# Bayesian inference in a class of partially identified models

BRENDAN KLINE
Department of Economics, University of Texas at Austin


ELIE TAMER
Department of Economics, Harvard University

This paper develops a Bayesian approach to inference in a class of partially identified econometric models. Models in this class are characterized by a known mapping between a point identified reduced-form parameter $\mu$ and the identified set for a partially identified parameter $\theta$. The approach maps posterior inference about $\mu$ to various posterior inference statements concerning the identified set for $\theta$, without the specification of a prior for $\theta$. Many posterior inference statements are considered, including the posterior probability that a particular parameter value (or a set of parameter values) is in the identified set. The approach applies also to functions of $\theta$. The paper develops general results on large sample approximations, which illustrate how the posterior probabilities over the identified set are revised by the data, and establishes conditions under which the Bayesian credible sets also are valid frequentist confidence sets. The approach is computationally attractive even in high-dimensional models, in that the approach avoids an exhaustive search over the parameter space. The performance of the approach is illustrated via Monte Carlo experiments and an empirical application to a binary entry game involving airlines.

KEYWORDS. Partial identification, identified set, criterion function, Bayesian inference.

JEL CLASSIFICATION. C10, C11.

## 1. INTRODUCTION

This paper considers the problem of Bayesian inference in a class of partially identified models. These models are characterized by a known mapping between a point identified reduced-form parameter $\mu$ and the identified set for a partially identified parameter $\theta$. This set exhausts the information concerning $\theta$ contained in the data. Often $\mu$ can be viewed as directly observable characteristics of the data and $\theta$ can be viewed as the

parameter of an underlying econometric model. The parameter of interest is either $\theta$ or some function of $\theta$. For example, if $\theta$ is a parameter of an underlying econometric model and $\mu$ are statistics concerning the data, then the identified set mapping is the set of $\theta^*$ such that the underlying econometric model evaluated at $\theta^*$ generates $\mu$.

Since $\mu$ is point identified, there is a significant literature concerning the posterior $\mu|X$, where $X$ is the data. This paper takes the existence of a posterior $\mu|X$ as given. When establishing the theoretical results, the main condition this paper requires about $\mu|X$ is that it is approximately normally distributed in large samples, which is implied by "Bernstein–von Mises"-like results. In particular, such results are available even in the absence of finite-dimensional distributional assumptions about $X$. However, some of the theoretical results in this paper do not depend on the assumption that $\mu|X$ is approximately normally distributed in large samples, and the inference approach can be applied without that condition.

Then, given a posterior $\mu|X$ and the mapping from $\mu$ to the identified set for $\theta$, it is possible to construct various posterior probabilities concerning the identified set for $\theta$ without specifying a prior for $\theta$. One possibility is the posterior probability that a particular parameter value (or set of parameter values) is in the identified set, which concerns the question of whether a particular parameter value (or set of parameter values) could have generated the data. Another possibility is the posterior probability that all of the parameter values in the identified set have some property, which concerns the question of whether the parameter that generated the data necessarily has some property. Yet another possibility is the posterior probability that at least one of the parameter values in the identified set has some property, which concerns the question of whether the parameter that generated the data could have some property. Further, by checking the posterior probability that the identified set is nonempty, it is possible to do "specification testing." It is possible to make similar posterior probability statements concerning essentially any function of the identified set, including subvector inference.

For example, in many structural econometric models $\theta$ characterizes the utility functions of the decision makers and $\mu$ summarizes the observed behavior of the decision makers. Particularly in the case of models involving multiple decision makers, often $\theta$ is only partially identified, in which case it is not possible to uniquely recover the utility functions from the data. The identified set for $\theta$ exhausts the information in the data concerning the utility functions. In this setting, the posterior probabilities addressed in this paper answer empirically relevant questions including, "Are the data consistent with a particular specification of the utility functions?," "Do all utility functions consistent with the data possess a certain property?" (e.g., is it possible to conclude on the basis of the data that a certain observed explanatory variable has a positive effect on utility?), and "Are the data consistent with the utility function possessing a certain property?" (e.g., is it consistent with the data for a certain observed explanatory variable to have a positive effect on utility, or has the data ruled out that possibility?). See for example Manski (2007) or Tamer (2010) for further motivation for the identified set as the object of interest.

Prior results on inference in partially identified models has tended to follow other approaches. The frequentist approach (e.g., Imbens and Manski (2004), Rosen (2008),

Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), and Andrews and Barwick (2012)) generally requires working with discontinuous-in-parameters asymptotic (repeated sampling) approximations to test statistics. In contrast, the Bayesian approach is based only on the finite sample of data observed by the econometrician, and thereby avoids repeated sampling distributions.

Moreover, existing frequentist approaches are often difficult to implement computationally, especially in high-dimensional models, and especially as concerns the need to use a "exhaustive search" grid search (or "guess and verify" approach) to determine the set of parameter values belonging to the confidence set. In contrast, the Bayesian approach in this paper can use the developed literature on simulation of posterior distributions for point identified parameters, and also can use a variety of analytic and computational simplifications concerning the identified set mapping, implying that it is not necessary to use such an exhaustive search grid search. This is because there is separation between the "inference" problem, which concerns the posterior $\mu|X$ (not the whole parameter space), and the remaining computational problem of determining the identified set for $\theta$ evaluated at a particular value of $\mu$.

Because the inference concerns the identified set, the approach in this paper can be viewed as a sort of Bayesian analogue to the frequentist "random sets" approach (e.g., Beresteanu and Molinari (2008) and Beresteanu, Molchanov, and Molinari (2011, 2012)), in the sense that the posterior concerns the random set that arises due to uncertainty about the identified set.[1]

However, from the Bayesian perspective, it is possible to further revise the posterior inference concerning $\theta$ by introducing a prior over $\theta$. Such prior information would influence "conventional" posterior inference statements concerning $\theta$ even asymptotically (e.g., Poirier (1998)). In contrast, the typical situation with point identified parameters is that prior information does not influence posterior inference statements asymptotically. This issue with Bayesian inference in partially identified models causes the typical "asymptotic equivalence" between Bayesian and frequentist inference to fail to hold in partially identified models. Moon and Schorfheide (2012) establish that the Bayesian credible set for a partially identified parameter will tend to be *contained in* the identified set, whereas a frequentist confidence set for a partially identified parameter will tend to *contain* the identified set.[2]

Recently, a few alternative approaches to Bayesian inference in partially identified models have been proposed. The robust Bayes results of Kitagawa (2012) establish the

---

[1]However, there are some differences beyond simply Bayesian versus frequentist inference. In one formulation of the prior random sets approach, each observation in the data maps to a random set, and the identified set is the "average" (or some other random set operation) of those random sets. In other formulations, the econometric model evaluated at any specification of the parameters implies a certain random set that the observables must be "contained in," in a suitable sense. See also Beresteanu, Molchanov, and Molinari (2012). In contrast, the random set approach in this paper arises due to the mapping between the uncertainty concerning $\mu$ and uncertainty concerning the identified set. Kaido and White (2014) and Shi and Shum (2015) have addressed certain questions about improving frequentist inference in similar model frameworks.

[2]Woutersen and Ham (2014) study another nonstandard inference problem (where delta method arguments fail), and show that a certain proposed bootstrap method for constructing confidence intervals has a Bayesian interpretation and fails to provide valid frequentist inference. See also Freedman (1999).

"bounds" on the posterior for a partially identified parameter due to considering a class of priors, and shows a sense in which this robust Bayes approach reconciles Bayesian and frequentist inference for a partially identified parameter, in the sense that a credible set from the robust Bayes perspective also is a valid frequentist confidence set. Kitagawa (2012) establishes those results in a different model framework based on a standard likelihood with a partially identified parameter, with a standard prior specified only over the "sufficient parameter," and a class of priors specified over the remaining parameters. Intuitively, the sufficient parameter is a point identified reparameterization of the likelihood.[3] Norets and Tang (2014) study Bayesian inference in partially identified dynamic binary choice models. Similar to the approach in this paper, Norets and Tang (2014) relate the Bayesian inference on point identified quantities (i.e., conditional choice probabilities and transition probabilities) to partially identified quantities, but due to a different focus of the paper, do not address the same posterior inference questions concerning the identified set, and do not formally derive the theoretical properties of their proposed inference approach that would be analogous to the results derived in this paper. Liao and Simoni (2012) study Bayesian inference on the support function of a convex identified set, particularly in the context of an identified set characterized by inequality constraints, and show that under appropriate conditions, the associated credible sets are valid frequentist confidence sets. Convex sets are uniquely characterized by their support functions, but it may not be straightforward how to map inference on the support function to the posterior probability statements addressed in this paper. Further comparison is elaborated in Remark 4.

By focusing on posterior probability statements concerning the *identified set* rather than the *partially identified parameter*, this paper establishes a method for Bayesian inference that results in posterior inference statements that do not depend on the prior asymptotically. Indeed, this approach does not even require the specification of any prior for the partially identified parameter, and hence is a starting point that summarizes the information about $\theta$ given the data and the model.[4] See Section 3 and particularly Remark 2 for a discussion of the role of priors and posteriors in this approach. Intuitively, the identified set in a partially identified model is itself a point identified quantity, and therefore large sample approximations to posterior probability statements concerning the identified set do not depend on the prior, which is similar to the "typical" situation with point identified parameters in general.

---

[3]The sufficient parameter is the mapping of the parameter of the likelihood to the "sufficient parameter space," with two values of the parameter of the likelihood mapping to the same value of the sufficient parameter if and only if the likelihood function is the same evaluated at those two values of the parameter. Kitagawa (2012, p. 9) describes the sufficient parameter: it "carries all the information for the structural parameters through the value of the likelihood function."

[4]Broadly, the approach of not specifying a prior for the partially identified parameter is shared also by Kline (2011). Kline (2011) focuses on comparing Bayesian and frequentist inference on testing inequality hypotheses concerning a moment of a multivariate distribution, which can be interpreted to provide some limited results on posterior probability statements about whether a specified value of the parameter is in the identified set (because it satisfies the moment inequality conditions). However, already at the level of model framework, Kline (2011) differs substantially from this paper, with the consequence that the main contributions of the approach in this paper are not present in Kline (2011).

One consequence is that, under certain regularity conditions, in large samples the posterior probabilities associated with true statements concerning the identified set are approximately 1, and the posterior probabilities associated with false statements concerning the identified set are approximately 0. The behavior for statements that are "on the boundary" is complicated, but can be derived analytically. See Section 4.

Another consequence is that, under certain necessary and sufficient conditions, the $(1 - \alpha)$-level Bayesian credible set for the identified set is also an exact $(1 - \alpha)$-level frequentist confidence set for the identified set. This result means that there is an "asymptotic equivalence" between Bayesian and frequentist approaches to partially identified models, if the focus is on inference concerning the *identified set* rather than the *partially identified parameter*, which was the focus in other results including Moon and Schorfheide (2012). These results concern pointwise, but not necessarily uniform, validity of the resulting frequentist inference.

The remainder of the paper is organized as follows. Section 2 sets up the class of models considered in this paper, and provides examples. Section 3 sets up the posterior probabilities over the identified set that concern the question of whether a certain value of the partially identified parameter is in the identified set, and derives the large sample approximations to that posterior probability. Section 4 sets up the further posterior probabilities over the identified set that concern other questions about the identified set, and derives the large sample approximations to those posterior probabilities. Section 5 establishes the frequentist coverage properties of the Bayesian credible sets. Section 6 describes the computational implementation. Section 7 reports Monte Carlo experiments. Section 8 provides an empirical example of estimating a binary entry game with airline data. Section 9 concludes. Moreover, the Supplement, available in files on the journal website, http://qeconomics.org/supp/399/supplement.pdf and http://qeconomics.org/supp/399/code_and_data.zip, contains additional material.[5]

## 2. Model and examples

The model is characterized by a point identified reduced-form finite-dimensional parameter $\mu$, a partially identified finite-dimensional parameter $\theta$, and a known mapping between $\mu$ and the identified set for $\theta$. Often, $\mu$ can be viewed as statistics concerning the observable data (e.g., moments) and $\theta$ can be viewed as the parameter of an underlying econometric model. The parameter space for $\mu$ is $M$ and the parameter space for $\theta$ is $\Theta$. The parameter space $M$ is a subspace of $\mathbb{R}^{d_\mu}$, endowed with the subspace topology, where $d_\mu$ is the dimension of $\mu$. The parameter space $\Theta$ is a subset of $\mathbb{R}^{d_\theta}$, where $d_\theta$ is the dimension of $\theta$. The unknown true value of $\mu$ is $\mu_0$.

The defining property of this class of models is the existence of a known mapping from $\mu$ to the identified set for $\theta$. For example, this mapping might give the set of parameter values $\theta^*$ such that the underlying econometric model evaluated at $\theta^*$ generates $\mu$. This mapping often arises as an obvious implication of the specification of the underlying econometric model. Examples are provided below. The mapping can equivalently

---

[5]Section S1 provides further examples of the model framework, Section S2 provides results on measurability, and Section S3 provides further Monte Carlo experiments.

be expressed as a level set of a known criterion function of $\theta$ and $\mu$, or as a known set-valued mapping of $\mu$. In either case, this mapping gives the set of $\theta$ consistent with $\mu$, and thus the identified set for $\theta$.

Under the criterion function approach, there is a function $Q(\theta, \mu) \geq 0$ that summarizes the relationship between $\mu$ and the identified set for $\theta$. The criterion function is a function of the point identified parameter (which essentially substitutes for the data) and the partially identified parameter, which differs from the prior literature (e.g., Chernozhukov, Hong, and Tamer (2007), and Romano and Shaikh (2008, 2010)) where the criterion function depends on the data and the (potentially) partially identified parameter.

By construction, the identified set for $\theta$ can be expressed as

$$\Theta_I \equiv \Theta_I(\mu_0) \equiv \{\theta \in \Theta : Q(\theta, \mu_0) = 0\}.$$

Further, the identified set for $\theta$ that would arise at any parameter value $\mu^*$ is

$$\Theta_I(\mu^*) \equiv \{\theta \in \Theta : Q(\theta, \mu^*) = 0\}.$$

Therefore, $\Theta_I$ is the true identified set, whereas $\Theta_I(\mu)$ is the identified set as a mapping of $\mu$. If the model is point identified, then $\Theta_I(\mu)$ is a singleton for all $\mu \in M$.

It is allowed that $\Theta_I(\mu)$ is a "potentially nonsharp" specification of the identified set, in the sense that it potentially contains values of the partially identified parameter that are not consistent with the data summarized by $\mu$ and the assumptions of the underlying econometric model. All inference statements on the identified set are relative to the specification of $\Theta_I(\mu)$. In many applications, $\Theta_I(\mu)$ will be a "sharp" specification of the identified set, and therefore the inference will "fully exploit" the assumptions of the underlying econometric model. If $\Theta_I(\mu)$ is a potentially nonsharp specification of the identified set, then inference will be valid relative to that specification, but will not necessarily fully exploit the assumptions of the underlying econometric model.

Let the inverse identified set be $\mu_I(\theta) \equiv \{\mu : Q(\theta, \mu) = 0\}$. It follows that $\mu_I(\theta)$ is the set of $\mu$ consistent with $\theta$ being in the identified set evaluated at $\mu$. Therefore, the statement that $\mu \in \mu_I(\theta)$ is equivalent to the statement that $\theta \in \Theta_I(\mu)$.

Finally, let $\Delta(\cdot)$ be a function defined on $\Theta$. Suppose that $\delta$ is the partially identified parameter of interest, defined by $\delta \equiv \Delta(\theta)$. For example, if $\Delta(\theta) = \theta_1$, then the first component of $\theta$ is the parameter of interest, resulting in subvector inference. Alternatively, if $\Delta(\theta) = \theta$, then the entirety of $\theta$ is the parameter of interest. Then $\Delta(\Theta)$ is the induced parameter space for $\delta$, $\Delta_I \equiv \Delta(\Theta_I)$ is the induced true identified set for $\delta$, and $\Delta_I(\mu) \equiv \Delta(\Theta_I(\mu))$ is the induced identified set for $\delta$ as a mapping of $\mu$. The parameter space $\Delta(\Theta)$ is a subset of $\mathbb{R}^{d_\delta}$, where $d_\delta$ is the dimension of $\delta$.

The following paragraphs give a few examples of models that fit this framework. The Supplement discusses further examples, including moment inequality models.

