Liao, Yuan; Simoni, Anna

Working Paper

Bayesian inference for partially identified convex models: Is it valid for frequentist inference?

Working Paper, No. 2016-07

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Liao, Yuan; Simoni, Anna (2016) : Bayesian inference for partially identified convex models: Is it valid for frequentist inference?, Working Paper, No. 2016-07, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/162929

Bayesian inference for partially identified convex models: is it valid for frequentist inference?*

Yuan Liao† University of Maryland

Anna Simoni‡ CREST-Ensae and CNRS

June 7, 2016

Abstract

Inference on partially identified models plays an important role in econometrics. This paper proposes novel Bayesian procedures for these models when the identified set is closed and convex and so is completely characterized by its support function. We shed new light on the connection between Bayesian and frequentist inference for partially identified convex models. We construct Bayesian credible sets for the identified set and uniform credible bands for the support function, as well as a Bayesian procedure for marginal inference, where we may be interested in just one component of the partially identified parameter. Importantly, our procedure is shown to be an asymptotically valid frequentist procedure as well. It is computationally efficient, and we describe several algorithms to implement it. We also construct confidence sets for the partially identified parameter by using the posterior distribution of the support function and show that they have correct frequentist coverage asymptotically. In addition, we establish a local linear approximation of the support function which facilitates set inference and numerical implementation of our method, and allows us to establish the Bernstein-von Mises theorem of the posterior distribution of the support function.

JEL Classification: C11
Key words: Bernstein-von Mises theorem, partial identification, Bayesian credible sets, support function, moment inequality models

*This is a substantially revised version of earlier manuscripts entitled "Semi-parametric Bayesian partially identified models based on support function".

†University of Maryland at College Park, MD 20742, USA. Email: yuanliao@umd.edu

‡CREST, 15, Boulevard Gabriel Péri, 92240 Malakoff, France. Email: simoni.anna@gmail.com

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 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.

Moon and Schorfheide (2012) was one of the first papers that constructed the BCS for the partially identified parameter, denoted by θ in this paper. They impose a prior on θ and show that when such a prior has a non-degenerate support on 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 about the partially identified parameter even asymptotically. Their study motivates two important questions:

Question (1) : What about the identified set? More precisely, does the Bayesian credible set for the identified set have a correct frequentist coverage?

Question (2) : Can we use the posterior distribution of the identified set to construct a credible set for the partially identified parameter which also has a correct frequentist coverage?

Our major contribution in this paper is to provide affirmative answers to both these questions when the identified set is convex. In particular, in answering the second question, we aim to construct a confidence set for θ using the posterior of the identified set instead of the posterior of the partially identified parameter, in sharp contrast to Moon and Schorfheide (2012)'s approach.

To fulfill these goals, we study in detail convex identified sets defined by moment restrictions. Because of the convexity of the identified set the major tool in our analysis is the *support function*, which has been successfully used in the frequentist literature on partially identified convex models (e.g., Beresteanu and Molinari (2008); Kaido and Santos (2014)). Let $\mathcal{C}$ be a convex and closed set of interest. Its support function is defined as

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

Any closed and convex set is uniquely determined by its support function: $\theta \in \mathcal{C}$ if and only if for all $\|\nu\| = 1$, $\theta^T \nu \leq S_{\mathcal{C}}(\nu)$. When the identified set is convex, studying the support function greatly facilitates our analysis. To take advantages of the support function

approach, we focus on one-dimensional models and multi-dimensional models with convex restrictions. Specifically, we assume that the identified set is characterized by (in)equalities involving a convex function Ψ, and is given by

$$\Theta(\phi) := \{\theta \in \Theta : \Psi(\theta, \phi) \leq 0\} \tag{1.1}$$

where Θ denotes a vector space that contains the unknown set of interest. The function Ψ, provided by the econometric model, is known up to a finite dimensional parameter $\phi \in \Phi \subset \mathbb{R}^{d_\phi}$, which is point identified. Models of this type, though rule out some important cases, are still reasonably general and cover several important models, such as the multi-dimensional instrumental variable (IV) regression with interval outcomes (*e.g.*, Bontemps et al. (2011)), two-player entry games, and one-dimensional partially identified models (as e.g., Imbens and Manski (2004)). Even for these relatively simple models, the Bayesian-frequentist correspondence on the coverage properties is not sufficiently clear.

1.1 Overview of our results.

We develop a Bayesian procedure for set inferences. Differently from Moon and Schorfheide (2012), our method does not require a prior on the partially identified parameter. Instead, the prior is only imposed on the identified set. We can divide our results in two groups: one concerning the construction of BCS and their coverage properties with the aim of answering questions (1) and (2), and a second group of results concerning asymptotic properties of the posterior of the support function.

Coverage properties.

As for question (1) raised above, we propose a simple way to construct a BCS for the set $\Theta(\phi)$. An important theoretical result is that the constructed BCS not only has a correct Bayesian coverage but also covers the true set $\Theta(\phi_0)$ with correct frequentist probability (asymptotically). Therefore, while Moon and Schorfheide (2012)'s BCS for the partially identified parameter does not have a correct frequentist coverage, our constructed BCS for the identified set does.

As for question (2), we also find that the answer is affirmative as long as we use the posterior of the identified set, instead of the posterior of the partially identified parameter θ, to construct a credible set for θ. Specifically, 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 credible set as the collection of all the "accepted" θ's, and show that it also has a desired frequentist coverage for the partially identified parameter. Therefore, our results complement those of Moon and

Schorfheide (2012), and provide a complete picture of the connection between Bayes and frequentist procedures for partially identified convex models.

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. In sharp contrast, Moon and Schorfheide (2012) impose the prior directly on the partially identified parameter. Then the posterior tends to be supported only within the identified set, leading to a smaller nominal frequentist coverage.

In addition, we also show that it is straightforward to make marginal inferences using our procedure. This is appealing when we are interested in just one component of the partially identified parameter.

We emphasize that in this paper we do not aim to compare with frequentist inference procedures for partially identified models, or claim any advantage of using a Bayesian approach for set inferences. Instead, we aim to provide a complete picture of the coverage properties of Bayesian partial identification approaches, based on the support functions and convex identified sets. In addition, we also aim to provide fast algorithms that can be useful to Bayesian researchers for studying partially identified models.

Studies on the posterior of the support function.

The support function plays a central role in our analysis. The support function of $\Theta(\phi)$ is indexed by ϕ and we write: $S_\phi(\nu) := S_{\Theta(\phi)}(\nu)$. We put a prior on $S_\phi(\cdot)$ (and on $\Theta(\phi)$) via the prior on ϕ. In multi-dimensional models, the support function may not have a closed form or may depend on ϕ in a complicated way. Therefore, it is not immediate to translate the posterior properties of ϕ to the posterior of $S_\phi(\cdot)$. In this paper we show the frequentist asymptotic properties of the support function. More precisely, by denoting with ϕ_0 the true value of ϕ that generates the data, we show that, for every ν: (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). In order to show these results we need to assume that the posterior distribution of the point identified parameter ϕ has suitable asymptotic properties (which is well known to be verified in many parametric and semiparametric models). Moreover, we need to assume that the support function $\phi \mapsto S_\phi(\nu)$ can be well approximated by a linear function of ϕ in a shrinking neighborhood of ϕ_0: namely, there is a shrinking neighborhood

B of ϕ_0, and a continuous vector function $A(\nu)$ such that uniformly in $\phi_1, \phi_2 \in B$,

$$\sup_{\nu \in \mathbb{S}^d} \left| (S_{\phi_1}(\nu) - S_{\phi_2}(\nu)) - A(\nu)^T(\phi_1 - \phi_2) \right| = o(\|\phi_1 - \phi_2\|), \quad \text{as } \|\phi_1 - \phi_2\| \to 0 \tag{1.2}$$

where $\mathbb{S}^d$ denotes the unit sphere in $\mathbb{R}^d$. Roughly speaking, this means that if two sets are "close", so should be their support functions. This local linear approximation (LLA) assumption can be directly verified when the support function admits a closed-form. When this is not the case, or verification of this assumption is too complicated, we provide primitive conditions under which (1.2) holds for two cases: (i) the one-dimensional case where the identified set is a closed interval and (ii) a more general multi-dimensional case. Case (ii) requires further assumptions and, under these assumptions, we prove that Assumption 4.1 holds by exploiting the fact that $S_\phi(\nu)$ is the optimal value of an ordinary convex program, and then using the envelop theorem of Milgrom and Segal (2002). In this case, the particular form of the vector function $A(\nu)$ allows to see that the asymptotic variance of the posterior distribution of $S_\phi(\nu)$ achieves the semiparametric efficiency bound for estimating the support function derived by Kaido and Santos (2014).

The LLA of the support function also allows to considerably simplify computations and statistical inferences of the support function, especially when the latter is highly nonlinear in ϕ, or when its closed form is not available. In particular, when one wants to simulate from the posterior of the support function, the LLA avoids solving a numerical maximization problem in each step of the MCMC. With this respect, we introduce an efficient algorithm to calculate the critical values fast, which is based on Monte Carlo methods.

