BAYESIAN INFERENCE IN A CLASS OF PARTIALLY IDENTIFIED MODELS

BRENDAN KLINE AND ELIE TAMER
UNIVERSITY OF TEXAS AT AUSTIN AND HARVARD UNIVERSITY

Abstract. This paper develops a Bayesian approach to inference in an empirically relevant class of partially identified econometric models. Models in this class are characterized by a known mapping between a point identified reduced-form parameter μ, and the identified set for a partially identified parameter θ. The approach maps posterior inference about μ to various posterior inference statements concerning the identified set for θ, without the specification of a prior for θ. 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 θ. 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 credible sets also are valid frequentist confidence sets. The approach is computationally attractive even in high-dimensional models: the approach tends to avoid an exhaustive search over the parameter space. The approach is illustrated via Monte Carlo experiments and an empirical application to a binary entry game involving airlines.

JEL codes: C10, C11

Keywords: partial identification, identified set, criterion function, Bayesian inference

1. Introduction

This paper considers the problem of Bayesian inference in an empirically relevant class of partially identified models. These models are characterized by a known mapping between a point identified reduced-form parameter μ, and the identified set for a partially identified parameter θ. The identified set for θ exhausts the information concerning θ contained in the data. Often, μ can be viewed as directly observable characteristics of the data and θ can be viewed as the parameter of an econometric model. The parameter of interest is either θ, or some function of θ. For example, if θ is a parameter of an econometric model and μ are statistics concerning the data, then the identified set mapping is the set of θ^* such that the econometric model evaluated at θ^* generates μ.

First Version: November 2012. This Version: February 11, 2015. We thank G. Chamberlain, R. Moon, W. Newey, and A. de Paula for useful discussion, and participants at the 2013 Asian Meeting of the Econometric Society, the CRETE 2013, the NSF CEME at Stanford, OSU, and the Texas Econometrics Camp 2013 for comments. We also thank the co-editor and three anonymous referees for helpful comments and suggestions. Thanks also to S. Sinha for excellent research assistance. Any errors are ours. Email: brendan.kline@austin.utexas.edu and elietamer@fas.harvard.edu.

Since μ 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. 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.

Then, given a posterior $\mu|X$ and the mapping from μ to the identified set for θ, it is possible to construct various posterior probabilities concerning the identified set for θ, without specifying a prior for θ. 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 non-empty, 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 θ characterizes the utility functions of the decision makers and μ summarizes the observed behavior of the decision makers. Particularly in the case of models involving multiple decision makers, often θ is only partially identified, in which case it is not possible to uniquely recover the utility functions from the data. The identified set for θ 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: Is 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)? Is 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 θ evaluated at a particular value of μ.

The identified set for θ is the target of inference in this paper because, by construction, the identified set for θ exhausts the information in the data concerning θ. This implies that it is not possible to distinguish between values of θ within the identified set, on the basis of the data, without additional prior information. 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), Beresteanu et al. (2011) and Beresteanu et al. (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 θ by introducing a prior over θ. Such prior information would influence "conventional" posterior inference statements concerning θ 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 credible set for a partially identified parameter will tend to be "contained in" the identified set,

[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 et al. (2012). In contrast, the "random set" approach in this paper arises due to the mapping between the uncertainty concerning μ and uncertainty concerning the identified set. Kaido and White (2014) and Shi and Shum (2012) have addressed certain questions about improving frequentist inference in similar model frameworks.

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 "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 re-parametrization 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.[4] See section 3 and particularly remark 2

[2] Woutersen and Ham (2014) study another non-standard 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).

[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

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.

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 credible set for the identified set is also an $(1-\alpha)$-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).

1.1. **Outline.** 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, an online supplement contains additional material.[5]

2. Model and examples

This section sets up the class of models, and provides examples.

The model is characterized by a point identified reduced-form finite-dimensional parameter μ, a partially identified finite-dimensional parameter θ, and a known mapping between μ and the identified set for θ. Often, μ can be viewed as statistics concerning the observable data (e.g., moments) and θ can be viewed as the parameter of an econometric model. The parameter space for μ is M and the parameter space for θ is Θ. The

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).

[5]Section B.1 provides further examples of the model framework, section B.2 provides results on measurability, section B.3 discusses inference under misspecification, section B.4 provides further Monte Carlo experiments, and section B.5 provides further results in the context of the empirical application.

parameter space M is a subspace of $\mathbb{R}^{d_\mu}$, endowed with the subspace topology, where d_μ is the dimension of μ. The parameter space Θ is a subset of $\mathbb{R}^{d_\theta}$, where d_θ is the dimension of θ. The unknown true value of μ is μ_0.

The defining property of this class of models is the existence of a known mapping from μ to the identified set for θ. For example, this mapping might give the set of parameter values θ^* such that the econometric model evaluated at θ^* generates μ. This mapping often arises as an obvious implication of the specification of the econometric model. Examples are provided below. The mapping can equivalently be expressed as a level set of a known criterion function of θ and μ, or as a known set-valued mapping of μ. In either case, this mapping gives the set of θ consistent with μ, and thus the identified set for θ.

Under the criterion function approach, there is a function $Q(\theta, \mu) \geq 0$ that summarizes the relationship between μ and the identified set for θ. The criterion function is a function of the point identified parameter (which essentially substitutes for the data) and the partially identified parameter, a distinction from the prior literature (e.g., Chernozhukov et al. (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 θ 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 θ that would arise at any parameter value μ^* is

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

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

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 μ consistent with θ being in the identified set evaluated at μ. 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 Θ. Suppose that δ is the partially identified parameter of interest, defined by $\delta = \Delta(\theta)$. For example, if $\Delta(\theta) = \theta_1$, then the first component of θ is the parameter of interest, resulting in subvector inference. Alternatively, if $\Delta(\theta) = \theta$, then the entirety of θ is the parameter of interest. Then, $\Delta(\Theta)$ is the induced parameter space for δ, $\Delta_I \equiv \Delta(\Theta_I)$ is the induced true identified set for δ, and $\Delta_I(\mu) \equiv \Delta(\Theta_I(\mu))$ is the induced identified set for δ as a mapping of μ. The parameter space $\Delta(\Theta)$ is a subset of $\mathbb{R}^{d_\delta}$, where d_δ is the dimension of δ.

The following gives a few examples of models that fit this framework. The online supplement discusses further examples, including moment inequality models.

Example 1 (Intersection bounds)**.** Suppose that μ is a $d_\mu \times 1$ "regularly estimable"[6] parameter vector, perhaps moments of a distribution. Suppose μ_0 is the true value. Suppose that the identified set for θ 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 μ contribute to the lower and upper bounds for θ. See also, for example, Chernozhukov et al. (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 μ 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 μ 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 θ, and μ 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 μ 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 θ to the distribution of the observables μ, 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 θ 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,b) for further details of various models of this general form, each of which imply a certain form for $f(\cdot)$.

[6] "Regularly estimable" means, essentially, that the Bernstein-von Mises theorem applies to $\mu|X$.

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

3. Posterior probabilities over the identified set

3.1. **Setup.** Since μ is point identified, let $\Pi(\mu|X)$ be a posterior for μ 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 probability statements concerning Δ_I. This section addresses the posterior probability statements concerned with answering questions related to: Could δ^* have generated the data? Could each $\delta^* \in \Delta^*$ have generated the data? Section 4 addresses posterior probability statements concerned with answering other questions. Section 5 addresses credible sets for the identified set. Section 6 addresses the computational implementation.

Definition 1. Based on the posterior for μ, define the following posterior probability statements:

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

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

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

$$\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\Delta^* \subseteq \Delta(\Theta_I(\mu))|X) \equiv \Pi(\mu \in \cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$$

The posterior probability statements on the left correspond to statements concerning the "posterior uncertainty" about Δ_I. These are then expressed in terms of the posterior for μ. The non-trivial identities in definition 1 are justified by lemma 2.

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

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

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 $co(A)$. 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 μ.

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

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

Assumption 2 (Posterior for μ consistent at μ_0)**.** *Along almost all sample sequences, for any open neighborhood U of μ_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 μ has support on a neighborhood of μ_0 (e.g., the prior for μ has support on the entire parameter space).

Assumption 3 (Large sample normal posterior for μ)**.** *There is a function of the data $\mu_n(X)$ and a convariance matrix Σ_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 Σ_0 to be the inverse Fisher information matrix.[9] This assumption can also hold, for example, for the Bayesian bootstrap for non-parametric estimation of moments of an unknown distribution under a suitably flat Dirichlet process prior, taking $\mu_n(X)$ to be the sample average and Σ_0 to be the covariance of the unknown distribution (e.g., Ferguson (1973), Rubin (1981), Lo (1987), Gasparini (1995), and Choudhuri (1998)).

[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, i.e., $\mathcal{B}(M) = \{A \cap M : A \in \mathcal{B}(\mathbb{R}^{d_\mu})\}$. Note in particular that if $M \in \mathcal{B}(\mathbb{R}^{d_\mu})$, then $\mathcal{B}(M) = \{A \in \mathcal{B}(\mathbb{R}^{d_\mu}) : A \subseteq M\} \subseteq \mathcal{B}(\mathbb{R}^{d_\mu})$.

[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 1.3 of theorem 1, or parts 3.3 and 3.6 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.

See also for example Kline (2011) for a connection to a different (more limited) way of pointwise testing of moment inequality conditions from a Bayesian perspective.

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 Δ^* such that $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$ does not exist because it corresponds to a non-measurable event. Consequently, $\mathcal{M}_1$ is introduced as the subsets such that for $\Delta^* \in \mathcal{M}_1$, $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$ 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 Δ^*. Lemma 3 in the online supplement shows that if the criterion function is continuous, $\Delta(\cdot)$ is continuous, and Θ 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).

Theorem 1. *Under assumptions 1 and 2, for any Δ^* 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:*

(1.1) if $\mu_0 \in int\left(\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$, then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 1$.

(1.2) if $\mu_0 \in \cup_{\delta \in \Delta^*:\{\delta\} \in \mathcal{M}_1} \left(ext\left(\cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)\right)$, then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 0$.

Under the additional assumption 3:

(1.3) $\left|\Pi(\Delta^* \subseteq \Delta_I|X) - P_{N(0,\Sigma_0)}\left(\sqrt{n}\left(\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\right)\right)\right| \to 0$.

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 $\delta \in int(\Delta_I)$ then $\mu_0 \in int(\cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. For all $\delta \in \mathbb{R}^{d_\delta}$, $\cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed. For any open $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is open.*

The first part of this assumption requires that if δ is in the interior of Δ_I, then there is a neighborhood of μ_0 such that δ is also in the identified sets $\Delta_I(\mu)$ for all μ in that neighborhood. The second part of this assumption requires that the set of μ such that $\delta \in \Delta_I(\mu)$ is closed. The third part of this assumption requires that the set of μ such that $\Delta_I(\mu) \subseteq \Delta^*$ for open Δ^* is open.