EXAMPLE 1 (Intersection Bounds). Suppose that $\mu$ is a $d_\mu \times 1$ parameter vector whose estimation satisfies "standard regularity conditions,"[6] perhaps moments of a distribu-

---

[6]"Standard regularity conditions" means, essentially, that the conclusions of the Bernstein–von Mises theorem applies to $\mu|X$, as characterized by Assumption 3. See references following Assumption 3 for sufficient conditions.

tion. Suppose $\mu_0$ is the true value. Suppose that the identified set for $\theta$ is the interval $[\max_{j \in L} \mu_{0j}, \min_{j \in U} \mu_{0j}]$. The sets $L$ and $U$ are a partition of $\{1, 2, \ldots, d_\mu\}$ that determine which of the elements of $\mu$ contribute to the lower and upper bounds for $\theta$. See also, for example, Chernozhukov, Lee, and Rosen (2013). Then one possible specification of the criterion function is $Q(\theta, \mu) = (\max_{j \in L} \mu_j - \theta)_+ + (\theta - \min_{j \in U} \mu_j)_+$. The identified set at $\mu$ is $\Theta_I(\mu) = \{\theta : \max_{j \in L} \mu_j \leq \theta \leq \min_{j \in U} \mu_j\}$. Note that $\Theta_I(\mu) = \emptyset$ when $\max_{j \in L} \mu_j > \min_{j \in U} \mu_j$. The inverse identified set is $\mu_I(\theta) = \{\mu : \max_{j \in L} \mu_j \leq \theta \leq \max_{j \in U} \mu_j\}$.

In particular, "simple interval identified parameters" concerns $d_\mu = 2$, and arises in the context of missing data and general "selection problems" (e.g., Manski (2003)) and best response functions in games (e.g., Kline and Tamer (2012)).

EXAMPLE 2 (Discrete-Support Models). Suppose that $X$ has discrete support, and let $\mu$ be a parameter vector that characterizes the distribution of $X$. ($X$ comprises all of the data, not just the "explanatory variables.") Then for any such model, $f(\theta)$ can be the discrete distribution of the data implied by the econometric model at the parameter $\theta$, and $\mu$ can be the actual distribution of the data. Evaluated at the truth, $\mu_0 = f(\theta_0)$, so one possible specification of the criterion function is $Q(\theta, \mu) = \|\mu - f(\theta)\|$. The identified set at $\mu$ is $\Theta_I(\mu) \equiv \{\theta : \mu = f(\theta)\}$. The inverse identified set is $\mu_I(\theta) = \{\mu : \mu = f(\theta)\}$. This shows that essentially any partially identified model, with discretized observables, fits the framework of this paper.[7]

In particular, consider the example of a discrete game involving $N$ players, such that the actions available to player $i$ are $\mathcal{A}_i \equiv \{0, 1, \ldots, A_i\}$ for some finite $A_i$. Then the observables are the outcomes of the game $Y \in \prod \mathcal{A}_i$, and possibly discretized covariates $Z$. The game theory model implies that there is some function from unknown parameters $\theta$ to the distribution of the observables $\mu$, where $\mu = \{P(Y = y | Z = z)\}_{y,z}$, so that the model has the form that $f(\theta) = \mu$ for some function $f$ that is implied by the game theory model. See the Monte Carlo experiments in Section 7.1 and the empirical application in Section 8 for specifications of $f(\cdot)$. The parameters in $\theta$ can include parameters characterizing how the utility functions depend on the covariates, parameters characterizing the distribution(s) of the unobservables, and parameters characterizing the selection mechanisms over regions of multiple equilibrium outcomes. See for example Tamer (2003), Berry and Tamer (2006), or Kline (2015a, 2015b) for further details of various models of this general form, each of which imply a certain form for $f(\cdot)$.

## 3. POSTERIOR PROBABILITIES OVER THE IDENTIFIED SET

### 3.1 *Setup*

Since $\mu$ is point identified, let $\Pi(\mu | X)$ be a posterior for $\mu$ after observing the data $X$. This paper takes $\Pi(\mu | X)$ as given, only supposing that it satisfies standard regularity conditions elaborated later in this section. The posterior $\Pi(\mu | X)$ induces posterior

---

[7]This is a minimum distance approach to inference in models with discrete data, but the approach allows $\theta$ to be non-point identified.

probability statements concerning $\Delta_I$. This section addresses the posterior probability statements concerned with answering questions related to "Could $\delta^*$ have generated the data?" and "Could each $\delta^* \in \Delta^*$ have generated the data?".

DEFINITION 1. Based on the posterior for $\mu$, define the following posterior probability statements:

(i) For a singleton $\delta^* \in \Delta(\Theta)$,

$$\Pi\big(\delta^* \in \Delta_I | X\big) \equiv \Pi\big(\delta^* \in \Delta(\Theta_I(\mu)) | X\big) = \Pi\bigg(\mu \in \bigcup_{\{\theta : \Delta(\theta) = \delta^*\}} \mu_I(\theta) \big| X\bigg).$$

(ii) For a set $\Delta^* \subseteq \Delta(\Theta)$,

$$\Pi\big(\Delta^* \subseteq \Delta_I | X\big) \equiv \Pi\big(\Delta^* \subseteq \Delta(\Theta_I(\mu)) | X\big) = \Pi\bigg(\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) \big| X\bigg).$$

The posterior probability statements on the left correspond to statements concerning the "posterior uncertainty" about $\Delta_I$. These are then expressed in terms of the posterior for $\mu$. The nontrivial identities in Definition 1 are proved by Lemma 2.

The posterior $\Pi(\delta^* \in \Delta_I | X)$ answers an important question about the identified set: "Does a specified $\delta^*$ belong to the identified set?". It answers this question by giving the posterior probability that $\delta^*$ is in the identified set. This can be used to check whether $\delta^*$ could have generated the data. The posterior $\Pi(\Delta^* \subseteq \Delta_I | X)$ answers another important question: "Is a specified set $\Delta^*$ contained in the identified set?". It answers this question by giving the posterior probability that $\Delta^*$ is contained in the identified set. This can be used to check whether all parameter values in $\Delta^*$ could have generated the data.

These posterior probability statements concerning $\Delta_I$ do not address questions relating to the actual "true value" of $\delta$ that generated the data. In partially identified models, the data reveal only that the true value of $\delta$ is contained in $\Delta_I$, suggesting that $\Delta_I$ rather than $\delta$ should be the target of inference.

In the context of a simple interval identified parameter, the following example illustrates the approach to inference.

EXAMPLE 3 (Posterior Probabilities for the Simple Interval Identified Parameter). Suppose $\theta$ is a simple interval identified parameter, as in Example 1, so $\Theta_I(\mu) = [\mu_L, \mu_U]$, where $\mu = (\mu_L, \mu_U)$. In this example, $\Delta(\theta) = \theta$, so $\delta \equiv \theta$. Therefore, $\{\theta : \Delta(\theta) = \delta\} = \{\delta\}$, so essentially all expressions involving $\delta$ can be "replaced" by $\theta$. Suppose $\Theta^* = [a, b]$ is a finite interval, possibly with $a = b$ so $\Theta^*$ is a singleton.

Consider $\Pi(\Theta^* \subseteq \Theta_I | X)$. This is the posterior probability that each of the values in $\Theta^*$ is contained in the identified set, or equivalently the posterior probability that each of the values in $\Theta^*$ could have generated the data. Note that $\bigcap_{\theta \in \Theta^*} \mu_I(\theta) = \bigcap_{\theta \in \Theta^*} \{\mu : \mu_L \leq \theta \leq \mu_U\} = \{\mu : \mu_L \leq a, \mu_U \geq b\}$, so $\Pi(\Theta^* \subseteq \Theta_I | X) = \Pi(\{\mu : \mu_L \leq a, \mu_U \geq b\} | X)$. Consequently, $\Pi(\Theta^* \subseteq \Theta_I | X)$ is the posterior probability of the set $\{\mu : \mu_L \leq a, \mu_U \geq b\}$. Equivalently, $\Pi(\Theta^* \subseteq \Theta_I | X)$ is the posterior probability of the set of $\mu$ such that the identified set evaluated at $\mu$ does indeed contain $\Theta^*$.

Similarly, note that before the econometrician observes the data, $\Pi(\Theta^* \subseteq \Theta_I)$ would be the prior probability of the set $\{\mu : \mu_L \leq a, \mu_U \geq b\}$. In that sense, as discussed in more detail in Remark 2, this approach to inference implicitly entails the specification of a prior over the identified set.

Some of the main theoretical results in this paper concern large sample approximations to posterior probability statements about the identified set. Intuitively, these are derived from the large sample approximations of the posterior $\mu|X$, via the identified set mapping $\Theta_I(\mu)$. In this example, the large sample approximation to the posterior probability that $\Theta^* \subseteq \Theta_I$ is derived from the large sample approximation to the posterior probability of the set $\{\mu : \mu_L \leq a, \mu_U \geq b\}$ according to $\mu|X$. For example, if $\mu_0$ is such that $\Theta^*$ is contained in the interior of $\Theta_I(\mu_0)$, then $\mu_{0L} < a$ and $\mu_{0U} > b$, so consistency of the posterior $\mu|X$ implies that the posterior probability of the set $\{\mu : \mu_L \leq a, \mu_U \geq b\}$ is approximately 1 in large samples, and therefore that the posterior probability that $\Theta^* \subseteq \Theta_I$ is approximately 1 in large samples.

Section 3.2 formalizes this intuition and establishes the properties of the large sample approximations. One technical consideration is the necessity to establish that posterior probability statements concerning the identified set are equivalent to posterior probability statements concerning measurable sets of $\mu$.

### 3.2 *Large sample approximations*

This section establishes the regularity conditions under which there is a large sample approximation to $\Pi(\Delta^* \subseteq \Delta_I|X)$. Intuitively, because the identified set is a point identified quantity, under regularity conditions $\Pi(\Delta^* \subseteq \Delta_I|X)$ does not depend on the prior asymptotically. The results establish that, in many cases, in large samples $\Pi(\Delta^* \subseteq \Delta_I|X)$ equals either 1 or 0 depending on whether $\Delta^* \subseteq \Delta_I$ is true or false.

DEFINITION 2 (Topological Terminology). This paper uses standard topological terminology. For a given subset $A$ of $B$, the interior of $A$ is int($A$). The exterior of $A$ is ext($A$), which is the complement of the closure of $A$. The boundary of $A$ is bd($A$). The complement of $A$ is $A^C$. The convex hull of $A$ is co($A$). The subset $A$ is a convex polytope if $A$ is convex and compact, and has finitely many extreme points (i.e., $A$ is the convex hull of finitely many points).

The first regularity condition concerns the probability space for the posterior for $\mu$.

ASSUMPTION 1 (Regularity Condition for $\Pi(\mu|X)$). *The parameter space for $\mu$ (i.e., $M$) is a subspace of the Euclidean space $\mathbb{R}^{d_\mu}$ endowed with the subspace topology. The posterior distribution for $\mu$, $\Pi(\mu|X)$, is a probability measure defined on the Borel[8] $\sigma$-algebra of $M$, $\mathcal{B}(M)$.*

---

[8]The Borel sets of $M$ are the Borel sets corresponding to the subspace topology on $M$ viewed as a subspace of a Euclidean space, that is, $\mathcal{B}(M) = \{A \cap M : A \in \mathcal{B}(\mathbb{R}^{d_\mu})\}$. Note in particular that if $M \in \mathcal{B}(\mathbb{R}^{d_\mu})$, then $\mathcal{B}(M) = \{A \in \mathcal{B}(\mathbb{R}^{d_\mu}) : A \subseteq M\} \subseteq \mathcal{B}(\mathbb{R}^{d_\mu})$.

Also, the results suppose the following regularity conditions on the large sample behavior of the posterior for $\mu$.

ASSUMPTION 2 (Posterior for $\mu$ Consistent at $\mu_0$). *Along almost all sample sequences, for any open neighborhood $U$ of $\mu_0$ it holds that $\Pi(\mu \in U|X) \to 1$.*

Posterior consistency for a point identified parameter holds under very general conditions, for example by Doob's theorem. This requires, in particular, that the prior for $\mu$ has support on a neighborhood of $\mu_0$ (e.g., the prior for $\mu$ has support on the entire parameter space).

ASSUMPTION 3 (Large Sample Normal Posterior for $\mu$). *There is a function of the data $\mu_n(X)$ and a covariance matrix $\Sigma_0$ such that along almost all sample sequences, $\sqrt{n}(\mu - \mu_n(X))|X$ converges in total variation to $N(0, \Sigma_0)$.*

This assumption is essentially the conclusion of the various Bernstein–von Mises-like theorems for a point identified parameter (e.g., Van der Vaart (1998), Shen (2002), or Bickel and Kleijn (2012)), taking $\mu_n(X)$ to be the maximum likelihood estimator and $\Sigma_0$ to be the inverse Fisher information matrix.[9] This assumption can also hold, for example, for the Bayesian bootstrap for nonparametric estimation of moments of an unknown distribution under a suitably flat Dirichlet process prior, taking $\mu_n(X)$ to be the sample average and $\Sigma_0$ to be the covariance of the unknown distribution (e.g., Ferguson (1973), Rubin (1981), Lo (1987), Gasparini (1995), and Choudhuri (1998)). See also for example Kline (2011) for a connection to a different (more limited) way to pointwise test moment inequality conditions from a Bayesian perspective. Note that some of the theoretical results in this paper do not depend on Assumption 3, and that the inference approach can be applied without Assumption 3.

REMARK 1 (Technical Consideration: Measurability). It is not immediate that posterior probabilities over the identified set exist, because it is possible that there are subsets $\Delta^*$ such that $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$ does not exist because it corresponds to a nonmeasurable event. Consequently, $\mathcal{M}_1$ is introduced as the subsets such that for $\Delta^* \in \mathcal{M}_1$, $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$ corresponds to a measurable event. The theoretical analysis of the posterior probabilities over the identified set necessarily restrict attention to assigning posterior probabilities to those $\Delta^*$. Lemma 3 in the Supplement shows that if the criterion function is continuous, $\Delta(\cdot)$ is continuous, and $\Theta$ is closed, then $\mathcal{M}_1$ contains all the Borel sets. Therefore, although measurability could potentially be a problem in some settings, measurability is not a problem for assigning posterior probabilities concerning "nice" sets (i.e., Borel sets) in "nice" models (i.e., continuous $Q(\cdot)$ and $\Delta(\cdot)$ and closed parameter space).

---

[9]Depending on the topological "complexity" of $\mu_I(\cdot)$ and the posterior probability under study, it is possible to relax this assumption to require only convergence in distribution and an application of Pólya's theorem or similar results to get uniform convergence over the relevant subsets of the parameter space $M$. (See the proof of part (iii) of Theorem 1, or parts (iii) and (vi) of Theorem 3 for the relevant considerations.) For example, see Rao (1962), Billingsley and Topsøe (1967), or Bickel and Millar (1992) for the cases including convex subsets.

Theorem 1. *Under Assumptions* 1 *and* 2, *for any* $\Delta^*$ *such that* $\Pi(\Delta^* \subseteq \Delta_I | X)$ *is defined (i.e.,* $\Delta^* \in \mathcal{M}_1$; *see Remark* 1), *along almost all sample sequences the following statements hold*:

  (i) *If* $\mu_0 \in \mathrm{int}(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, *then* $\Pi(\Delta^* \subseteq \Delta_I | X) \to 1$.

  (ii) *If* $\mu_0 \in \bigcup_{\delta \in \Delta^*:\{\delta\} \in \mathcal{M}_1} (\mathrm{ext}(\bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)))$, *then* $\Pi(\Delta^* \subseteq \Delta_I | X) \to 0$.

  *Under the additional Assumption* 3,

  (iii) $|\Pi(\Delta^* \subseteq \Delta_I | X) - P_{N(0,\Sigma_0)}(\sqrt{n}(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)))| \to 0$.

It is possible to simplify the statement of Theorem 1 under the assumption of "continuity" of the identified set.

Assumption 4 (Continuity of the Identified Set). *For all* $\delta \in \mathbb{R}^{d_\delta}$, *if* $\delta \in \mathrm{int}(\Delta_I)$, *then* $\mu_0 \in \mathrm{int}(\bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. *For all* $\delta \in \mathbb{R}^{d_\delta}$, $\bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ *is closed. For any open* $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ *is open.*

The first part of this assumption requires that if $\delta$ is in the interior of $\Delta_I$, then there is a neighborhood of $\mu_0$ such that $\delta$ is also in the identified sets $\Delta_I(\mu)$ for all $\mu$ in that neighborhood. The second part of this assumption requires that the set of $\mu$ such that $\delta \in \Delta_I(\mu)$ is closed. The third part of this assumption requires that the set of $\mu$ such that $\Delta_I(\mu) \subseteq \Delta^*$ for open $\Delta^*$ is open.