1.2 Related literature.

In the Bayesian partial identification literature, Moon and Schorfheide (2012) studied the coverage properties for BCSs of partially identified parameters. Kline and Tamer (forthcoming) independently wrote a paper that also provides BCS for the identified set that have the correct frequentist coverage asymptotically.[1] They also require the set to depend on a finite dimensional point-identified parameter ϕ but, differently from us, they do not discuss as to extend this assumption to allow for an infinite dimensional ϕ (see our Remark 5.1). Their Bayesian procedure differs from ours in that it does not require the convexity of the identified set and hence is not based on the support function. On the other side, they do

[1] We would like to mention that the first version of our manuscript was circulated on arXiv in December 2012 under the title "Semi-parametric Bayesian Partially Identified Models based on Support Function". This was about the same time of the first circulated version of Kline and Tamer (forthcoming) as mentioned in their paper.

not provide a general procedure for computing the critical values which instead have to be found case by case. Moreover, via the support function, we provide frequentist asymptotic results for the whole distribution of the identified set and not only for the probabilities of some specific sets. In our simulation study for missing data we show that the constructed BCS still has the correct coverage even if the identified set shrinks to a singleton. At the best of our knowledge, other Bayesian procedures tend to be conservative in this case, see (Kline and Tamer, forthcoming, Remark 8). Besides, we also study the coverage properties of the partially identified parameter, and provide an affirmative answer to the previously raised question (2). In addition, we show that it is straightforward to make marginal inferences using our procedure. Finally, the support function approach is also very attractive in linear models (*e.g.*, Bontemps et al. (2011)). The (quasi-) Bayesian literature on partial identification also includes the following contributions whose approaches are substantially different from our, *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. In addition to the references mentioned above, a partial list includes Manski (2003); Manski and Tamer (2002), Imbens and Manski (2004); Chernozhukov et al. (2007, 2015, 2013), Beresteanu and Molinari (2008); Beresteanu et al. (2012); Andrews and Guggenberger (2009); Romano and Shaikh (2010); Andrews and Soares (2010); Canay (2010); Stoye (2009); Rosen (2008); Bugni (2010); Pakes et al. (2015), 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), 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), Norets (2015).

The paper is organized as follows. Section 2 presents the model, examples, and discusses the prior on ϕ and the assumptions on it. Section 3 constructs Bayesian credible sets and bands for $\Theta(\phi)$ and $S_\phi(\nu)$, respectively, and frequentist confidence sets for θ and provides the computational MCMC-algorithms to implement this inference. Moreover, it provides marginal set inference methods for subsets of θ and $\Theta(\phi)$. Section 4 establishes the connection between Bayesian and frequentist inference by showing the frequentist validity of our Bayesian procedure. Section 5 demonstrates the LLA of the support function for the multi-dimensional case. Frequentist asymptotic properties of the posterior of the support

function are established in Section 6. Numerical simulations are in Section 7 and Section 8 concludes. All the proofs are in a Supplementary Appendix. Throughout the paper, the prior distribution will be denoted by π while the posterior distribution will be denoted by p. The argument of these functions will make clear to which variables they refer.

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 $\Psi : \Theta \times \Phi \to \mathbb{R}^k$ be a known and continuous vector-function of (θ, ϕ) that is convex in θ for every $\phi \in \Phi$. We are interested in making inference on the set

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

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

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

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. The support function plays a crucial role for our Bayesian inference on $\Theta(\phi)$. We point out that the characterization of $\Theta(\phi)$ through the vector-function Ψ is not necessary for the validity of our general results, for which only convexity and closedness of the set (plus Assumptions 2.1 and 4.1 below) is required. However, we characterize the set in this way to be coherent with the representation of $\Theta(\phi)$ in Sections 5 where this characterization is required.

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 $\phi = (\phi_1, \text{vec}(\phi_2), \phi_3)$ where $(\phi_1^T, \phi_3^T) = (E(ZY_1)^T, E(ZY_2)^T)$ and $\phi_2 = E(Zx^T)$. It then follows from (3.1) that θ belongs to the following set

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

□

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, equation (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) showed that every SDF M_{t+1} that satisfies (2.2) should have $\sigma^2 \geq \phi_1\mu^2 - 2\phi_2\mu + \phi_3$, 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 of 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. We also put

a nonparametric prior, either on F directly, or on an infinite-dimensional density function. Importantly, the identified set depends on F only through ϕ, and thus by "integrating out" F (or the density function), the posteriors of the identified set $\Theta(\phi)$ and the support function $S_\phi(\cdot)$ are deduced from the posterior $p(\phi|D_n)$. We illustrate below three different scenarios concerning the knowledge of F, the prior on it and the relation between ϕ and F.

When illustrating asymptotic properties of our Bayesian procedure we adopt a frequentist point of view, and denote by F_0 the true value of F and by ϕ_0 the true value of ϕ that generate the data. Hence, the true set and its support function will be denoted by $\Theta(\phi_0)$ and $S_{\phi_0}(\cdot)$, respectively. Moreover, P_{D_n} will denote the probability measure in the frequentist sense (*i.e.* based on the true DGP F_0).

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 proposed by Florens and Simoni (2011).

Semi-parametric prior. Write $F = F_{\phi,\eta}$, where η is 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 $\mathcal{P}$ in 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 $X = (ZY_1^T, ZY_2^T, \text{vec}(Zx^T)^T)^T$ 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 density function η. Then the likelihood is given by $l_n(\phi, \eta) = \prod_{i=1}^n \eta(X_i - \phi)$. Examples of the prior on the infinite dimensional density parameter η includes, 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 $X = (ZY_1^T, ZY_2^T, \text{vec}(Zx^T)^T)^T$ is normally distributed. Then we can parametrize 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, under very mild conditions, it is well known that 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. Moreover, we

denote by $\tilde{\psi}$ the (semiparametric) efficient influence function for estimating ϕ_0.

Assumption 2.1. *(i) The marginal posterior of ϕ is such that, for some constant $C > 0$,*

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

where $C_n = (\log n)^{1/2}$.
(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 $I_0^{-1} := E[\tilde{\psi}\tilde{\psi}^T]$.
(iii) There exists a regular estimator $\hat{\phi}$ of ϕ that satisfies

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

Remark 2.1. The quantities ℓ_{ϕ_0} and I_0 in the previous assumption have to be calculated case by case since their precise definition relies on the model under consideration, that is, on the link between ϕ and F. Both quantities are well known for many classical problems, see *e.g.* Bickel et al. (1993). In regular parametric models, where the likelihood is fully parametric, ℓ_{ϕ_0} and I_0 are the usual score function (first derivative of the likelihood) and Fisher information matrix, respectively. In semiparametric models, the matrix I_0 is the semiparametric efficient Fisher information matrix, which is implicitly assumed to be nonsingular, and $\tilde{\psi}$ is the semiparametric efficient influence function for estimating ϕ_0. In many semiparametric problems, the d_ϕ-vector ℓ_{ϕ_0} is given by the first derivative of the log-integrated likelihood of the model, where the integration is with respect to the nonparametric prior of the infinite dimensional parameter. A 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 in (semi) parametric Bayesian literature and are in general satisfied, under mild restrictions, for both nonparametric and semiparametric prior on (ϕ, F). For nonparametric priors, and $\phi := \phi(F)$ then the notation used in part *(i)* of this assumption is a shorthand for $P(\|\phi(F) - \phi(F_0)\| \leq Cn^{-1/2}(\log n)^{1/2}|D_n) \to^p 1$ and similarly for part *(ii)*. In most of the parametric settings, the sequence C_n in Assumption 2.1 *(i)* can be any diverging sequence so that the $\log n$ term is avoided. Instead, in

the semiparametric setting, the sequence C_n can be any diverging sequence under stronger primitive conditions, for instance Lemma 6.1 in Bickel and Kleijn (2012) proves Assumption 2.1 *(i)* for any $C_n \to \infty$. We refer to Belloni and Chernozhukov (2009), Bickel and Kleijn (2012), and Rivoirard and Rousseau (2012) for primitive conditions for this assumption in semiparametric models.

Assumption 2.1 *(iii)* requires the existence of a Gaussian efficient regular estimator. It holds, under mild conditions, for many estimators such as the posterior mean, mode and the maximum likelihood estimator.

Remark 2.2. Assumption 2.1 can be weakened by allowing a general limiting distribution and a rate of convergence different from $n^{-1/2}$ as long as the two limiting distributions in Assumption 2.1 *(ii)* and *(iii)* are the same. Thus, Assumption 2.1 is a special case of the following, more general, assumption.

Assumption 2.1′. *Let r_n be a deterministic sequence converging to zero, $r_n = o(1)$.*
(i) The marginal posterior of ϕ is such that, for some constant $C > 0$,

$$P(\|\phi - \phi_0\| \leq C r_n C_n | D_n) \to^p 1$$

where $C_n \to \infty$.
(ii) Let $P_{r_n^{-1}(\phi-\phi_0)|D_n}$ denote the posterior distribution of $r_n^{-1}(\phi-\phi_0)$ and Q be some probability distribution. We assume

$$\|P_{r_n^{-1}(\phi-\phi_0)|D_n} - Q\|_{TV} \to^p 0.$$

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

$$r_n^{-1}(\hat{\phi} - \phi_0) \to^d Q,$$

where Q is the same probability distribution as in (ii).

Under this assumption, all the results of Sections 3-6 still hold with more involved notation and minor changes. In particular, this assumption allows to treat models with non-regular estimators like the ones in Manski and Tamer (2002) and Chernozhukov et al. (2013). For simplicity of exposition we present our results and proofs under Assumption 2.1.

3 Bayesian credible sets and credible bands

Our major objective is to conduct Bayesian inference for the unknown set $\Theta(\phi)$ by constructing a Bayesian Credible Set (BCS) for it, and to make a connection with the frequentist

inference asymptotically. The support function plays a central role in our construction. We shall also construct a Uniform Bayesian Credible Band (UBCB) for the support function.