Lemma 3 in the online 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

[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

by convexity of $\Delta_I(\mu)$ for all μ and inner semicontinuity of $\Delta_I(\mu)$ at μ_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:*

(2.1) if $\Delta^* \subseteq int(\Delta_I)$ and Δ^* is a convex polytope such that $\Delta_I(\mu) \cap \Delta^*$ is convex for all μ in a neighborhood of μ_0, then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 1$.

(2.2) 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 2.1 shows that if $\Delta^* \subseteq int(\Delta_I)$ and Δ^* is not too "complex" then $\Pi(\Delta^* \subseteq \Delta_I|X) \to 1$. Part 2.1 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 Δ^* 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 μ_0, because convexity of $\Delta_I(\mu) \cap \Delta^*$ is a weaker condition than convexity of $\Delta_I(\mu)$.

Part 2.2 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 μ 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 revises 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.

The following illustrates the results in the context of an interval identified parameter.

Example 3 (Posterior probabilities for the simple interval identified parameter)**.** Suppose θ is a simple interval identified parameter, as in example 1, so $\Theta_I(\mu) = [\mu_L, \mu_U]$, where $\mu = (\mu_L, \mu_U)$ are often moments of a distribution. In this example, the function

function $Q(\theta, \mu_0)$ equals zero for all θ 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 θ 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 θ and all $\mu \neq \mu_0$, suggesting that this assumption is reasonable.

$\Delta(\theta) = \theta$, so $\delta \equiv \theta$. Therefore, $\{\theta : \Delta(\theta) = \delta\} = \delta$, so essentially all expressions involving δ can be "replaced" by θ. Suppose $\Theta^* = [a, b]$ is a finite interval, possibly with $a = b$ so Θ^* is a singleton.

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

Similarly, note that, prior to observing 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, per remark 2, this approach to inference implicitly entails the specification of a prior over the identified set.

The following discusses the implications of theorem 1.

Case 1: Suppose that $[a, b] \subset int(\Theta_I) = (\mu_{0L}, \mu_{0U}) \subset \Theta_I = [\mu_{0L}, \mu_{0U}]$. This implies $\mu_{0L} < a \leq b < \mu_{0U}$. Then, $\mu_0 \in int(\cap_{\theta\in\Theta^*}\mu_I(\theta))$, so by part 1.1 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 $\mu_{0L} \leq \mu_{0U}$ (so that the identified set is non-empty). 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, $\mu_0 \in int(\mu_I(a)^C) = ext(\mu_I(a))$ or $\mu_0 \in int(\mu_I(b)^C) = ext(\mu_I(b))$, respectively, so by part 1.2 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 5 in the online supplement.

4. Further posterior probabilities over the identified set

4.1. **Setup.** The posterior $\Pi(\mu|X)$ also induces posterior probability statements concerning Δ_I that answer questions not already addressed in section 3. This section addresses the posterior 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? Do none of the parameter values in the identified set have some property?

Definition 3. Based on the posterior for μ, define[11] the following posterior probability statements:

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

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

[11]As in section 3, the posterior probability statements on the left correspond to statements concerning the "posterior uncertainty" about Δ_I which are then expressed in terms of the posterior for μ. The non-trivial identities in definition 3 are justified by lemma 2.

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

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

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

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

$\Pi(\Delta_I \subseteq \Delta^* | X)$ 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 Δ^*. This can be used to check whether *all* parameter values that could have generated the data have the property defined by Δ^*. For example, if δ 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 non-negative. If θ 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 θ, in the sense that $\Theta_I(\mu)$ is just a singleton, so $\Pi(\Theta_I \subseteq \Theta^* | X)$ is simply the posterior probability that $\theta \in \Theta^*$.

$\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X)$ 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 non-empty intersection with Δ^*. 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 Δ^*. For example, if δ is a scalar and $\Delta^* = [0, \infty)$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X)$ is the posterior probability that at least one non-negative δ could have generated the data. In particular, taking $\Delta^* = \Delta(\Theta)$,

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

is the posterior probability that the identified set Δ_I is non-empty, which can be interpreted to be a conservative (but implementable) measure of the posterior probability that the model is not misspecified. It is conservative because the fact that the identified is non-empty does not imply that the model is correctly specified. But, if the identified set is empty, then the model must be misspecified.

$\Pi(\Delta_I \cap \Delta^* = \emptyset | X)$ 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 Δ^*. This can be used to check whether *none* of the parameter values that could have generated the data has the property defined by Δ^*. For example, if δ is a scalar and $\Delta^* = [0, \infty)$, then $\Pi(\Delta_I \cap \Delta^* = \emptyset | X)$ is the posterior probability that no non-negative δ could have generated the data.

Remark 3 (Technical consideration: measurability)**.** As in section 3, it is not immediate that posterior probabilities over the identified set exist. $\mathcal{M}_2$ are the subsets such that for $\Delta^* \in \mathcal{M}_2$, $\Pi(\Delta_I \subseteq \Delta^* | X) \equiv \Pi(\mu \in \cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C | X)$ corresponds to a measurable event. $\mathcal{M}_3$ are the subsets such that for $\Delta^* \in \mathcal{M}_3$, $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X) \equiv \Pi(\mu \in \cup_{\delta \in \Delta^*} \cup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) | X)$ corresponds to a measurable event. Lemma 3 in the

online supplement shows that if the criterion function is continuous, $\Delta(\cdot)$ is continuous, and Θ is closed, then $\mathcal{M}_2$ and $\mathcal{M}_3$ contain all the Borel sets.

Theorem 3. *Under assumptions 1 and 2, for any Δ^* 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:*

(3.1) if $\mu_0 \in int\left(\cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C\right)$, then $\Pi(\Delta_I \subseteq \Delta^*|X) \to 1$.

(3.2) if $\mu_0 \in \cup_{\delta \in (\Delta^*)^C:\{\delta\} \in \mathcal{M}_1} \left(ext\left(\cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C\right)\right)$, then $\Pi(\Delta_I \subseteq \Delta^*|X) \to 0$.

Under the additional assumption 3:

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

Under assumptions 1 and 2, for any Δ^* 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:

(3.4) if $\mu_0 \in int\left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \to 1$.

(3.5) if $\mu_0 \in ext\left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \to 0$.

Under the additional assumption 3:

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

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 Δ^* 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:*

(4.1) if $\Delta_I \subseteq int(\Delta^*)$, then $\Pi(\Delta_I \subseteq \Delta^*|X) \to 1$.

(4.2) if $int(\Delta_I) \not\subseteq \Delta^*$, then $\Pi(\Delta_I \subseteq \Delta^*|X) \to 0$.

Under the same assumptions, for any Δ^* 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:

(4.3) If $\Delta_I(\mu) \cap \Delta^* \neq \emptyset$ for all μ in a neighborhood of μ_0, then $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \to 1$.

(4.4) If $\Delta_I(\mu) \cap \Delta^* = \emptyset$ for all μ in a neighborhood of μ_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: there is a likelihood and ϕ 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 η, as a function of ϕ. A class of priors is specified over the partially identified parameter, and bounds are derived for the posterior for η due to specifying a class of priors. Very roughly, ϕ is analogous to μ in this paper, and $H(\phi)$ is analogous to $\Delta_I(\mu)$ in this paper. However, these quantities are distinct in critical ways: ϕ and $H(\phi)$ arise implicitly

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

from the specification of a likelihood, whereas μ and $\Delta_I(\mu)$ are explicitly specified by the econometrician. The difference between an identified set implicitly arising from flat-spots in a likelihood versus an explicit specification of an identified set becomes clear when considering an interval identified parameter or moment inequality model. The difference between ϕ and μ becomes clear when considering a structural econometric model (e.g., the entry game models considered in this paper) where the prior is either placed on the "sufficient parameter" ϕ of the model likelihood or the "summary statistics" μ concerning the data that is generated by the underlying econometric model. These differences result in further differences along other dimensions, for example the computational results in this paper depend on the separation between standard Bayesian inference on μ and computation of the identified set as a known mapping of μ.

Kitagawa (2012) shows that, under appropriate conditions, the smallest posterior probability that can be assigned to a set D of the parameter space for η 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 η is the posterior probability under $\phi|X$ of the event $H(\phi) \cap D \neq \emptyset$. Thus, subject to caveats about differences in model frameworks, the posterior probability statements concerning the identified set in this paper can be interpreted also as bounds on the possible posteriors for the partially identified parameter.

5. Frequentist properties of the credible sets

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

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

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

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 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)), 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 coverage. 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).[13] 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 a valid 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

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

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

The latter quantity, $\tilde{\Delta}^{-1,\Delta_I}_{1-\alpha}(X)$, is the set of μ consistent with $\Delta_I(\mu) \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)$. Theorem 3 implies that $P_{N(0,\Sigma_0)}\left(\tilde{\Delta}^{-1,\Delta_I}_{1-\alpha}(X)\right) \approx 1-\alpha$ for each large dataset, since $\mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ 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)**.**

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

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 $\tilde{\Delta}^{-1,\Delta_I}_{1-\alpha}(X)$ 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:

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

[13] 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.

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.** To get the main result in this section, we also require regularity conditions about the repeated sampling behavior of the estimator $\mu_n(X)$.

Assumption 6 (Repeated sampling behavior of the estimator of μ)**.** *The estimator $\mu_n(X)$ appearing in assumption 3 has the property that $\sqrt{n}(\mu_0 - \mu_n(X))$ converges in total variation to $N(0, \Sigma_0)$.*

This is essentially the "frequentist" version of assumption 3. The fact that the asymptotic covariances in assumption 3 and 6 are the same is part of the conclusion of the various "Bernstein-von Mises"-like theorems.[14]

The following theorem establishes the frequentist coverage properties of $\mathcal{C}_{1-\alpha}^{\Delta_I}(X)$.

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*

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

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

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

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 "non-standard 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 δ 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 δ is an interval in many important cases, including the case where the identified set for θ is convex, and δ is a scalar element of θ. 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 have the credible set property. Suppose that for all μ in a neighborhood of μ_0, $\Delta_I(\mu) \neq \emptyset$. This essentially requires that the identified set be non-empty in a sufficiently

[14] As with convergence in total variation assumed for the large sample approximation to the posterior $\mu|X$, it is possible to relax the assumption of convergence in total variation if the relevant collection of sets is sufficiently small. See the proof of theorem 5 for the relevant considerations.

small neighborhood around the true μ.[15] Then supposing that $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ satisfy the regularity conditions of the (Bayesian) delta method, in a neighborhood of μ_0, with positive definite covariance, 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 μ from the delta method rules out kinks in $\Delta_{IL}(\cdot)$ and $\Delta_{IU}(\cdot)$ at μ_0, for example intersection bounds with multiple simultaneously binding constraints at μ_0.[16]

The credible set $\mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ can be computed by first 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 $\mu|X$.