Lemma 3 in the Supplement shows that a sufficient condition for the second and third parts of the assumption is continuity of the criterion function, continuity of $\Delta(\cdot)$, and compactness of the parameter space. Unfortunately, continuity of the criterion function does not imply the first part of the assumption; however, this assumption is satisfied in typical models.[10] In particular, the first part of this assumption is implied by convexity of $\Delta_I(\mu)$ for all $\mu$ and inner semicontinuity of $\Delta_I(\mu)$ at $\mu_0$ viewed as a mapping between Euclidean spaces (e.g., Rockafellar and Wets (2009, Theorem 5.9)).

Under Assumption 4, the statement of the large sample approximation results simplifies substantially. (Some parts of Theorem 1 do not change with the addition of Assumption 4, and so are not displayed in Corollary 2.)

Corollary 2. *Under Assumptions* 1, 2, *and* 4, *along almost all sample sequences the following statements are satisfied*:

  (i) *If* $\Delta^* \subseteq \mathrm{int}(\Delta_I)$ *and* $\Delta^*$ *is a convex polytope such that* $\Delta_I(\mu) \cap \Delta^*$ *is convex for all* $\mu$ *in a neighborhood of* $\mu_0$, *then* $\Pi(\Delta^* \subseteq \Delta_I | X) \to 1$.

---

[10]The following is a counterexample, that illustrates the seeming "strangeness" of models that would violate this assumption. Suppose that $\Delta(\cdot)$ is the identity function, and suppose that the criterion function $Q(\theta, \mu_0)$ equals zero for all $\theta$ in [0, 1]. Therefore, all points in (0, 1) are in the interior of the identified set. It is consistent with $Q$ being continuous that $Q(\theta, \mu) > 0$ for all $\theta$ and all $\mu \neq \mu_0$, which would violate the first part of the assumption. However, models like the interval identified parameter model share this basic structure, but do satisfy the assumption since in that model it would not happen that $Q(\theta, \mu) > 0$ for all $\theta$ and all $\mu \neq \mu_0$, suggesting that this assumption is reasonable.

(ii)  *If $\Delta^* \not\subseteq \Delta_I$, then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 0$.*

Essentially, Corollary 2 shows that $\Pi(\Delta^* \subseteq \Delta_I|X)$ is approximately 1 (respectively, 0) in large samples if $\Delta^* \subseteq \Delta_I$ is true (respectively, false).

Part (i) shows that if $\Delta^* \subseteq \text{int}(\Delta_I)$ and $\Delta^*$ is not too complex, then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 1$. Part (i) can be applied to finitely many convex polytopes in the interior of the identified set, so by "piecing together" an approximation of the interior of the identified set by convex polytopes, in models with sufficiently "simple" identified sets, each compact subset $\Delta^*$ of the interior of the identified set will have the property that $\Pi(\Delta^* \subseteq \Delta_I|X) \to 1$. It is not necessary that $\Delta_I(\mu)$ is convex in a neighborhood of $\mu_0$, because convexity of $\Delta_I(\mu) \cap \Delta^*$ is a weaker condition than convexity of $\Delta_I(\mu)$.

Part (ii) shows that if $\Delta^* \not\subseteq \Delta_I$, then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 0$.

REMARK 2 (The Role of Prior Information).  This approach to inference entails the implicit specification of a prior over the identified set, in the same sense that this approach results in a posterior over the identified set. This is because a prior for $\mu$ implies a prior for the identified set by the same logic as appears in Definition 1, dropping conditioning on $X$. The key distinction between this approach and "conventional" Bayesian approaches concerns the inferential object (identified set versus the partially identified parameter) and how the data revise the "prior" over the inferential object. There is "no prior" for the partially identified parameter in the same sense that no (conventional) posterior for the partially identified parameter results.

In the context of a simple interval identified parameter, the following example discusses the implications of Theorem 1.

EXAMPLE 4 (Posterior Probabilities for the Simple Interval Identified Parameter).  This example continues the discussion from Example 3.

*Case 1*: Suppose that $[a, b] \subset \text{int}(\Theta_I) = (\mu_{0L}, \mu_{0U}) \subset \Theta_I = [\mu_{0L}, \mu_{0U}]$. This implies $\mu_{0L} < a \le b < \mu_{0U}$. Then $\mu_0 \in \text{int}(\bigcap_{\theta \in \Theta^*} \mu_I(\theta))$, so by part (i) of Theorem 1, $\Pi([a, b] \subseteq \Theta_I|X) \to 1$. Therefore, in large samples, there will essentially be posterior certainty assigned to the (true) statement that $[a, b]$ is contained in the identified set.

*Case 2*: Conversely, suppose that $[a, b] \not\subseteq \Theta_I$. Suppose also that indeed $\mu_{0L} \le \mu_{0U}$ (so that the identified set is nonempty). Therefore, either $\mu_{0L} > a$ or $\mu_{0U} < b$. Note that $\mu_I(\theta)^C = \{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\}$. Therefore, $\mu_0 \in \text{int}(\mu_I(a)^C) = \text{ext}(\mu_I(a))$ or $\mu_0 \in \text{int}(\mu_I(b)^C) = \text{ext}(\mu_I(b))$, respectively, so by part (ii) of Theorem 1, $\Pi([a, b] \subseteq \Theta_I|X) \to 0$. Therefore, in large samples, there will essentially be no posterior probability assigned to the (false) statement that $[a, b]$ is contained in the identified set.

Further discussion of this example is in Example 6 in the Supplement.

## 4. FURTHER POSTERIOR PROBABILITIES OVER THE IDENTIFIED SET

### 4.1 *Setup*

The posterior $\Pi(\mu|X)$ also induces posterior probability statements concerning $\Delta_I$ that answer questions not already addressed in Section 3. This section addresses the poste-

rior probability statements concerned with answering questions related to "Do all parameter values in the identified set have some property?," "Does at least one parameter value in the identified set have some property?," and "Do none of the parameter values in the identified set have some property?".

DEFINITION 3. Based on the posterior for $\mu$, define[11] the following posterior probability statements:

(i) For a set $\Delta^* \subseteq \Delta(\Theta)$,

$$\Pi\big(\Delta_I \subseteq \Delta^*|X\big) \equiv \Pi\big(\Delta\big(\Theta_I(\mu)\big) \subseteq \Delta^*|X\big) = \Pi\bigg(\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X\bigg).$$

(ii) For a set $\Delta^* \subseteq \Delta(\Theta)$,

$$\Pi\big(\Delta_I \cap \Delta^* \neq \emptyset|X\big) \equiv \Pi\big(\Delta\big(\Theta_I(\mu)\big) \cap \Delta^* \neq \emptyset|X\big) = \Pi\bigg(\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X\bigg).$$

(iii) For a set $\Delta^* \subseteq \Delta(\Theta)$,

$$\Pi\big(\Delta_I \cap \Delta^* = \emptyset|X\big) = 1 - \Pi\big(\Delta_I \cap \Delta^* \neq \emptyset|X\big).$$

The posterior $\Pi(\Delta_I \subseteq \Delta^*|X)$ answers the question, "Do all parameter values in the identified set have some property?". It answers this question by giving the posterior probability that the identified set is contained in $\Delta^*$. This can be used to check whether *all* parameter values that could have generated the data have the property defined by $\Delta^*$. For example, if $\delta$ is a scalar and $\Delta^* = [0, \infty)$, then $\Pi(\Delta_I \subseteq \Delta^*|X)$ is the posterior probability that all parameter values that could have generated the data are nonnegative. If $\theta$ is point identified for all $\mu \in M$ and $\Delta(\theta) \equiv \theta$, then $\Pi(\Theta_I \subseteq \Theta^*|X)$ is the ordinary posterior for $\theta$, in the sense that $\Theta_I(\mu)$ is just a singleton, so $\Pi(\Theta_I \subseteq \Theta^*|X)$ is simply the posterior probability that $\theta \in \Theta^*$.

The posterior $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X)$ answers the question, "Does at least one parameter value in the identified set have some property?". It answers this question by giving the posterior probability that the identified set has nonempty intersection with $\Delta^*$. This can be used to check whether *at least one* of the parameter values that could have generated the data has the property defined by $\Delta^*$. For example, if $\delta$ is a scalar and $\Delta^* = [0, \infty)$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X)$ is the posterior probability that at least one nonnegative $\delta$ could have generated the data. In particular, taking $\Delta^* = \Delta(\Theta)$,

$$\Pi\big(\Delta_I \cap \Delta^* \neq \emptyset|X\big) = \Pi\big(\Delta_I \neq \emptyset|X\big) \equiv \Pi\bigg(\mu \in \bigcup_{\delta \in \Delta(\Theta)} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X\bigg)$$

is the posterior probability that the identified set $\Delta_I$ is nonempty, which can be interpreted to be a conservative (but implementable) measure of the posterior probability

---

[11]As in Section 3, the posterior probability statements on the left correspond to statements concerning the "posterior uncertainty" about $\Delta_I$ which are then expressed in terms of the posterior for $\mu$. The nontrivial identities in Definition 3 are proved by Lemma 2.

that the model is not misspecified. It is conservative because the fact that the identified set is nonempty does not imply that the model is correctly specified. But if the identified set is empty, then the model must be misspecified.

The posterior $\Pi(\Delta_I \cap \Delta^* = \emptyset | X)$ answers the question, "Do none of the parameter values in the identified set have some property?". It answers this question by giving the posterior probability that the identified set has empty intersection with $\Delta^*$. This can be used to check whether *none* of the parameter values that could have generated the data has the property defined by $\Delta^*$. For example, if $\delta$ is a scalar and $\Delta^* = [0, \infty)$, then $\Pi(\Delta_I \cap \Delta^* = \emptyset | X)$ is the posterior probability that no nonnegative $\delta$ could have generated the data.

### 4.2 *Large sample approximations*

REMARK 3 (Technical Consideration: Measurability).  As in Section 3, it is not immediate that posterior probabilities over the identified set exist. The collection $\mathcal{M}_2$ are the subsets such that for $\Delta^* \in \mathcal{M}_2$, $\Pi(\Delta_I \subseteq \Delta^* | X) \equiv \Pi(\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C | X)$ corresponds to a measurable event. The collection $\mathcal{M}_3$ are the subsets such that for $\Delta^* \in \mathcal{M}_3$, $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X) \equiv \Pi(\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) | X)$ corresponds to a measurable event. Lemma 3 in the Supplement shows that if the criterion function is continuous, $\Delta(\cdot)$ is continuous, and $\Theta$ is closed, then $\mathcal{M}_2$ and $\mathcal{M}_3$ contain all the Borel sets.

THEOREM 3.  *Under Assumptions 1 and 2, for any $\Delta^*$ such that $\Pi(\Delta_I \subseteq \Delta^* | X)$ is defined (i.e., $\Delta^* \in \mathcal{M}_2$; see Remark 3), along almost all sample sequences, the following statements hold:*

   (i)  *If $\mu_0 \in \mathrm{int}(\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C)$, then $\Pi(\Delta_I \subseteq \Delta^* | X) \to 1$.*
   (ii)  *If $\mu_0 \in \bigcup_{\delta \in (\Delta^*)^C : \{\delta\} \in \mathcal{M}_1} (\mathrm{ext}(\bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C))$, then $\Pi(\Delta_I \subseteq \Delta^* | X) \to 0$.*

   *Under the additional Assumption 3,*

   (iii)  $|\Pi(\Delta_I \subseteq \Delta^* | X) - P_{N(0, \Sigma_0)}(\sqrt{n}(\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C - \mu_n(X)))| \to 0.$

   *Under Assumptions 1 and 2, for any $\Delta^*$ such that $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X)$ is defined (i.e., $\Delta^* \in \mathcal{M}_3$; see Remark 3), along almost all sample sequences, the following statements hold:*

   (iv)  *If $\mu_0 \in \mathrm{int}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta))$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X) \to 1$.*
   (v)  *If $\mu_0 \in \mathrm{ext}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta))$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X) \to 0$.*

   *Under the additional Assumption 3,*

   (vi)  $|\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X) - P_{N(0, \Sigma_0)}(\sqrt{n}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) - \mu_n(X)))| \to 0.$

It is possible to simplify the statement of Theorem 3, under the assumption of "continuity" of the identified set.

COROLLARY 4.  *Under Assumptions 1, 2, and 4, for any $\Delta^*$ such that $\Pi(\Delta_I \subseteq \Delta^* | X)$ is defined (i.e., $\Delta^* \in \mathcal{M}_2$; see Remark 3), along almost all sample sequences, the following statements hold:*

(i)  *If $\Delta_I \subseteq \text{int}(\Delta^*)$, then $\Pi(\Delta_I \subseteq \Delta^*|X) \to 1$.*

(ii)  *If $\text{int}(\Delta_I) \nsubseteq \Delta^*$, then $\Pi(\Delta_I \subseteq \Delta^*|X) \to 0$.*

*Under the same assumptions, for any $\Delta^*$ such that $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X)$ is defined (i.e., $\Delta^* \in \mathcal{M}_3$; see Remark 3), along almost all sample sequences, the following statements hold*:

(iii)  *If $\Delta_I(\mu) \cap \Delta^* \neq \emptyset$ for all $\mu$ in a neighborhood of $\mu_0$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \to 1$.*

(iv)  *If $\Delta_I(\mu) \cap \Delta^* = \emptyset$ for all $\mu$ in a neighborhood of $\mu_0$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \to 0$.*

Essentially, Corollary 4 shows that the posterior probability of a true (respectively, false) statement concerning the identified set is approximately 1 (respectively, 0) in large samples.

REMARK 4 (Relation to Robust Bayesian Inference of Kitagawa (2012)). The model framework for Kitagawa (2012)[12] is essentially as follows: There is a likelihood and $\phi$ is a "sufficient parameter" for the likelihood, with a prior specified, resulting in a posterior $\phi|X$, and $H(\phi)$ is the identified set for the partially identified parameter of interest $\eta$, as a function of $\phi$. A class of priors is specified over the partially identified parameter, and bounds are derived for the posterior for $\eta$ due to specifying a class of priors. Very roughly, $\phi$ is analogous to $\mu$ in this paper, and $H(\phi)$ is analogous to $\Delta_I(\mu)$ in this paper. Despite this analogy, these two frameworks place different requirements on the econometrician: $\phi$ and $H(\phi)$ arise implicitly from the specification of a likelihood, whereas $\mu$ and $\Delta_I(\mu)$ are explicitly specified by the econometrician.[13] The differences in model framework result in further differences: for example, the computational approach proposed in this paper depends on the separation between standard Bayesian inference on $\mu$ and computation of the identified set as a known mapping of $\mu$. Kitagawa (2012) shows that, under appropriate conditions, the smallest posterior probability that can be assigned to a set $D$ of the parameter space for $\eta$ is the posterior probability under $\phi|X$ of the event $H(\phi) \subseteq D$. Also, the largest posterior probability that can be assigned to a set $D$ of the parameter space for $\eta$ is the posterior probability under $\phi|X$ of the event $H(\phi) \cap D \neq \emptyset$. Therefore, if an underlying econometric model fits both model frameworks, then the posterior probability statements concerning the identified set in this paper can be interpreted as bounds on the possible posteriors for the partially identified parameter. However, the frameworks differ in their compatibility with underlying econometric models. For example, it can be difficult to specify the likelihood for incomplete structural models (e.g., models of games as in Example 2) or moment inequality models (e.g., Example 5 in the Supplement).

---

[12]See also Giacomini and Kitagawa (2014).

[13]Therefore, in particular, for an underlying econometric model that is compatible with both frameworks, the frameworks differ in the specifics of the priors. For example, in structural econometric models, a prior is either placed on the "sufficient parameter" $\phi$ of the underlying likelihood or the "summary statistics" $\mu$ that is generated by the underlying econometric model. Despite a (possibly difficult to characterize) one-to-one correspondence between $\phi$ and $\mu$, because those parameters have different direct interpretations there is a practical difference between specifying a prior on $\phi$ and $\mu$.

## 5. Frequentist properties of the credible sets

A credible set for $\Delta_I$ is a set $\mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ that satisfies the following definition.