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. 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 and the UBCB as $\Theta(\hat{\phi})^{\epsilon_\tau}$ and $\{S_\phi(\cdot) : S_\phi(\nu) \in [S_{\hat{\phi}}(\nu) \pm \tilde{\epsilon}_\tau], \forall \nu \in \mathbb{S}^d\}$, respectively, such that:

Bayesian coverage:

$$P(\Theta(\phi) \subset \Theta(\hat{\phi})^{\epsilon_\tau} | D_n) \geq 1 - \tau, \tag{3.1}$$

and

$$P\left(\sup_{\|\nu\|=1} |S_\phi(\nu) - S_{\hat{\phi}}(\nu)| \leq \tilde{\epsilon}_\tau \middle| D_n \right) \geq 1 - \tau, \tag{3.2}$$

where $\hat{\phi}$ is some point estimator of ϕ. We are particularly interested in the asymptotic frequentist properties of the constructed BCS and UBCB. Under Assumption 2.1, we prove in Theorem 4.1 below that our constructed BCS and UBCB satisfy:

Frequentist coverages:

$$P_{D_n}(\Theta(\phi_0) \subset \Theta(\hat{\phi})^{\epsilon_\tau}) \geq 1 - \tau + o_P(1), \tag{3.3}$$

and

$$P_{D_n}\left(\sup_{\|\nu\|=1} |S_{\phi_0}(\nu) - S_{\hat{\phi}}(\nu)| \leq \tilde{\epsilon}_\tau \right) \geq 1 - \tau + o_P(1). \tag{3.4}$$

Hence, both $\Theta(\hat{\phi})^{\epsilon_\tau}$ and the set $\{S_\phi(\cdot) : S_\phi(\nu) \in [S_{\hat{\phi}}(\nu) \pm \tilde{\epsilon}_\tau], \forall \nu \in \mathbb{S}^d\}$ 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 (3.1), $\Theta(\phi)$ is the random set (with respect to the posterior of ϕ), while $\Theta(\hat{\phi})^{\epsilon_\tau}$ is treated as fixed. On the contrary, in the frequentist coverage (3.3), $\Theta(\hat{\phi})^{\epsilon_\tau}$ is the random set (with respect to the sampling distribution of $\hat{\phi}$), while $\Theta(\phi_0)$ is the true and fixed set object.

Therefore, our result complements the discoveries of Moon and Schorfheide (2012): while the BCS for the partially identified parameter may not have a correct frequentist coverage, the BCS for the identified set does. The intuition behind the latter asymptotic equivalence is that the prior is imposed on the identified set $\Theta(\phi)$ directly through ϕ, and because the identified set is "identified", the classical large sample Bayesian-frequentist equivalence is still valid. Thus, its posterior will asymptotically concentrate on a "neighborhood" of the true identified set, resulting in a correct frequentist coverage of the BCS.

In the contrary, when studying inference for the partially identified parameter, Moon and

Schorfheide (2012) placed a prior on the parameter itself with support equal to the identified set. Then, the posterior tends to be supported only inside the identified set, leading to a smaller nominal frequentist coverage.

Because of the duality between support function and Hausdorff distance and between confidence sets and test, our results on UBCB for the support function can be used to test hypothesis on the identified set $\Theta(\phi_0)$. In particular, the critical value for the UBCB for the support function, obtained with our procedure described in Section 3.1, can be used to implement the test based on the Hausdorff distance that was proposed in Beresteanu and Molinari (2008). Therefore, one of the possible uses of UBCB for the support function is to construct testing procedures alternative to the ones in Beresteanu and Molinari (2008). In addition, in Section 3.2 we will show that we can use the critical value of the UBCB to construct confidence sets for the partially identified parameter θ.

3.1 Critical Values for BCS and UBCB

Recall the definition of $\Theta(\hat{\phi})^\epsilon$ as the ϵ-expansion of $\Theta(\hat{\phi})$. Here $\hat{\phi}$ can be any regular $\sqrt{n}$-consistent estimator of the true ϕ_0 that is asymptotically normal. The support function has the following property: for any $\epsilon > 0$,

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

Therefore, we can find both the critical values ϵ_τ and $\tilde{\epsilon}_\tau$, used to construct the BCS and the UBCB, through the posterior 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

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

respectively, so that

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

For these q_τ and $\tilde{q}_\tau$, the following theorem holds.

Theorem 3.1 (Bayesian coverage)**.** *Suppose $\Theta(\phi)$ is convex for every ϕ in its parameter space. For every sampling sequence D_n, and any $\tau \in (0,1)$,*

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

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 requires a Monte Carlo procedure as follows. Uniformly generate $\{\nu_j\}_{j\leq G}$ such that $\|\nu_j\| = 1$. In addition, sample $\{\phi_i\}_{i\leq M}$ 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 $i = 1, ..., M$.

Then q_τ and $\tilde{q}_\tau$ are respectively calculated as the $1-\tau$ quantiles of

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

Generating $\{\nu_j\}$ 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, our Theorem 5.1 below shows that uniformly over ϕ in a neighborhood of $\hat{\phi}$,

$$S_\phi(\nu) - S_{\hat{\phi}}(\nu) \approx \lambda(\nu, \hat{\phi})^T \nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})[\phi - \hat{\phi}], \tag{3.5}$$

where: $\approx$ means "equal up to $o_p(1)$", $\theta^*(\nu)$ 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 (5.1), and $\nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})$ is the partial gradient of Ψ with respect to ϕ. Importantly, both θ_* and the KT-vector are automatically obtained in the outer-loop. The approximation error in (3.5) is smaller than the first-order statistical error $n^{-1/2}$. As a result, $S_{\phi_i}(\nu_j)$ can be approximated by: for $\theta_j^* := \theta^*(\nu_j)$,

$$S_{\phi_i}(\nu_j) \approx S_{\hat{\phi}}(\nu_j) + \lambda(\nu_j, \hat{\phi})^T \nabla_\phi \Psi(\theta_j^*, \hat{\phi})[\phi_i - \hat{\phi}],$$

avoiding solving M optimization problems in the inner-loop.

We summarize 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 $\{\phi_i\}_{i\leq M}$ 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 $\{\nu_j\}_{j\leq G}$ such that $\|\nu_j\| = 1$ for each j.

2. (outer-loop): For each $j \leq G$, solve the following constrained convex problem

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

and obtain $\theta_j^* = \arg\max_\theta \{\nu_j^T \theta : \Psi(\theta, \hat{\phi}) \leq 0\}$ (if the latter is a set then θ_j^* is one element of this set) and the corresponding KT-vector $\lambda(\nu_j, \hat{\phi})$. [2]

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

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

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

As a result, for each generated ν_j, we only need to solve the constraint optimization once (in the outer-loop for $S_{\hat{\phi}}(\nu_j)$). The maximizations in the inner-loop are very easy as they involve optimizations on a finite set $j \in \{1, ..., G\}$. After obtaining the critical values, we can further approximate the BCS for the set using a Monte Carlo method: uniformly generate $\{\tilde{\theta}_i\}_{i \leq B}$ from the space Θ, and approximate $\Theta(\hat{\phi})^{q_\tau/\sqrt{n}}$ by:

$$\left\{ \tilde{\theta}_i : d(\tilde{\theta}_i, \Theta(\hat{\phi})) \leq \frac{q_\tau}{\sqrt{n}}, \quad i \leq B \right\}.$$

3.2 Coverage of the partially identified parameter using the posterior distribution

Another important question is whether we can construct a confidence set for the partially identified parameter $\theta \in \Theta(\phi)$ using the posterior distribution of $\Theta(\phi)$ which has a desired frequentist coverage. In this subsection we show that the answer is affirmative and we construct it by using the support function and its Bayesian confidence band. Consider a Bayesian testing problem

$$H_0(\theta) : \theta \in \Theta(\phi), \tag{3.6}$$

for a fixed and known $\theta \in \Theta$, where $\Theta(\phi)$ is drawn from the posterior distribution of ϕ. The well known duality of Bayesian credible sets and hypothesis tests shows that

$$\inf_{\theta \in \Theta(\phi_0)} P_{D_n}(\theta : H_0(\theta) \text{ is "accepted"}) \geq 1 - \tau. \tag{3.7}$$

[2] The Matlab function `fmincon` provides both the minimizer and the KT-vector.

Therefore, we can build a Bayesian test statistic, and construct a frequentist confidence set as the collection of all the accepted θ's.

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 ϵ_τ, we can define

$$\Omega_\tau(\phi) := \left\{\theta : \theta^T\nu \leq S_\phi(\nu) + \frac{\epsilon_\tau}{\sqrt{n}}, \forall \|\nu\| = 1\right\}. \tag{3.8}$$

Intuitively, $\theta \in \Omega_\tau(\phi)$ means θ is "close" to $\Theta(\phi)$. In fact, it can be shown that $\Omega_\tau(\phi) = \Theta(\phi)^{\epsilon_\tau/\sqrt{n}}$. Therefore, we construct a Bayesian test for (3.6) by investigating whether θ is "posteriorly covered" with a high probability:

$$\text{accept } H_0(\theta) \Leftrightarrow P(\theta \in \Theta(\phi)^{\epsilon_\tau/\sqrt{n}}|D_n) \geq 1-\tau, \tag{3.9}$$

for the confidence level $1-\tau$. Here $P(\theta \in \Theta(\phi)^{\epsilon_\tau/\sqrt{n}}|D_n)$ is the posterior probability with respect to the posterior distribution of ϕ, treating θ as fixed. Combining (3.7)-(3.9), we construct the frequentist confidence set for θ as:

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

which depends on the critical value ϵ_τ. We will choose $\epsilon_\tau = 2\tilde{q}_\tau$, where $\tilde{q}_\tau$ is the critical value for the UBCB for the support function defined in Section 3.1.