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 μ in a neighborhood of μ_0, $\Delta_I(\mu) \neq \emptyset$. Then assumption 5 is satisfied for $\mathcal{C}^{\Delta_I}_{1-\alpha}(X) = [\Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}]$, where $c_{1-\alpha}(X)$ is chosen to have the credible set property.*

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 α. This cannot be expected to have a Bayesian interpretation, even though it is a valid frequentist confidence set.[17] 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 δ^* 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 δ) which have at least $1-\alpha$ posterior probability of generating the same distribution of the data as

[15] Note that in many models this rules out point identification, since in many models if $\Delta_I(\mu_0)$ is a singleton, then some μ in any neighborhood of μ_0 result in $\Delta_I(\mu) = \emptyset$.

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

[17] 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.

the true data generating process. Or, $\mathcal{C}^{\delta}_{1-\alpha}(X)$ 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 α and/or situations of (near) point identification. Consider the limiting situation of point identification. Then, $\delta^* \in \Delta_I$ is equivalent to δ^* being the singleton "true value" of δ. Often, there will not be high posterior probability that any particular δ^* is the "true value" of δ (e.g., if the "posterior for δ" 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 δ, in the sense that $\mathcal{R}^{\delta}_{r}(\mathrm{X})$ is the set of all values δ^* 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 θ rather than some δ 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 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 examples 3 and 5. See also Kline (2011) for cases where it is not true.

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

Remark 8 (An alternative credible set)**.** Another approach to constructing a credible set for the identified set is to project a credible set for μ onto the space of subsets of $\Delta(\Theta)$. That is, for any credible set $\mathcal{C}^{\mu}_{1-\alpha}(X)$ for μ, $\Delta_I(\mathcal{C}^{\mu}_{1-\alpha}(X)) = \{\delta : \exists \mu \in \mathcal{C}^{\mu}_{1-\alpha}(X) \text{ s.t. } \delta \in \Delta_I(\mu)\}$ is a credible set for the identified set, such that $\Pi(\Delta_I(\mu) \subseteq \Delta_I(\mathcal{C}^{\mu}_{1-\alpha}(X))|X) \geq 1-\alpha$. Moreover, because per lemma 2, $\Delta_I(\mu) \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ is logically equivalent to

$\mu \in \cap_{\delta \in (\mathcal{C}^{\Delta_I}_{1-\alpha}(X))^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, any $1-\alpha$ credible set for the identified set can be associated with a $1-\alpha$ credible set for μ: $\cap_{\delta \in (\mathcal{C}^{\Delta_I}_{1-\alpha}(X))^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$. Under the condition that the credible set for μ 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 μ 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 μ, in general a $1-\alpha$ credible set for μ 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:

(1) Generate a large sample $\{\Delta(\Theta_I(\mu^{(s)}))\}_{s=1}^S$ according to:
   (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 $\mu^{(s)}$.
(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 θ evaluated at a particular value of μ, 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 (b) involves getting the set $\Theta_I(\mu)$ for a given draw of μ from the posterior $\mu|X$, which is the problem of finding all solutions in θ to $Q(\theta, \mu) = 0$ for a given μ. 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 (b) depends on the application.

One approach involves "guessing and verifying:" guessing values of θ 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 μ 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 $\Delta(\theta) = \theta_k$, the k-th element of θ. 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 B.4.2 of the online 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 Θ. 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

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

be the corresponding un-normalized density. There are many methods for simulating draws from an un-normalized density: among these methods are Metropolis-Hastings sampling and slice sampling. See for example Gamerman and Lopes (2006) for a textbook on related methods. The "recommendations" of this paper based on our own experimentation are discussed below.

In some cases, especially when $\Theta_I(\mu)$ has empty interior, that (un-normalized) 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 un-normalized density

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

where $T > 0$ is a small tuning parameter.[18] 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 $\Theta_I(\mu) \approx \{\theta : \hat{f}(\theta) > 1 - \epsilon\}$ for small $\epsilon > 0$, 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 "non-sharp" identified sets). Another possibility is to check that each of the draws from $\tilde{f}_{\Theta_I(\mu),T}(\theta)$ at least approximately satisfy the condition that the criterion function evaluated at the draw equals zero,[19] which will sharpen the numerical approximation of the identified set.

There are many methods for drawing from such un-normalized densities in the Markov chain Monte Carlo literature. From the perspective of ease of computational implementation, based on our own experimentation we suggest using slice sampling (e.g., Neal (2003)). Slice sampling is implemented in many computational and statistical software packages, but one caveat is that some implementations require an initial "guess" for θ in the identified set (i.e., a guess for where the "density" is non-zero). This can be accomplished in a first step of finding one solution to $Q(\theta,\mu) = 0$ by a standard optimization method. One useful feature of slice sampling is that it does not require the specification of auxiliary distributions (e.g., a proposal distribution) required by some other methods like Metropolis-Hastings sampling, which can be difficult to determine in some applications. Overall, the advantage of this approach is the limited programming required because of pre-implemented slice sampling routines. Of course, the complete implementation details depend on the software package, but generically it is enough to program the criterion function $Q(\mu,\theta)$, so that there is a function that can be called to evaluate $Q(\mu,\theta)$ at a desired value of the arguments, program the function that is the density $\tilde{f}_{\Theta_I(\mu)}(\theta)$ or $\tilde{f}_{\Theta_I(\mu),T}(\theta)$, which will call the $Q(\mu,\theta)$ function, and then apply the pre-implemented slice sampling routine to that density.

7. Monte Carlo experiments

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

[18] 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).

[19] 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.

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 game described in table 1:

		Player 2	
		0	1
Player 1	0	0, 0	0, $\beta_2 + \epsilon_2$
	1	$\beta_1 + \epsilon_1$, 0	$\beta_1 + \Delta_1 + \epsilon_1$, $\beta_2 + \Delta_2 + \epsilon_2$

TABLE 1. Payoff matrix for the binary entry game

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 Δ_1 and Δ_2 are both negative, and that players play a pure strategy Nash equilibrium. This game admits two pure strategy Nash equilibria when $-\beta_i \leq \epsilon_i \leq -\beta_i - \Delta_i$, $i = 1, 2$: 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 ϵ_1 and ϵ_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 known that the econometrician correctly knows the sign of the parameters.

There are six parameters: the β_1, β_2, Δ_1, Δ_2, ρ, 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 μ 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 μ to the identified set for θ results from the assumptions made on the game, as follows.

The criterion function is $Q(\theta, \mu) = (P_{11} - P_{11}(\theta))^2 + (P_{10} - P_{10}(\theta))^2 + (P_{01} - P_{01}(\theta))^2 + (P_{00} - P_{00}(\theta))^2 + \min\{|s(\theta, \mu)|, |s(\theta, \mu) - 1|\}(1 - 1[0 \leq s(\theta, \mu) \leq 1])$, where $P_{00}(\theta) = P(\epsilon_1 \leq -\beta_1, \epsilon_2 \leq -\beta_2)$ and $P_{11}(\theta) = P(\epsilon_1 \geq -\beta_1 - \Delta_1, \epsilon_2 \geq -\beta_2 - \Delta_2)$ correspond to the model-predicted probabilities of the outcomes that occur only as a unique equilibrium, at θ. The $s(\theta, \mu)$ term is the candidate equilibrium selection probability at θ and μ, described below.

$P_{10}(\theta)$ 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,

$$\begin{aligned} P_{01}(\theta) \;=\; & P(-\beta_1 \leq \epsilon_1 \leq -\beta_1 - \Delta_1, \epsilon_2 \geq -\beta_2 - \Delta_2) + P(\epsilon_1 \leq -\beta_1, \epsilon_2 \geq -\beta_2) \\ & + s \times P(-\beta_1 \leq \epsilon_1 \leq -\beta_1 - \Delta_1, -\beta_2 \leq \epsilon_2 \leq -\beta_2 - \Delta_2), \end{aligned}$$

where the parameter s represents the equilibrium selection probability (of choosing the $(0,1)$ equilibrium) in the region of multiple equilbiria. Since it must be that $P_{01} = P_{01}(\theta)$ in the identified set, there is a unique candidate value for s after fixing θ and μ, given by $s(\theta,\mu) = \frac{P_{01}-(P(-\beta_1\leq\epsilon_1\leq-\beta_1-\Delta_1,\epsilon_2\geq-\beta_2-\Delta_2)+P(\epsilon_1\leq-\beta_1,\epsilon_2\geq-\beta_2))}{P(-\beta_1\leq\epsilon_1\leq-\beta_1-\Delta_1,-\beta_2\leq\epsilon_2\leq-\beta_2-\Delta_2)}$. For this to be a valid probability, it must be that $0 \leq s(\theta,\mu) \leq 1$, 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 in the region of multiple equilibria, but this is not known by the econometrician.