Definition 4.  For some $\alpha \in (0, 1)$,

$$\mathcal{C}^{\Delta_I}_{1-\alpha}(X) \text{ has the property that } \Pi\big(\Delta_I \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)|X\big) = 1 - \alpha.$$

Under a set of minimal regularity conditions, this section establishes necessary and sufficient conditions for $\mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ to be a valid *exact* frequentist confidence set for the identified set, in the sense that $P(\Delta_I \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)) \approx 1 - \alpha$ in repeated large samples. In general, the definition of a confidence set allows conservative coverage, $P(\Delta_I \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)) \gtrapprox 1 - \alpha$. Based on previous results comparing Bayesian and frequentist inference under partial identification (i.e., Moon and Schorfheide (2012)), but for the partially identified parameter rather than the identified set, the leading concern appears to be the opposite case: a Bayesian credible set that does not even achieve at least the required frequentist coverage. Nevertheless, the results in this section do not address the possibility of a Bayesian credible set that has conservative frequentist coverage.[14] The computation of the credible set is discussed in Remark 5, and in Section 6, alongside other discussion of computational implementation.

Under the sufficient conditions, these results reveal an "asymptotic equivalence" between Bayesian and frequentist inference in partially identified models, implying that $\mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ can also be used by frequentist econometricians, even for functions of the partially identified parameter (without conservative projection methods).[15] However, it is worth noting that the frequentist coverage may not be uniform, an important problem addressed in the frequentist literature (see prior references). It is also worth noting that the necessary and sufficient condition can be false, in which case the credible set fails to be an exact frequentist confidence set. But even in those cases, the credible set is valid from the Bayesian perspective, which has been the main focus of this paper. These properties are highlighted in the Monte Carlo experiments in Section 7.

### 5.1 *Asymptotic independence of the credible set*

The proof of Theorem 5 establishes that in repeated samples,

$$P\big(\Delta_I \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)\big)$$
$$= P\bigg( \sqrt{n}\big(\mu_0 - \mu_n(X)\big) \in \sqrt{n}\bigg( \bigcap_{\delta \in (\mathcal{C}^{\Delta_I}_{1-\alpha}(X))^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C - \mu_n(X) \bigg) \bigg).$$

---

[14]Related results reconciling Bayesian and frequentist inference in point identified models (i.e., the literature on the Bernstein–von Mises theorem) is analogous, in the sense that it generally shows that the Bayesian posterior distribution is asymptotically the same as the frequentist sampling distribution, and therefore that a confidence set at the $1 - \alpha$ significance level is asymptotically the same as a credible set at the $1 - \alpha$ credibility level. See for example Freedman (1999).

[15]One caveat to claims about exact credible sets concerns computation of the credible set. Some computationally attractive methods for computing the credible set may result in slight "overcoverage," but in principle, with sufficient computing time, exact posterior probabilities are possible.

Use the notation that $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X) = \sqrt{n}(\bigcap_{\delta \in (\mathcal{C}_{1-\alpha}^{\Delta_I}(X))^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C - \mu_n(X))$. Therefore, it is necessary to make an assumption concerning the joint sampling distribution of $\sqrt{n}(\mu_0 - \mu_n(X))$ and $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$.

The set $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ is the set of $\sqrt{n}(\mu - \mu_n(X))$ consistent with $\Delta_I(\mu) \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)$. Theorem 3 implies that $P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)) \approx 1 - \alpha$ for each large data set, since $\mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ is a credible set. Further, under reasonable conditions on $\mu_n(X)$ (see Assumption 6), the repeated large sample distribution of $\sqrt{n}(\mu_0 - \mu_n(X))$ is $N(0, \Sigma_0)$. However, those properties do not necessarily uniquely characterize the joint sampling distribution.

Use the notation that $F_n(A) = P(\sqrt{n}(\mu_0 - \mu_n(X)) \in A)$ for any Borel set $A$.

ASSUMPTION 5 (Asymptotic Independence of Credible Sets). *It holds that*

$$\left| P\big(\sqrt{n}(\mu_0 - \mu_n(X)) \in \tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\big) - E\big(F_n\big(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\big)\big) \right| \to 0 \quad as\ n \to \infty.$$

This asymptotic independence assumption concerns repeated sampling behavior, and therefore is inherently a frequentist (and non-Bayesian) concept. It is motivated by and related to an assumption that, in sampling distribution, $\sqrt{n}(\mu_0 - \mu_n(X))$ and $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ are independent for all sufficiently large sample sizes. Under that independence assumption, the condition in Assumption 5 holds with equality in sufficiently large sample sizes:

$$P\big(\sqrt{n}(\mu_0 - \mu_n(X)) \in \tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\big)$$
$$= E\big(1\big[\sqrt{n}(\mu_0 - \mu_n(X)) \in \tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\big]\big)$$
$$= E_{\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)} E_{\sqrt{n}(\mu_0 - \mu_n(X))}\big(1\big[\sqrt{n}(\mu_0 - \mu_n(X)) \in \tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\big]\big)$$
$$= E\big(F_n\big(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\big)\big).$$

Therefore, Assumption 5 can be understood to be an assumption that requires that in sampling distribution, $\sqrt{n}(\mu_0 - \mu_n(X))$ and $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ are "almost" independent for sufficiently large sample sizes.

### 5.2 *Characterization of the frequentist properties of the credible set*

ASSUMPTION 6 (Repeated Sampling Behavior of the Estimator of $\mu$). *The estimator* $\mu_n(X)$ *appearing in Assumption 3 satisfies one of the following properties*:

(a) *The distribution $\sqrt{n}(\mu_0 - \mu_n(X))$ converges in total variation to $N(0, \Sigma_0)$.*

(b) *The distribution $\sqrt{n}(\mu_0 - \mu_n(X))$ converges in distribution to $N(0, \Sigma_0)$ for nonsingular $\Sigma_0$ and $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ is a finite union[16] of disjoint convex Borel sets.*

---

[16]There must be a number $K$ such that $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ is the union of at most $K$ disjoint convex Borel sets, for all realizations of the data.

This is essentially the "frequentist" version of Assumption 3. The fact that the asymptotic covariances in Assumptions 3 and 6 are the same is part of the conclusion of the various Bernstein–von Mises-like theorems referenced after the statement of Assumption 3. Central limit theorems establishing convergence in total variation are available (e.g., Van der Vaart (1998, Theorem 2.31)), and more generally Scheffé's lemma (e.g., Van der Vaart (1998, Corollary 2.30)) relates convergence of densities to convergence in total variation. If $\mu_n(X)$ has a discrete sampling distribution, then it cannot converge in total variation to the continuously distributed $N(0, \Sigma_0)$. If the identified set mapping and credible set is of sufficiently low topological "complexity" so that $\tilde{\Delta}_{1-\alpha}^{-1, \Delta_I}(X)$ is a finite union of disjoint convex Borel sets, then Assumption 6 requires only convergence in distribution. For example, that condition holds in the case of a simple interval identified parameter and interval credible set, as illustrated in Example 6 in the Supplement. More generally, because Assumption 6 is used only in one place in the proof of Theorem 5 to establish a convergence related to sets related to the credible sets, other conditions that also establish that convergence could be substituted for Assumption 6.

The following theorem establishes the frequentist coverage properties of $\mathcal{C}_{1-\alpha}^{\Delta_I}(X)$. This theorem can be viewed as extending the Bernstein–von Mises results from the point identified parameter $\mu$ to the identified set for the partially identified parameter $\delta$.[17]

THEOREM 5. *Suppose that for all realizations of the data $X$, $\mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ is a credible set for the identified set, in the sense that*

$$\Pi\big(\Delta_I \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)|X\big) = 1 - \alpha.$$

*Suppose also that Assumptions 1, 3, and 6 obtain. Assumption 5 obtains if and only if $\mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ are exact frequentist confidence sets*:

$$P\big(\Delta_I \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)\big) \to 1 - \alpha.$$

In general, it is necessary to study Assumption 5 on a case-by-case basis, as it depends on the model-specific structure of the identified set, similar to how inference in "nonstandard models" tends to proceed on a case-by-case basis. However, an important *sufficient condition* for Assumption 5 is discussed in Remark 5 below, with the result collected in Lemma 1 that follows.

REMARK 5 (Sufficient Condition: Smooth Interval Identified Set). Suppose that the identified set for $\delta$ is an interval $\Delta_I(\mu) = [\Delta_{IL}(\mu), \Delta_{IU}(\mu)]$, where $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ are functions that may not be explicitly known by the econometrician. The identified set for $\delta$ is an interval in many important cases, including the case where the identified set for $\theta$ is convex and $\delta$ is a scalar element of $\theta$. Suppose that the credible set has the form $\mathcal{C}_{1-\alpha}^{\Delta_I}(X) = [\Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}]$, where $c_{1-\alpha}(X)$ is chosen to

---

[17]In particular, in the case of a point identified $\delta$ with $\delta \equiv \Delta_I(\mu)$, where $\Delta_I(\cdot)$ satisfies the regularity conditions of the (Bayesian) delta method, then arguments similar to the proof of Lemma 1 establish that Assumption 5 is satisfied for the credible set $\mathcal{C}_{1-\alpha}^{\Delta_I}(X) = [\Delta_I(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_I(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}]$, where $c_{1-\alpha}(X)$ is chosen to have the credible set property, given that Assumptions 1, 3, and 6 obtain.

have the credible set property. Suppose that for all $\mu$ in a neighborhood of $\mu_0$, $\Delta_I(\mu) \neq \emptyset$. This essentially requires that the identified set be nonempty in a sufficiently small neighborhood around $\mu_0$.[18] Then if $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ satisfy the regularity conditions of the (Bayesian) delta method, in a neighborhood of $\mu_0$, with positive definite covariance, then Assumption 5 is satisfied. This result is formalized in Lemma 1.

The existence of derivatives of $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ with respect to $\mu$ from the delta method rules out kinks in $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ at $\mu_0$, for example intersection bounds with multiple simultaneously binding constraints at $\mu_0$.[19] If the functions $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ are explicitly known by the econometrician (e.g., Example 3), then existence of derivatives can be checked directly. If the functions $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ are only implicitly known by the econometrician, then other methods are required. In particular, in some models it is possible to write $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ as the optimal value functions of an optimization problem with parameterized constraints. For example, if the criterion function is $Q(\theta, \mu) = \|f(\theta) - \mu\|$ and $\Delta(\theta) = \theta_k$ (e.g., Example 2), then $\Delta_{IL}(\mu)$ is the solution to minimizing $\theta_k$ subject to the parameterized constraints $f(\theta) = \mu$ and $\Delta_{IU}(\mu)$ is the solution to maximizing $\theta_k$ subject to the parameterized constraints $f(\theta) = \mu$. Sufficient conditions for the differentiability of these optimal value functions are provided in the optimization literature (e.g., Fiacco and McCormick (1990, Section 2.4)). The requirement of a positive definite covariance rules out identified sets such that $\Delta_{IL}(\cdot)$ and/or $\Delta_{IU}(\cdot)$ have zero derivative at $\mu_0$. In particular, the requirement of a positive definite covariance rules out identified sets such that one or both of $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ are functions of a scalar element of $\mu$, and are nonmonotonic at $\mu_0$ (which implies $\Delta_{IL}(\cdot)$ and/or $\Delta_{IU}(\cdot)$ have zero derivative at $\mu_0$).[20] Since the frequentist coverage is not necessarily uniform, frequentist inference based on the Bayesian credible set can have poor performance in small samples if these conditions are "almost" violated.

The credible set $\mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ can be computed by computing an "estimate" of the identified set (i.e., $[\Delta_{IL}(\mu_n(X)), \Delta_{IU}(\mu_n(X))]$) and then symmetrically "expanding" from that estimate outward until the credible set achieves the required Bayesian credibility level. The identified set is "estimated" by computing the identified set at $\mu_n(X)$ rather than a draw from the posterior $\mu|X$. Using the approach discussed in Remark 4, Kitagawa (2012) provides a computationally attractive method for computing a shortest-width interval.

LEMMA 1. *Suppose that Assumptions 1, 3, and 6 obtain. Suppose also that the setup in this remark obtains: both the Bayesian and frequentist delta methods (e.g., Bernardo and Smith (2009, Section 5.3)) apply to $(\Delta_{IL}(\mu), \Delta_{IU}(\mu))$ with the same full rank covariance, and for all $\mu$ in a neighborhood of $\mu_0$, $\Delta_I(\mu) \neq \emptyset$. Then Assumption 5 is satisfied for $\mathcal{C}_{1-\alpha}^{\Delta_I}(X) = [\Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}]$, where $c_{1-\alpha}(X)$ is chosen to have the credible set property.*

---

[18]Note that in many models this rules out point identification, since in many models if $\Delta_I(\mu_0)$ is a singleton, then some $\mu$ in any neighborhood of $\mu_0$ results in $\Delta_I(\mu) = \emptyset$.

[19]The reconciliation between robust Bayes credible sets and frequentist confidence sets, in Kitagawa (2012), also tends to not hold in this sort of setting.

[20]For example, suppose that $\mu = (\mu_L, \mu_U)$ and $\Delta_{IL}(\mu) = \mu_L^2$. Then $\Delta_{IL}(\cdot)$ is nonmonotonic at $\mu_{0L} = 0$ and has zero derivative at $\mu_{0L} = 0$.

A generic converse of Theorem 5, a result that says that any frequentist confidence set can be interpreted as an approximation (in large samples) to a Bayesian credible set, is not available. For example, one $(1 - \alpha)$-level confidence set is the entire parameter space with probability $1 - \alpha$ and the empty set with probability $\alpha$. This cannot be expected to have a Bayesian interpretation, even though it is a valid frequentist confidence set.[21] One method to "nudge" a desired frequentist confidence set to have at least a minimal Bayesian interpretation is to compute that frequentist confidence set as usual, compute the Bayesian credible set proposed in this paper, and then report the union of those two sets. This will inherit all of the coverage properties of both underlying approaches, although of course it can be "conservative" from one or both perspectives.

REMARK 6 (Frequentist Properties of the Credible Set for the Partially Identified Parameter).   A credible set for the partially identified parameter is $\mathcal{C}^\delta_{1-\alpha}(X) \equiv \{\delta^* : \Pi(\delta^* \in \Delta_I | X) \geq \alpha\}$. Roughly, since $\delta^* \in \Delta_I$ means that the model specification with $\delta^*$ generates the same distribution of the data as does the true data generating process, $\mathcal{C}^\delta_{1-\alpha}(X)$ can be viewed as collecting all model specifications (i.e., specifications of $\delta$) that have at least $1 - \alpha$ posterior probability of generating the same distribution of the data as the true data generating process. Alternatively, $\mathcal{C}^\delta_{1-\alpha}(X)$ can be viewed as collecting all model specifications for which there is at least a minimal amount of evidence (in the above sense). It is a necessary implication of this definition that it is possible that $\mathcal{C}^\delta_{1-\alpha}(X)$ is the empty set, particularly for large $\alpha$ and/or situations of (near) point identification. Consider the limiting situation of point identification. Then $\delta^* \in \Delta_I$ is equivalent to $\delta^*$ being the singleton "true value" of $\delta$. Often there will not be high posterior probability that any particular $\delta^*$ is the "true value" of $\delta$ (e.g., if the "posterior for $\delta$" is an ordinary density), in which case $\mathcal{C}^\delta_{1-\alpha}(X)$ may be the empty set.

A related possibility is to report the set $\mathcal{R}^\delta_r(X) \equiv \{\delta^* \in \Delta(\Theta) : \Pi(\delta^* \in \Delta_I | X) \geq r \max_\delta \Pi(\delta \in \Delta_I | X)\}$ for some $r \in (0, 1)$. This is a *highest relative odds set* for $\delta$, in the sense that $\mathcal{R}^\delta_r(X)$ is the set of all values $\delta^*$ that are at least $r$ times as likely to be in the identified set as the most likely parameter value. In some but not all cases $\mathcal{C}^\delta_{1-\alpha} \approx \mathcal{R}^\delta_\alpha(X)$, because in some but not all cases $\max_\delta \Pi(\delta \in \Delta_I | X) \approx 1$.