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

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

To explain this in words: if we define $\widehat{\Omega}$ as the set of all the "accepted" θ's (covered by $\Theta(\phi)^{2\tilde{q}_\tau/\sqrt{n}}$ with a 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 very efficiently 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. In addition, uniformly generate $\{\tilde{\theta}_b\}_{b\leq B}$ from the parameter space Θ.

2. For each $b = 1, ..., B$ and for a $\tau \in (0, 1)$, if

$$\frac{1}{M}\sum_{i=1}^{M} 1\left\{d(\tilde{\theta}_b, \Theta(\phi_i)) \leq \frac{2\tilde{q}_\tau}{\sqrt{n}}\right\} \geq 1 - \tau, \tag{3.10}$$

then accept $\tilde{\theta}_b$; otherwise discard $\tilde{\theta}_b$. The critical value $\tilde{q}_\tau$ is obtained in Algorithm 1.

3. Collect all the accepted $\tilde{\theta}_b$'s as a set $\widehat{\Omega}^*$.

We see that (3.10) is an MCMC approximation of the event $\tilde{\theta}_b \in \widehat{\Omega}$, since:

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

Therefore, $\widehat{\Omega}^*$ is an approximation of $\widehat{\Omega}$.

3.3 Marginal set inferences

Suppose we are particularly interested in just one component, θ_k, of the partially identified parameter, for some $k \leq d$, where $d = \dim(\theta)$. Our method provides a simple procedure to construct a BCS for the marginal identified set of θ_k and a Bayesian credible interval for θ_k. For any $(d-1)$-dimensional vector θ_{-k}, we use (θ_k, θ_{-k}) to denote a d-vector, whose k-th component is θ_k, and the remaining components are those of θ_{-k}.

Let e_k be a vector in $\mathbb{R}^d$ with a one in the k-th coordinate and zeros elsewhere. It is easy to show that (see Appendix A.2) when $\Theta(\phi)$ is convex, the marginal identified set for θ_k is:

$$\Theta(\phi)_k := [\theta_k : \text{ there is } \theta_{-k} \text{ such that } (\theta_k, \theta_{-k}) \in \Theta(\phi)] = [-S_\phi(-e_k), S_\phi(e_k)].$$

Thus, it is straightforward to find the support function for $\Theta(\phi)_k$: $\tilde{S}_\phi(\nu) = S_\phi(\nu e_k)$, for $\nu \in \{-1, +1\}$.

Marginal identified set: Let $c_{\tau,k}$ be the $1-\tau$ quantile of the posterior of

$$H(\phi) := \sqrt{n}\max_{\nu=\pm 1}\left(S_\phi(\nu e_k) - S_{\hat{\phi}}(\nu e_k)\right),$$

that is, $P(\sqrt{n}\max_{\nu=\pm1}[S_\phi(\nu e) - S_{\hat{\phi}}(\nu e)] \leq c_{\tau,k}|D_n) = 1-\tau$. Then, Theorem 3.1 immediately implies:

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

Therefore, the interval $[-S_{\hat{\phi}}(-e_k) - \frac{c_{\tau,k}}{\sqrt{n}}, S_{\hat{\phi}}(e_k) + \frac{c_{\tau,k}}{\sqrt{n}}]$ is a $1-\tau$ BCS for the marginal identified

set of θ_k.

Marginal partially identified parameter: Similarly to the construction in Section 3.2, the critical value for the marginal confidence interval of θ_k is also simple to compute. Let $\tilde{c}_{\tau,k}$ be the $1-\tau$ quantile of the posterior of

$$\bar{H}(\phi) := \sqrt{n} \max_{\nu=\pm 1} |S_{\hat{\phi}}(\nu e_k) - S_{\phi}(\nu e_k)|.$$

Define

$$\widehat{\Omega}_k := \{\theta_k : P(\theta_k \in [-S_\phi(-e_k) - \frac{2\tilde{c}_{\tau,k}}{\sqrt{n}}, S_\phi(e_k) + \frac{2\tilde{c}_{\tau,k}}{\sqrt{n}}] | D_n) \geq 1-\tau\}.$$

We shall show in Section 4 that $\widehat{\Omega}_k$ is an asymptotically valid confidence set for the marginal parameter θ_k.

Importantly, we directly construct our marginal sets for the individual component instead of projecting from the set for the full vector of the partially identified parameter. This is an appealing feature also computationally. Indeed, both $H(\phi)$ and $\bar{H}(\phi)$ can be approximated using the local linear approximation derived in Theorem 5.1 below:

$$\begin{aligned} H(\phi) &\approx \sqrt{n} \max_{\nu=\pm 1} \left\{ \lambda(\nu e_k, \hat{\phi})^T \nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})[\phi - \hat{\phi}] \right\}, \\ \bar{H}(\phi) &\approx \sqrt{n} \max_{\nu=\pm 1} \left\{ |\lambda(\nu e_k, \hat{\phi})^T \nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})[\phi - \hat{\phi}]| \right\}. \end{aligned}$$

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 is straightforward using the MCMC samples.

Algorithm 3 (marginal inference for $\Theta(\phi)_k$)

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

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

   and obtain $\theta^*(\nu) = \arg\max_\theta \{\nu e_k^T \theta : \Psi(\theta, \hat{\phi}) \leq 0\}$ (if the latter is a set then θ_j^* is one element of this set) and the corresponding Kuhn-Tucker vector $\lambda(\nu e_k, \hat{\phi})$ (respectively for $\nu = \pm 1$).
3. For each $i \leq M$, let

$$H_i = \sqrt{n} \max_{\nu=\pm 1} \left\{ \lambda(\nu e_k, \hat{\phi})^T \nabla_\phi \Psi(\theta^*(\nu), \hat{\phi})[\phi_i - \hat{\phi}] \right\},$$

$$\bar{H}_i \;=\; \sqrt{n}\max_{\nu=\pm 1}\left\{|\lambda(\nu e_k,\hat{\phi})^T\nabla_\phi\Psi(\theta^*(\nu),\hat{\phi})[\phi_i-\hat{\phi}]|\right\}.$$

4. Let $c_{\tau,k}$ and $\tilde{c}_{\tau,k}$ be the $(1-\tau)$-th quantile of $\{H_i\}_{i\leq M}$ and $\{\bar{H}_i\}_{i\leq M}$, respectively. Then the BCS for $\Theta(\phi)_k$ is
$$[-S_{\hat{\phi}}(-e_k)-\frac{c_{\tau,k}}{\sqrt{n}},S_{\hat{\phi}}(e_k)+\frac{c_{\tau,k}}{\sqrt{n}}].$$

Algorithm 3' (marginal inference for θ_k)

1. Obtain $\tilde{c}_{\tau,k}$ from the above algorithm.
2. Uniformly generate $\{\tilde{\theta}_b\}_{b\leq B}$ from the marginal parameter space of θ_k.
3. For each $b=1,...,B$, if
$$\frac{1}{M}\sum_{i=1}^{M}1\left\{\tilde{\theta}_b\in[-S_{\phi_i}(-e_k)-\frac{2\tilde{c}_{\tau,k}}{\sqrt{n}},S_{\phi_i}(e_k)+\frac{2\tilde{c}_{\tau,k}}{\sqrt{n}}]\right\}\geq 1-\tau, \tag{3.11}$$
then set $\theta_b^*:=\tilde{\theta}_b$; otherwise discard $\tilde{\theta}_b$.
4. Approximate $\widehat{\Omega}_k$ by the following interval:
$$[\min\{\theta_b^*\},\max\{\theta_b^*\}].$$

Remark 3.1. Since we only need to find the minimum and the maximum θ_b such that (3.11) holds, step 3 of Algorithm 3' can be simplified and replaced by the following step. Thus, we do not need to evaluate (3.11) for all the $b\leq B$.

Step 3': Re-arrange $\tilde{\theta}_{(1)}\leq ...\leq\tilde{\theta}_{(B)}$. Starting from $\theta_{(1)}$ to gradually increase, find the smallest $\theta_{(b)}$ such that (3.11) holds, and set it to be $\min\{\theta_b^*\}$. Starting from $\theta_{(B)}$ to gradually decrease, find the largest $\theta_{(b)}$ such that (3.11) holds, and set it to be $\max\{\theta_b^*\}$.

3.4 Connections with frequentist confidence sets

In this subsection, we make a connection, from a computational point of view, between our constructed BCS and the frequentist confidence set (FCS) constructed based on the support function. Connections from the asymptotic coverage point of view are made in Section 4.

The support function is also used in the frequentist literature to construct FCS for the set. For instance, Beresteanu and Molinari (2008) constructed FCS 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 bootstrap samples. Then $\bar{c}_\tau$ is determined as the $(1-\tau)$th quantile of (let $S_{\mathcal{C}}(\cdot)$ denote the support function for a generic set $\mathcal{C}$)

$$\{\sqrt{n}\sup_{\|\nu\|=1}|S_{\widehat{\Theta}_i}(\nu)-S_{\widehat{\Theta}}(\nu)|: \quad i\leq M\}. \tag{3.12}$$

Computing (3.12) 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, avoiding solving the optimization in $S_{\widehat{\Theta}_i}$ for each of the bootstrap sample. Our approach is similar to theirs when $\widehat{\Theta}_i$ can be parametrized 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 are computationally efficient algorithms and 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.

