < -\beta_1; -\beta_2 < \epsilon_{i2} < -(\beta_2 - \gamma_2))\\
&\phi_{01} \le P(\epsilon_{i1} < -(\beta_1 - \gamma_1); \epsilon_{i2} > -\beta_2).
\end{aligned} \tag{4.3}$$

Given ϕ, (4.3) defines the joint identified set for the parameters $(\beta_1, \gamma_1, \beta_2, \gamma_2, \varrho)$. Suppose we are interested in the marginal identified set $\Theta_\beta(\phi)$ for β_1 and the marginal identified set $\Theta_\gamma(\phi)$ for γ_1. The inference for (β_2, γ_2) follows from the same argument. The marginal identified sets can then be characterized as:

$$\Theta_\beta(\phi) = \{\beta_1 \in [\underline{\beta}, \bar{\beta}] : \text{there are } \gamma_1, \gamma_2, \beta_2, \varrho \text{ such that (4.3) holds}\}$$
$$\Theta_\gamma(\phi) = \{\gamma_1 \in [0, \bar{\gamma}] : \text{there are } \gamma_2, \beta_1, \beta_2, \varrho \text{ such that (4.3) holds}\} \tag{4.4}$$

We make the following assumption.

Assumption 4.3. $F(x, y, \varrho)$ is twice differentiable in (x, y, ϱ). In addition, for any given Borel subset $\mathcal{B} \subset \mathbb{R}$, $P(\epsilon_{i1} \leq x; \epsilon_{i2} \in \mathcal{B})$ is strictly increasing in x, and $P(\epsilon_{i1} \in \mathcal{B}; \epsilon_{i2} \leq y)$ is strictly increasing in y.

The following lemma shows that the marginal identified sets are closed intervals that satisfy Assumption 4.2.

Lemma 4.1. *Suppose Assumption 4.3 holds. There are continuously differentiable functions $h_1(\phi), h_2(\phi), g_1(\phi), g_2(\phi)$, so that the marginal identified sets $\Theta_\beta(\phi)$ and $\Theta_\gamma(\phi)$ characterized in (4.4) are equal to*

$$\Theta_\beta(\phi) = [h_1(\phi), h_2(\phi)], \quad \Theta_\gamma(\phi) = [g_1(\phi), g_2(\phi)].$$

Therefore the LLA condition Assumption 4.1 is satisfied. In the special case where ϵ_{i1} and ϵ_{i2} are independent bivariate standard normal,

$$h_1(\phi) = \Phi_N^{-1}(\phi_{10} + \phi_{11}), \quad h_2(\phi) = -\Phi_N^{-1}\left(\frac{\phi_{00}\Phi_N(-(\underline{\beta} - \bar{\gamma}))}{\Phi_N(-(\underline{\beta} - \bar{\gamma})) - \phi_{01}}\right)$$
$$g_1(\phi) = \underline{\beta} + \Phi_N^{-1}\left(\frac{\phi_{01}\Phi_N(-\underline{\beta})}{\Phi_N(-\underline{\beta}) - \phi_{00}}\right),$$
$$g_2(\phi) = \bar{\beta} - \Phi_N^{-1}\left(\frac{-\phi_{11}\Phi_N(\bar{\beta})}{\phi_{10} - \Phi_N(\bar{\beta})}\right).$$

Here Φ_N denotes the standard normal distribution.

In the presence of general correlated shocks ($\varrho \neq 0$), these functions do not have analytic forms. But they are uniquely determined by the implicit function theorem applied to the joint distribution function $F(x, y, \varrho)$, and thus can still be easily computed. For the definitions of h_1, h_2, g_1, g_2 in the general case we refer to the proof of Lemma 4.1 in the Supplementary Appendix.

4.1.2. Marginal inference for the random design interval censored regression

Consider the interval censored regression model. We have

$$Y = x^T\theta + \epsilon, \quad Ex\epsilon = 0$$

with Y censored in the interval $[Y_1, Y_2]$ and x and θ d-vectors; x has positive support. Let $\phi_1 := Ex^TY_1$, $\phi_2 := Ex^TY_2$ and $\phi_3 := (Exx^T)^{-1}$. Then the identified set of θ is characterized by

$$\phi_1 \leq \phi_3^{-1}\theta \leq \phi_2.$$

In particular, we let ϕ_3 denote the inverse of the covariance matrix of x (assumed to exist). Using a similar argument as in Bontemps et al. (2011), it can be shown that the support function has a closed form:

$$S_\phi(\nu) = \frac{1}{2}\nu^T\phi_3(\phi_1 + \phi_2) + \frac{1}{2}\left|\nu^T\phi_3\right|(\phi_2 - \phi_1), \tag{4.5}$$

where the absolute value is taken coordinatewise. Because of the presence of the vector absolute value, $S_\phi(\nu)$ is not differentiable with respect to ϕ_3 for certain directions ν. But the proposed approach allows for marginal inferences, as we show below.

Specifically, suppose we are interested in inference about the k th component of θ, denoted by θ_k. Then the LLA condition only depends on the differentiability of a particular direction $\nu = e_k$, where e_k is the kth unit vector (with one on the kth component). Hence, it is satisfied as long as $S_\phi(e_k)$ is differentiable with respect to ϕ in a neighborhood of the true value. Let $\phi_{3,k}$ denote the kth column of ϕ_3, and $\phi_{3,k,j}$ denote the jth component of the vector $\phi_{3,k}$. Let $(\phi_{3,k}^0, \phi_{3,k,j}^0)$ denote the true value of $(\phi_{3,k}, \phi_{3,k,j})$. In addition, define the following sets:

$$\mathcal{S}_+ = \{j : \phi_{3,k,j}^0 > 0\}, \quad \mathcal{S}_- = \{j : \phi_{3,k,j}^0 < 0\},$$

which are deterministic sets because they are defined using the true values. Let $\phi_{1,j}$ and $\phi_{2,j}$ respectively denote the jth component of ϕ_1 and ϕ_2.

When all the components of the true value $\phi_{3,k}^0$ are nonzero, the following lemma shows that the marginal identified set for θ_k is a closed interval that satisfies Assumption 4.2.

Lemma 4.2. *Suppose all the components of the true value $\phi_{3,k}^0$ are nonzero. Then the marginal identified set for θ_k is given by the closed interval $[h_1(\phi), h_2(\phi)]$, where*

$$h_1(\phi) = \phi_{3,k}^T \left(\frac{\phi_1 + \phi_2}{2}\right) - M(\phi),$$

$$h_2(\phi) = \phi_{3,k}^T \left(\frac{\phi_1 + \phi_2}{2}\right) + M(\phi),$$

$$M(\phi) = \frac{1}{2}\sum_{j\in\mathcal{S}_+} \phi_{3,k,j}(\phi_{2,j} - \phi_{1,j}) - \frac{1}{2}\sum_{j\in\mathcal{S}_-} \phi_{3,k,j}(\phi_{2,j} - \phi_{1,j}).$$

Both $h_1(\phi)$ and $h_2(\phi)$ are continuously differentiable on a neighborhood of the true value for ϕ, so Assumption 4.2 is satisfied.

It is straightforward to extend the above result to the inference for a general linear functional $g^T\theta$ where g is a known vector. Then Lemma 4.2 holds with $\phi_{3,k}$ replaced by $\phi_3 g$.

In Lemma 4.2, it is crucial to assume that all the components of $\phi_{3,k}^0$ are nonzero. Below we present a relatively general example that satisfies this condition.