For this to be a valid frequentist confidence set, considering $\theta$ rather than some $\delta$ of interest for simplicity, it must be that for any $\theta^* \in \Theta_I$ that in repeated large samples $P(\theta^* \in \mathcal{C}^\theta_{1-\alpha}(X)) \geq 1 - \alpha$, or equivalently that $P(\Pi(\theta^* \in \Theta_I | X) \geq \alpha) \geq 1 - \alpha$, or equivalently $P(\Pi(\theta^* \in \Theta_I | X) < \alpha) \leq \alpha$. Therefore, essentially, it must be that $\Pi(\theta^* \in \Theta_I | X)$ has the $U[0, 1]$ distribution in repeated large samples, or stochastically dominates the $U[0, 1]$ distribution, or equivalently it must be that $\Pi(\theta^* \in \Theta_I | X)$ can be interpreted as a (possibly conservative) $p$-value for the null hypothesis that $\theta^* \in \Theta_I$. By the large sample approximation in Theorem 1, for fixed realization of the data $X$, $\Pi(\theta^* \in \Theta_I | X) \approx P_{N(0, \Sigma_0)}(\sqrt{n}(\mu_I(\theta^*) - \mu_n(X))) = P_{N(0, \Sigma_0)}(\sqrt{n}(\mu_I(\theta^*) - \mu_0 + \mu_0 - \mu_n(X)))$. In repeated large samples, this is distributed approximately as $P_{N(0, \Sigma_0)}(\sqrt{n}(\mu_I(\theta^*) - \mu_0) + N(0, \Sigma_0))$. So the credible set for the partially identified parameter is a valid frequentist confidence

---

[21] In point identified models, the "obvious" frequentist confidence set to study is the confidence set based on inverting the Wald test based on the asymptotic approximation to $\mu_n(X)$, but that is not sensible in partially identified models.

set whenever $P_{N(0,\Sigma_0)}(\sqrt{n}(\mu_I(\theta^*) - \mu_0) + N(0, \Sigma_0))$ is (or stochastically dominates) the $U[0, 1]$ distribution. (Obviously, this is only a heuristic argument as $n$ appears in the "limiting" distribution.) For example, this is true in the important special case of an interval identified parameter, without point identification, from Example 1. See also Kline (2011) for cases where it is not true.

REMARK 7 (Measurability of $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$). The discussion in this section treats $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ essentially as a random variable. This is understood to be justified based on the underlying measurability of the random variables that characterize the set $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$: $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ is "equivalent" to the bundle of random variables that characterize $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ plugged into the functional form for $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$.

REMARK 8 (An Alternative Credible Set). Another approach to constructing a credible set for the identified set is to project a credible set for $\mu$ onto the space of subsets of $\Delta(\Theta)$. That is, for any credible set $\mathcal{C}_{1-\alpha}^{\mu}(X)$ for $\mu$, $\Delta_I(\mathcal{C}_{1-\alpha}^{\mu}(X)) = \{\delta : \exists \mu \in \mathcal{C}_{1-\alpha}^{\mu}(X) \text{ s.t. } \delta \in \Delta_I(\mu)\}$ is a credible set for the identified set, such that $\Pi(\Delta_I(\mu) \subseteq \Delta_I(\mathcal{C}_{1-\alpha}^{\mu}(X))|X) \geq 1 - \alpha$. Moreover, because per Lemma 2, $\Delta_I(\mu) \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ is logically equivalent to $\mu \in \bigcap_{\delta \in (\mathcal{C}_{1-\alpha}^{\Delta_I}(X))^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C$, any $1 - \alpha$ credible set for the identified set can be associated with a $1 - \alpha$ credible set for $\mu$: $\bigcap_{\delta \in (\mathcal{C}_{1-\alpha}^{\Delta_I}(X))^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C$. Under the condition that the credible set for $\mu$ is also a valid frequentist confidence set under Bernstein–von Mises-like conditions, then also this credible set for the identified set will be a valid frequentist confidence set for the identified set, in the sense of having at least the required coverage probability. However, as with projection methods in general, such an approach is likely to be conservative (from both the Bayesian and frequentist perspectives), unless the credible set for $\mu$ is somehow constructed in a special way to avoid conservativeness under the projection. That is, even though every $1 - \alpha$ credible set for the identified set can be associated with a $1 - \alpha$ credible set for $\mu$, in general a $1 - \alpha$ credible set for $\mu$ will project as a greater than $1 - \alpha$ credible set for the identified set. This sort of approach is mentioned in Moon and Schorfheide (2009).

## 6. Computational implementation

An important feature of this approach is that it is computationally attractive even in high-dimensional models. In general, inference is accomplished by the following sampler that can be used to approximate the posterior probabilities:

Step 1. Generate a large sample $\{\Delta(\Theta_I(\mu^{(s)}))\}_{s=1}^S$ according to the following procedure:

(a) Draw $\mu^{(s)} \sim \mu|X$ by any method that is appropriate for $\Pi(\mu|X)$.

(b) Compute $\Delta(\Theta_I(\mu^{(s)}))$, the identified set at $\mu^{(s)}$.

Step 2. Based on $\{\Delta(\Theta_I(\mu^{(s)}))\}_{s=1}^S$, compute an approximation to the desired posterior probability.

For example, $\Pi(\Delta^* \subseteq \Delta_I | X)$ is the percentage of the draws $\{\Delta(\Theta_I(\mu^{(s)}))\}_{s=1}^S$ such that indeed $\Delta^* \subseteq \Delta(\Theta_I(\mu^{(s)}))$, and a credible set (i.e., Definition 4) is a set that contains $1 - \alpha$ percent of the draws $\{\Delta(\Theta_I(\mu^{(s)}))\}_{s=1}^S$.

By separating the "inference" problem, which concerns the posterior $\mu | X$ (not the whole parameter space), from the remaining computational problem of determining the identified set for $\theta$ evaluated at a particular value of $\mu$, which admits a variety of analytic and computational simplifications, it is possible to avoid in general the sorts of "exhaustive search" grid search (or "guess and verify") procedures that are commonly used to construct frequentist confidence sets.

## 6.1 *Computational approaches*

Step 1(b) involves getting the set $\Theta_I(\mu)$ for a given draw of $\mu$ from the posterior $\mu | X$, which is the problem of finding all solutions in $\theta$ to $Q(\theta, \mu) = 0$ for a given $\mu$. The computational difficulty is increased due to the necessity of finding the *set* of solutions, rather than just one of the solutions. The best approach to Step 1(b) depends on the application.

One approach involves "guessing and verifying": guessing values of $\theta$ and verifying whether $Q(\theta, \mu) = 0$. That will always work, but often there are much faster approaches.

In some models, $\Theta_I(\mu)$ has a known expression as a function of $\mu$ that is computationally simpler than checking whether each $\theta \in \Theta$ satisfies $\theta \in \Theta_I(\mu)$. For example, in a simple interval identified parameter model, $\Theta_I(\mu) = [\mu_L, \mu_U]$. This is computationally simpler than computing the identified set by guessing and verifying based on the definition that $\Theta_I(\mu) \equiv \{\theta : Q(\theta, \mu) = 0\}$.

In some other models, and for some $\Delta(\cdot)$, it is possible to simplify the computation of $\Delta(\Theta_I(\mu))$. For example, suppose that $\Theta_I(\mu)$ is a compact and convex set, and that $\Delta(\theta) = \theta_k$ is the $k$th element of $\theta$. Then $\Delta(\Theta_I(\mu))$ is a finite closed interval in $\mathbb{R}$. Consequently, $\Delta(\Theta_I(\mu))$ can be computed by computing $\min_{\theta \in \Theta_I(\mu)} \theta_k$ and $\max_{\theta \in \Theta_I(\mu)} \theta_k$, which can be computationally simpler than guessing and verifying by computing $\Theta_I(\mu)$ and then checking whether each $\delta \in \Delta(\Theta)$ satisfies $\delta \in \Delta(\Theta_I(\mu))$. This is demonstrated by example in Section S3.2 of the Supplement in a Monte Carlo experiment involving interval data on the outcome in a linear regression model.

## 6.2 *Markov chain Monte Carlo approximation*

It may only be known that $\Theta_I(\mu) \equiv \{\theta : Q(\theta, \mu) = 0\}$, without any known analytic simplifications as above. If so, then some numerical method must be applied to compute $\Theta_I(\mu)$. One approach is based on simulating a random variable whose support is the identified set.

Let

$$f_{\Theta_I(\mu)}(\theta) = \frac{1[Q(\theta, \mu) = 0]}{\lambda(\Theta_I(\mu))}$$

be the ordinary Lebesgue density of the uniform distribution on $\Theta_I(\mu)$, where $\lambda(\cdot)$ is Lebesgue measure on $\Theta$. If $\Theta_I(\mu)$ is measurable and bounded with positive Lebesgue

measure, then $f_{\Theta_I(\mu)}$ is well defined and has support on $\Theta_I(\mu)$. Consequently, any method that can simulate draws from the density $f_{\Theta_I(\mu)}$ can be used to numerically approximate $\Theta_I(\mu)$, by taking the approximation of $\Theta_I(\mu)$ to be *the support* of the simulated draws from $f_{\Theta_I(\mu)}$. However, the normalizing constant $\lambda(\Theta_I(\mu))$ is difficult to determine, because it is difficult to explicitly characterize $\Theta_I(\mu)$. Therefore, let

$$\tilde{f}_{\Theta_I(\mu)}(\theta) = 1\big[Q(\theta,\mu) = 0\big]$$

be the corresponding unnormalized density. There are many methods for simulating draws from an unnormalized density: among these methods are Metropolis–Hastings sampling and slice sampling. See for example Gamerman and Lopes (2006) for a textbook on related methods.

In some cases, especially when $\Theta_I(\mu)$ has empty interior, that (unnormalized) density may not perform well because the density is supported on a lower-dimensional subspace. In those cases, it is possible to use the alternative unnormalized density

$$\tilde{f}_{\Theta_I(\mu),T}(\theta) = \exp\bigg(\frac{-Q(\theta,\mu)}{T}\bigg),$$

where $T > 0$ is a small tuning parameter.[22] Then $\tilde{f}_{\Theta_I(\mu),T}(\theta) = 1$ on $\Theta_I(\mu)$, and $\tilde{f}_{\Theta_I(\mu),T}(\theta) \approx 0$ far from $\Theta_I(\mu)$ (i.e., when $Q(\theta,\mu) \gg 0$ and/or $T$ is small). Therefore, $\Theta_I(\mu)$ can be simulated as $\Theta_I(\mu) \approx \{\theta : \hat{f}(\theta) > 1 - \varepsilon\}$ for small $\varepsilon > 0$, where $\hat{f}(\theta)$ is the density of the simulated draws from $\tilde{f}_{\Theta_I(\mu),T}(\theta)$. In practice, it seems reasonable to take $\Theta_I(\mu)$ to be the support of the draws from $\tilde{f}_{\Theta_I(\mu),T}$. This will potentially result in a numerical approximation of the identified set that is "too big," but that is generally acceptable in the literature on partially identified models (as "nonsharp" identified sets). Another possibility is to check that each of the draws from $\tilde{f}_{\Theta_I(\mu),T}(\theta)$ at least approximately satisfies the condition that the criterion function evaluated at the draw equals zero,[23] which will sharpen the numerical approximation of the identified set.

There are many methods for drawing from unnormalized densities in the Markov chain Monte Carlo literature. Particularly from the perspective of the difficulty of the computational implementation, slice sampling (e.g., Neal (2003)) is recommended. Specifically, the Monte Carlo experiments and empirical application are based on the `slicesample` implementation that is provided in MATLAB. More generally, slice sampling is implemented in many computational and statistical software packages. Some implementations require an initial "guess" for $\theta$ in the identified set (i.e., a guess for where the "density" is nonzero). This can be accomplished by finding one solution to $Q(\theta,\mu) = 0$ by a standard optimization method. One useful feature of slice sampling is

---

[22]It can be shown under certain conditions that as $T \to 0$, the limit of the sequence $\tilde{f}_{\Theta_I(\mu),T}$ is supported on the set of minimizers (the identified set). Consequently, as discussed in the text, with small $T$, most draws from the density $\tilde{f}_{\Theta_I(\mu),T}$ will be close to $\Theta_I(\mu)$. See Hwang (1980).

[23]In some models, it may not be desirable to require that the criterion function evaluated at the draw equals exactly zero. For example, if the evaluation of the criterion function itself involves a complicated numerical problem (like evaluating a multivariate normal cumulative distribution function) that is subject to numerical error, a "numerical error tolerance" may be desired.

that it does not require the specification of auxiliary distributions (e.g., a proposal distribution) required by some other methods like Metropolis–Hastings sampling. Overall, the advantage of this approach is the low difficulty of the programming required, because of built-in slice sampling implementations. Generically, it is enough to program the criterion function $Q(\mu, \theta)$ and the density $\tilde{f}_{\Theta_I(\mu)}(\theta)$ or $\tilde{f}_{\Theta_I(\mu), T}(\theta)$, and then apply the slice sampling implementation to that density.

## 7. Monte Carlo experiments

This section reports Monte Carlo experiments that illustrate the behavior of this approach to inference. The Supplement provides further Monte Carlo experiments in the context of moment inequality models (a simple interval identified parameter and regression with interval data).

### 7.1 *Binary entry game*

This section reports the results of a Monte Carlo experiment in the context of a simple version of a binary entry game. A related model will be estimated with real data in Section 8. For the experiment, consider the standard specification of a binary entry game described in Table 1.

In each cell, the first entry is the payoff to player 1, and the second entry is the payoff to player 2. It is assumed that $\Delta_1$ and $\Delta_2$ are both negative, and that players play a pure strategy Nash equilibrium. This game admits two pure strategy Nash equilibria when $-\beta_i \leq \varepsilon_i \leq -\beta_i - \Delta_i$, $i = 1, 2$: in this region, there are no assumptions on equilibrium selection. The true parameters are set at $\Delta_{01} = -0.5 = \Delta_{02}$ and $\beta_{01} = 0.2 = \beta_{02}$, and $\varepsilon_1$ and $\varepsilon_2$ are jointly normally distributed with variance 1 and correlation $\rho_0 = 0.5$, and this correlation is constrained by the econometrician to be positive. It is assumed to be known that the econometrician correctly knows the sign of the parameters.

There are six parameters: $\beta_1, \beta_2, \Delta_1, \Delta_2, \rho$, and the equilibrium selection probability for the region of multiple equilibria. The equilibrium selection probability is "profiled out," as described below when defining the criterion function. The point identified parameter $\mu$ is the vector of choice probabilities $\mu = (P_{11}, P_{10}, P_{01}, P_{00})$, where $P_{a_1 a_2}$ is the probability that player 1 takes action $a_1$ and player 2 takes action $a_2$, and the partially identified parameter is $\theta = \delta = (\beta_1, \Delta_1, \beta_2, \Delta_2, \rho)$. The mapping that links $\mu$ to the identified set for $\theta$ results from the assumptions made on the game, as follows.

Table 1. Payoff matrix for the binary entry game.

|  |  | Player 2 | | | |
|---|---|---|---|---|---|
|  |  | 0 | | 1 | |
| Player 1 | 0 | 0 | 0 | 0 | $\beta_2 + \varepsilon_2$ |
|  | 1 | $\beta_1 + \varepsilon_1$ | 0 | $\beta_1 + \Delta_1 + \varepsilon_1$ | $\beta_2 + \Delta_2 + \varepsilon_2$ |

The criterion function is $Q(\theta, \mu) = (P_{11} - P_{11}(\theta))^2 + (P_{10} - P_{10}(\theta))^2 + (P_{01} - P_{01}(\theta))^2 + (P_{00} - P_{00}(\theta))^2 + \min\{|s(\theta, \mu)|, |s(\theta, \mu) - 1|\}(1 - 1[0 \le s(\theta, \mu) \le 1])$, where $P_{00}(\theta) = P(\varepsilon_1 \le -\beta_1, \varepsilon_2 \le -\beta_2)$ and $P_{11}(\theta) = P(\varepsilon_1 \ge -\beta_1 - \Delta_1, \varepsilon_2 \ge -\beta_2 - \Delta_2)$ correspond to the model-predicted probabilities of the outcomes that occur only as a unique equilibrium, at $\theta$. The $s(\theta, \mu)$ term is the candidate equilibrium selection probability at $\theta$ and $\mu$, described below.