On the other hand, the proposed confidence set $\widehat{\Omega}$ for the partially identified parameter is computationally different from all the existing frequentist confidence set to the best of our knowledge. While it is also computationally simple, we do not claim that it is advantegous over any existing method in the literature. Instead, we use it to clearly show that with the help of our Bayesian analysis on the support function, a frequentist confidence set 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 Frequentist Coverages

We now show that the BCSs and UBCB constructed in Section 3 for both the identified set and the partially identified parameter, and for the support function, have correct frequentist coverage probability asymptotically. The general result relies on a local linear expansion of the support function, which is presented in this section as a high-level condition.

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.

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

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

we have, for $r_n = \sqrt{(\log n)/n}$, as $n \to \infty$,

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

Theorem 4.1. *Suppose $\Theta(\phi)$ is convex for every ϕ in its parameter space. 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,*

(i) $P_{D_n}\left(\Theta(\phi_0) \subset \Theta(\hat{\phi})^{q_\tau/\sqrt{n}}\right) \geq 1 - \tau + o_P(1)$;[3]

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

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

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

Concerning the subset inference, we have the following result for the asymptotic frequentist coverage of the Bayesian sets: let $\Theta(\phi)_k$ be the marginal identified set for the k-th component θ_k of θ, e_k be a vector in $\mathbb{R}^d$ with a one in the k-th coordinate and zeros elsewhere.

Theorem 4.2. *Suppose the assumptions of Theorem 4.1 hold. Then, for any $\tau \in (0,1)$ and for $c_{\tau,k}$ and $\tilde{c}_{\tau,k}$ as defined in Section 3.3,*

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

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

(ii) Frequentist coverage of the marginal identified set:

$$P_{D_n}\left(\Theta(\phi_0)_k \subset [-S_{\hat{\phi}}(-e_k) - \frac{c_{\tau,k}}{\sqrt{n}}, S_{\hat{\phi}}(e_k) + \frac{c_{\tau,k}}{\sqrt{n}}]\right) \geq 1 - \tau + o_P(1);$$

[3]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 $P_{D_n}(\Theta(\phi_0) \subset \Theta(\hat{\phi})^{q_\tau/\sqrt{n}}) \geq 1 - \tau + \Delta(D_n)$.

(iii) Frequentist coverage of the marginal parameter:

$$\inf_{\theta_k \in \Theta(\phi_0)_k} P_{D_n}\left(\theta_k \in \widehat{\Omega}_k\right) \geq 1 - \tau - o_P(1),$$

$$\textit{where } \widehat{\Omega}_k = \left\{\theta_k : P\left(\theta_k \in [-S_\phi(-e_k) - \tfrac{2\tilde{c}_{\tau,k}}{\sqrt{n}}, S_\phi(e_k) + \tfrac{2\tilde{c}_{\tau,k}}{\sqrt{n}}]\big|D_n\right) \geq 1 - \tau\right\}.$$

Remark 4.1. The results presented here are pointwise and valid for a fixed DGP. In principle it is possible to achieve the coverage results uniformly over a set of DGPs, allowing sequences of DGP that converge to the point identification. This is true as long as Assumptions 2.1 and 4.1 hold uniformly in both ϕ and a class of $\Theta(\phi)$. However, we expect that it might be technically difficult to verify the LLA and the Bernstein-von-mises theorem uniformly, and hence do not pursue them in this paper.

Therefore, it remains to verify the high-level Assumption 4.1. We shall verify this condition in two setups: (1) in Section 4.1 below, we verify it in the one-dimensional case where the identified set is a closed interval; (2) in Section 5 below, 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

$$\Theta(\phi) = [g_1(\phi), g_2(\phi)] \subset \Theta \subset \mathbb{R},$$

where g_1, g_2 are known functions taking values in $\mathbb{R}$. This is the case for the one-dimensional partially identified model (e.g., Imbens and Manski (2004)). In particular, we allow $\sup_\phi |g_1(\phi) - g_2(\phi)| = o(1)$, and with $g_1(\cdot) = g_2(\cdot)$ as a special case. Hence, the identified set can shrink to a singleton.

It is easy to verify that the support function is given by: $S_\phi(1) = g_2(\phi)$, and $S_\phi(-1) = -g_1(\phi)$. Hence, the critical values are obtained from the posteriors of

$$J(\phi) := \sqrt{n} \sup_{\|\nu\|=1} \left(S_\phi(\nu) - S_{\hat{\phi}}(\nu)\right) = \sqrt{n} \max\{g_2(\phi) - g_2(\hat{\phi}), g_1(\hat{\phi}) - g_1(\phi)\}$$

and

$$\tilde{J}(\phi) := \sqrt{n} \sup_{\|\nu\|=1} |S_\phi(\nu) - S_{\hat{\phi}}(\nu)| = \sqrt{n} \max\{|g_2(\phi) - g_2(\hat{\phi})|, |g_1(\hat{\phi}) - g_1(\phi)|\}.$$

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

Assumption 4.2. *(i) g_1 and g_2 are twice differentiable.*
(ii) Let $H_1(\phi)$ and $H_2(\phi)$ be the Hessian matrices of g_1 and g_2. Then there is $C > 0$, so that $\sup_{\phi \in B(\phi_0, r_n)} \|H_1(\phi)\| + \sup_{\phi \in B(\phi_0, r_n)} \|H_2(\phi)\| < C$.

We have the following proposition.

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

Remark 4.2. Sometimes the functions g_1 and g_2 may be only partially known, up to an additional infinite-dimensional parameter η, representing the unknown (but identifiable) distribution of the DGP. Then we can write them as $g_1(\phi, \eta), g_2(\phi, \eta)$, or $g_1(F), g_2(F)$, 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, a LLA similar to that of Assumption 4.1 can still be verified. See Remark 5.1 below for more discussions.

The one-dimensional case, though simple, contains several partially identified models, and has been extensively studied in the literature. We shall provide a detailed numerical study of the missing-data model in Section 7. Below, we present a two-player entry game model, also studied from a Bayesian perspective by Moon and Schorfheide (2012), as an illustrating example.

4.1.1 A Two-Player Entry Game

We consider the entry game in Ciliberto and Tamer (2009); Moon and Schorfheide (2012), and show that the marginal identified set for the parameter of interest 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:

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

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 $\phi = [\phi_{10}, \phi_{11}, \phi_{00}]$, and $\sum_{ij} \phi_{ij} = 1$, to denote the probabilities of observing a monopoly, no entry, or the entry 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 $\hat{\phi} = [\hat{\phi}_{10}, \hat{\phi}_{11}, \hat{\phi}_{00}]$, we have $\sqrt{n}(\hat{\phi} - \phi_0) \to^d \mathcal{N}(0, V)$, whose covariance matrix V is easy to obtain.

Assume $\epsilon_{ij} \sim N(0, 1)$ and $\gamma_j \geq 0$, $j = 1, 2$. The probabilities that firm j in market i is profitable are: as monopolist $P(\pi_{ij}^M > 0) = \Phi_N(\beta_j)$, and as duopolist $P(\pi_{ij}^D > 0) = \Phi_N(\beta_j - \gamma_j)$ for $j = 1, 2$, where Φ_N denotes the CDF of the standard normal distribution. It is well known that the pure strategy Nash equilibrium implies:

$$\begin{aligned}
&\phi_{11} = \Phi_N(\beta_1 - \gamma_1)\Phi_N(\beta_2 - \gamma_2), \quad \phi_{00} = \Phi_N(-\beta_1)\Phi_N(-\beta_2), \\
&\phi_{10} \geq \Phi_N(\beta_1)\Phi_N(-\beta_2) + \Phi_N(\beta_1 - \gamma_1)(\Phi_N(\beta_2) - \Phi_N(\beta_2 - \gamma_2)), \\
&\phi_{10} \leq \Phi_N(\beta_1)(1 - \Phi_N(\beta_2 - \gamma_2))
\end{aligned} \tag{4.1}$$

The first two equations imply that β_2, γ_2 are uniquely determined by ϕ, β_1, γ_1. Hence the free partially identified parameters are (β_1, γ_1). Then (4.1) defines the joint identified set for (β_1, γ_1), given ϕ.

Suppose, we are interested in the marginal identified set $\Theta(\phi)$ for $\theta = \beta_1$, which is a common parameter in the two profit functions. The marginal identified set can be characterized as:

$$\Theta(\phi) = \{\beta_1 \in \Theta : \text{there are } \gamma_1, \gamma_2, \beta_2 \text{ such that (4.1) hold}\}$$

where Θ is a bounded set. Moon and Schorfheide (2012) describe the marginal identified set using a projection approach. Here, we show that it is a closed interval that satisfies Assumption 4.2.

We assume $P(\pi_{i1}^D > 0) \in [\underline{d}, \bar{d}] \subset (0, 1)$ with a known parameter space $[\underline{d}, \bar{d}]$ (e.g., $\underline{d} = \epsilon$ and $\bar{d} = 1 - \epsilon$), and $P(\pi_{i2}^D > 0) < 1$ for all $i = 1, \ldots, n$. Then, the following lemma holds.

Lemma 4.1. *We have:* $\Theta(\phi) = [g_1(\phi), g_2(\phi)]$, *where*

$$g_1(\phi) = \Phi_N^{-1}\left(\frac{\bar{d}\phi_{10}}{\bar{d} - \phi_{11}}\right), \quad g_2(\phi) = \Phi_N^{-1}\left(\frac{\phi_{10} - \underline{d} + \phi_{11} + \phi_{00}\underline{d}}{\phi_{00} + \phi_{10} - \underline{d} + \phi_{11}}\right),$$

and Assumption 4.2 is satisfied.