Example 4.1 (*Factor Models*)**.** Suppose x_i is generated from the following factor model:

$$x_i = \Lambda f_i + u_i, \quad \dim(x_i) = d,$$

where f_i is a single latent factor with $Ef_i = 0$; $\Lambda = (\lambda_1, \dots, \lambda_d)'$ is a $d \times 1$ vector of factor loadings, and the components of u_i are independent, satisfying $E(u_i|f_i) = 0$. This is then a strict factor model. We now show that all the components of ϕ_3^0 are nonzero as long as $\lambda_j \neq 0$ for all $j \leq d$, which means the common factor f_i determines all the components of x_i. The covariance of x_i is given by

$$Ex_i x_i' = \Lambda \mathrm{var}(f_i)\Lambda' + \mathrm{var}(u_i),$$

where $\mathrm{var}(f_i)$ denotes the variance of f_i and $\mathrm{var}(u_i)$ denotes the $d \times d$ covariance matrix of u_i which is assumed to be diagonal. Then by the matrix Sherman–Morrison–Woodbury identity,

$$\phi_3^0 = (Ex_i x_i')^{-1} = \mathrm{var}(u_i)^{-1} - a\mathrm{var}(u_i)^{-1}\Lambda\Lambda'\mathrm{var}(u_i)^{-1}$$

where $a = 1/(\mathrm{var}(f_i)^{-1} + \Lambda'\mathrm{var}(u_i)^{-1}\Lambda)$. Then for any $k, j \leq d$, it is straightforward to calculate that

$$\phi_{3,k,j}^0 = v_{jk} - av_{jj}v_{kk}\lambda_j\lambda_k, \quad v_{jk} = (\mathrm{var}(u_i)^{-1})_{jk}.$$

Because $v_{jk} = 0$ if and only if $j \neq k$, it is easy to see that $\phi_{3,k,j}^0 \neq 0$ as long as $\lambda_j \neq 0$ for all $j = 1, \dots, d$.

If $\phi_{3,k}^0$ contains zero components, $M(\phi)$ would have an additional non-differentiable term $\frac{1}{2}\sum_{\phi_{3,k,j}^0=0}|\phi_{3,k,j}|(\phi_{2,j} - \phi_{1,j})$. The good thing is that this is a testable statement as ϕ_3 is point identified. While this assumption should be generally satisfied, it is possible to *go around* this assumption when fixed design is considered. In this case the data for x is non-random and we can allow arbitrary $(\frac{1}{n}\sum_{i=1}^n x_i x_i^T)^{-1}$, as long as it exists. See Section 6.2.2 for detailed derivations.

4.2. LLA in the multi-dimensional case

In this section we 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. We allow the set of interest $\Theta(\phi)$ to be characterized by both equality and inequality restrictions as follows:

$$\begin{aligned}\Theta(\phi) = \{\theta \in \Theta :\ & \Psi_i(\theta, \phi) \leq 0 \text{ for } i = 1, \dots, k_1, \\ & \Psi_i(\theta, \phi) = 0 \text{ for } i = k_1 + 1, \dots, k_1 + k_2\}\end{aligned}$$

with $k_1 + k_2 = k$ and where $\{\Psi_i\}_{i=1,\dots,k}$ are known functions. Below we present a novel approach based on the implicit function theorem. Alternatively, a different approach to verify the LLA condition is the Lagrange multiplier method, employed by Kaido and Santos (2014). We present this alternative approach in the appendix.

Define the "equality boundary" as

$$\begin{aligned}\overline{\partial\Theta(\phi)} := \{\theta \in \Theta(\phi);\ & \Psi_i(\theta, \phi) = 0 \text{ for some } i \leq k_1, \text{ and} \\ & \Psi_i(\theta, \phi) = 0 \text{ for } i = k_1 + 1, \dots, k_1 + k_2\}.\end{aligned}$$

Our analysis relies on the following key fact, which holds in most applications. For every $C > 0$, $\phi \in B(\phi_0, Cn^{-1/2})$ and $\nu \in \mathbb{S}^d$, if $\Theta(\phi)$ is non-empty, it holds that

$$S_\phi(\nu) := \sup\{\nu^T\theta;\ \theta \in \Theta(\phi)\} = \sup\{\nu^T\theta;\ \theta \in \overline{\partial\Theta(\phi)}\}. \tag{4.6}$$

Eq. (4.6) says that the supremum defining the support function is achieved on the "equality boundary", that is, the support function is computed based on only equality constraints and binding inequality constraints. By Theorem 13.1 of Rockafellar (1970), $\nu^T\theta < S_\phi(\nu)$ if and only if θ belongs to the interior of the convex set $\Theta(\phi)$. Therefore, Eq. (4.6) is guaranteed to hold under the following assumption:

Assumption 4.4. For every $C > 0$, $\phi \in B(\phi_0, Cn^{-1/2})$ and $\nu \in \mathbb{S}^d$, if $\Theta(\phi)$ is non-empty, it holds that

$$\overline{\partial\Theta(\phi)} \subset \partial\Theta(\phi),$$

where $\partial\Theta(\phi)$ is the regular notation of "boundary" for a connected set.[2]

Assumption 4.4 says that the equality boundary should be a subset of the usual boundary, that is, the inequalities are binding only on the boundary. This condition is not stringent and is satisfied by many examples of interest, like for instance the interval regression model, examples where we are interested in a one-dimensional transformation of the partially identified parameter, the missing data example, the two-player entry game mode, and the financial asset pricing model of Example 2.2. The latter is discussed in more detail in Appendix B.

Above all, Eq. (4.6) allows us to use the implicit function theorem (IFT) to derive the local linear approximation for the support function. Define

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

as the intersection of the equality boundary and the *support set* of $\Theta(\phi)$. For every $\nu \in \mathbb{S}^d$ and every $\phi \in \Phi$, define the subset $\mathcal{I}_{\nu,\phi}$ of indices of the constraints that are active for this particular direction ν and this particular ϕ (this includes both binding inequality constraints and equality constraints):

$$\mathcal{I}_{\nu,\phi} := \{i \in \{1,2,\ldots,k\};\ \Psi_i(\theta,\phi) = 0, \forall\theta \in \Xi(\nu,\phi)\}$$

and by $d_{\nu,\phi} := card(\mathcal{I}_{\nu,\phi})$ the cardinality of $\mathcal{I}_{\nu,\phi}$. We notice that if there are equality constraints then $d_{\nu,\phi} \geq k_2$. Hence, by definition of $\mathcal{I}_{\nu,\phi}$, we have:

$$\forall\nu \in \mathbb{S}^d,\ \forall\phi \in \Phi, \qquad S_\phi(\nu) = \sup\{\nu^T\theta;\ \bigcap_{i\in\mathcal{I}_{\nu,\phi}}\{\theta;\ \Psi_i(\theta,\phi) = 0\}\}.$$

Denote by $(\tilde\theta,\phi)$ a solution of the system of equations $\{\Psi_i(\theta,\phi) = 0,\ \forall i \in \mathcal{I}_{\nu,\phi}\}$ such that $\tilde\theta \in \Xi(\nu,\phi)$, that is, $\tilde\theta \in \overline{\partial\Theta(\phi)}$. The idea is that the feasible set of the support function $S_\phi(\nu)$ is now completely characterized by these $d_{\nu,\phi}$ equality constraints. We can then apply the implicit function theorem. This theorem characterizes $d_{\nu,\phi}$ elements of $\tilde\theta$ as functions of ϕ and the other components of $\tilde\theta$. Without loss of generality we solve the $d_{\nu,\phi}$ equations $\{\Psi_i(\theta,\phi) = 0,\ \forall i \in \mathcal{I}_{\nu,\phi}\}$ for the first $d_{\nu,\phi}$ components of $\tilde\theta$. Therefore, it is useful to introduce the notation $\tilde\theta_{\mathcal{I}_{\nu,\phi}} := (\tilde\theta_1,\ldots,\tilde\theta_{d_{\nu,\phi}})^T$ and $\tilde\theta_{-\mathcal{I}_{\nu,\phi}} := (\tilde\theta_{d_{\nu,\phi}+1},\ldots,\tilde\theta_d)^T$. By applying the implicit function theorem to $\{\Psi_i(\theta,\phi) = 0,\ \forall i \in \mathcal{I}_{\nu,\phi}\}$, there exist $d_{\nu,\phi}$ unique functions $h_1,\ldots,h_{d_\nu}$ defined in a neighborhood of $(\tilde\theta_{-I_{\nu,\phi}},\phi)$ such that