The terms $P_{10}(\theta)$ and $P_{01}(\theta)$ are more complicated, as they correspond to the model-predicted probabilities of outcomes that occur in the region of multiple equilibria. By the law of total probability and using the definition of pure strategy Nash equilibrium,

$$P_{01}(\theta) = P(-\beta_1 \le \varepsilon_1 \le -\beta_1 - \Delta_1, \varepsilon_2 \ge -\beta_2 - \Delta_2) + P(\varepsilon_1 \le -\beta_1, \varepsilon_2 \ge -\beta_2)$$

$$+ s \times P(-\beta_1 \le \varepsilon_1 \le -\beta_1 - \Delta_1, -\beta_2 \le \varepsilon_2 \le -\beta_2 - \Delta_2),$$

where the parameter $s$ represents the equilibrium selection probability (of choosing the $(0, 1)$ equilibrium) in the region of multiple equilibria. Since it must be that $P_{01} = P_{01}(\theta)$ in the identified set, there is a unique candidate value for $s$ after fixing $\theta$ and $\mu$, given by $s(\theta, \mu) = \frac{P_{01} - (P(-\beta_1 \le \varepsilon_1 \le -\beta_1 - \Delta_1, \varepsilon_2 \ge -\beta_2 - \Delta_2) + P(\varepsilon_1 \le -\beta_1, \varepsilon_2 \ge -\beta_2))}{P(-\beta_1 \le \varepsilon_1 \le -\beta_1 - \Delta_1, -\beta_2 \le \varepsilon_2 \le -\beta_2 - \Delta_2)}$. For this to be a valid probability, it must be that $0 \le s(\theta, \mu) \le 1$, explaining that part of the criterion function. The expression for $P_{10}(\theta)$ is similar (and is uniquely determined by the others since probabilities sum to 1). When simulating data from the game, $(1, 0)$ and $(0, 1)$ are actually chosen with equal probability whenever the game is in the region of multiple equilibria, but this is not known by the econometrician.

So as to compute the identified set, the slice sampler is used to sample from the "density" $\tilde{f}_{\Theta_I(\mu)}(\theta) = 1[Q(\theta, \mu) = 0]$, as described in Section 6.2.[24] The support of draws from $\tilde{f}_{\Theta_I(\mu)}(\theta)$ is taken to be the identified set for $\theta$ evaluated at that value of $\mu$, which is then used in the sampler described at the beginning of Section 6. Moreover, the identified set evaluated at that value of $\mu$, for any function $\Delta(\cdot)$ of $\theta$, can be taken to be $\Delta(\cdot)$ applied to that computed identified set. In particular, the identified sets for subvectors of $\theta$ can be easily computed by "ignoring" the other elements of $\theta$. By computing the identified set at each draw $\mu^{(s)}$ from a sample of draws from the posterior $\mu|X$, it is possible to simulate draws from the posterior distribution "over the identified set." Based on numerical approximation, the parameters are not point identified (which is not surprising since there are four equations (one of which is redundant) and six unknowns). The true marginal identified sets for $\Delta_1$ and $\Delta_2$ are each approximately $[-1.50, -0.04]$,[25]

---

[24]So as to account for numerical error in the computation of the multivariate normal cumulative distribution function, actually a small tolerance is allowed; that is, the criterion function can be very slightly above zero. The tolerance implies that in practice the "density" is $1[Q(\theta, \mu) \le 0.0015]$.

[25]By numerical approximation, 0 is not in the identified sets. This is also possible to see analytically. Suppose that indeed $(\Delta_1, \Delta_2) = (0, 0)$. Then it must be that $\beta_i = -\Phi^{-1}(P(y_i = 0))$. For this data generating process, $P(y_i = 0) > \frac{1}{2}$, so $\beta_i < 0$. Further, $P_{01} = P(\varepsilon_1 \le \beta_1, \varepsilon_2 \ge -\beta_2) + P(\beta_1 \le \varepsilon_1 \le 0, \varepsilon_2 \ge -\beta_2) + P(0 \le \varepsilon_1 \le -\beta_1, \varepsilon_2 \ge -\beta_2)$ and $P_{00} = P(\varepsilon_1 \le \beta_1, \varepsilon_2 \le \beta_2) + P(\beta_1 \le \varepsilon_1 \le 0, \varepsilon_2 \le \beta_2) + P(0 \le \varepsilon_1 \le -\beta_1, \varepsilon_2 \le \beta_2) + P(\varepsilon_1 \le -\beta_1, \beta_2 \le \varepsilon_2 \le -\beta_2)$. By the rotational symmetry property of the multivariate normal distribution, $P_{00} - P_{01} = P(\varepsilon_1 \le \beta_1, \varepsilon_2 \le \beta_2) - P(\varepsilon_1 \le \beta_1, \varepsilon_2 \ge -\beta_2) + P(\varepsilon_1 \le -\beta_1, \beta_2 \le \varepsilon_2 \le -\beta_2) \ge P(\varepsilon_1 \le \beta_1, \varepsilon_2 \le \beta_2) - P(\varepsilon_1 \le \beta_1, \varepsilon_2 \ge -\beta_2)$ since some terms cancel. This is nonnegative since $\rho \ge 0$. But for this data generating process, this is actually false (albeit numerically close to being true). So it cannot be that $(\Delta_1, \Delta_2) = (0, 0)$ is in the identified set.

(a) $\Delta_1$

(b) $\beta_1$

(c) $(\beta_1, \Delta_1)$

(d) Identified set for $(\beta_1, \Delta_1)$
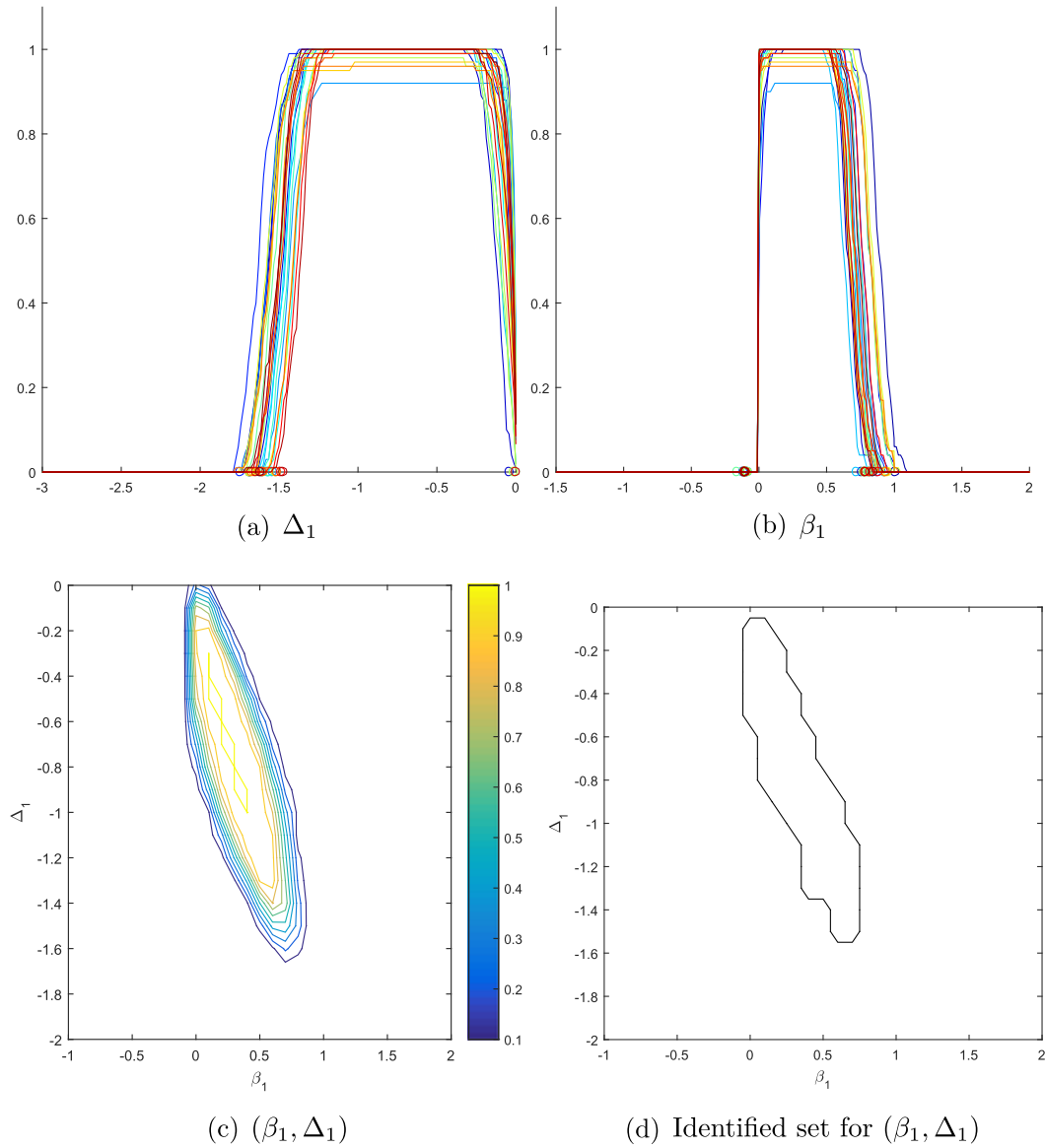
FIGURE 1. Posteriors $\Pi(\cdot|X)$ for various parameters.

while the true identified sets for $\beta_1$ and $\beta_2$ are each approximately $[0, 0.75]$. Further, the data appear to be uninformative about the correlation coefficient, in the sense that the identified set is essentially the entire parameter space.

Figure 1 displays posterior probabilities that various values of the parameters belong to the identified set based on samples of size $N = 500$ from this data generating process. Each posterior "curve" of a different gray shade in panels 1(a) and 1(b) corresponds to a different draw from the data generating process. The $\mu$ parameters are multinomial,

so an uninformative conjugate Dirichlet prior is used, implying a Dirichlet posterior for $\mu|X$.

Panel 1(a) displays the posterior probabilities that various values of $\Delta_1$ belong to the identified set. Panel 1(b) does the same for $\beta_1$. Panel 1(c) displays the posterior probabilities that various values of $(\beta_1, \Delta_1)$ belong to the identified set, whereas panel 1(d) displays the true identified set for $(\beta_1, \Delta_1)$, computed by numerical approximation. Panel 1(c) displays a "contour plot" of the posterior, with the legend showing the interpretation of the level curves: points inside the outermost level curve have at least posterior probability 0.1 of being in the identified set, points inside the middle level curve have at least posterior probability 0.6 of being in the identified set, and points inside the innermost level curve have posterior probability approximately 1 of being in the identified set. Unlike the graphs in the first row, the posterior displayed in panel 1(c) corresponds to just one draw from the data generating process, as it would be too cluttered to try to show the results across draws. It is interesting to note from panel 1(d) that the joint identified set for $(\beta_1, \Delta_1)$ lies on a diagonal, that is, large values of $\Delta_1$ are associated with small values of $\beta_1$ and vice versa, and that this is indeed reflected in the posterior over the identified set for this pair of parameters. In all of the panels, the posterior "curve" closely approximates an indicator function for the true identified set, as expected based on the theoretical results. The results corresponding to $(\beta_2, \Delta_2)$ are similar and so are not reported.

The circles along the horizontal axis in panels 1(a) and 1(b) are the endpoints of the 95% credible sets for the identified sets, for each draw from the data generating process, and the corresponding parameter. The credible set of a given shade corresponds to the same draw of $X$ as the posterior "curve" displayed in the same shade. In approximately 92.8% of the draws from the data generating process, the 95% credible set for the identified set for $\beta_1$ indeed does contain the true identified set for $\beta_1$, and in approximately 92.0% of the draws from the data generating process, the 95% credible set for the identified set for $\Delta_1$ indeed does contain the true identified set for $\Delta_1$, with similar results for $\beta_2$ and $\Delta_2$, so the credible sets are also valid frequentist confidence sets. As also discussed above, since these credible sets/confidence sets concern functions of the partially identified parameter, other frequentist approaches might require conservative projection methods. The credible sets throughout this paper are computed as described in Remark 5. In particular, the identified set is "estimated" by computing the identified set using the slice sampling routine, evaluating the criterion function at the sample choice probabilities rather than a draw from the posterior $\mu|X$, and then expanded outward until it achieves the required Bayesian credibility level.

## 8. Empirical illustration: Estimating a binary entry game

This section reports the results of applying this approach to inference to a real data application. The model is a binary entry game (similar to that used in Section 7.1), applied to data from airline markets. The data come from the second quarter of the 2010 Airline Origin and Destination Survey (DB1B). The data contain 7882 markets, which are formally defined as trips between two airports irrespective of intermediate stops. The

empirical question concerns the entry behavior of two kinds of firms: LCC (low cost carriers)[26] and OA (other airlines). A firm that is not an LCC is by definition an OA. Essentially the question is, "What explains the decision of these firms to enter each market?" or, equivalently, "What explains the decision of an airline to provide service between two airports?". The unconditional choice probabilities are $(0.16, 0.61, 0.07, 0.15)$, which are respectively the probabilities that both OA and LCC serve the market, that OA and not LCC serve the market, that LCC and not OA serve the market, and finally that neither serves the market.

The model is essentially the same as that in Section 7.1 except that explanatory variables are introduced to the utility functions. For the purposes of mapping the data to a binary entry game, the airlines are aggregated into two firms: LCC and OA. So, firm LCC (resp. OA) enters the market if any low cost carrier (resp. other airline) serves that market. The payoff to firm $i$ from entering market $m$ is

$$\beta_i^{\mathrm{cons}} + \beta_i^x x_{im} + \Delta_i y_{3-i} + \varepsilon_{im},$$

which essentially results in the payoff matrix in Section 7.1 except that $\beta_i^{\mathrm{cons}} + \beta_i^x x_{im}$ replaces $\beta_i$. This implies that the "nonstrategic" terms (that part of utility that does not depend on the action of the opponent) varies across firms and markets. The variables $y_{im}$ indicate whether firm $i$ enters market $m$. As in Section 7.1, the unobservables are assumed to be normally distributed with variance 1 and unknown correlation.

The analysis considers two explanatory variables: *market presence* and *market size*. The first explanatory variable is market presence, which is a market- and airline-specific variable: for each airline and for each airport, compute the number of markets that airline serves from that airport and divide by the total number of markets served from that airport by any airline. The market presence variable for a given market and airline is the average of these ratios (excluding the one market under consideration) at the two endpoints of the trip, providing some proxy for an airline's presence in the airports associated with that market. See also Berry (1992). This variable is important because it is an *excluded* regressor: the market presence for firm $i$ enters only firm $i$'s payoffs. Since the airlines are aggregated into two firms (LCC and OA), the market presence variable must also be aggregated: the market presence for the LCC firm (resp. OA firm) is the maximum among the actual airlines in the LCC category (resp. OA category). The second explanatory variable is market size, which is a market-specific variable (but shared by all airlines in that market), which is defined as the population at the endpoints of the trip. The market size and market presence variables actually used in the empirical application are discretized binary variables based on the continuous variables just described. They take the value of 1 if the variable is higher than its median value and 0 otherwise.

The point identified parameter $\mu$ is a vector of choice probabilities conditional on the explanatory variables, and the partially identified parameter $\theta$ is the vector that characterizes the payoff functions and the correlation in the unobservables, as in Section 7.1. The link between $\mu$ and $\theta$ uses the assumptions that players are playing a pure strategy

---

[26]The low cost carriers are AirTran, Allegiant Air, Frontier, JetBlue, Midwest Air, Southwest, Spirit, Sun Country, USA3000, and Virgin America.

Nash equilibrium, and that the $\Delta$ parameters are both negative. However, the approach can handle a weakening of either of these assumptions.

The link between $\mu$ and $\theta$ is based on moment equalities that match the model-predicted probabilities of the outcomes (conditional on the explanatory variables) to the observed probabilities, similar to those used in Section 7.1.[27] The criterion function is the "sum" of the criterion functions in Section 7.1 across the types of market defined by the explanatory variables (the "nonstrategic" term varies across different types of markets). The computation otherwise parallels that in Section 7.1.