5 LLA in 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 equalities and inequalities. More

precisely, we consider the structure of the identified set given in the following assumption. For some $\delta > 0$, recall that $B(\phi_0, \delta) := \{\phi \in \Phi : \|\phi - \phi_0\| \leq \delta\}$.

Assumption 5.1. *The identified set $\Theta(\phi)$ in (1.1) is defined as*

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

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

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

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

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

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

We make the following further assumptions to derive the LLA condition in Assumption 4.1 for the identified set of Assumption 5.1. Denote by $\nabla_\phi \Psi(\theta, \phi)$ the $k \times d_\phi$ matrix of partial derivatives of Ψ with respect to ϕ, and by $\nabla_\theta \Psi_i(\theta, \phi)$ the d-vector of partial derivatives of Ψ_i with respect to θ for each $i \leq k$. Their existence and continuity is assumed in the Assumption 5.4.

Assumption 5.2. *(i) The true parameter value for ϕ_0 is in the interior of Φ;*
(ii) The parameter space $\Theta \subset \mathbb{R}^d$ is convex, compact and has nonempty interior (relative to $\mathbb{R}^d$).

Assumption 5.3. *For any $\theta \in \Theta(\phi_0)$, the gradient vectors $\{\nabla_\theta \Psi_i(\theta, \phi_0)\}_{i \in Act(\theta, \phi_0)}$ are linearly independent.*

Assumption 5.4. *There is $\delta > 0$ such that for all $\phi \in B(\phi_0, \delta)$, we have:*
(i) the matrix $\nabla_\phi \Psi(\theta, \phi)$ exists and is continuous in $(\theta, \phi) \in \Theta(\phi) \times B(\phi_0, \delta)$;
(ii) $\Theta(\phi) \neq \emptyset$ and $\Theta(\phi)$ is contained in the interior of Θ (relative to $\mathbb{R}^d$);
(iii) the vector $\nabla_\theta \Psi_i(\theta, \phi)$ exists and is continuous in $(\theta, \phi) \in \Theta(\phi) \times B(\phi_0, \delta)$ for every $i \leq k$.

Consider the ordinary convex problem that defines the support function: $S_\phi(\nu) := \sup_\theta \{\nu^T \theta; \theta \in \Theta(\phi)\}$ where $\Theta(\phi)$ is characterized as in Assumption 5.1, and assume that this optimal value is finite. Assumption 5.3 guarantees: existence of a unique Kuhn-Tucker (KT) vector for this problem and that the strong duality holds, so that the KT-conditions are necessary and sufficient optimality conditions. Therefore, under the previous assumptions and if $S_\phi(\cdot) < \infty$, the support function admits a Lagrangian representation, see (Rockafellar, 1970, Theorem 28.2): $\forall \phi \in B(\phi_0, \delta)$ with δ as in Assumptions 5.1, and $\forall \nu \in \mathbb{S}^d$

$$S_\phi(\nu) = \sup_{\theta \in \Theta} \left\{ \nu^T \theta - \lambda(\nu, \phi)^T \Psi(\theta, \phi) \right\}, \tag{5.1}$$

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

$$\Xi(\nu, \phi) := \arg\max_{\theta \in \Theta} \{\nu^T \theta : \Psi(\theta, \phi) \leq 0\} \tag{5.2}$$

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

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

and the maximizers in (5.2) consist of the boundary points of $\Theta(\phi)$ at which the set $\Theta(\phi)$ is tangent to the hyperplane $\{\theta \in \Theta; \nu^T \theta = S_\phi(\nu)\}$.

Discussion of the assumptions. Our imposed assumptions are very similar to those of Kaido and Santos (2014)'s. Assumption 5.1 is more general in the sense that we allow (linear) equality constraints while they do not. We also place the same type of restrictions on the convexity of the maps $\theta \mapsto \Psi_i(\theta, \phi)$, $i = 1, \ldots, k_1$. Assumption 5.1 also requires the slopes a_i

of the moment functions to be know. When the slope of a linear constraint is unknown (depending on the unknown distribution of the DGP), there might be no asymptotically linear regular estimators, leading to a violation of the assumption of existence of an asymptotically normal estimator for ϕ_0.

Assumption 5.3 requires that the active inequality and equality gradients $\nabla_\theta \Psi_i(\theta, \phi_0)$ be linearly independent, which guarantees that the strong duality holds. Even though the strong duality can be guaranteed under weaker assumptions (e.g., like the Slater's condition, see *e.g.* (Rockafellar, 1970, Theorem 28.3)), Assumption 5.3 also ensures the uniqueness of the KT-vector which we need in our proofs. Moreover, as remarked by Kaido and Santos (2014), it is possible to construct testing procedures to detect cases where Assumption 5.3 does not hold.

Assumption 5.4 *(i)* and *(iii)* are used to prove directional differentiability of the function $\phi \mapsto S_\phi(\nu)$. Assumption 5.4 *(ii)* means that the boundary of $\Theta(\phi)$ is determined by the inequalities/equalities and not by the parameter space Θ. Assumptions very similar to Assumption 5.4 are made also in Kaido and Santos (2014).

5.1 LLA for the support function

Assumptions 5.1-5.4 imply that the support function of the closed and convex set $\Theta(\phi)$ admits directional derivatives in ϕ and that it is differentiable at ϕ. The next theorem exploits this fact and states that the support function can be locally approximated by a linear function of ϕ establishing, in this way, Assumption 4.1.

Theorem 5.1 (LLA)**.** *If Assumptions 5.2, 5.3 hold and Assumptions 5.1 and 5.4 hold with $\delta = r_n$ for some $r_n = o(1)$, then for all large n, there exist: (i) a real function $f(\phi_1, \phi_2)$ defined for every $\phi_1, \phi_2 \in B(\phi_0, r_n)$, (ii) a vector function of KT multipliers $\lambda(\cdot, \cdot) : \mathbb{S}^d \times \mathbb{R}^{d_\phi} \to \mathbb{R}_+^{k_1} \times \mathbb{R}^{k_2}$, and (iii) a Borel measurable mapping $\theta_*(\cdot) : \mathbb{S}^d \to \Theta$ satisfying $\theta_*(\nu) \in \Xi(\nu, \phi_0)$ for all $\nu \in \mathbb{S}^d$, such that :*

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

and

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

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

We remark that the functions λ and θ_* do not depend on the specific choice of ϕ_1 and ϕ_2 inside $B(\phi_0, r_n)$, but only on ν and the true value ϕ_0. The linear expansion can also be

viewed as stochastic when ϕ_1, ϕ_2 are interpreted as random variables associated with the posterior distribution $p(\phi|D_n)$. This interpretation is particularly useful to understand the proofs of Theorems 6.1 and 6.2 below.

The linear approximation given in Theorem 5.1 is particularly helpful to implement our procedure. In fact, as shown in Section 3 in numerical simulations and implementations, one can compute the support function by using a simple linear transformation, as stated in Theorem 5.1, instead of solving an optimization problem which might be challenging in some cases.

The LLA of the support function provided in Theorem 5.1 is similar to the one proved in Kaido and Santos (2014) to characterize the influence function of the support function, and our proof of Theorem 5.1 rests in many places on the proof of Kaido and Santos (2014). However, there are some differences in the two results: (1) we state this result for every ϕ in a shrinking ball centred on the true ϕ_0 while their result is stated at $\hat{\phi}$ (by using our notation); (2) our LLA is valid on a ball of radius r_n, for any $r_n = o(1)$, while their result is established for a rate $r_n = n^{-1/2}$; (3) we establish the LLA of the support function when the set is more generally characterized by moment inequalities and equalities, while moment equalities are not allowed in Kaido and Santos (2014). Point (2) is particularly important in Bayesian analysis as, for asymptotic results, the ball of radius r_n must be the ball on which the posterior distribution of ϕ puts all its mass as $n \to \infty$. The radius r_n differs depending on the model and the prior, and therefore for results as the ones in Theorems 6.2-6.3, it is important that the LLA holds for all $r_n = o(1)$.

Remark 5.1. A local linear approximation similar to the one given in Theorem 5.1 can be obtained in the more general case where ϕ is the data distribution, $\phi := F$, if Assumption 2.1 *(i)-(ii)* is replaced by Assumption 2.1″ below. In this case, the identified set $\Theta(\phi)$, as defined in Assumption 5.1, depends on the unknown distribution ϕ of X through an index function that depends on θ, contrarily to the previous case where the index function did not depend on θ. More precisely, for two known real-valued functions $\Psi^1_{s,i}$ and $\Psi^2_{s,i}$, the identified set writes (by omitting the equality constraints for simplicity)

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

where $\{a_i\}_{i=1}^{k_1}$ are as defined in Assumption 5.1 and $\theta \mapsto \Psi^1_{s,i} \left(\int \Psi^2_{s,i}(x, \theta) d\phi(x) \right)$ is strictly convex in θ, for every $i = 1, \ldots, k_1$ (or strictly convex in a subvector of θ if it depends only on this subvector). This case is the general case considered in Kaido and Santos (2014). With this characterization of $\Theta(\phi)$ the prior on ϕ has to be nonparametric. Let us denote the vectors of indices $g(\phi, \theta) := \int \Psi^2_s(x, \theta) d\phi(x)$ and $g_0(\theta) := g(\phi_0, \theta)$ where $\phi_0 = F_0$.