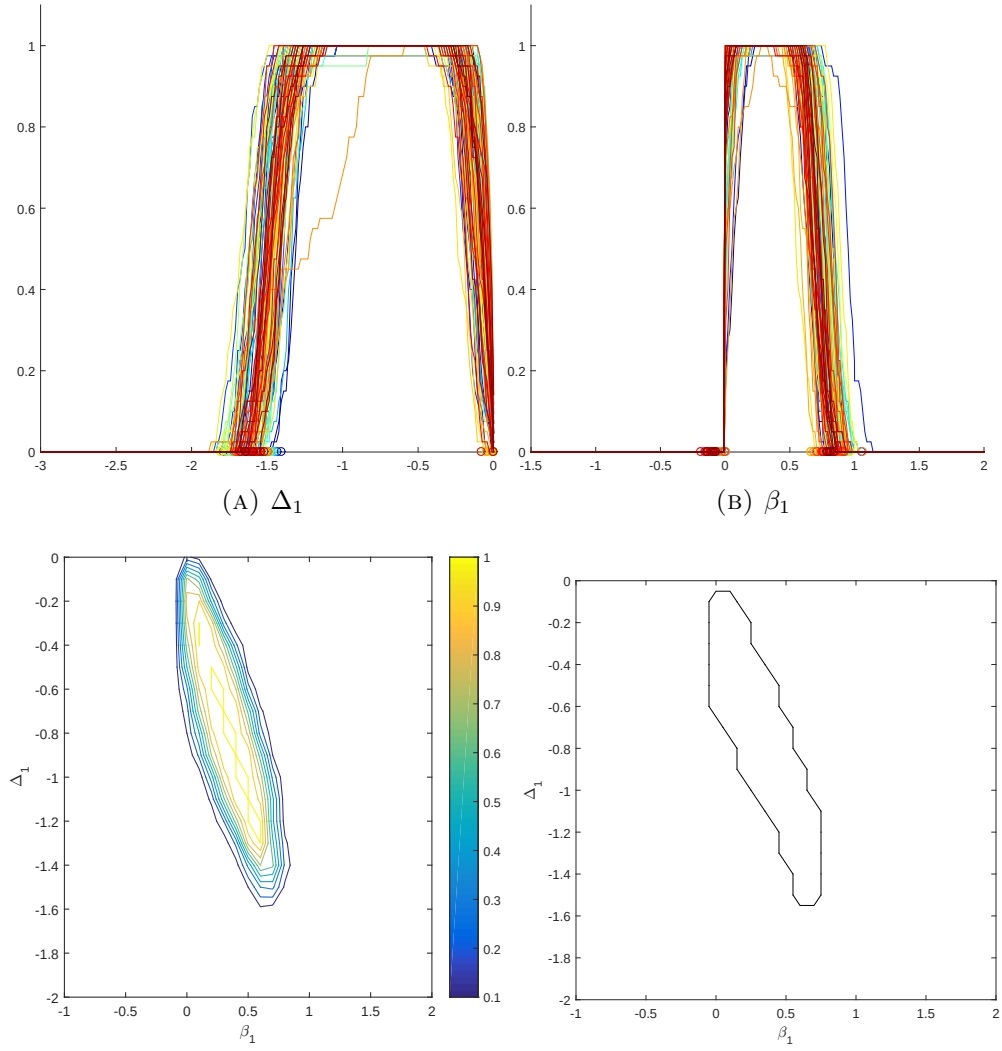
In order to compute the identified set, the slice sampler is used to sample from the "density" $\tilde{f}_{\Theta_I(\mu)}(\theta) = 1[Q(\theta,\mu) = 0]$, as described in section 6.2.[20] The support of draws from $\tilde{f}_{\Theta_I(\mu)}(\theta)$ is taken to be the identified set for θ evaluated at that value of μ, which is then used in the sampler described at the beginning of section 6. Moreover, the identified set evaluated at that value of μ, for any function $\Delta(\cdot)$ of θ, can be taken to be $\Delta(\cdot)$ applied to that computed identified set. In particular, the identified sets for subvectors of θ can be easily computed by "ignoring" the other elements of θ. By computing the identified 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 4 equations [one of which is redundant] and 6 unknowns). The true marginal identified sets for Δ_1 and Δ_2 are each approximately $[-1.49, -0.05]$,[21] while the true identified sets for β_1 and β_2 are each approximately $[0, 0.72]$. Further, the data appears 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 color in panels 1a and 1b corresponds to a different draw from the data generating process. The μ parameters are multinomial, so an uninformative conjugate Dirichlet prior is used, implying a Dirichlet posterior for $\mu|X$.

Panel 1a displays the posterior probabilities that various values of Δ_1 belong to the identified set. Panel 1b does the same for β_1. Panel 1c displays the posterior probabilities

[20] In order to account for numerical error in the computation of the multivariate normal cumulative distribution function, actually a small tolerance is allowed, i.e., the criterion function can be very slightly above zero. The tolerance implies that in practice the "density" is $1[Q(\theta,\mu) \leq 0.001]$.

[21] 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, $P_{01} = P(\epsilon_1 \leq \beta_1, \epsilon_2 \geq -\beta_2) + P(\beta_1 \leq \epsilon_1 \leq 0, \epsilon_2 \geq -\beta_2) + P(0 \leq \epsilon_1 \leq -\beta_1, \epsilon_2 \geq -\beta_2)$ and $P_{00} = P(\epsilon_1 \leq \beta_1, \epsilon_2 \leq \beta_2) + P(\beta_1 \leq \epsilon_1 \leq 0, \epsilon_2 \leq \beta_2) + P(0 \leq \epsilon_1 \leq -\beta_1, \epsilon_2 \leq \beta_2) + P(\epsilon_1 \leq -\beta_1, \beta_2 \leq \epsilon_2 \leq -\beta_2)$. By the rotational symmetry property of the multivariate normal distribution, $P_{00} - P_{01} = P(\epsilon_1 \leq \beta_1, \epsilon_2 \leq \beta_2) - P(\epsilon_1 \leq \beta_1, \epsilon_2 \geq -\beta_2) + P(\epsilon_1 \leq -\beta_1, \beta_2 \leq \epsilon_2 \leq -\beta_2) \geq P(\epsilon_1 \leq \beta_1, \epsilon_2 \leq \beta_2) - P(\epsilon_1 \leq \beta_1, \epsilon_2 \geq -\beta_2)$ since some terms cancel. This is non-negative since $\rho \geq 0$. 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) Δ_1 (B) β_1

(C) (β_1, Δ_1) (D) Identified set for (β_1, Δ_1)

FIGURE 1. $\Pi(\cdot|X)$ for various parameters

that various values of (β_1, Δ_1) belong to the identified set, whereas panel 1d displays the true identified set for (β_1, Δ_1), computed by numerical approximation. The figure in panel 1c is a "contour plot" of the posterior, with the legend on the right showing the numerical interpretation of the level curves. For example, any point inside of the green region has posterior probability of being in the identified set of at least 0.6. Unlike the graphs in the first row, the posterior displayed in panel 1c 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 1d that the joint identified set for (β_1, Δ_1) lies on a diagonal, i.e., large values of Δ_1 are associated with small values of β_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 (β_2, Δ_2) are similar, and so are not reported.

The circles along the horizontal axis in panels 1a and 1b 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 color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 94.0% of the draws from the data generating process, the 95% credible set for the identified set for β_1 indeed does contain the true identified set for β_1, and in approximately 91.0% of the draws from the data generating process, the 95% credible set for the identified set for Δ_1 indeed does contain the true identified set for Δ_1, 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 comes from the second quarter of the 2010 Airline Origin and Destination Survey (DB1B). The data contains 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 (or, low cost carriers)[22] and OA (or, 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

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

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

which essentially results in the payoff matrix in section 7.1 except that "$\beta_i^{cons} + \beta_i^x x_{im}$" replaces "$\beta_i$." This implies that the "non-strategic" 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, divided 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 Berry (1992).) This variable is important since 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 one if the variable is higher than its median value and zero otherwise.

The analysis entails estimating three versions of the model: the first contains only the market presence variable, the second contains only the market size variable, and the third contains both the market presence and market size variables.

For all of the models, the point identified parameter μ is a vector of choice probabilities conditional on the explanatory variables, and the partially identified parameter θ is the vector that characterizes the payoff functions and the correlation in the unobservables, as in section 7.1. The link between μ and θ uses the assumptions that: players are playing a pure strategy Nash equilibrium, and that the Δ parameters are both negative. However, the approach can handle a weakening of either of these assumptions.

The link between μ and θ 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.[23] 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 "non-strategic" term varies across different types of markets). The computation otherwise parallels that in section 7.1.

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

The following concerns the specification including both explanatory variables. The online supplement includes results for two more specifications, each corresponding to including just one of the explanatory variables.

8.1. **Model with market presence and market size.** This specification has both binary explanatory variables: *market presence* and *market size*. The payoff of firm LCC if it enters market m is

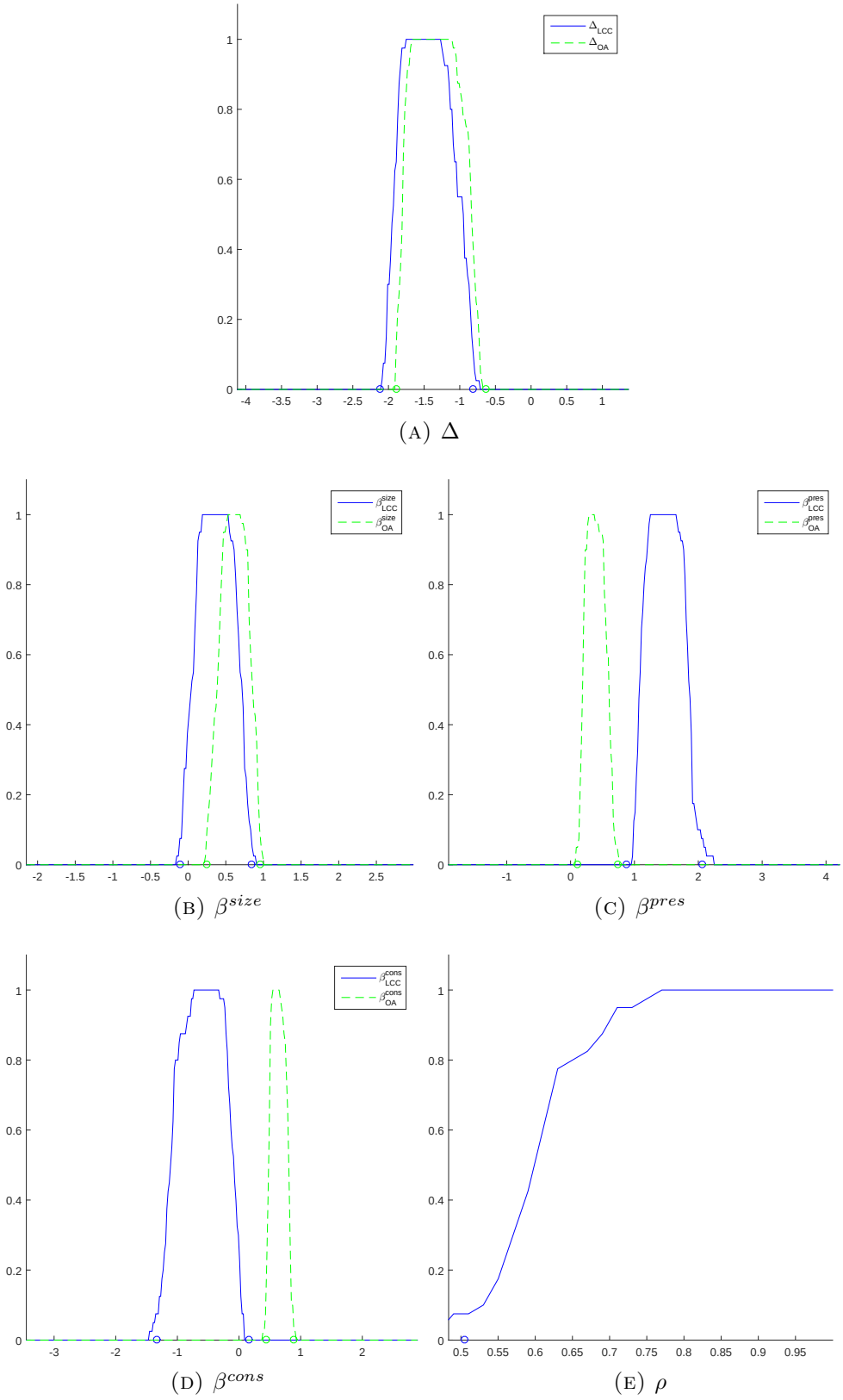
$$\beta_{LCC}^{cons} + \beta_{LCC}^{size} X_{m,size} + \beta_{LCC}^{pres} X_{LCCm,pres} + \Delta_{LCC} y_{OAm} + \epsilon_{LCCm}$$

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

$$\beta_{OA}^{cons} + \beta_{OA}^{size} X_{m,size} + \beta_{OA}^{pres} X_{OAm,pres} + \Delta_{OA} y_{LCCm} + \epsilon_{OAm}.$$

The variable $X_{im,pres}$ 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 $X_{m,size}$ 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, μ is a 32-dimensional vector of conditional choice probabilities (because there are three binary explanatory variables per market resulting in 8 types of markets and each type of market is characterized by four choice probabilities). The partially identified parameter θ is 9-dimensional. The equilibrium selection function (which is a function of the explanatory variables) is profiled out for a given θ, 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 Δ parameters, and the β^{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) seems to be smaller for LCC firms compared to OA firms. And the "curve" of posterior probabilities associated with ρ is basically flat and equal to one for values of ρ 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.