$$\underbrace{\begin{pmatrix}\tilde\theta_1\\ \vdots\\ \tilde\theta_{d_{\nu,\phi}}\end{pmatrix}}_{=:\tilde\theta_{\mathcal{I}_{\nu,\phi}}} = \underbrace{\begin{pmatrix}h_1(\tilde\theta_{-\mathcal{I}_{\nu,\phi}},\phi)\\ \vdots\\ h_{d_{\nu,\phi}}(\tilde\theta_{-\mathcal{I}_{\nu,\phi}},\phi)\end{pmatrix}}_{=:h(\tilde\theta_{-\mathcal{I}_{\nu,\phi}},\phi)}, \tag{4.7}$$

so that $\Psi_i(\tilde\theta,\phi) = 0$ for all $i \in \mathcal{I}_{\nu,\phi}$, if and only if $\tilde\theta = (h(\tilde\theta_{-\mathcal{I}_{\nu,\phi}},\phi),\tilde\theta_{-\mathcal{I}_{\nu,\phi}})$. By substituting these functions into the corresponding elements of $\tilde\theta$ in (4.6) we transform the constrained optimization problem into an unconstrained optimization problem:

$$\begin{aligned} S_\phi(\nu) &= \sup\{\nu^T\tilde\theta;\quad \tilde\theta \in \overline{\partial\Theta(\phi)}\}\\ &= \sup\{\nu^T\tilde\theta;\quad \Psi_i(\tilde\theta,\phi) = 0,\ \forall i \in \mathcal{I}_{\nu,\phi}\}\\ &= \sup\{\nu^T\tilde\theta;\quad \tilde\theta = (h(\tilde\theta_{-\mathcal{I}_{\nu,\phi}},\phi),\tilde\theta_{-\mathcal{I}_{\nu,\phi}}) \text{ for some } \tilde\theta_{-\mathcal{I}_{\nu,\phi}}\}\\ &= \sup_m[\nu^T_{1:d_{\nu,\phi}}h(m,\phi) + \nu^T_{d_{\nu,\phi}+1:d}m], \end{aligned} \tag{4.8}$$

where $h := (h_1,\ldots,h_{d_{\nu,\phi}})^T$, $\nu_{1:d_{\nu,\phi}}$ and $\nu_{d_{\nu,\phi}+1:d}$ respectively denote the first $d_{\nu,\phi}$ and last $d - d_{\nu,\phi}$ components of ν. This simplifies the local analyses of the corresponding value function. In the case $d_{\nu,\phi} \geq d$ then, for a given (ν,ϕ), $\tilde\theta$ is uniquely

2 $\overline{\partial\Theta(\phi)}$ is defined as the set of points in the closure of $\Theta(\phi)$ not belonging to $int(\Theta(\phi))$, the interior of $\Theta(\phi)$. To show (4.6) under this assumption, let $x \in \overline{\partial\Theta(\phi)}$ be such that $\nu^Tx = \sup\{\nu^T\theta : \theta \in \overline{\partial\Theta(\phi)}\}$, if $\nu^Tx < S_\phi(\nu)$ then Theorem 13.1 of Rockafellar (1970) implies $x \in int(\Theta(\phi))$, contradicting with Assumption 4.4.

determined as a function of ϕ and $d_{\nu,\phi} - d$ restrictions are redundant. Therefore, one can eliminate the $d_{\nu,\phi} - d$ redundant constraints (when $d_{\nu,\phi} > d$) and apply the implicit function theorem to solve for $\tilde{\theta}$.

We need the following assumptions for the implicit function theorem. Denote $r_n := n^{-1/2}$ and $\Psi(\theta,\phi) := (\Psi_1(\theta,\phi),\ldots,\Psi_k(\theta,\phi))^T$.

Assumption 4.5. For every $C > 0$, $\nu \in \mathbb{S}^d$, and $\phi \in B(\phi_0, Cr_n)$ so that $\Theta(\phi)$ is non-empty, let $(\tilde{\theta},\phi)$ denote a solution to the system of equations $\{\Psi_i(\theta,\phi) = 0, \text{ for all } i \in \mathcal{I}_{\nu,\phi}\}$ such that $\tilde{\theta} \in \Xi(\nu,\phi)$. For all such $(\tilde{\theta},\phi)$ and ν the following holds:

(i) there exists a closed neighborhood of $(\tilde{\theta},\phi)$ on which $\Psi_i(\cdot,\cdot)$ is continuously twice differentiable for all $i \in \mathcal{I}_{\nu,\phi}$;
(ii) the following Jacobian determinant is nonzero:

$$|J(\phi,\nu)| := \left|\left\{\frac{\partial \Psi_i(\tilde{\theta},\phi)}{\partial \theta^T_{\mathcal{I}_{\nu,\phi}}},\ i \in \mathcal{I}_{\nu,\phi}\right\}\right| \neq 0. \tag{4.9}$$

This assumption can be easily checked since the functions Ψ_i are known. Conditions *(i)* and *(ii)* in the assumption are the classical conditions in the statement of the implicit function theorem. If $\{\Psi_i(\cdot,\cdot)\}_{i=1}^k$ are continuously twice differentiable on $\Theta \times \Phi$ then Assumption 4.5 *(i)* is automatically verified. Moreover, to check Assumption 4.5 *(ii)* one can just consider the values of $\tilde{\theta}$ for which the Jacobian determinant is zero and check whether they are solutions to the system of equations $\{\Psi_i(\theta,\phi) = 0, \text{ for all } i \in \mathcal{I}_{\nu,\phi}\}$.

Under Assumption 4.5 the implicit function theorem holds, which is used in the proof of Theorem 4.4 together with the following assumption.

Assumption 4.6.

(i) The true parameter value 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 $\mathbb{R}^d$).
(iii) The sets $\Theta(\phi)$ and $\{\theta \in \Theta : \Psi_i(\theta,\phi) \le 0, \text{ for } i = 1,\ldots,k_1\}$ are convex;
(iv) The functions $(\theta,\phi) \mapsto \Psi(\theta,\phi)$ are continuous $\forall \phi \in B(\phi_0, r_n)$;
(v) $\mathcal{I}_{\nu,\phi} = \mathcal{I}_{\nu,\phi_0}$ for all $\nu \in \mathbb{S}^d$ and $\forall \phi \in B(\phi_0, r_n)$.

Convexity of the set $\{\theta \in \Theta : \Psi_i(\theta,\phi) \le 0, \text{ for } i = 1,\ldots,k_1\}$ (Assumption 4.6 *(iii)*) is used to prove that the correspondence $\phi \mapsto \Theta(\phi)$ is lower hemicontinuous at any $\phi \in B(\phi_0, r_n)$, which we need to prove an intermediate step in the proof of Theorem 4.4. Moreover, convexity of $\Theta(\phi)$, together with the continuity Assumption 4.6 *(iv)*, guarantees that $\Theta(\phi)$ is completely characterized by the support function of $\Theta(\phi)$. Finally, condition (v) requires that the index of binding inequality constraints and equality constraints, $\mathcal{I}_{\nu,\phi}$, does not depend on ϕ in the shrinking neighborhood of $\phi = \phi_0$, so that this index set is locally "continuous" with respect to ϕ. This assumption is needed to ensure that the support function is locally smooth with respect to ϕ.