Assumption 2.1 *(i)* and *(ii)* has to be replaced by the following one:

Assumption 2.1″**.** *(i) The posterior of* ϕ *is such that, for some constant* $C > 0$ *and for every* $\nu \in \mathbb{S}^d$ *and* $\theta_*(\nu) \in \Xi(\nu, \phi_0)$*,*

$$P(\|g(\phi, \theta_*(\nu)) - g_0(\phi_0, \theta_*(\nu))\| \leq Cn^{-1/2}C_n|D_n) \to^p 1$$

where $C_n = (\log n)^{1/2}$*.*
(ii) For a given $\nu \in \mathbb{S}^d$*, let* $P_{\sqrt{n}(g_*-g_0)|D_n}$ *denote the posterior distribution of* $\sqrt{n}(g(\phi, \theta_*(\nu)) - g_0(\theta_*(\nu)))$ *with* $\theta_*(\nu) \in \Xi(\nu, \phi_0)$*. We assume that, for every* $\nu \in \mathbb{S}^d$ *and* $\theta_*(\nu) \in \Xi(\nu, \phi_0)$*,*

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

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

We remark that $\Delta_{n,g_0}(\nu)$ and $I_0^{-1}(\nu)$ in Assumption 2.1″ depend on ν. This is because the index g depends on θ which in turn is taken to be an element of $\Xi(\nu, \phi_0)$. Moreover, by slightly modifying Assumptions 5.1-5.4 the linear expansion in Theorem 5.1 holds with $\nabla_\phi \Psi(\theta_*(\nu), \phi_0)[\phi_1 - \phi_2]$ replaced by

$$\nabla\Psi_s^1\left(\int \Psi_s^2(x, \theta_*(\nu))d\phi_0(x)\right)\left[\int \Psi_s^2(x, \theta_*(\nu))d\phi_2(x) - \int \Psi_s^2(x, \theta_*(\nu))d\phi_1(x)\right] \quad (5.3)$$

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

6 Asymptotic properties of the posterior of $S_\phi(\nu)$

In this section we state posterior consistency, asymptotic normality and, recover the concentration rate of the posterior distribution of $S_\phi(\nu)$. These are known as frequentist asymptotic properties because they assume the existence of a true value ϕ_0, F_0 that generates the data. These properties for the support function hinge on similar properties for the posterior distribution of ϕ. In a semiparametric model where ϕ is identified, the posterior distribution of ϕ is known to contract at the parametric rate (or, in some models, at a parametric rate up to a logarithmic term) and to satisfy the Bernstein-von Mises (BvM)

theorem under suitable prior conditions. We stress that, even though the asymptotic properties for the posterior of ϕ are assumed to hold (Assumption 2.1), recovering the asymptotic properties for the posterior of $S_\phi(\nu)$ is not trivial as it could seem. In fact, due to the particular structure of the mapping between ϕ and $S_\phi(\nu)$, the posterior of $S_\phi(\nu)$ is linked to the posterior of ϕ in a complicated way.

The next theorem gives the contraction rate for the posterior of the support function.

Theorem 6.1 (Posterior concentration)**.** *Under Assumptions 2.1 (i) and 4.1, for some $C > 0$,*

$$P\left(\sup_{\nu\in\mathbb{S}^d} |S_\phi(\nu) - S_{\phi_0}(\nu)| \leq C(\log n)^{1/2} n^{-1/2} \middle| D_n\right) \to^p 1.$$

Remark 6.1. The result of Theorem 6.1 holds for both nonparametric and semiparametric prior on (ϕ, F). The concentration rate, as given in the theorem, is nearly parametric: $\sqrt{\frac{\log n}{n}}$ and is the same as the rate in Assumption 2.1 *(i)*. When the posterior for ϕ concentrates at the parametric rate $n^{-1/2}$, the same holds for the posterior of the support function. Note that under stronger primitive conditions, it is possible to replace the rate in Assumption 2.1 *(i)* and Theorem 6.1 with $C_n n^{-1/2}$ for any sequence $C_n \to \infty$ (see also the discussion after Assumption 2.1). Here, we fix $C_n = \sqrt{\log n}$ for simplicity. This is common in the posterior concentration rate literature (see *e.g.* Ghosal et al. (2000) and Shen and Wasserman (2001)). For estimation of the identified set, the same rate of convergence $\sqrt{\frac{\log n}{n}}$ has been achieved in the frequentist perspective by Chernozhukov et al. (2007) and Kaido and Santos (2014), among others.

The next theorem that we state is the 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 6.2 (BvM)**.** *(I) Let Assumptions 2.1 (i)-(ii) and 4.1 hold. Then for any $\nu \in \mathbb{S}^d$,*

$$\|P_{\sqrt{n}(S_\phi(\nu) - S_{\phi_0}(\nu))|D_n} - \mathcal{N}(\bar{\Delta}_{n,\phi_0}(\nu), \bar{I}_0^{-1}(\nu))\|_{TV} \to^p 0, \tag{6.1}$$

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.
(II) If, in addition to Assumption 2.1 (i)-(ii), Assumptions 5.2, 5.3 hold and Assumptions 5.1 and 5.4 hold with $\delta = r_n = \sqrt{(\log n)/n}$, then (6.1) holds with $\bar{\Delta}_{n,\phi_0}(\nu) :=$

$\lambda(\nu,\phi_0)^T\nabla_\phi\Psi(\theta_*(\nu),\phi_0)\Delta_{n,\phi_0}$, $\theta_*(\nu)\in\Xi(\nu,\phi_0)$ *and*

$$\bar{I}_0^{-1}(\nu) := \lambda(\nu,\phi_0)^T\nabla_\phi\Psi(\theta_*(\nu),\phi_0)I_0^{-1}\nabla_\phi\Psi(\theta_*(\nu),\phi_0)^T\lambda(\nu,\phi_0).$$

The asymptotic mean and covariance matrix of part (II) of the theorem can be estimated by replacing ϕ_0 by any consistent estimator $\hat{\phi}$. Thus, $\theta_*(\nu)$ will be replaced by any element $\hat{\theta}_*(\nu)\in\Xi(\nu,\hat{\phi})$ and an estimate of $\lambda(\nu,\phi_0)$ will be obtained by numerically solving the ordinary convex program in (5.1) with ϕ_0 replaced by $\hat{\phi}$.

Remark 6.2. The asymptotic covariance matrix of the posterior distribution may differ in parts (I) and (II) of Theorem 6.2 depending on the expression of $A(\nu)$. Under the assumptions of part (II) of the theorem, Kaido and Santos (2014) have derived the semiparametric efficiency bound for estimating the support function. Thus, under these assumptions, 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 $\bar{\Delta}_{n,\phi_0}(\cdot) = \lambda(\cdot,\phi_0)^T\nabla_\phi\Psi(\theta_*(\cdot),\phi_0)\Delta_{n,\phi_0}$ and covariance operator with kernel: $\forall\nu_1,\nu_2\in\mathbb{S}^d$

$$\bar{I}_0^{-1}(\nu_1,\nu_2) = \lambda(\nu_1,\phi_0)^T\nabla_\phi\Psi(\theta_*(\nu_1),\phi_0)I_0^{-1}\nabla_\phi\Psi(\theta_*(\nu_2),\phi_0)^T\lambda(\nu_2,\phi_0).$$

We then have the following theorem which we directly state under the assumptions of part (II) of Theorem 6.2. 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 $\|\cdot\|_\infty$).

Theorem 6.3 (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 6.2 (II),*

$$P_{\sqrt{n}(S_\phi(\cdot)-S_{\phi_0}(\cdot))|D_n}(B) \to^p \mathbb{G}(B), \tag{6.2}$$

for all $B\in\mathcal{B}$.

We remark that under the weaker assumptions of part (I) of Theorem 6.2 the weak convergence in (6.2) holds with $\mathbb{G}$ 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.

7 Simulations

7.1 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$(\alpha_2 + n_2, \beta_2 + n_1 - n_2)$.

7.1.1 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

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

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, $B = 1,000$ posterior draws $\{\phi^i\}_{i=1}^B$ are sampled from the posterior Beta distribution. For each of them, compute $J(\phi^i)$ (resp. $\tilde{J}(\phi^i)$) and set $q_{0.05}$ (resp. $\tilde{q}_{0.05}$) as the 95% upper quantile of $\{J(\phi^i)\}_{i=1}^B$ (resp. $\{\tilde{J}(\phi^i)\}_{i=1}^B$) to obtain the critical values. Each simulation is repeated for 2,000 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 our Theorem 4.1. 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 the nominal level. But the critical value for the partially identified parameter is conservative, making the coverage probability lower bounded by the nominal level.

Table 1: Frequentist coverages over 2,000 replications, $1 - \tau = 0.95$, prior for ϕ_1, ϕ_2 is Beta(α, β).

		$P_{D_n}(\Theta(\phi_0) \subset \Theta(\hat{\phi})^{q_\tau/\sqrt{n}})$			$\inf_{\theta \in \Theta(\phi_0)} P_{D_n}(\theta \in \widehat{\Omega})$	
α	β	$n = 50$	$n = 100$	$n = 500$	$n = 50$	$n = 100$
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

7.1.2 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. As discussed in Section 4.1, our results still hold when $P(Y = 1)$ is nearly identifiable.

The frequency of coverage over 2,000 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. On the other hand, the coverage of the partially identified parameter is as conservative as in the "nonshrinking" case.

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 $\Delta_n = 0$.