(A) Δ

(B) β^{size}

(C) β^{pres}

(D) β^{cons}

(E) ρ

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

9. Conclusions

This paper has developed a Bayesian[24] approach to inference in partially identified models. The approach results in posterior probability statements concerning the identified set, which is the quantity about which the data is 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., μ and θ 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 non-parametric version of the Bernstein-von Mises theorem that could replace assumption 3.

[24]There is some disagreement in the overall statistical literature concerning the appropriate meaning of "Bayesian," for example Good (1971) has identified the existence of 46656 varieties of Bayesians. Since 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 this 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.

Appendix A. Proofs

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

Proof of lemma 2. $\Delta^* \subseteq \Delta(\Theta_I(\mu))$ is equivalent to $\delta \in \Delta(\Theta_I(\mu))$ for all $\delta \in \Delta^*$. $\delta \in \Delta(\Theta_I(\mu))$ 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 θ such that $\delta = \Delta(\theta)$. And that is equivalent to $\mu \in \cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)$.

$\Delta(\Theta_I(\mu)) \subseteq \Delta^*$ 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 θ such that $\delta = \Delta(\theta)$. And that is equivalent to $\mu \in \cap_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)^C$.

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

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

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

Proof of theorems 1 and 3. For 1.1, since $\mu_0 \in int\left(\cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$, there is an open neighborhood U of μ_0 such that $U \subseteq int\left(\cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$. Therefore, since $\mu|X$ is consistent by assumption 2, $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \cap_{\delta\in\Delta^*}\cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For 1.2, let $\delta^* \in \Delta^*$ be such that $\mu_0 \in int\left(\left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)\right)^C\right)$. Then, it follows $\Pi(\Delta^* \subseteq \Delta_I|X) \leq \Pi(\delta^* \in \Delta_I|X) = \Pi(\mu \in \cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)|X) = 1 - \Pi(\mu \in \left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)\right)^C|X)$. Since $\mu_0 \in int\left(\left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)\right)^C\right)$, there is an open neighborhood U of μ_0 such that $U \subseteq int\left(\left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)\right)^C\right)$. Therefore, since $\mu|X$ is consistent by assumption 2, $\Pi(\mu \in \left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)\right)^C|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For 3.1, note that $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X)$. Therefore, by the same arguments as in the proof of 1.1, but applied to $\cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$, the result follows.

Similarly, for 3.2, let $\delta^* \in (\Delta^*)^C$ be such that $\mu_0 \in int\left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)\right)$. Then, $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X) \leq \Pi(\mu \in \cap_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)^C|X)$. Then, by the same arguments as in the proof of 1.2, but applied to $\cap_{\{\theta:\Delta(\theta)=\delta^*\}}\mu_I(\theta)^C$, the result follows.

For 3.4, since $\mu_0 \in int\left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$, there is an open neighborhood U of μ_0 such that $U \subseteq int\left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$. Therefore, since $\mu|X$ is consistent by assumption 2, $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \equiv \Pi(\mu \in \cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For 3.5, since $\mu_0 \in ext\left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)$, there is an open neighborhood U of μ_0 such that $U \subseteq int\left(\left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)^C\right)$. Therefore, since $\mu|X$ is consistent by assumption 2, it follows that $\Pi(\Delta_I \cap \Delta^* = \emptyset|X) = \Pi(\mu \in \left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)^C|X) \geq \Pi(\mu \in U|X) \to 1$ along almost all sample sequences.

For 1.3, again $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$, so

$$
\begin{aligned}
&\left|\Pi(\Delta^* \subseteq \Delta_I|X) - P_{N(0,\Sigma_0)}\left(\sqrt{n}\left(\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\right)\right)\right| \\
&= \left|\Pi\left(\mu \in \cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X\right) - P_{N(0,\Sigma_0)}\left(\sqrt{n}\left(\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\right)\right)\right| \\
&= \left|\Pi\left(\sqrt{n}(\mu - \mu_n(X)) \in \sqrt{n}\left(\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\right) \mid X\right) - P_{N(0,\Sigma_0)}\left(\sqrt{n}\left(\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) - \mu_n(X)\right)\right)\right| \to 0
\end{aligned}
$$

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

The proof of 3.3 is similar, except applied to the posterior $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X)$. The proof of 3.6 is similar, except applied to the posterior $\Pi(\Delta_I \cap \Delta^* \neq \emptyset|X) \equiv \Pi(\mu \in \cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$. □

Proof of corollaries 2 and 4. For 2.1, the event $\Delta^* \subseteq \Delta_I(\mu)$, which is equivalent to the event that $\mu \in \cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ 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 Δ^* be S. Also, let the neighborhood of μ_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, $\Delta^* = co(S) \subseteq \Delta_I(\mu) \cap \Delta^* \subseteq \Delta_I(\mu)$. Consequently, $\Pi(\Delta^* \subseteq \Delta_I, \mu \in U|X) \geq \Pi(S \subseteq \Delta_I, \mu \in U|X)$.

Since $\Delta^* \subseteq int(\Delta_I)$, in particular $S \subseteq int(\Delta_I)$. Therefore, for each $\delta \in S$, by assumption 4, $\mu_0 \in int(\cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta))$. Therefore, $\mu_0 \in \cap_{\delta \in S} int(\cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta))$.

Since S is finite, equivalently $\mu_0 \in int(\cap_{\delta \in S} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$. Then, by the same arguments as in the proof of part 1.1 of theorem 1, $\Pi(S \subseteq \Delta_I, \mu \in U|X) \to 1$ along almost all sample sequences, which establishes the claim.

For 2.2, since $\Delta^* \not\subseteq \Delta_I$, there is δ^* such that $\delta^* \in \Delta^*$ and $\delta^* \notin \Delta_I$. In particular, therefore $\mu_0 \notin \cup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)$, which is equivalent to $\mu_0 \in \left(\cup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)\right)^C$, which is an open set by assumption 4. Therefore, part 1.2 of theorem 1 applies, which establishes the claim.

For 4.1, let $\tilde{\Delta}^* = int(\Delta^*)$ 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, $\cap_{\delta \in \tilde{\Delta}^{*C}} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is open. Since $\Delta_I \subseteq \tilde{\Delta}^*$, by lemma 2, $\mu_0 \in \cap_{\delta \in \tilde{\Delta}^{*C}} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$. And therefore there is an open neighborhood U of μ_0 such that $U \subseteq \cap_{\delta \in \tilde{\Delta}^{*C}} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$. Therefore, part 3.1 of theorem 3 applies, so $\Pi(\Delta_I \subseteq \tilde{\Delta}^*|X) \to 1$, which establishes the claim.

For 4.2, let $\delta^* \in (\Delta^*)^C \cap int(\Delta_I)$. Then by assumption 4, $\mu_0 \in int(\cup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta))$. Then part 3.2 of theorem 3 establishes the claim.

For 4.3, $\Delta_I(\mu) \cap \Delta^* \neq \emptyset$ for all μ in an open neighborhood of μ_0 is equivalent, by lemma 2, to the statement that all such μ satisfy $\mu \in \cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$, which implies that $\mu_0 \in int(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, so part 3.4 of theorem 3 establishes the claim.

For 4.4, $\Delta_I(\mu) \cap \Delta^* = \emptyset$ for all μ in an open neighborhood of μ_0 is equivalent, by lemma 2, to the statement that all such μ satisfy $\mu \in \left(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)\right)^C$, which implies that $\mu_0 \in ext(\cup_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta))$, so part 3.5 of theorem 3 establishes the claim. □

Proof of theorem 5. Note that, per lemma 2, $\Delta_I(\mu) \subseteq \mathcal{C}^{\Delta_I}_{1-\alpha}(X)$ is logically equivalent to $\mu \in \cap_{\delta \in (\mathcal{C}^{\Delta_I}_{1-\alpha}(X))^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C \equiv \Delta^{-1,\Delta_I}_{1-\alpha}(X)$.

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

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

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

By assumption 6, for any given $\epsilon > 0$, there is a minimal sample size N'_ϵ such that for any sample size $n \geq N'_\epsilon$, $F_n(A) \in P_{N(0,\Sigma_0)}(A) + [-\epsilon, \epsilon]$ for all Borel sets

A. Therefore, $E\left(F_n\left(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\right)\right) \in E\left(P_{N(0,\Sigma_0)}\left(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\right)\right) + [-\epsilon, \epsilon]$. So, because $E\left(P_{N(0,\Sigma_0)}\left(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\right)\right) \to 1-\alpha$ from above, $E\left(F_n\left(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\right)\right) \to 1-\alpha$.