The model specification has two binary explanatory variables: *market presence* and *market size*. The payoff of firm LCC if it enters market $m$ is

$$\beta_{\text{LCC}}^{\text{cons}} + \beta_{\text{LCC}}^{\text{size}} X_{m,\text{size}} + \beta_{\text{LCC}}^{\text{pres}} X_{\text{LCC}m,\text{pres}} + \Delta_{\text{LCC}} y_{\text{OA}m} + \varepsilon_{\text{LCC}m};$$

similarly, the payoff of firm OA if it enters market $m$ is

$$\beta_{\text{OA}}^{\text{cons}} + \beta_{\text{OA}}^{\text{size}} X_{m,\text{size}} + \beta_{\text{OA}}^{\text{pres}} X_{\text{OA}m,\text{pres}} + \Delta_{\text{OA}} y_{\text{LCC}m} + \varepsilon_{\text{OA}m}.$$

The variable $X_{im,\text{pres}}$ is a binary firm- and market-specific variable that is equal to 1 if market presence for firm $i$ in market $m$ is larger than the median market presence for firm $i$. The variable $X_{m,\text{size}}$ is a binary market-specific variable that is equal to 1 if market size for market $m$ is larger than the median market size. In this specification, $\mu$ is a 32-dimensional vector of conditional choice probabilities (because there are three binary explanatory variables per market resulting in eight types of markets and each type of market is characterized by four choice probabilities). The partially identified parameter $\theta$ is 9-dimensional. The equilibrium selection function (which is a function of the explanatory variables) is profiled out for a given $\theta$, as in Section 7.1.

Figure 2 reports the posterior probabilities that various parameter values belong to the identified set. The posterior probabilities over the identified sets for the $\Delta$ parameters and the $\beta^{\text{size}}$ parameters seem similar across the two types of firms. The effect of market presence seems to be greater for LCC firms compared to OA firms, since it seems that the identified set for the LCC firms is disjoint from and greater than the identified set for the OA firms. The monopoly profits associated with a market with below-median size and below-median market presence (i.e., the constant terms) seem to be smaller for LCC firms compared to OA firms. And the "curve" of posterior probabilities associated with $\rho$ is basically flat and equal to 1 for values of $\rho$ greater than approximately 0.7, implying that any sufficiently high correlation almost certainly could have generated the data. The circles along the horizontal axes in Figure 2 are the endpoints of the 95% credible sets for the identified set for the corresponding parameter.

## 9. Conclusions

This paper has developed a Bayesian[28] approach to inference in partially identified models. The approach results in posterior probability statements concerning the iden-

---

[27]An alternative is moment inequalities similar to those used in Ciliberto and Tamer (2009), but with only two firms, the approach is to use moment equalities that "profile out" the selection probabilities.

[28]There is some disagreement in the overall statistical literature concerning the appropriate meaning of "Bayesian"; for example, Good (1971) has identified the existence of 46,656 varieties of Bayesians. Since

(a) $\Delta$

(b) $\beta^{\text{size}}$

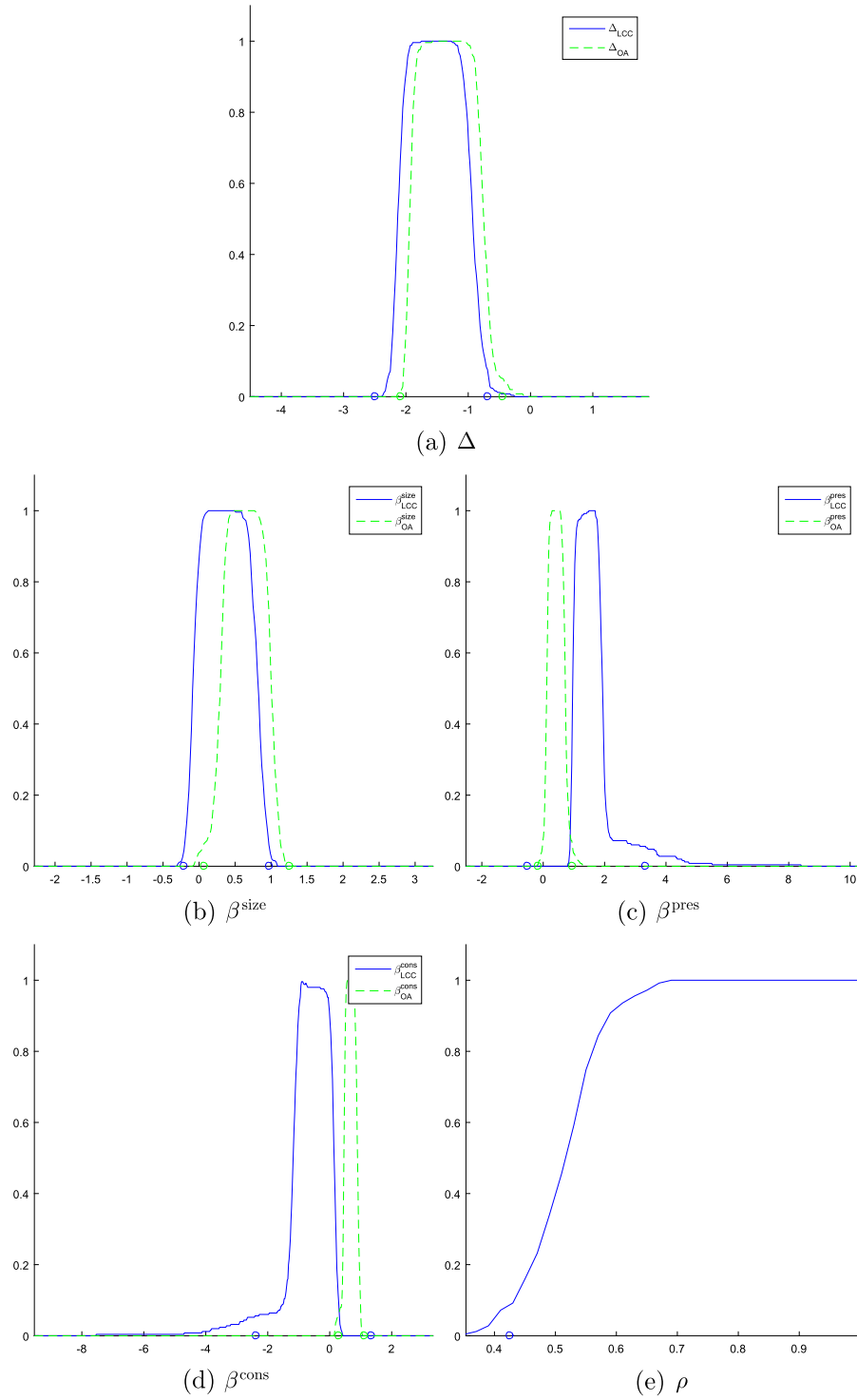(c) $\beta^{\text{pres}}$

(d) $\beta^{\text{cons}}$

(e) $\rho$

FIGURE 2. Posterior probabilities that various parameter values belong to the identified set in the model with market presence and market size.

tified set, which is the quantity about which the data are informative, without the specification of a prior for the partially identified parameter. The resulting posterior probability statements have intuitive interpretations and answer empirically relevant questions, are revised by the data, require no asymptotic repeating sampling approximations, can accommodate inference on functions of the partially identified parameters, and are computationally attractive even in high-dimensional models. Also, this paper establishes conditions under which the credible sets for the identified set also are valid frequentist confidence sets for the identified set, providing an "asymptotic equivalence" between Bayesian and frequentist inference in partially identified models. The approach works well in Monte Carlo experiments and in an empirical illustration.

This paper has restricted attention to *finite-dimensional* models (i.e., $\mu$ and $\theta$ are in finite-dimensional Euclidean spaces), consistent with much of the literature on partially identified models. However, nothing about the approach in this paper fundamentally relies on the fact that the parameters are finite-dimensional. A formal extension to models with infinite-dimensional parameters would involve recent work in Bayesian statistics. Just to give one recent example, Castillo and Nickl (2013) prove a nonparametric version of the Bernstein–von Mises theorem that could replace Assumption 3.

## Appendix: Proofs

Lemma 2. *The event $\Delta^* \subseteq \Delta_I(\mu)$ is equivalent to the event $\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. The event $\Delta_I(\mu) \subseteq \Delta^*$ is equivalent to the event $\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, which is equivalent to the event $\mu \in \bigcap_{\delta \in (\Delta^*)^C \cap \Delta(\Theta)} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$. The event $\Delta_I(\mu) \cap \Delta^* \neq \emptyset$ is equivalent to the event $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$, which is equivalent to the event $\mu \in \bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$.*

Proof. The relation $\Delta^* \subseteq \Delta(\Theta_I(\mu))$ is equivalent to $\delta \in \Delta(\Theta_I(\mu))$ for all $\delta \in \Delta^*$. The relation $\delta \in \Delta(\Theta_I(\mu))$ is equivalent to the existence of $\theta \in \Theta_I(\mu)$ such that $\delta = \Delta(\theta)$, which in turn is equivalent to $\mu \in \mu_I(\theta)$ for some $\theta$ such that $\delta = \Delta(\theta)$, and that is equivalent to $\mu \in \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$.

The relation $\Delta(\Theta_I(\mu)) \subseteq \Delta^*$ is equivalent to $\delta \notin \Delta(\Theta_I(\mu))$ for all $\delta \in (\Delta^*)^C$. $\delta \notin \Delta(\Theta_I(\mu))$ is equivalent to the nonexistence of $\theta \in \Theta_I(\mu)$ such that $\delta = \Delta(\theta)$, which in turn is equivalent to $\mu \in \mu_I(\theta)^C$ for all $\theta$ such that $\delta = \Delta(\theta)$, and that is equivalent to $\mu \in \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$.

It is immediate that if $\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, then $\mu \in \bigcap_{\delta \in (\Delta^*)^C \cap \Delta(\Theta)} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$. Suppose that $\mu \in \bigcap_{\delta \in (\Delta^*)^C \cap \Delta(\Theta)} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, and let $\delta^* \in (\Delta^*)^C$ and $\theta^*$ such that $\delta^* = \Delta(\theta^*)$ be given. Then it must be that $\delta^* \in \Delta(\Theta)$. Therefore, if $\mu \in \bigcap_{\delta \in (\Delta^*)^C \cap \Delta(\Theta)} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, then $\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$.

---

the approach to inference in this paper does not result in a conventional posterior over the parameters, this approach does not satisfy the requirements of all varieties of Bayesianism. However, it does satisfy the following definition: "It seems to me [I. J. Good, in Good (1965)] that the essential defining property of a Bayesian is that he regards it as meaningful to talk about the probability $P(H|E)$ of a hypothesis $H$, given evidence $E$." The approach to inference in this paper talks about hypotheses concerning the identified set.

The relation $\Delta(\Theta_I(\mu)) \cap \Delta^* \neq \emptyset$ is equivalent to the existence of some $\delta \in \Delta^*$ such that $\delta \in \Delta(\Theta_I(\mu))$. The relation $\delta \in \Delta(\Theta_I(\mu))$ is equivalent to $\mu \in \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ from above, so $\Delta(\Theta_I(\mu)) \cap \Delta^* \neq \emptyset$ is equivalent to $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$.

It is immediate that if $\mu \in \bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$, then $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. Suppose that $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. Then it must be that there is $\delta^* \in \Delta^*$ and $\theta^*$ such that $\delta^* = \Delta(\theta^*)$ and $\mu \in \mu_I(\theta^*)$. Therefore, it must be that $\delta^* \in \Delta(\Theta)$, and, therefore, if $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$, then $\mu \in \bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. $\qquad\square$

PROOF OF THEOREMS 1 AND 3. For Theorem 1(i), since $\mu_0 \in \mathrm{int}(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, there is an open neighborhood $U$ of $\mu_0$ such that $U \subseteq \mathrm{int}(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. Therefore, since $\mu|X$ is consistent by Assumption 2, $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For Theorem 1(ii), let $\delta^* \in \Delta^*$ be such that $\mu_0 \in \mathrm{int}((\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))^C)$. Then it follows $\Pi(\Delta^* \subseteq \Delta_I|X) \leq \Pi(\delta^* \in \Delta_I|X) = \Pi(\mu \in \bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)|X) = 1 - \Pi(\mu \in (\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))^C|X)$. Since $\mu_0 \in \mathrm{int}((\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))^C)$, there is an open neighborhood $U$ of $\mu_0$ such that $U \subseteq \mathrm{int}((\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))^C)$. Therefore, since $\mu|X$ is consistent by Assumption 2, $\Pi(\mu \in (\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))^C|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For Theorem 3(i), note that $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X)$. Therefore, by the same arguments as in the proof of Theorem 1(i), but applied to $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, the result follows.

Similarly, for Theorem 3(ii), let $\delta^* \in (\Delta^*)^C$ be such that $\mu_0 \in \mathrm{int}(\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))$. Then $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X) \leq \Pi(\mu \in \bigcap_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)^C|X)$. Then, by the same arguments as in the proof of Theorem 1(ii), but applied to $\bigcap_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)^C$, the result follows.

For Theorem 3(iv), since $\mu_0 \in \mathrm{int}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, there is an open neighborhood $U$ of $\mu_0$ such that $U \subseteq \mathrm{int}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. Therefore, since $\mu|X$ is consistent by Assumption 2, $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \equiv \Pi(\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For Theorem 3(v), since $\mu_0 \in \mathrm{ext}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, there is an open neighborhood $U$ of $\mu_0$ such that $U \subseteq \mathrm{int}((\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))^C)$. Therefore, since $\mu|X$ is consistent by Assumption 2, it follows that $\Pi(\Delta_I \cap \Delta^* = \emptyset|X) = \Pi(\mu \in (\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))^C|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For Theorem 1(iii), again $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$, so

$$\left| \Pi\big(\Delta^* \subseteq \Delta_I|X\big) - P_{N(0,\Sigma_0)}\Big(\sqrt{n}\Big(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\Big)\Big) \right|$$

$$= \left| \Pi\Big(\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X\Big) \right.$$

$$\left. - P_{N(0,\Sigma_0)}\Big(\sqrt{n}\Big(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\Big)\Big) \right|$$

$$= \left| \Pi\left( \sqrt{n}(\mu - \mu_n(X)) \in \sqrt{n}\left( \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X) \right) \Big| X \right) \right.$$

$$\left. - P_{N(0,\Sigma_0)}\left( \sqrt{n}\left( \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X) \right) \right) \right|$$

$$\to 0.$$

The second equality follows from the fact that $\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is equivalent to $\sqrt{n}(\mu - \mu_n(X)) \in \sqrt{n}(\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X))$. The claimed limit holds along almost all sample sequences, by Assumption 3.

The proof of Theorem 3(iii) is similar, except applied to the posterior $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X)$. The proof of Theorem 3(vi) is similar, except applied to the posterior $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \equiv \Pi(\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$.  □

PROOF OF COROLLARIES 2 AND 4. For Corollary 2(i), the event $\Delta^* \subseteq \Delta_I(\mu)$, which is equivalent to the event that $\mu \in \bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ by Lemma 2, is a measurable event by Assumption 4, since it is the intersection of closed sets. Let the set of finitely many extreme points of $\Delta^*$ be $S$. Also, let the neighborhood of $\mu_0$ where $\Delta_I(\mu) \cap \Delta^*$ is convex be $U$. Then $\Pi(\Delta^* \subseteq \Delta_I|X) = \Pi(\Delta^* \subseteq \Delta_I, \mu \in U|X) + \Pi(\Delta^* \subseteq \Delta_I, \mu \in U^C|X) \geq \Pi(\Delta^* \subseteq \Delta_I, \mu \in U|X)$.

Suppose that for $\mu \in U$, $S \subseteq \Delta_I(\mu) \cap \Delta^*$, which is implied by $S \subseteq \Delta_I(\mu)$. Then since $\Delta_I(\mu) \cap \Delta^*$ is convex, $\Delta^* = \text{co}(S) \subseteq \Delta_I(\mu) \cap \Delta^* \subseteq \Delta_I(\mu)$. Consequently, $\Pi(\Delta^* \subseteq \Delta_I, \mu \in U|X) \geq \Pi(S \subseteq \Delta_I, \mu \in U|X)$.

Since $\Delta^* \subseteq \text{int}(\Delta_I)$, in particular $S \subseteq \text{int}(\Delta_I)$. Therefore, for each $\delta \in S$, by Assumption 4, $\mu_0 \in \text{int}(\bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. Therefore, $\mu_0 \in \bigcap_{\delta \in S}(\text{int}(\bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)))$. Since $S$ is finite, equivalently $\mu_0 \in \text{int}(\bigcap_{\delta \in S} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. Then, by the same arguments as in the proof of Theorem 1(i), $\Pi(S \subseteq \Delta_I, \mu \in U|X) \to 1$ along almost all sample sequences, which establishes the claim.