Theorem 4.4. *Let Assumptions 4.4–4.6 be satisfied, and denote $\nu_{1:d_{\nu,\phi_0}} := (\nu_1,\ldots,\nu_{d_{\nu,\phi_0}})^T$. Moreover, for every $\theta \in \Xi(\nu,\phi_0)$ let $M_1(\theta,\nu)$ be the $d_{\nu,\phi_0} \times d_{\nu,\phi_0}$ matrix, whose elements are given by:*

$$\frac{\partial \Psi_i(\theta,\phi_0)}{\partial \theta_{\mathcal{I}_{\nu,\phi_0},i'}}$$

for $(i,i') \in \mathcal{I}_{\nu,\phi_0}$, and let $M_2(\theta,\nu)$ be a $d_{\nu,\phi_0} \times d_\phi$ matrix whose elements are given by

$$\frac{\partial \Psi_i(\theta,\phi_0)}{\partial \phi_j}$$

for $i \in \mathcal{I}_{\nu,\phi_0}$ and $j = 1,\ldots,d_\phi$. Then, Assumption 4.1 holds with

$$A(\nu) = \max_{\theta \in \Xi(\nu,\phi_0)} -\nu^T_{1:d_{\nu,\phi_0}} M_1(\theta,\nu)^{-1} M_2(\theta,\nu).$$

Remark 4.2. To illustrate this theorem and explain in detail the notations introduced, in Section 6.2.2, we study the fixed design interval regression model as a concrete example.

The following algorithm describes how one can compute $A(\nu)$ as defined in the previous theorem.

Algorithm 4: Computing $A(\nu_j)$ as an Alternative to the Lagrange Multiplier Approach

1. Fix a prior $\pi(\phi)$, and construct the posterior of ϕ. Let $\{\phi_i\}_{i\le M}$ be the MCMC draws from the posterior of ϕ. Let $\hat{\phi} = \frac{1}{M}\sum_{i=1}^M \phi_i$. In addition, uniformly generate $\{\nu_j\}_{j\le G}$ such that $\|\nu_j\| = 1$ for each j.
2. (outer-loop):

2.1. For each $j \leq G$, solve one of the following constrained convex problems (depending on whether $k_2 = 0$ or $k_2 \neq 0$)

- if $k_2 = 0$, $\max_{\theta} \nu_j^T \theta$ subject to $\Psi_i(\theta, \hat{\phi}) \leq 0$ for $i = 1, \ldots, k_1$
- if $k_2 \neq 0$ $\max_{\theta \in \Theta(\hat{\phi})} \nu_j^T \theta$

and obtain $\Xi(\nu_j, \hat{\phi})$ as the set of the maximizers and the set $\mathcal{I}_{\nu_j, \hat{\phi}}$ of indices of binding inequality constraints and equality constraints at this particular value $(\nu_j, \hat{\phi})$.

2.2. Uniformly generate $\{\theta_b\}_{b \leq B} \in \Xi(\nu_j, \hat{\phi})$. For each $b \leq B$, calculate

$$A(\theta_b, \nu_j) := -\nu_{1:d_{\nu_j,\hat{\phi}}}^T \widehat{M}_1(\theta_b, \nu_j)^{-1} \widehat{M}_2(\theta_b, \nu_j)$$

where $\widehat{M}_1(\theta_b, \nu_j)$ is the $d_{\nu,\hat{\phi}} \times d_{\nu,\hat{\phi}}$ matrix, whose elements are given by:

$$\left. \frac{\partial \Psi_i(\theta, \hat{\phi})}{\partial \theta_{\mathcal{I}_{\nu,\hat{\phi}}, i'}} \right|_{\theta = \theta_b}, \qquad \text{for } (i, i') \in \mathcal{I}_{\nu,\hat{\phi}},$$

and $\widehat{M}_2(\theta_b, \nu_j)$ is the $d_{\nu,\hat{\phi}} \times d_\phi$ matrix whose elements are given by

$$\left. \frac{\partial \Psi_i(\theta, \hat{\phi})}{\partial \phi_j} \right|_{\theta = \theta_b}, \qquad \text{for } i \in \mathcal{I}_{\nu,\hat{\phi}} \text{ and } j = 1, \ldots, d_\phi.$$

Take

$$A(\nu_j) = \max_{b \leq B} A(\theta_b, \nu_j).$$

If $\Xi(\nu_j, \hat{\phi})$ is a singleton, then this step only requires to be calculated $B = 1$ times.

3. (inner-loop): For each $i \leq M$, let

$$J_i = \sqrt{n} \max_{j \leq G} \left\{ A(\nu_j)^T [\phi_i - \hat{\phi}] \right\}, \quad \tilde{J}_i = \sqrt{n} \max_{j \leq G} \left\{ |A(\nu_j)^T [\phi_i - \hat{\phi}]| \right\}.$$

4. Let q_τ and $\tilde{q}_\tau$ be the $(1 - \tau)$th quantile of $\{J_i\}_{i \leq M}$ and $\{\tilde{J}_i\}_{i \leq M}$, respectively.

5. Bernstein–von mises theorem of the posterior of $S_\phi(\nu)$

In this section we state the Bernstein–von Mises (BvM) theorem for the posterior distribution of the support function. It establishes convergence, in TV norm, of the posterior of the support function to a normal distribution as $n \to \infty$. This theorem is valid under the assumption that a BvM theorem holds for the posterior distribution of the identified parameter ϕ (Assumption 2.1 *(ii)*). We denote by $P_{\sqrt{n}(S_\phi(\nu) - S_{\phi_0}(\nu))|D_n}$ the posterior distribution of $\sqrt{n}(S_\phi(\nu) - S_{\phi_0}(\nu))$.

Theorem 5.1 (*BvM*)**.** *Let Assumption 2.1 (i)-(ii) and 4.1 hold. Then for any $\nu \in \mathbb{S}^d$,*

$$\|P_{\sqrt{n}(S_\phi(\nu) - S_{\phi_0}(\nu))|D_n} - \mathcal{N}(\bar{\Delta}_{n,\phi_0}(\nu), \bar{I}_0^{-1}(\nu))\|_{TV} \to^P 0, \tag{5.1}$$

as $n \to \infty$, where $\bar{\Delta}_{n,\phi_0}(\nu) := A(\nu)^T \Delta_{n,\phi_0}$, $\bar{I}_0^{-1}(\nu) := A(\nu)^T I_0^{-1} A(\nu)$ and $\nu \mapsto A(\nu)$ is as defined in Assumption 4.1.

The quantities in Theorem 5.1 can be estimated by replacing ϕ_0 and $A(\nu)$ by any consistent estimator $\hat{\phi}$ and $\hat{A}(\nu)$. If one adopts the Lagrange multipliers approach described in Appendix E to verify Assumption 4.1, then the semiparametric efficiency bound for estimating the support function is known for this case and has been established by Kaido and Santos (2014). Thus, in this case it can be seen that our Bayesian estimation of the support function is asymptotically semiparametric efficient in the sense of Bickel et al. (1993), since the posterior asymptotic variance $\bar{I}_0^{-1}$ achieves the semiparametric efficiency bound derived in Kaido and Santos (2014).