7.2 Marginal inference for fixed design interval regression

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

$$X^T(EY_1) \leq X^T X\theta \leq X^T(EY_2),$$

where X is a $n \times p$ fixed design matrix and Y_1 and Y_2 are $n \times 1$ vectors. Suppose the full parameter θ is high-dimensional, and we are interested in the first component θ_1. Let $\phi_1 = \frac{1}{n}X^T EY_1$ and $\phi_2 = \frac{1}{n}X^T EY_2$. Then the identified set is given by $\{\theta : \phi_1 \leq \frac{1}{n}X^T X\theta \leq \phi_2\} = \{(\frac{1}{n}X^T X)^{-1}\zeta : \phi_1 \leq \zeta \leq \phi_2\}$, where we assume $\frac{1}{n}X^T X$ is nonsingular. Using a similar argument as in Bontemps et al. (2011), it can be shown that the support function has a closed form: write $\phi := (\phi_1^T, \phi_2^T)^T$,

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

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

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

7.2.1 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 $Z_{ji} = X_i Y_{ji}$, and $V = \frac{1}{n} X^T X$. Then $\frac{1}{n}\sum_{i=1}^n Z_{ji} \sim \mathcal{N}(\phi_j, \frac{1}{n}V)$, 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 measure the informativeness of the prior. Then it is well known that the posterior of ϕ_j is also Gaussian with mean $\sigma_0^2(\sigma_0^2 I + \frac{1}{n}V)^{-1}\frac{1}{n}\sum_{i=1}^n Z_{ji}$ and covariance $\sigma_0^2(\sigma_0^2 I + \frac{1}{n}V)^{-1}\frac{1}{n}V$.

The BCSs are constructed using Algorithms 3'. Because the support function has a closed form, our algorithms run very fast. The results are reported in Table 4. The coverage probabilities are generally close to the nominal level. To compare between the two types of coverages, the coverage of the marginal partially identified parameter is slightly more conservative. To compare between the prior choices, the less informative prior (larger σ_0^2) in general yields higher frequentist coverage probabilities. This observation is not surprising, as less informative prior often results in wider credible intervals.

Table 3: Frequentist coverages of the first component over 1,000 replications, $1 - \tau = 0.95$; the prior variance is σ_0^2.

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

8 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

Table 4: Frequentist coverages of the first component over 1,000 replications, $1-\tau=0.95$.

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

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 and convex.

One of the main contributions of this paper is to shed light on the connection between Bayesian and frequentist inference for partially identified convex models and complements the important results in Moon and Schorfheide (2012). While Moon and Schorfheide (2012) show that a BCS for the partially identified parameter θ does not have a correct frequentist coverage even asymptotically, we instead construct a BCS for the partially identified set and a UBCB for the support function and demonstrate that they have asymptotically correct frequentist coverage. Moreover, we 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. We also describe the computational algorithms to implement our inference procedures.

While in the paper we use relatively simple assumptions about the prior and posterior of ϕ, we discuss how it is possible to modify them in order to deal with models like Manski and Tamer (2002) and Chernozhukov et al. (2013). Having a finite dimensional ϕ is not necessary: an infinite dimensional ϕ can be allowed in some cases even though a more involved study of the nonparametric prior of ϕ would have to be carried out in such situations.

Finally, an important open question that we leave for future research is the uniform validity of our inference procedure. Our asymptotic coverage results hold under point identification since both the asymptotic normality of ϕ and the LLA of the support function continue to hold in the point identified case. However, to achieve the coverage results uniformly over a set of DGPs, allowing sequences of DGP that converge to the point identification, requires that both the asymptotic normality of ϕ and the LLA of the support function hold uniformly in both ϕ and a large class of DGP. This is technically difficult to verify and it might not hold in some cases.

References

Andrews, D. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119–157.

Andrews, D. W. and Guggenberger, P. (2009). Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. *Econometric Theory* **25** 669–709.

Belloni, A. and Chernozhukov, V. (2009). On the computational complexity of mcmc-based estimators in large samples. *Annals of Statistics* **37** 1591–2082.

Beresteanu, A., Molchanov, I. and Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica* **79** 1785–1821.

Beresteanu, A., Molchanov, I. and Molinari, F. (2012). Partial identification using random set theory. *Journal of Econometrics* **166** 17–32.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica* **76** 763–814.

Bickel, P., Klassen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semi Parametric Models*. The first edition ed. Springer, New York.

Bickel, P. and Kleijn, B. (2012). The semiparametric Bernstein-Von Mises theorem. *Annals of Statistics* **40** 206–237.

Bontemps, C., Magnac, T. and Maurin, E. (2011). Set identified linear models. *Econometrica* **80** 1129–1155.

Bontemps, D. (2011). Bernstein-von Mises theorems for gaussian regression with increasing number of regressors. *Annals of Statistics* **39** 2557–2584.

Bugni, F. A. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* **78** 735–753.

Canay, I. A. (2010). El inference for partially identified models: Large deviations optimality and bootstrap validity. *Journal of Econometrics* **156** 408–425.

Castillo, I. and Rousseau, J. (2015). A Bernstein-von Mises theorem for smooth functionals in semiparametric models. *Annals of Statistics* **56** 2353–2383.

Chandrasekhar, A., Chernozhukov, V., Molinari, F. and Schrimpf, P. (2012). Inference for best linear approximations to set identified functions. Tech. rep., Cemmap Working Papers, CWP43/12.

Chang, I. H., Kim, B. H. and Mukerjee, R. (2009). Bayesian and frequentist confidence intervals via adjusted likelihoods under prior specification on the interest parameter. *Statistics* **43** 203–211.

Chernozhukov, V., Hong, H. and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* **75** 1243–1284.

Chernozhukov, V., Kocatulum, E. and Menzel, K. (2015). Inference on sets in finance. *Quantitative Economics* **6** 309–358.

Chernozhukov, V., Lee, S. and Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica* **81** 667–737.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica* **77** 1791–1828.

Florens, J.-P. and Simoni, A. (2011). Bayesian identification and partial identification. Tech. rep., Toulouse School of Economics.

Ghosal, S., Ghosh, J. and Ramamoorthi, R. (1999a). Posterior consistency of dirichlet mixtures in density estimation. *Annals of Statistics* **27** 143–158.

Ghosal, S., Ghosh, J. and van der Vaart, A. (1999b). Consistent semiparametric bayesian inference about a location parameter. *Journal of Statistical Planning and Inference* **77** 181–193.

Ghosal, S., Ghosh, J. and van der Vaart, A. (2000). Convergence rates of posterior distributions. *Annals of Statistics* **28** 500–531.

Gospodinov, N., Kan, R. and Robotti, C. (2010). On the hansen-jagannathan distance with a no-arbitrage constraint. Tech. rep., Working Paper, Federal Reserve Bank of Atlanta.

Guntuboyina, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Annals of Statistics* **40** 385–411.

Gustafson, P. (2012). On the behaviour of bayesian credible intervals in partially identified models. *Electronic Journal of Statistics* **6** 2107–2124.

Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy* **99** 225–262.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica* **72** 1845–1857.

Ishwaran, H. and James, L. F. (2002). Approximate dirichlet process computing in finite normal mixtures: Smoothing and prior information. *Journal of Computational and Graphical Statistics* **11** 508–532.

Kaido, H. and 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. *Annals of Statistics* **41** 2359–2390.

Kitagawa, T. (2012). Estimation and inference for set-identified parameters using posterior lower probability. Tech. rep., UCL.

Kline, B. and Tamer, E. (forthcoming). Bayesian inference in a class of partially identified models. *Quantitative Economics* .

Leahu, H. et al. (2011). On the bernstein-von mises phenomenon in the gaussian white noise model. *Electronic Journal of Statistics* **5** 373–404.

Liao, Y. and Jiang, W. (2010). Bayesian analysis in moment inequality models. *Annals of Statistics* **38** 275–316.

Mammen, E., Marron, J., Turlach, B. and Wand, M. (2001). A general projection framework for constrained smoothing. *Statistical Science* **16** 232–248.

Manski, C. (2003). *Partial identification of probability distributions*. The first edition ed. Springer, New York.

Manski, C. and Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica* **70** 519–547.

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

Moon, R. and Schorfheide, F. (2012). Bayesian and frequentist inference in partially-identified models. *Econometrica* **80** 755–782.

Norets, A. (2015). Bayesian regression with nonparametric heteroskedasticity. *Journal of Econometrics* **185** 409–419.

Norets, A. and Tang, X. (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies* **81** 1229–1262.

Pakes, A., Porter, J., Ho, K. and Ishii, J. (2015). Moment inequalities and their application. *Econometrica* **83** 315–334.

Poirier, D. (1998). Revising beliefs in nonidentified models. *Econometric Theory* **14** 483–509.

Rivoirard, V. and Rousseau, J. (2012). Bernstein-von mises theorem for linear functionals of the density. *Annals of Statistics* **40** 1489–1523.

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

Romano, J. P. and Shaikh, A. M. (2010). Inference for the identified set in partially identified econometric models. *Econometrica* **78** 169–211.

Rosen, A. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics* **146** 107–117.

Sethuraman, J. (1994). A constructive definition of dirichlet priors. *Statistica sinica* 639–650.

Severini, T. (1991). On the relationship between bayesian and non-bayesian interval estimates. *Journal of the Royal Statistical Society, Series B* **53** 611–618.

Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Annals of Statistics* **29** 687–714.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica* **77** 1299–1315.

Sweeting, T. J. (2001). Coverage probability bias, objective bayes and the likelihood principle. *Biometrika* **88** 657–675.

van der Vaart, A. (2002). *"Semiparametric Statistics" in Ecole d'Eté de St Flour 1999 (eds. P. Bernard)*. The first edition ed. Springer Verlag, New York.

Walker, S., Lijoi, A. and Prünster, I. (2007). On rates of convergence for posterior distributions in infinite dimensional models. *Annals of Statistics* **35** 738–746.

Wan, Y. (2013). An integration-based approach to moment inequality models. Tech. rep., University of Toronto.