For Corollary 2(ii), since $\Delta^* \not\subseteq \Delta_I$, there is $\delta^*$ such that $\delta^* \in \Delta^*$ and $\delta^* \notin \Delta_I$. In particular, therefore $\mu_0 \notin \bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)$, which is equivalent to $\mu_0 \in (\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))^C$, which is an open set by Assumption 4. Therefore, Theorem 1(ii) applies, which establishes the claim.

For Corollary 4(i), let $\tilde{\Delta}^* = \text{int}(\Delta^*)$ and note that since $\Delta_I \subseteq \tilde{\Delta}^* \subseteq \Delta^*$, it follows that $\Pi(\Delta_I \subseteq \Delta^*|X) \geq \Pi(\Delta_I \subseteq \tilde{\Delta}^*|X)$. The event that $\Delta_I(\mu) \subseteq \Delta^*$ is measurable by assumption. The event that $\Delta_I(\mu) \subseteq \tilde{\Delta}^*$ is measurable, since by Assumption 4, $\bigcap_{\delta \in \tilde{\Delta}^{*C}} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is open. Since $\Delta_I \subseteq \tilde{\Delta}^*$, by Lemma 2, $\mu_0 \in \bigcap_{\delta \in \tilde{\Delta}^{*C}} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, and therefore there is an open neighborhood $U$ of $\mu_0$ such that $U \subseteq \bigcap_{\delta \in \tilde{\Delta}^{*C}} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$. Therefore, part (i) of Theorem 3 applies, so $\Pi(\Delta_I \subseteq \tilde{\Delta}^*|X) \to 1$, which establishes the claim.

For Corollary 4(ii), let $\delta^* \in (\Delta^*)^C \cap \text{int}(\Delta_I)$. Then by Assumption 4, $\mu_0 \in \text{int}(\bigcup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))$. Then part (ii) of Theorem 3 establishes the claim.

For Corollary 4(iii), $\Delta_I(\mu) \cap \Delta^* \neq \emptyset$ for all $\mu$ in an open neighborhood of $\mu_0$ is equivalent, by Lemma 2, to the statement that all such $\mu$ satisfy $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$,

which implies that $\mu_0 \in \text{int}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, so part (iv) of Theorem 3 establishes the claim.

For Corollary 4(iv), $\Delta_I(\mu) \cap \Delta^* = \emptyset$ for all $\mu$ in an open neighborhood of $\mu_0$ is equivalent, by Lemma 2, to the statement that all such $\mu$ satisfy $\mu \in (\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))^C$, which implies that $\mu_0 \in \text{ext}(\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, so part (v) of Theorem 3 establishes the claim.                                                                                    $\square$

PROOF OF THEOREM 5.  Note that, per Lemma 2, $\Delta_I(\mu) \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ is logically equivalent to $\mu \in \bigcap_{\delta \in (\mathcal{C}_{1-\alpha}^{\Delta_I}(X))^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C \equiv \Delta_{1-\alpha}^{-1,\Delta_I}(X)$.

By Assumption 3, for any given $\varepsilon > 0$, there is a set of sample sequences for the data $X$ with probability at least $1 - \varepsilon$ under the true data generating process and a minimal sample size $N_\varepsilon$ such that, for any sample size $n \geq N_\varepsilon$ (and for all such sample sequences resulting in an $X$), $\|\Pi(\sqrt{n}(\mu - \mu_n(X)) \in \cdot | X) - P_{N(0,\Sigma_0)}(\cdot)\|_{\text{TV}} < \varepsilon$.

Applying this to $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X) \equiv \sqrt{n}(\Delta_{1-\alpha}^{-1,\Delta_I}(X) - \mu_n(X))$, it follows that $P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)) \in \Pi(\sqrt{n}(\mu - \mu_n(X)) \in \sqrt{n}(\Delta_{1-\alpha}^{-1,\Delta_I}(X) - \mu_n(X))|X) + [-\varepsilon, \varepsilon]$.

Note that $\Pi(\sqrt{n}(\mu - \mu_n(X)) \in \sqrt{n}(\Delta_{1-\alpha}^{-1,\Delta_I}(X) - \mu_n(X))|X) = \Pi(\mu \in \Delta_{1-\alpha}^{-1,\Delta_I}(X)|X) = 1 - \alpha$, by definition of a credible set for the identified set. That implies $P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)) \in [1 - \alpha - \varepsilon, 1 - \alpha + \varepsilon]$. That implies $P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)) \to^{a.s.} 1 - \alpha$. Finally, that implies $E(P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))) \to 1 - \alpha$.

By Assumption 6, for any given $\varepsilon > 0$, there is a minimal sample size $N'_\varepsilon$ such that for any sample size $n \geq N'_\varepsilon$, $F_n(A) \in P_{N(0,\Sigma_0)}(A) + [-\varepsilon, \varepsilon]$ either for all Borel sets $A$ (in case of part (a)) or all finite unions of disjoint convex sets $A$ (in case of part (b), in which case also $\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)$ is a finite union of disjoint convex sets, after application of Rao (1962, Theorem 4.2) or Bickel and Millar (1992, Example 4.2)). Therefore, $E(F_n(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))) \in E(P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))) + [-\varepsilon, \varepsilon]$. So, because $E(P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))) \to 1 - \alpha$ from above, $E(F_n(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))) \to 1 - \alpha$.

But also $P(\sqrt{n}(\mu_0 - \mu_n(X)) \in \sqrt{n}(\Delta_{1-\alpha}^{-1,\Delta_I}(X) - \mu_n(X))) = P(\mu_0 \in \Delta_{1-\alpha}^{-1,\Delta_I}(X)) = P(\Delta_I \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X))$, since $\mu_0 \in \Delta_{1-\alpha}^{-1,\Delta_I}(X)$ is logically equivalent to $\Delta_I \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)$ by Lemma 2. Therefore, $P(\Delta_I \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)) \to 1 - \alpha$ if and only if

$$\left| P\left(\sqrt{n}(\mu_0 - \mu_n(X)) \in \sqrt{n}\left(\Delta_{1-\alpha}^{-1,\Delta_I}(X) - \mu_n(X)\right)\right) - E\left(F_n\left(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\right)\right) \right| \to 0,$$

which is Assumption 5.                                                                      $\square$

PROOF OF LEMMA 1. In large samples, $\Pi(\Delta_I(\mu) \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X)|X) \approx \Pi(\Delta_{IL}(\mu) \geq \Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu) \leq \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}|X)$, since $\Delta_I(\mu) \neq \emptyset$ with posterior probability approaching 1 in large samples by Assumption 2. Then $\Pi(\Delta_{IL}(\mu) \geq \Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu) \leq \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}|X) = \Pi(\sqrt{n}(\Delta_{IL}(\mu) - \Delta_{IL}(\mu_n(X))) \geq -c_{1-\alpha}(X), \sqrt{n}(\Delta_{IU}(\mu) - \Delta_{IU}(\mu_n(X))) \leq c_{1-\alpha}(X)|X)$. Let $\Delta'_I(\mu) = (\Delta'_{IL}(\mu)\Delta'_{IU}(\mu))$ be the $d_\mu \times 2$ matrix of derivatives of $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ with respect to the elements of $\mu$. By the Bayesian delta method (e.g., Bernardo and Smith (2009, Section 5.3)), the posterior for $(\sqrt{n}(\Delta_{IL}(\mu) - \Delta_{IL}(\mu_n(X))), \sqrt{n}(\Delta_{IU}(\mu) - \Delta_{IU}(\mu_n(X))))$

is approximately $N(0, (\Delta_I'(\mu_0))^T \Sigma_0 \Delta_I'(\mu_0))$ in large samples. Because the covariance is full rank (i.e., $(\Delta_I'(\mu_0))^T \Sigma_0 \Delta_I'(\mu_0)$ is positive definite), $c_{1-\alpha}(X)$ must converge to the unique constant $c_{1-\alpha}$ that solves $P_{N(0,(\Delta_I'(\mu_0))^T \Sigma_0 \Delta_I'(\mu_0))}(\mu_L \geq -c_{1-\alpha}, \mu_U \leq c_{1-\alpha}) = 1 - \alpha$. Therefore, $P(\sqrt{n}(\mu_0 - \mu_n(X)) \in \tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)) = P(\Delta_I(\mu_0) \subseteq \mathcal{C}_{1-\alpha}^{\Delta_1}(X)) = P(\Delta_{IL}(\mu_0) \geq \Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu_0) \leq \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}) = P(\sqrt{n}(\Delta_{IL}(\mu_0) - \Delta_{IL}(\mu_n(X))) \geq -c_{1-\alpha}(X), \sqrt{n}(\Delta_{IU}(\mu_0) - \Delta_{IU}(\mu_n(X))) \leq c_{1-\alpha}(X)) \to 1 - \alpha$, since by the delta method, $(\sqrt{n}(\Delta_{IL}(\mu_0) - \Delta_{IL}(\mu_n(X))), \sqrt{n}(\Delta_{IU}(\mu_0) - \Delta_{IU}(\mu_n(X))))$ is distributed $N(0, (\Delta_I'(\mu_0))^T \Sigma_0 \Delta_I'(\mu_0))$ in repeated large samples. Moreover, $P_{N(0,\Sigma_0)}(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)) \to 1 - \alpha$ by Theorem 3, so (as established in the proof of Theorem 5 without using Assumption 5), also $E(F_n(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))) \to 1 - \alpha$, establishing Assumption 5. $\qquad \square$

## References

Andrews, D. W. and P. J. Barwick (2012), "Inference for parameters defined by moment inequalities: A recommended moment selection procedure." *Econometrica*, 80 (6), 2805–2826. [331]

Andrews, D. W. K. and P. Guggenberger (2009), "Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities." *Econometric Theory*, 25 (3), 669–709. [331]

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78 (1), 119–157. [331]

Beresteanu, A., I. Molchanov, and F. Molinari (2011), "Sharp identification regions in models with convex moment predictions." *Econometrica*, 79 (6), 1785–1821. [331]

Beresteanu, A., I. Molchanov, and F. Molinari (2012), "Partial identification using random set theory." *Journal of Econometrics*, 166 (1), 17–32. [331]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76 (4), 763–814. [331]

Bernardo, J. M. and A. F. Smith (2009), *Bayesian Theory*. John Wiley & Sons, New York. [347, 362]

Berry, S. and E. Tamer (2006), "Identification in models of oligopoly entry." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Vol. II (R. Blundell, W. K. Newey, and T. Persson, eds.), Econometric Society Monographs, Chapter 2, 46–85, Cambridge University Press, Cambridge. [335]

Berry, S. T. (1992), "Estimation of a model of entry in the airline industry." *Econometrica*, 60 (4), 889–917. [356]

Bickel, P. J. and B. J. K. Kleijn (2012), "The semiparametric Bernstein–von Mises theorem." *The Annals of Statistics*, 40 (1), 206–237. [338]

Bickel, P. J. and P. W. Millar (1992), "Uniform convergence of probability measures on classes of functions." *Statistica Sinica*, 2 (1), 1–15. [338, 362]

Billingsley, P. and F. Topsøe (1967), "Uniformity in weak convergence." *Probability Theory and Related Fields*, 7 (1), 1–16. [338]

Bugni, F. A. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78 (2), 735–753. [331]

Canay, I. A. (2010), "EL inference for partially identified models: Large deviations optimality and bootstrap validity." *Journal of Econometrics*, 156 (2), 408–425. [331]

Castillo, I. and R. Nickl (2013), "Nonparametric Bernstein–von Mises theorems in Gaussian white noise." *The Annals of Statistics*, 41 (4), 1999–2028. [359]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75 (5), 1243–1284. [334]

Chernozhukov, V., S. Lee, and A. M. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81 (2), 667–737. [335]

Choudhuri, N. (1998), "Bayesian bootstrap credible sets for multidimensional mean functional." *The Annals of Statistics*, 26 (6), 2104–2127. [338]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77 (6), 1791–1828. [357]

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1 (2), 209–230. [338]

Fiacco, A. V. and G. P. McCormick (1990), *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Classics in Applied Mathematics, Vol. 4. SIAM, Philadelphia, PA. [347]

Freedman, D. (1999), "Wald lecture: On the Bernstein–von Mises theorem with infinite-dimensional parameters." *The Annals of Statistics*, 27 (4), 1119–1141. [331, 344]

Gamerman, D. and H. F. Lopes (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman & Hall/CRC, Boca Raton, FL. [351]

Gasparini, M. (1995), "Exact multivariate Bayesian bootstrap distributions of moments." *The Annals of Statistics*, 23 (3), 762–768. [338]

Giacomini, R. and T. Kitagawa (2014), "Inference about non-identified SVARs." Working paper. [343]

Good, I. J. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press, Cambridge. [359]

Good, I. J. (1971), "46656 varieties of Bayesians." *The American Statistician*, 25 (5), 62–63. [357]

Hwang, C.-R. (1980), "Laplace's method revisited: Weak convergence of probability measures." *The Annals of Probability*, 8 (6), 1177–1182. [351]

Imbens, G. W. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72 (6), 1845–1857. [330]

Kaido, H. and H. White (2014), "A two-stage procedure for partially identified models." *Journal of Econometrics*, 182 (1), 5–13. [331]

Kitagawa, T. (2012), "Estimation and inference for set-identified parameters using posterior lower probability." Working paper. [331, 332, 343, 347]

Kline, B. (2011), "The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean." *Journal of Statistical Planning and Inference*, 141 (9), 3131–3141. [332, 338, 349]

Kline, B. (2015a), "The empirical content of games with bounded regressors." Working paper. [335]

Kline, B. (2015b), "Identification of complete information games." *Journal of Econometrics*, 189 (1), 117–131. [335]

Kline, B. and E. Tamer (2012), "Bounds for best response functions in binary games." *Journal of Econometrics*, 166 (1), 92–105. [335]

Liao, Y. and A. Simoni (2012), "Semi-parametric Bayesian partially identified models based on support function." Working paper. [332]

Lo, A. Y. (1987), "A large sample study of the Bayesian bootstrap." *The Annals of Statistics*, 1 (15), 360–375. [338]

Manski, C. F. (2003), *Partial Identification of Probability Distributions*. Springer, New York. [335]

Manski, C. F. (2007), *Identification for Prediction and Decision*. Harvard University Press, Cambridge, MA. [330]

Moon, H. R. and F. Schorfheide (2009), "Bayesian and frequentist inference in partially identified models." Working paper. [349]

Moon, H. R. and F. Schorfheide (2012), "Bayesian and frequentist inference in partially identified models." *Econometrica*, 80 (2), 755–782. [331, 333, 344]

Neal, R. M. (2003), "Slice sampling." *The Annals of Statistics*, 31 (3), 705–741. [351]

Norets, A. and X. Tang (2014), "Semiparametric inference in dynamic binary choice models." *Review of Economic Studies*, 81 (3), 1229–1262. [332]

Poirier, D. J. (1998), "Revising beliefs in nonidentified models." *Econometric Theory*, 14 (4), 483–509. [331]

Rao, R. R. (1962), "Relations between weak and uniform convergence of measures with applications." *The Annals of Mathematical Statistics*, 33 (2), 659–680. [338, 362]

Rockafellar, R. T. and R. J.-B. Wets (2009), *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Vol. 317. Springer, Dordrecht. [339]

Romano, J. P. and A. M. Shaikh (2008), "Inference for identifiable parameters in partially identified econometric models." *Journal of Statistical Planning and Inference*, 138 (9), 2786–2807. [334]

Romano, J. P. and A. M. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78 (1), 169–211. [334]

Rosen, A. M. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146 (1), 107–117. [330]

Rubin, D. B. (1981), "The Bayesian bootstrap." *The Annals of Statistics*, 9 (1), 130–134. [338]

Shen, X. (2002), "Asymptotic normality of semiparametric and nonparametric posterior distributions." *Journal of the American Statistical Association*, 97 (457), 222–235. [338]

Shi, X. and M. Shum (2015), "Simple two-stage inference for a class of partially identified models." *Econometric Theory*, 31 (3), 493–520. [331]

Stoye, J. (2009), "More on confidence regions for partially identified parameters." *Econometrica*, 77 (4), 1299–1315. [331]

Tamer, E. (2003), "Incomplete simultaneous discrete response model with multiple equilibria." *Review of Economic Studies*, 70 (1), 147–165. [335]

Tamer, E. (2010), "Partial identification in econometrics." *Annual Review of Economics*, 2 (1), 167–195. [330]

Van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambrige. [338, 346]

Woutersen, T. and J. C. Ham (2014), "Confidence sets for continuous and discontinuous functions of parameters." Working paper. [331]

---