Let $\mathcal{C}(\mathbb{S}^d)$ be the space of bounded continuous functions on $\mathbb{S}^d$ equipped with the supremum norm $\|f\|_\infty := \sup_{\nu \in \mathbb{S}^d} |f(\nu)|$. When ϕ is interpreted as a random variable drawn from its posterior distribution, the support function $S_\phi(\cdot)$ is a stochastic process with realizations in $\mathcal{C}(\mathbb{S}^d)$. For this process, a *weak* BvM theorem holds with respect to the weak topology. More precisely, let $\mathbb{G}$ be a Gaussian measure on $\mathcal{C}(\mathbb{S}^d)$ with mean function $\bar{\Delta}_{n,\phi_0}(\cdot) = A(\cdot)^T \Delta_{n,\phi_0}$ and covariance operator with kernel: $\forall \nu_1, \nu_2 \in \mathbb{S}^d$

$$\bar{I}_0^{-1}(\nu_1, \nu_2) = A(\nu_1)^T I_0^{-1} A(\nu_2)$$

where $\nu \mapsto A(\nu)$ is as defined in Assumption 4.1. We then have the following theorem. For a set B in $\mathcal{C}(\mathbb{S}^d)$, denote by ∂B the boundary set of B, namely, the closure of B minus its interior (with respect to the metric $\|\cdot\|_\infty$).

Table 1
Frequentist coverages over 2000 replications, $1-\tau=0.95$, prior for ϕ_1,ϕ_2 is Beta(α,β).

α	β	$P_{D_n}(\Theta(\phi_0)\subset\Theta(\hat{\phi})^{q_\tau/\sqrt{n}})$ $n=50$	$n=100$	$n=500$	$\inf_{\theta\in\Theta(\phi_0)}P_{D_n}(\theta\in\widehat{\Omega})$ $n=50$	$n=100$
1	1	0.929	0.948	0.950	0.988	0.984
1	0.1	0.912	0.950	0.956	0.980	0.984
0.1	1	0.916	0.948	0.950	0.992	0.988
0.1	0.1	0.938	0.944	0.952	0.980	0.980

Theorem 5.2 (*Weak BvM*). *Let $\mathcal{B}$ be the class of Borel measurable sets in $\mathcal{C}(\mathbb{S}^d)$ such that $\mathbb{G}(\partial B)=0$. Under the assumptions of Theorem 5.1, for each fixed $B\in\mathcal{B}$,*

$$P_{\sqrt{n}(S_\phi(\cdot)-S_{\phi_0}(\cdot))|D_n}(B)\to^P\mathbb{G}(B). \tag{5.2}$$

The difference between the convergence results in the previous Theorem 5.1 and in Theorem 5.2 is that in the latter the support function is considered as a stochastic process. On the other hand, Theorem 5.1 is stronger because the convergence is in the TV norm while the result in Theorem 5.2 only establishes weak convergence.

6. Simulations

6.1. Missing data

This section illustrates the coverage of the BCS's constructed in Section 3 in the missing data problem. Let Y be a binary variable, indicating whether a treatment is successful ($Y=1$) or not ($Y=0$). The variable Y is observed subject to missing. We write $M=0$ if Y is missing, and $M=1$ otherwise. Hence, we observe (M,MY). The parameter of interest is $\theta=P(Y=1)$. The identified parameters are denoted by $\phi_1=P(M=1)$ and $\phi_2=P(Y=1|M=1)$. Let $\phi_0=(\phi_{10},\phi_{20})$ be the true value of $\phi=(\phi_1,\phi_2)$. Then, without further assumptions on $P(Y=1|M=0)$, θ is only partially identified on $\Theta(\phi)=[\phi_1\phi_2,\phi_1\phi_2+1-\phi_1]$. The support function is

$$S_\phi(1)=\phi_1\phi_2+1-\phi_1,\qquad S_\phi(-1)=-\phi_1\phi_2.$$

Suppose we observe i.i.d. data $\{(M_i,Y_iM_i)\}_{i=1}^n$, and define $\sum_{i=1}^n M_i=:n_1$ and $\sum_{i=1}^n Y_iM_i=:n_2$. The likelihood function is given by $l_n(\phi)\propto\phi_1^{n_1}(1-\phi_1)^{n-n_1}\phi_2^{n_2}(1-\phi_2)^{n_1-n_2}$.

We place independent beta priors, Beta(α_1,β_1) and Beta(α_2,β_2), on (ϕ_1,ϕ_2). Then the posterior of (ϕ_1,ϕ_2) is a product of Beta$(\alpha_1+n_1,\beta_1+n-n_1)$ and Beta$(\alpha_2+n_2,\beta_2+n_1-n_2)$.

6.1.1. Bayesian credible sets

We now construct the BCS for $\Theta(\phi)$. The estimator $\hat{\phi}$ is taken to be the posterior mode: $\hat{\phi}_1=(n_1+\alpha_1-1)/(n+\alpha_1+\beta_1-2)$, and $\hat{\phi}_2=(n_2+\alpha_2-1)/(n_1+\alpha_2+\beta_2-2)$. Then $J(\phi)=\sqrt{n}\max\left\{\phi_1\phi_2-\phi_1-\hat{\phi}_1\hat{\phi}_2+\hat{\phi}_1,-\phi_1\phi_2+\hat{\phi}_1\hat{\phi}_2\right\}$. Let q_τ be the $1-\tau$ quantile of the posterior of $J(\phi)$, which can be obtained by simulating from the Beta distributions. The $1-\tau$ level BCS for $\Theta(\phi)$ is

$$\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}=\left[\hat{\phi}_1\hat{\phi}_2-q_\tau/\sqrt{n},\hat{\phi}_1\hat{\phi}_2+1-\hat{\phi}_1+q_\tau/\sqrt{n}\right],$$

which is also the asymptotic $1-\tau$ frequentist confidence set of the true $\Theta(\phi_0)$.

We can also construct the confidence set for θ based on the posterior of ϕ, using Algorithm 2. Here we present a simple simulated example, where the true ϕ is $\phi_0=(0.7,0.5)$. This implies the true identified interval to be $\Theta(\phi_0)=[0.35,0.65]$ and about thirty percent of the simulated data are "missing". We set $\alpha_1=\alpha_2=:\alpha$, $\beta_1=\beta_2=:\beta$ in the prior. In addition, $B=1000$ posterior draws $\{\phi_i\}_{i=1}^B$ are sampled from the posterior Beta distribution. For each of them, compute $J(\phi_i)$ and set $q_{0.05}$ as the 95% upper quantile of $\{J(\phi_i)\}_{i=1}^B$ to obtain the critical values. Each simulation is repeated 2000 times.

Table 1 presents the results for different values of α, β and n. We see that the frequentist coverage probability for the set is close to the desired 95% when sample size increases. This confirms the results of our Theorem 4.2. In addition, the frequentist coverage of θ is significantly higher than the nominal level. This result is expected: the critical value for the set is exact, making the coverage probability approximately equal to the nominal level. But the critical value for the partially identified parameter is conservative, leading to conservative coverages.

6.1.2. When the set parameter "shrinks" to a singleton

We now illustrate the case when the identified set "shrinks" to a singleton. Let the true ϕ_{10} be $\phi_{10}=1-\Delta_n$ with $\Delta_n\to 0$, that is, the probability of missing is close to zero. We set $\phi_{20}=0.5$. This case is interesting because, given that $\Theta(\phi)=[\phi_1\phi_2,\phi_1\phi_2+1-\phi_1]$ and ϕ_1 represents the probability of "non-missing", letting the length of the identified set shrink to zero corresponds to letting ϕ_1, the probability of non-missing, converging to one. Our results still hold when $P(Y=1)$ is nearly identifiable.

Table 2
Frequentist coverages under near identifiability, $1-\tau=0.95$, prior for ϕ_1,ϕ_2 is Beta(α,β).