But also, $P\left(\sqrt{n}(\mu_0 - \mu_n(X)) \in \sqrt{n}\left(\Delta_{1-\alpha}^{-1,\Delta_I}(X) - \mu_n(X)\right)\right) = P(\mu_0 \in \Delta_{1-\alpha}^{-1,\Delta_I}(X)) = P(\Delta_I \subseteq \mathcal{C}_{1-\alpha}^{\Delta_I}(X))$, 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. □

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 μ. By the Bayesian delta method, 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, $P\left(\sqrt{n}(\mu_0 - \mu_n(X)) \in \tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X)\right) = P\left(\Delta_I(\mu_0) \subseteq \mathcal{C}_{1-\alpha}^{\Delta_1}(X)\right) = P(\Delta_{IL}(\mu_0) \geq \Delta_{IL}(\mu_n(X)) - \frac{c_{1-\alpha}(X)}{\sqrt{n}}, \Delta_{IU}(\mu_0) \leq \Delta_{IU}(\mu_n(X)) + \frac{c_{1-\alpha}(X)}{\sqrt{n}}) = P(\sqrt{n}(\Delta_{IL}(\mu_0) - \Delta_{IL}(\mu_n(X))) \geq -c_{1-\alpha}(X), \sqrt{n}(\Delta_{IU}(\mu_0) - \Delta_{IU}(\mu_n(X))) \leq c_{1-\alpha}(X)) \to 1-\alpha$, 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 $E\left(F_n(\tilde{\Delta}_{1-\alpha}^{-1,\Delta_I}(X))\right) \to 1-\alpha$, establishing assumption 5. □

Appendix B. Online supplement

This online supplement contains additional material. B.1 provides further examples of the model framework, B.2 provides results on measurability, B.3 discusses inference under misspecification, B.4 provides further Monte Carlo experiments, and B.5 provides further results in the context of the empirical application.

B.1. **Further examples of model framework.**

Example 4 (Moment inequalities)**.** Suppose that θ is known to satisfy the moment inequality conditions $E_{\mathbb{P}_0}(m(X,\theta)) \geq 0$, where $m(X,\theta)$ is a known vector-valued "moment function" of the data, X, and the parameter, θ. The expectation is taken with respect to the true unknown data generating process $\mathbb{P}_0$ for X. Suppose that (perhaps just as an approximation) the random vector X has a discrete distribution with J support points $(x_1,\ldots,x_J)$, such that $\mathbb{P}(X = x_j) = p_j$. In this model, the parameter μ is equal to a specification of $(p_1,\ldots,p_J)$, and so the identified set at μ is $\Theta(\mu) = \{\theta \in \Theta : \sum_{j=1}^J m(x_j,\theta)p_j \geq 0\}$.

More generally, the identified set is $\Theta_I(\mathbb{P}_0) \equiv \{\theta : E_{\mathbb{P}_0}(m(X,\theta)) \geq 0\}$. The identified set that would arise if the data generating process for X equaled $\mathbb{P}$ would similarly be $\Theta_I(\mathbb{P}) \equiv \{\theta : E_{\mathbb{P}}(m(X,\theta)) \geq 0\}$. Suppose that the structure of the moment function $m(\cdot)$ is such that there is a point identified parameter $\mu(\mathbb{P})$ (e.g., moments of functions of X) and a mapping $\Theta_I(\mu)$ such that $\Theta_I(\mu(\mathbb{P})) = \{\theta : E_{\mathbb{P}}(m(X,\theta)) \geq 0\} = \Theta_I(\mathbb{P})$. Then, the point identified parameter is μ, the identified set at μ is $\Theta_I(\mu)$, and the inverse identified set is $\mu_I(\theta) = \{\mu : \theta \in \Theta_I(\mu)\}$.

The existence of $\mu(\mathbb{P})$ and $\Theta_I(\mu)$ is satisfied if the moment function satisfies the property that $m(X,\theta) = \sum_{j=1}^J m_{j1}(X)m_{j2}(\theta)$. Then, $\mu(\mathbb{P}) = \{E_{\mathbb{P}}(m_{j1}(X))\}_j$ and $\Theta_I(\mu) = \{\theta : \sum_{j=1}^J \mu_j m_{j2}(\theta) \geq 0\}$. Many empirically relevant moment inequalities conditions satisfy this property, particularly including various moment inequality conditions based on linear regression.[25] If the moment inequality conditions do not satisfy this property, by discretization the approximation in example 2 can be used.

Example 5 (Posterior probabilities for the simple interval identified parameter, continued)**.** This example continues the discussion in example 3.

Consider $\Pi(\Theta_I \subseteq \Theta^*|X)$. This is the posterior probability that all values in the identified set are contained in Θ^*, or equivalently the posterior probability that all values of the parameter that could have generated the data are contained in Θ^*. Note that $\cap_{\theta\in(\Theta^*)^C}\mu_I(\theta)^C = \cap_{\theta\in(-\infty,a)\cup(b,\infty)}\{\mu : \mu_L \leq \theta \leq \mu_U\}^C = \cap_{\theta\in(-\infty,a)\cup(b,\infty)}\{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\} = \{\mu : \mu_L > \mu_U \text{ or } a \leq \mu_L \leq \mu_U \leq b\} \supset \{\mu : a \leq \mu_L, \mu_U \leq b\}$.[26] So,

[25] See for example section B.4.2 concerning regression with interval data.

[26] The last equality follows: for the first direction, suppose that $\mu \in \{\mu : \mu_L > \mu_U \text{ or } a \leq \mu_L \leq \mu_U \leq b\}$. Suppose that $\mu_L > \mu_U$. Then let θ be any number. Then either $\theta < \mu_L$, or $\theta \geq \mu_L$ (and therefore $\theta > \mu_U$). So either $\theta < \mu_L$ or $\theta > \mu_U$. So, clearly $\mu \in \cap_{\theta\in(-\infty,a)\cup(b,\infty)}\{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\}$. Alternatively, suppose that $a \leq \mu_L \leq \mu_U \leq b$. Let $\theta \in (-\infty,a) \cup (b,\infty)$. If $\theta \in (-\infty,a)$, then $\mu_L \geq a > \theta$, so $\mu_L > \theta$.

$\Pi(\Theta_I \subseteq \Theta^*|X) \equiv \Pi(\{\mu : \mu_L > \mu_U \text{ or } a \leq \mu_L \leq \mu_U \leq b\}|X)$. In other words, this posterior probability is the posterior probability of the set of μ such that the identified set evaluated at μ is either empty, or non-empty but contained within Θ^*.

Or, consider $\Pi(\Theta_I \neq \emptyset|X)$. Note that $\cup_{\theta\in\Theta}\mu_I(\theta) = \cup_{\theta\in\Theta}\{\mu : \mu_L \leq \theta \leq \mu_U\} = \{\mu : \mu_L \leq \mu_U\}$. So, $\Pi(\Theta_I \neq \emptyset|X) \equiv \Pi(\{\mu : \mu_L \leq \mu_U\}|X)$. In other words, this posterior probability is the posterior probability of the set of μ such that the identified set evaluated at μ is non-empty.

First, consider again the large sample behavior of $\Pi(\Theta^* \subseteq \Theta_I|X)$.

Case 3: Consider the general case when $\mu|X$ has a large sample normal approximation as in assumption 3. Suppose that μ are the moments of some bivariate distribution. Suppose that $\mu_n(X)$ is the sample average and that Σ_0 is the covariance of the moments.[27] Then, in large samples, by part 1.3 of theorem 1, the posterior probability is approximately

$$\begin{aligned}\Pi(\Theta^* \subseteq \Theta_I|X) &\approx P_{N(0,\Sigma_0)}\left((\mu_L,\mu_U) : \sqrt{n}\left(\{\tilde{\mu} : \tilde{\mu}_L \leq a, \tilde{\mu}_U \geq b\} - \mu_n(X)\right)\right) \\ &= P_{N(0,\Sigma_0)}\left(\mu_L \leq \sqrt{n}(a - \mu_{nL}(X)), \mu_U \geq \sqrt{n}(b - \mu_{nU}(X))\right)\end{aligned}$$

This large sample approximation makes it possible to derive the repeated large sample behavior. The repeated large sample distribution in some cases is degenerate, in particular if $\mu_{0L} < a \leq b < \mu_{0U}$ or if either $\mu_{0L} > a$ or $\mu_{0U} < b$, as considered in previous cases. So, suppose for example that $\mu_{0L} = a$ and $\mu_{0U} > b$. Then, under suitable regularity conditions, $\sqrt{n}(a - \mu_{nL}(X)) \to^d N(0, \Sigma_{0,LL})$ and $\sqrt{n}(b - \mu_{nU}(X)) \to -\infty$ almost surely. Consequently, in repeated large samples, $\Pi(\Theta^* \subseteq \Theta_I|X) \to Uniform[0,1]$. The same result holds when $\mu_{0L} < a$ and $\mu_{0U} = b$. Consequently, if $\mu_{0L} < \mu_{0U}$, $\Pi(\mu_{0L} \subseteq \Theta_I|X) \to Uniform[0,1]$ and $\Pi(\mu_{0U} \subseteq \Theta_I|X) \to Uniform[0,1]$. So the boundary points of the identified set are "covered" with the same distribution as a p-value in repeated large samples, providing frequentist coverage properties. (See also section 5.)

Now, consider the large sample behavior of $\Pi(\Theta_I \subseteq \Theta^*|X)$.

Case 4: Suppose that $\Theta_I \subset (a,b) \subset [a,b]$ and $\Theta_I \neq \emptyset$. This implies that $a < \mu_{0L} \leq \mu_{0U} < b$. Then, $\mu_0 \in int\left(\cap_{\theta\in(\Theta^*)^C}\mu_I(\theta)^C\right)$, so by part 3.1 of theorem 3, $\Pi(\Theta_I \subseteq \Theta^*|X) \to 1$.

Alternatively, if $\theta \in (b,\infty)$, then $\mu_U \leq b < \theta$, so $\mu_U < \theta$. So, in either case, $\mu \in \cap_{\theta\in(-\infty,a)\cup(b,\infty)}\{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\}$. For the other direction, suppose that $\mu \in \cap_{\theta\in(-\infty,a)\cup(b,\infty)}\{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\}$. If $\mu_L > \mu_U$, then obviously $\mu \in \{\mu : \mu_L > \mu_U \text{ or } a \leq \mu_L \leq \mu_U \leq b\}$, so suppose that $\mu_L \leq \mu_U$. Suppose that it did not hold that $a \leq \mu_L \leq \mu_U \leq b$. Then either $\mu_L < a$ or $\mu_U > b$. Suppose that $\mu_L < a$. First, suppose that $\mu_L = \mu_U < a$. Then, let $\theta = \mu_L$. It must be that $\mu \in \{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\}$, but this is obviously impossible as then either $\mu_L > \theta = \mu_L$ or $\mu_U < \theta = \mu_U$. So assume that $\mu_L < \mu_U$, and let $\theta \in (\mu_L, \min\{a, \mu_U\})$, which exists as long as $\mu_L < \mu_U$. Since $\theta < a$, it must be that $\mu \in \{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\}$. But it cannot be that $\mu_L > \theta > \mu_L$, and also it cannot be that $\mu_U < \theta < \mu_U$, a contradiction. So it must be that $\mu_L \geq a$. Similarly, it must be that $\mu_U \leq b$.