		$P_{D_n}(\Theta(\phi_0)\subset\Theta(\hat{\phi})^{q_\tau/\sqrt{n}})$			$\inf_{\theta\in\Theta(\phi_0)}P_{D_n}(\theta\in\widehat{\Omega})$		
				Δ_n			
α	β	0.1	0.05	0	0.1	0.05	0
				$n=100$			
1	1	0.945	0.953	0.954	0.990	0.994	0.992
1	0.1	0.934	0.944	0.935	0.986	0.992	0.988
0.1	1	0.952	0.951	0.950	0.990	0.990	0.986
0.1	0.1	0.938	0.936	0.937	0.986	0.982	0.978
				$n=500$			
1	1	0.945	0.949	0.951	0.984	0.990	0.992
1	0.1	0.962	0.941	0.949	0.986	0.986	0.980
0.1	1	0.955	0.945	0.949	0.986	0.990	0.970
0.1	0.1	0.946	0.948	0.956	0.990	0.986	0.974

The length of the true identified set is Δ_n. The model achieves identifiability when $\Delta_n=0$. We see that the coverage of the identified set is nearly "exact", while the coverage of the partially identified parameter is conservative.

The frequency of coverage over 2000 replications are summarized in Table 2. The results continue to be as expected: the BCS with 95% credible level has the coverage probability for the true set $\Theta(\phi_0)$ close to 0.95 even for Δ_n very small. This case has also been considered in Chen et al. (2018) who also propose a non conservative procedure to construct confidence sets for $\Theta(\phi)$. On the other hand, the coverage of the partially identified parameter is as conservative as in the "nonshrinking" case.

6.2. Marginal inference for fixed design interval regression

We simulate a fixed design interval regression model. The model is given by linear constraints

$$X^T(EY_1)\le X^TX\theta\le X^T(EY_2),$$

where X is a $n\times d$ fixed design matrix and Y_1 and Y_2 are $n\times 1$ random vectors. Suppose the full parameter θ is high-dimensional, and we are interested in the first component θ_1. Let $\phi_1=\frac{1}{n}X^TEY_1$ and $\phi_2=\frac{1}{n}X^TEY_2$. Then the identified set for θ is given by $\{\theta:\phi_1\le\frac{1}{n}X^TX\theta\le\phi_2\}=\{(\frac{1}{n}X^TX)^{-1}\zeta:\phi_1\le\zeta\le\phi_2\}$, where we assume $\frac{1}{n}X^TX$ is nonsingular. Using a similar argument as in Bontemps et al. (2011), it can be shown that the support function has a closed form (proved in Lemma G.1 in the Supplementary Appendix): write $\phi:=(\phi_1^T,\phi_2^T)^T$,

$$S_\phi(\nu)=\frac{1}{2}\nu^T\left(\frac{1}{n}X^TX\right)^{-1}(\phi_1+\phi_2)+\frac{1}{2}\left|\nu^T\left(\frac{1}{n}X^TX\right)^{-1}\right|(\phi_2-\phi_1), \tag{6.1}$$

where the absolute value is taken coordinatewise. The support function is linear in ϕ and the LLA assumption (Assumption 4.1) is satisfied with

$$A(\nu)^T=\frac{1}{2}\left(\nu^T(\frac{1}{n}X^TX)^{-1}-|\nu^T(\frac{1}{n}X^TX)^{-1}|,\nu^T(\frac{1}{n}X^TX)^{-1}+|\nu^T(\frac{1}{n}X^TX)^{-1}|\right).$$

6.2.1. Simulation results

In the simulation below, we are interested in the first component θ_{01} but $\dim(\theta_0)>1$. The true (unknown) distribution for the DGP is $Y_{1i}\sim\mathcal{N}(0,1)$, $Y_{2i}=5+Y_{1i}$ and each component of X_i is generated uniformly from $[0,1]$. Define $Z_{ji}:=X_iY_{ji}$, and $\Sigma:=\frac{1}{n}X^TX$. Then, $\frac{1}{n}\sum_{i=1}^n Z_{ji}\sim\mathcal{N}(\phi_j,\frac{1}{n}\Sigma)$, where $j=1,2$. We impose a Gaussian prior $\phi_1,\phi_2\sim\mathcal{N}(0,I\sigma_0^2)$ where the pre-specified prior variance $\sigma_0^2>0$ measures the informativeness of the prior. Then it is well known that the posterior of ϕ_j is also Gaussian with mean $\sigma_0^2(\sigma_0^2I+\frac{1}{n}\Sigma)^{-1}\frac{1}{n}\sum_{i=1}^n Z_{ji}$ and covariance $\sigma_0^2(\sigma_0^2I+\frac{1}{n}\Sigma)^{-1}\frac{1}{n}\Sigma$.

The BCSs are constructed using Algorithms 3'. The support function has a closed form. Then even if $\dim(\theta)$ is large, for each drawn ϕ_i, it is sufficient and very easy to compute the quantile of

$$\{\sqrt{n}\max_{\nu=\pm1}\left(S_{\phi_i}(e_1\nu)-S_{\hat{\phi}_i}(e_1\nu)\right):i=1,\dots,B\}$$

based on the posterior draws $\{\phi_i\}_{i\le B}$, where $e_1=(1,0,\dots,0)^T$. The algorithms run very fast. On average, it took about eight seconds to finish 1000 replications on our computer with 2.3 GHz Intel Core i7 CPU.

Our numerical example of the interval regression model takes advantage in terms of computational burden of both the fact that it is about a marginal inference problem whose essential dimension is one and the fact that the support function has a closed form solution.

The results are reported in Table 3. We use $\widehat{\Omega}_1$ to denote the confidence interval for the first component of θ_0. In most cases, the coverage probabilities are close to the nominal level. We summarize the main numerical findings from this table as follows.

Table 3
Frequentist coverages of the first component of θ and of its identified set $G(\phi)$ over 1000 replications, $1-\tau=0.95$; the prior variance is σ_0^2.

	prior variance	$\dim(\theta_0)$	$n=50$	$n=100$	$n=300$
$P_{D_n}(G(\phi_0)\subset G(\hat{\phi})^{q_\tau/\sqrt{n}})$	5	2	0.943	0.939	0.950
		5	0.925	0.948	0.939
		10	0.905	0.938	0.944
$\inf_{\theta_1\in G(\phi_0)} P_{D_n}(\theta_1\in\widehat{\Omega}_1)$		2	0.962	0.960	0.955
		5	0.947	0.964	0.940
		10	0.938	0.952	0.947
$P_{D_n}(G(\phi_0)\subset G(\hat{\phi})^{q_\tau/\sqrt{n}})$	50	2	0.942	0.940	0.953
		5	0.945	0.951	0.948
		10	0.947	0.954	0.947
$\inf_{\theta_1\in G(\phi_0)} P_{D_n}(\theta_1\in\widehat{\Omega}_1)$		2	0.961	0.960	0.959
		5	0.955	0.959	0.953
		10	0.956	0.965	0.956

1. The less informative prior (larger σ_0^2) in general yields higher frequentist coverage probabilities. In particular, the identified set is slightly under-covered when the prior variance is small. This is shown by the calculated $P_{D_n}(G(\phi_0)\subset G(\hat{\phi})^{q_\tau/\sqrt{n}})$ when the prior variance equals 5. This observation is not surprising, as less informative prior often results in wider confidence intervals.
2. The coverage of the marginal partially identified parameter, as calculated by $\inf_{\theta_1\in G(\phi_0)} P_{D_n}(\theta_1\in\widehat{\Omega}_1)$, is slightly higher than the coverage of the identified set.
3. Even when the overall dimension is relatively large, e.g., $\dim(\theta_0)=10$, the constructed confidence intervals do not show a noticeable conservative coverage. This illustrates an appealing feature of our constructed marginal confidence intervals similar to Kaido et al. (2019), Bugni et al. (2017), Chen et al. (2018). As explained by Kaido et al. (2019), the usual projection procedure from a high-dimensional confidence interval often results in severe over-coverage.

6.2.2. *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. Formal derivations are given in Lemmas G.1–G.4 in the Supplementary Appendix.

First, define

$$m(\nu):=\Sigma^{-1}\nu,\quad \Sigma:=\frac{1}{n}X^TX.$$

Then the support function has an analytical form:

$$S_\phi(\nu)=\sup_{\theta\in\Psi(\theta,\phi)\le 0}\nu^T\theta=\frac{1}{2}m(\nu)^T(\phi_1+\phi_2)+\frac{1}{2}\left|m(\nu)^T\right|(\phi_2-\phi_1), \tag{6.2}$$

and the supremum is achieved by $\widetilde{\theta}(\nu,\phi,\alpha):=\Sigma^{-1}\xi^*_{\nu,\phi,\alpha}$ for any $\alpha\in\mathbb{R}^{\dim(\theta)}$, where

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

here $\circ$ is the component-wise product of two vectors. It is straightforward to see that indeed $\widetilde{\theta}(\nu,\phi,\alpha)$ belongs to the equality boundary $\overline{\partial\Theta(\phi)}$. To see this, note that $\widetilde{\theta}(\nu,\phi,\alpha)\in\overline{\partial\Theta(\phi)}$ if and only if there exists j so that the jth component of $\Sigma\widetilde{\theta}(\nu,\phi,\alpha)$ equals either ϕ_1 or ϕ_2. Indeed, we can find a component j so that $m_j(\nu)\neq 0$. Then the jth component of $\Sigma\widetilde{\theta}(\nu,\phi,\alpha)$ is the jth component of $\xi^*_{\nu,\phi,\alpha}$, which is either $\phi_{2,j}$ (if $m_j(\nu)>0$) or $\phi_{1,j}$ (if $m_j(\nu)<0$), so that indeed $\Psi_j(\widetilde{\theta}(\nu,\phi,\alpha),\phi)=0$. In addition, this immediately implies that the binding index $\mathcal{I}_{\nu,\phi}$ is given by:

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

where $I=\{1,\ldots,d\}$ and $II=\{d+1,\ldots,2d\}$ correspond to the indices of

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

There are three main features in this example. First, $S_\phi(\nu)$ as defined in (6.2), is continuously differentiable and linear in ϕ. Thus the LLA Assumption 4.1 is naturally satisfied with $f(\phi,\widetilde{\phi})=0$, $\forall\phi,\widetilde{\phi}\in\Phi$. Secondly, the binding index $\mathcal{I}_{\nu,\phi}$ is independent of ϕ. Third, $A(\nu)$ in the LLA expansion has an analytic form:

$$A(\nu)=\partial S_\phi(\nu)/\partial\phi\big|_{\phi=\phi_0}=\frac{1}{2}\left(m(\nu)^T-|m(\nu)^T|,m(\nu)^T+|m(\nu)^T|\right).$$

The implicit function theorem then ensures that components of $\widetilde{\theta}(\nu,\phi,\alpha)$ corresponding to $\mathcal{I}_{\nu,\phi}$, denoted by $\tilde{\theta}_{\mathcal{I}_{\nu,\phi}}$, can be determined as a function of other components, which is (4.7):

$$\tilde{\theta}_{\mathcal{I}_{\nu,\phi}} = h(\tilde{\theta}_{-\mathcal{I}_{\nu,\phi}},\phi).$$

To give an explicit expression for the function $h(\cdot)$, define the index sets $(m(\nu) > 0) := \{i : m_i(\nu) > 0\}$, and $(m(\nu) < 0) := \{i : m_i(\nu) < 0\}$. Let $\Sigma_{(m\neq 0)}$ be a square matrix that is formed by removing the rows and columns of Σ indexed by $\{j : m_j(\nu) = 0\}$. Similarly, let $\xi^*_{\nu,\phi}$ be the subvector of $\xi^*_{\nu,\phi,\alpha}$, removing the elements indexed by $\{j : m_j(\nu) = 0\}$. Hence, $\xi^*_{\nu,\phi}$ is independent of α. Then by permuting the rows and columns of Σ, and permuting the elements of $\widetilde{\theta}(\nu,\phi,\alpha)$, $\xi^*_{\nu,\phi,\alpha}$, we can obtain a permutation version of $\Sigma\widetilde{\theta}(\nu,\phi,\alpha) = \xi^*_{\nu,\phi,\alpha}$ such as:

$$\underbrace{\begin{pmatrix} \Sigma_{(m\neq 0)} & G_1 \\ G_1^T & G_2 \end{pmatrix}}_{\text{permutation of } \Sigma} \underbrace{\begin{pmatrix} \tilde{\theta}_{\mathcal{I}_{\nu,\phi}} \\ \tilde{\theta}_{-\mathcal{I}_{\nu,\phi}} \end{pmatrix}}_{\text{permutation of } \widetilde{\theta}(\nu,\phi,\alpha)} = \underbrace{\begin{pmatrix} \xi^*_{\nu,\phi} \\ \alpha_{(m(\nu)=0)} \end{pmatrix}}_{\text{permutation of } \xi^*_{\nu,\phi,\alpha}},$$

which leads to $\Sigma_{(m\neq 0)}\tilde{\theta}_{\mathcal{I}_{\nu,\phi}} + G_1\tilde{\theta}_{-\mathcal{I}_{\nu,\phi}} = \xi^*_{\nu,\phi}$. Solving for $\tilde{\theta}_{\mathcal{I}_{\nu,\phi}}$, we obtain

$$\tilde{\theta}_{\mathcal{I}_{\nu,\phi}} = \underbrace{\Sigma^{-1}_{(m\neq 0)}(\xi^*_{\nu,\phi} - G_1\tilde{\theta}_{-\mathcal{I}_{\nu,\phi}})}_{h(\tilde{\theta}_{-\mathcal{I}_{\nu,\phi}},\phi)}.$$

Finally, it is important to note that here we are concerned about the joint inference about θ (the entire vector). In this case assuming the fixed design is essential and ensures that the covariance $\Sigma := \frac{1}{n}X^TX$ is known. In random design regressions, the parameter $\phi_3 := (E\left[\frac{1}{n}X^TX\right])^{-1}$ has to be treated as an additional parameter, then $\phi = (\phi_1,\phi_2,\phi_3)$. In this case, the support function also has an analytic form:

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

However, the LLA no longer holds for $S_\phi(\nu)$, because $|\nu^T\phi_3|$ is not locally linearly approximable near the origin. But for models of random designs, we can still conduct marginal inference as we explained in Section 4.1.2.

7. Discussions

This paper proposes Bayesian inference for partially identified convex models based on the support function of the identified set. Our results have been described for a closed and convex identified set characterized by moment inequalities, but under the LLA our results hold more generally for identified sets characterized in other forms, such as the likelihood based models, as long as the set is closed, convex and with smooth boundaries.

We propose new algorithms to compute critical values used to construct BCS for inferences about the identified set, partially identified parameters, and scalar functions. While Moon and Schorfheide (2012) show that a BCS for θ does not have a correct frequentist coverage even asymptotically, we instead show that one can construct a frequentist confidence set for the partially identified parameter θ with the desired coverage once a prior is imposed directly on the identified set. Although it is intuitively clear that the Bayesian–frequentist large sample duality does not necessarily require the convexity of the identified set, we adopt a support function approach in this paper, which requires convexity. On the other hand, the convexity assumption has interesting implications for computations.

While in the paper we use relatively simple assumptions about the prior and posterior of ϕ, our analysis can be extended to cases where ϕ is infinite dimensional, though we expect that this analysis would require a more involved study of the nonparametric prior of ϕ.

Acknowledgments

This is a substantially revised version of an earlier manuscript entitled "Semi-parametric Bayesian partially identified models based on support function". The authors gratefully thank the Editors, an Associate Editor, and anonymous referees for their many constructive comments on the previous version of the paper. Anna Simoni gratefully acknowledges financial support from ANR-11-LABEX-0047 and from ANR-13-BSH1-0004 (IPANEMA).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.03.001.