[27] This arises for example if μ is the population mean of a normal distribution, or the population mean of an unknown distribution with suitably flat Dirichlet process prior.

Case 5: Conversely, suppose that $\Theta_I \not\subseteq [a,b]$ and $int(\Theta_I) \neq \emptyset$. This implies that $\mu_{0L} < \mu_{0U}$ and either $\mu_{0L} < a$ or $\mu_{0U} > b$. Consider the case that $\mu_{0L} < a$; the case that $\mu_{0U} > b$ is similar. Let θ^* be some point that is in $(\mu_{0L}, \min\{a, \mu_{0U}\})$. Note that if $a \leq \mu_{0U}$, then this interval is non-empty since $\mu_{0L} < a$. Alternatively, if $a > \mu_{0U}$, then this interval is non-empty since $\mu_{0L} < \mu_{0U}$. Then, $\mu_0 \in int(\mu_I(\theta^*)) = int(\{\mu : \mu_L \leq \theta^* \leq \mu_U\}) = \{\mu : \mu_L < \theta^* < \mu_U\}$, since $\mu_{0L} < \theta^* < \mu_{0U}$ by choice of θ^*. So by part 3.2 of theorem 3, $\Pi(\Theta_I \subseteq \Theta^*|X) \to 0$.

Case 6: And, in large samples, by part 3.3 of theorem 3, the posterior probability is approximately $\Pi(\Theta_I \subseteq \Theta^*|X) \approx P_{N(0,\Sigma_0)}(\sqrt{n}\left(\cap_{\theta \in (\Theta^*)^C} \mu_I(\theta)^C - \mu_n(X)\right))$.

Finally, consider the large sample behavior of $\Pi(\Theta_I \neq \emptyset|X)$.

Case 7: Suppose that $\mu_{0L} < \mu_{0U}$, so that the identified set is non-empty, and not a singleton. Then, $\mu_0 \in int\left(\cup_{\theta \in \Theta} \mu_I(\theta)\right)$. So by part 3.4 of theorem 3, $\Pi(\Theta_I \neq \emptyset|X) \to 1$.

Case 8: Conversely, suppose that $\mu_{0L} > \mu_{0U}$, so that the identified set is empty. Then, $\mu_0 \in ext\left(\cup_{\theta \in \Theta} \mu_I(\theta)\right)$. So by part 3.5 of theorem 3, $\Pi(\Theta_I \neq \emptyset|X) \to 0$.

Case 9: Finally, suppose that $\mu_{0L} = \mu_{0U}$, so that the identified set is a singleton. Then, in large samples, by part 3.6 of theorem 3,

$$\begin{aligned} \Pi(\Delta_I \neq \emptyset|X) &\approx P_{N(0,\Sigma_0)}\left((\mu_L, \mu_U) : \sqrt{n}\left(\{\tilde{\mu} : \tilde{\mu}_L \leq \tilde{\mu}_U\} - \mu_n(X)\right)\right) \\ &= P_{N(0,\Sigma_0)}\left(\mu_L - \mu_U \leq \sqrt{n}(\mu_{nU}(X) - \mu_{nL}(X))\right) \\ &= P_{N(0,\rho_0)}\left(\tilde{\mu} \leq \sqrt{n}(\mu_{nU}(X) - \mu_{nL}(X))\right), \end{aligned}$$

where $\rho_0 = \Sigma_{0,LL} + \Sigma_{0,UU} - 2\Sigma_{0,UL}$. Under regularity conditions, in repeated large samples, $\sqrt{n}(\mu_{nU}(X) - \mu_{nL}(X)) \to^d N(0, \rho_0)$, so in repeated large samples, $\Pi(\Delta_I \neq \emptyset|X) \to^d Uniform[0,1]$. So the posterior probability that the identified set is non-empty provides a consistent frequentist test of non-emptiness of the identified set.

B.2. **Measurability.** In order to establish the measurability of the events corresponding to the posterior probability statements, the following definitions are introduced relative to the measurable sets introduced in assumption 1.

Definition 5 (Measurable inverse included-in sets)**.** $\mathcal{M}_1$ is a collection of subsets of $\mathbb{R}^{d_\delta}$ such that for all $\Delta^* \in \mathcal{M}_1$, $\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a measurable subset of M, i.e., $\cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) \in \mathcal{B}(M)$.

These are the subsets such that $\Pi(\Delta^* \subseteq \Delta_I|X) \equiv \Pi(\mu \in \cap_{\delta \in \Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)|X)$ corresponds to a measurable event.

Definition 6 (Measurable inverse included sets)**.** $\mathcal{M}_2$ is a collection of subsets of $\mathbb{R}^{d_\delta}$ such that for all $\Delta^* \in \mathcal{M}_2$, $\cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is a measurable subset of M, i.e., $\cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C \in \mathcal{B}(M)$.

These are the subsets such that $\Pi(\Delta_I \subseteq \Delta^*|X) \equiv \Pi(\mu \in \cap_{\delta \in (\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C|X)$ corresponds to a measurable event.

Definition 7 (Measurable inverse intersection sets)**.** $\mathcal{M}_3$ is a collection of subsets of $\mathbb{R}^{d_\delta}$ such that for all $\Delta^* \in \mathcal{M}_3$, $\cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a measurable subset of M, i.e., $\cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) \in \mathcal{B}(M)$.

These are the subsets such that $\Pi(\Delta_I \cap \Delta^* \neq \emptyset | X) \equiv \Pi(\mu \in \cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) | X)$ corresponds to a measurable event.

Lemma 3 establishes a sufficient condition for the second and third parts of assumption 4, and establishes the measurability corresponding to definitions 5, 6, and 7. Lemma 3 shows that it is possible to establish measurability without assuming compactness of the parameter space, by using the fact that Euclidean spaces are σ-compact and somewhat subtle facts about Borel sets in metrizable spaces.

Lemma 3. *Suppose that $Q(\theta, \mu)$ is a continuous function, and that Θ is compact. Suppose that $\Delta(\cdot)$ is a continuous function. Then:*

(1) Δ_I is compact. $\Delta(\Theta)$ is compact.
(2) For any $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed.
(3) For any open $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\cap_{\delta\in(\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is open.
(4) For any closed $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed.

Suppose that $Q(\theta, \mu)$ is a continuous function, and that Θ is closed. Suppose that $\Delta(\cdot)$ is a continuous function. Then:

(5) $\mathcal{M}_1 = \mathcal{P}(\mathbb{R}^{d_\delta})$.
(6) $\mathcal{B}(\mathbb{R}^{d_\delta}) \subseteq \mathcal{M}_2$.
(7) $\mathcal{B}(\mathbb{R}^{d_\delta}) \subseteq \mathcal{M}_3$.

Proof of lemma 3. For 1, suppose that $\{\theta_n\}_n$ is a sequence in Θ_I that converges to some point $\theta^* \in \Theta$. Since $\theta_n \in \Theta_I$, $Q(\theta_n, \mu_0) = 0$. Since Q is continuous, $Q(\theta^*, \mu_0) = 0$, so $\theta^* \in \Theta_I$. Therefore, Θ_I is closed, and therefore compact since Θ is bounded. Consequently, $\Delta_I \equiv \Delta(\Theta_I)$ is compact. Similarly, since Θ is compact, $\Delta(\Theta)$ is compact.

For 2, suppose that $\{\mu_n\}_n$ is a sequence in $\cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)$ that converges to some point $\mu^* \in M$. Since $\mu_n \in \cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)$ there must be θ_n such that $\Delta(\theta_n) = \delta$ and $\mu_n \in \mu_I(\theta_n)$. Since Δ is a continuous function and Θ is compact, $\{\theta : \Delta(\theta) = \delta\}$ is compact. Therefore there is a convergent subsequence $\theta_{n_k} \to \theta^* \in \{\theta : \Delta(\theta) = \delta\}$. Since Q is continuous and $Q(\theta_{n_k}, \mu_{n_k}) = 0$ along this subsequence, also $Q(\theta^*, \mu^*) = 0$. So, $\mu^* \in \mu_I(\theta^*)$, and thus $\mu^* \in \cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)$. Therefore $\cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)$ is closed. And since an arbitrary intersection of closed sets is a closed set, for any $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed.

For 4, note by lemma 2 that $\cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) = \cup_{\delta\in\Delta^*\cap\Delta(\Theta)} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. Suppose that $\{\mu_n\}_n$ is a sequence in $\cup_{\delta\in\Delta^*\cap\Delta(\Theta)} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ that converges to some point $\mu^* \in M$. Since $\mu_n \in \cup_{\delta\in\Delta^*\cap\Delta(\Theta)} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ there must be $\delta_n \in \Delta^* \cap \Delta(\Theta)$ and $\theta_n \in \Theta$ such that $\Delta(\theta_n) = \delta_n$ and $\mu_n \in \mu_I(\theta_n)$. Since Θ is compact, and since $\Delta^* \cap \Delta(\Theta)$ is compact (since it is the intersection of a closed set and a compact set),

there is a convergent subsequence $\delta_{n_k} \to \delta^* \in \Delta^* \cap \Delta(\Theta)$ and $\theta_{n_k} \to \theta^* \in \Theta$. Since $\delta_{n_k} = \Delta(\theta_{n_k})$, $\delta^* = \Delta(\theta^*)$. Since Q is continuous and $Q(\theta_{n_k}, \mu_{n_k}) = 0$ along this subsequence, also $Q(\theta^*, \mu^*) = 0$. So, $\mu^* \in \mu_I(\theta^*)$, and thus $\mu^* \in \cup_{\{\theta:\Delta(\theta)=\delta^*\}} \mu_I(\theta)$. And so, $\mu^* \in \cup_{\delta\in\Delta^*\cap\Delta(\Theta)} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. Therefore $\cup_{\delta\in\Delta^*\cap\Delta(\Theta)} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed.

For 3, note by lemma 2 that $\left(\cap_{\delta\in(\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C\right)^C = \cup_{\delta\in(\Delta^*)^C} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) = \cup_{\delta\in(\Delta^*)^C\cap\Delta(\Theta)} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. By part 4, this is closed because $(\Delta^*)^C$ is closed. So, $\cap_{\delta\in(\Delta^*)^C} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is open.

For 5, let $\Theta_m = \Theta \cap \overline{B}(m)$, where $\overline{B}(m)$ is the closed ball of radius m. Since Θ is closed and $\overline{B}(m)$ is compact, Θ_m is compact. Then, $\cup_{m\geq 1}\Theta_m = \Theta$ and $\Theta_m \subseteq \Theta_{m+1}$, so Θ is the countable union of increasing compact sets.