References

Andrews, D.W., Guggenberger, P., 2009. Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. Econom. Theory 25 (03), 669–709.
Andrews, D., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–157.
Belloni, A., Chernozhukov, V., 2009. On the computational complexity of MCMC-based estimators in large samples. Ann. Statist. 37, 1591–2082.
Beresteanu, A., Molchanov, I., Molinari, F., 2011. Sharp identification regions in models with convex moment predictions. Econometrica 79 (6), 1785–1821.
Beresteanu, A., Molinari, F., 2008. Asymptotic properties for a class of partially identified models. Econometrica 76, 763–814.
Bickel, P., Klassen, C., Ritov, Y., Wellner, J., 1993. Efficient and Adaptive Estimation for Semi Parametric Models, first ed. Springer, New York.
Bickel, P., Kleijn, B., 2012. The semiparametric Bernstein-Von Mises theorem. Ann. Statist. 40, 206–237.
Bontemps, D., 2011. Bernstein-von mises theorems for gaussian regression with increasing number of regressors. Ann. Statist. 39, 2557–2584.
Bontemps, C., Magnac, T., 2017. Set identification, moment restrictions, and inference. Annu. Rev. Econ. 9 (1), 103–129.
Bontemps, C., Magnac, T., Maurin, E., 2011. Set identified linear models. Econometrica 80, 1129–1155.
Bugni, F.A., 2010. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. Econometrica 78 (2), 735–753.
Bugni, F.A., Canay, I.A., Shi, X., 2017. Inference for subvectors and other functions of partially identified parameters in moment inequality models. Quant. Econom. 8 (1), 1–38.
Canay, I.A., 2010. El inference for partially identified models: Large deviations optimality and bootstrap validity. J. Econometrics 156 (2), 408–425.
Casella, G., Berger, R.L., 2002. Statistical Inference, vol. 2. Duxbury Pacific Grove, CA.
Castillo, I., Rousseau, J., 2015. A Bernstein-von Mises theorem for smooth functionals in semiparametric models. Ann. Statist. 56, 2353–2383.
Chandrasekhar, A., Chernozhukov, V., Molinari, F., Schrimpf, P., 2012. Estimation and Inference for best linear approximations to set identified functions. Technical Report, Cemmap Working Papers, CWP43/12.
Chang, I.H., Kim, B.H., Mukerjee, R., 2009. Bayesian and frequentist confidence intervals via adjusted likelihoods under prior specification on the interest parameter. Statistics 43 (2), 203–211.
Chen, X., Christensen, T., Tamer, E.T., 2018. Monte Carlo confidence sets for identified sets. Econometrica 86 (6), 1965–2018.
Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243–1284.
Chernozhukov, V., Kocatulum, E., Menzel, K., 2015. Inference on sets in finance. Quant. Econom. 6 (2), 309–358.
Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77 (6), 1791–1828.
Florens, J.-P., Simoni, A., 2011. Bayesian identification and partial identification. Technical Report, Toulouse School of Economics.
Ghosal, S., Ghosh, J., Ramamoorthi, R., 1999a. Posterior consistency of Dirichlet mixtures in density estimation. Ann. Statist. 27, 143–158.
Ghosal, S., Ghosh, J., van der Vaart, A., 1999b. Consistent semiparametric Bayesian inference about a location parameter. J. Statist. Plann. Inference 77, 181–193.
Gospodinov, N., Kan, R., Robotti, C., 2010. On the Hansen-Jagannathan distance with a no-arbitrage constraint. Technical Report, Working Paper, Federal Reserve Bank of Atlanta.
Guntuboyina, A., 2012. Optimal rates of convergence for convex set estimation from support functions. Ann. Statist. 40, 385–411.
Gustafson, P., 2012. On the behaviour of bayesian credible intervals in partially identified models. Electron. J. Stat. 6, 2107–2124.
Hansen, L.P., Jagannathan, R., 1991. Implications of security market data for models of dynamic economies. J. Political Economy 99, 225–262.
Imbens, G.W., Manski, C.F., 2004. Confidence intervals for partially identified parameters. Econometrica 72, 1845–1857.
Kaido, H., Molinari, F., Stoye, J., 2019. Confidence intervals for projections of partially identified parameters. Econometrica (forthcoming), arXiv:1601.00934.
Kaido, H., Santos, A., 2014. Asymptotically efficient estimation of models defined by convex moment inequalities. Econometrica 82, 387–413.
Kato, K., 2013. Quasi-Bayesian analysis of nonparametric instrumental variables models. Ann. Statist. 41, 2359–2390.
Kitagawa, T., 2012. Estimation and inference for set-identified parameters using posterior lower probability. Technical Report, UCL.
Kline, B., Tamer, E., 2016. Bayesian inference in a class of partially identified models. Quant. Econom. 7 (2), 329–366.
Leahu, H., et al., 2011. On the Bernstein-von Mises phenomenon in the Gaussian white noise model. Electron. J. Stat. 5, 373–404.
Liao, Y., Jiang, W., 2010. Bayesian analysis in moment inequality models. Ann. Statist. 38, 275–316.
Mammen, E., Marron, J., Turlach, B., Wand, M., 2001. A general projection framework for constrained smoothing. Statist. Sci. 16, 232–248.
Manski, C., Tamer, E., 2002. Inference on regressions with interval data on a regressor or outcome. Econometrica 70, 519–547.
Moon, R., Schorfheide, F., 2009, Bayesian and frequentist inference in partially-identified models, NBER Working Paper No. w14882. Available at SSRN: https://ssrn.com/abstract=1391829.
Moon, R., Schorfheide, F., 2012. Bayesian and frequentist inference in partially-identified models. Econometrica 80, 755–782.
Norets, A., 2015. Bayesian regression with nonparametric heteroskedasticity. J. Econometrics 185, 409–419.
Norets, A., Tang, X., 2014. Semiparametric inference in dynamic binary choice models. Rev. Econom. Stud. 81, 1229–1262.
Poirier, D., 1998. Revising beliefs in nonidentified models. Econom. Theory 14, 483–509.
Rivoirard, V., Rousseau, J., 2012. Bernstein-von Mises theorem for linear functionals of the density. Ann. Statist. 40, 1489–1523.
Rockafellar, R., 1970. Convex Analysis, first ed. Princeton University Press, Princeton.
Romano, J.P., Shaikh, A.M., 2010. Inference for the identified set in partially identified econometric models. Econometrica 78 (1), 169–211.
Rosen, A., 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. J. Econometrics 146, 107–117.
Severini, T., 1991. On the relationship between Bayesian and non-Bayesian interval estimates. J. R. Stat. Soc. Ser. B Stat. Methodol. 53, 611–618.
Shen, X., 2002. Asymptotic normality of semiparametric and nonparametric posterior distributions. J. Amer. Statist. Assoc. 97 (457), 222–235.
Stoye, J., 2009. More on confidence intervals for partially identified parameters. Econometrica 77 (4), 1299–1315.
Sweeting, T.J., 2001. Coverage probability bias, objective Bayes and the likelihood principle. Biometrika 88 (3), 657–675.
Tamer, E., 2003. Incomplete simultaneous discrete response model with multiple equilibria. Rev. Econom. Stud. 70 (1), 147–165.
Van der Vaart, A.W., 2000. Asymptotic Statistics. In: Lectures on Probability Theory, Ecole d'Été de Probabilités de St. Flour XX.
van der Vaart, A., 2002. In: Bernard, P. (Ed.), "Semiparametric Statistics" in Ecole d'Eté de St Flour 1999, first ed. Springer Verlag, New York.
Walker, S., Lijoi, A., Prünster, I., 2007. On rates of convergence for posterior distributions in infinite dimensional models. Ann. Statist. 35, 738–746.
Wan, Y., 2013. An integration-based approach to moment inequality models. Technical Report, University of Toronto.