Also note that $\cup_{m\geq 1} \cap_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta) = \cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. If $\mu \in \cup_{m\geq 1} \cap_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$, then $\mu \in \cap_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$ for some $m \geq 1$, and therefore, for that m, for all $\delta \in \Delta^*$ there is $\theta \in \Theta_m$ such that $\Delta(\theta) = \delta$ and $\mu \in \mu_I(\theta)$, so $\mu \in \cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. Conversely, if $\mu \in \cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$, then for all $\delta \in \Delta^*$ there is θ such that $\Delta(\theta) = \delta$ and $\mu \in \mu_I(\theta)$. Since it must be that $\theta \in \Theta_m$ for all m large enough, then also $\mu \in \cup_{m\geq 1} \cap_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$.

The proof of part 2 also establishes: if Θ is closed but not necessarily compact: for any $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\cap_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed. So, $\cap_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) = \cup_{m\geq 1} \cap_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is the countable union of closed sets, so is a Borel set.

For 7, let Θ_m be defined as above, and also let $\Delta_m = \mathbb{R}^{d_\delta} \cap \overline{B}(m)$.[28] Since $\mathbb{R}^{d_\delta}$ is closed and $\overline{B}(m)$ is compact, Δ_m is compact. Also note that $\cup_{m\geq 1} \cup_{\delta\in\Delta^*\cap\Delta_m} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta) = \cup_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$. If $\mu \in \cup_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$, then there is $\delta \in \Delta^*$ and θ such that $\Delta(\theta) = \delta$ and $\mu \in \mu_I(\theta)$. Consequently, for large enough m, $\delta \in \Delta^* \cap \Delta_m$ and $\theta \in \Theta_m$, so $\mu \in \cup_{m\geq 1} \cup_{\delta\in\Delta^*\cap\Delta_m} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$. Conversely, if $\mu \in \cup_{m\geq 1} \cup_{\delta\in\Delta^*\cap\Delta_m} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$, then it is immediate that $\mu \in \cup_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$.

The proof of part 4 also establishes: if Δ^* is closed then $\cup_{\delta\in\Delta^*\cap\Delta_m} \cup_{\{\theta\in\Theta_m:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is closed. Consequently, $\cup_{\delta\in\Delta^*} \cup_{\{\theta\in\Theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is the countable union of closed sets, so is a Borel set. Suppose that Δ^* is either a countable union or a countable intersection of sets Δ_n^* such that $\cup_{\delta\in\Delta_n^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a Borel set for each Δ_n^*. In the case that Δ^* is a countable union, $\Delta^* = \cup_{n\geq 1}\Delta_n^*$. Therefore, $\cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta) = \cup_{n\geq 1} \cup_{\delta\in\Delta_n^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a countable union of Borel sets, so is a Borel set. In the case that Δ^* is a countable intersection, $\Delta^* = \cap_{n\geq 1}\Delta_n^*$. Therefore, $\cap_{\delta\in\Delta^*} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C = \cap_{n\geq 1} \cap_{\delta\in\Delta_n^*} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is the countable intersection of Borel sets, so is a Borel set. This is because $\cup_{\delta\in\Delta_n^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ are Borel sets, so also $\cap_{\delta\in\Delta_n^*} \cap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ are Borel sets. Consequently, $\cup_{\delta\in\Delta^*} \cup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a Borel set for any Borel set Δ^*. This is because the Borel sets of a metrizable space are contained in any collection of sets that has the property: all closed sets are elements of the collection, and the collection is

[28] Note that the dimension of the closed balls in the expressions for Θ_m and Δ_m may be different.

closed under countable unions and countable intersections. See for example Aliprantis and Border (2006, Corollary 4.18).

For 6, note that $\left(\cap_{\delta\in(\Delta^*)^C}\cap_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)^C\right)^C = \cup_{\delta\in(\Delta^*)^C}\cup_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)$ is a Borel set for any Borel set $(\Delta^*)^C$, by part 7. So, since the Borel sets are closed under complements, $\cap_{\delta\in(\Delta^*)^C}\cap_{\{\theta:\Delta(\theta)=\delta\}}\mu_I(\theta)^C$ is a Borel set for any Borel set Δ^*. □

It is worth noting that there exists other results establishing measurability of random sets, see for example Molchanov (2006) or Kitagawa (2012). Results similar to lemma 3 might be possible by establishing these (or similar) conditions on the criterion function and parameter space imply the sufficient conditions for the other measurability results.

Remark 9 (Restricting measurability to the parameter space $\Delta(\Theta)$)**.** Lemma 3 views the posterior probabilities as defined on $\mathbb{R}^{d_\delta}$, rather than restricted to $\Delta(\Theta)$. This is useful because it might be difficult to check whether a particular set of interest is a subset of the parameter space when $\Delta(\cdot)$ has a complicated functional form. However, it is relevant to know how measurability obtains when viewing the posterior probabilities as defined on $\Delta(\Theta)$ as a subspace of $\mathbb{R}^{d_\delta}$ with the subspace topology.

In this analysis, the Borel sets of $\Delta(\Theta)$ are the Borel sets corresponding to the subspace topology on $\Delta(\Theta)$ viewed as a subspace of a Euclidean space, i.e., $\mathcal{B}(\Delta(\Theta)) = \{A\cap\Delta(\Theta) : A \in \mathcal{B}(\mathbb{R}^{d_\delta})\}$. Note in particular that if $\Delta(\Theta) \in \mathcal{B}(\mathbb{R}^{d_\delta})$, then $\mathcal{B}(\Delta(\Theta)) = \{A \in \mathcal{B}(\mathbb{R}^{d_\delta}) : A \subseteq \Delta(\Theta)\} \subseteq \mathcal{B}(\mathbb{R}^{d_\delta})$. Therefore, essentially the same measurability results obtain when viewing the posterior probabilities as defined on $\Delta(\Theta)$.

Remark 10 (Measurability of $Q(\theta,\mu)$)**.** Because of the connection (by definition) of measurability of a function and the measurability of pre-images of measurable sets, it is tempting to ask for an analogue of lemma 3 that assumes only measurability of $Q(\theta,\mu)$. Unfortunately, such a result (in general) is not available. Suppose that $M = [0,1]$ and $\Theta = [0,1]$, and let A be any set in $\Theta \times M$ with the property that: A is Borel measurable, but the projection of A onto M is not Borel measurable. That such sets exists is the same as the existence of analytic but not Borel measurable sets. Let $Q(\theta,\mu)$ be one minus the characteristic function for A, which is measurable (by definition). Then, consider the set of μ such that $\Theta_I(\mu)\cap\Theta \neq \emptyset$, i.e., the set of μ corresponding to the posterior probability of a non-empty identified set, or equivalently quantity 2 in definition 3. That set of μ is the projection of A onto M, which is not Borel measurable by construction. That suggests we should not expect to be able to assign a posterior probability to non-emptiness of the identified set with this $Q(\theta,\mu)$, despite measurability of $Q(\theta,\mu)$.

B.3. **Misspecification.** A common concern in applied work is the behavior of an estimator under model misspecification. Many estimators in point identified models have the feature that even if the model is misspecified, the estimator still estimates a useful

parameter.[29] This section briefly describes the "expected" behavior under misspecification in the context of an interval identified parameter.

It is useful to consider the behavior of posterior probability statements that condition on the identified set being non-empty, which "imposes" the assumption of correct specification even when that assumption is false. These are the posterior probabilities of the form $\Pi(\cdot|X, \Theta_I \neq \emptyset)$. As shown in the theoretical results, the posterior probabilities in a misspecified model that do not condition on the identified set being non-empty (i.e., $\Pi(\cdot|X)$) will tend in large samples toward estimating that indeed the identified set is empty, which is perhaps the desired behavior under misspecification (since in that case indeed the identified set is empty). However, conditioning on $\Theta_I \neq \emptyset$ can be used as a way to get a "pseudo-true" identified set even under misspecification. This method amounts to "rejecting" draws from the posterior of $\mu|X$ that violate a certain condition (namely, non-emptiness of the identified set), which more generally has been implemented in other settings. For example, Uhlig (2005) imposes an identification constraint concerning sign restrictions on an impulse response function in VAR models by "rejecting" draws from the posterior that violate that restriction.

Essentially, based on $\Pi(\cdot|X, \Theta_I \neq \emptyset)$, the posterior should be expected to "estimate" the identified set corresponding to the value of μ that is "closest" (under a metric induced by the posterior for μ) to the true μ_0 among all values of μ that result in a non-empty identified set. In that sense, this approach estimates a "pseudo-true" identified set under misspecification. This can be most easily seen in the context of an interval identified parameter from example 1 summarized in figure 3, but the basic idea immediately generalizes to any model.

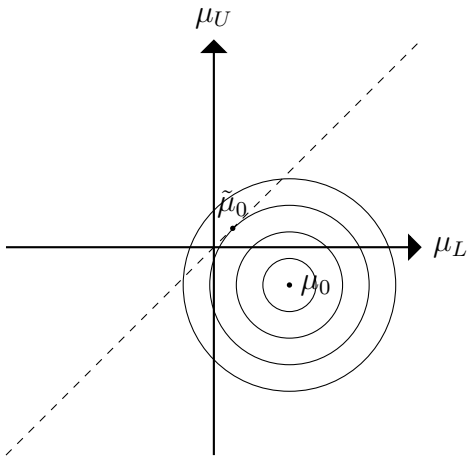


FIGURE 3. Behavior under misspecification

[29] For example, ordinary least squares regression estimates the minimum mean square error linear approximation to the conditional expectation, quantile regression estimates a certain weighted mean square error (i.e., Angrist et al. (2006)), and maximum likelihood estimates a "pseudo-true" parameter that minimizes the Kullback-Leibler divergence to the truth (e.g., White (1982)). The situation might be different in partially identified models (e.g., Ponomareva and Tamer (2011)).

This figure shows μ_L along the horizontal axis, and μ_U along the vertical axis. The dashed diagonal line corresponds to the values of μ such that θ is point identified, since along that line $\mu_L = \mu_U$, so it must be that $\theta = \mu_L = \mu_U$. Values of μ above the diagonal line correspond to values of μ such that θ is partially identified, since there $\mu_L < \mu_U$. Finally, values of μ below the diagonal line correspond to value of μ such that the model is misspecified, since there $\mu_L > \mu_U$. So, suppose that the true μ_0 indeed is below the diagonal line, as reflected in the figure. The posterior for μ is unaffected by the fact that the model is misspecified, so the posterior for μ still behaves like a conventional posterior for a point identified parameter. The concentric circles emanating from μ_0 represent the level sets of the posterior for μ, with values of μ closer to μ_0 more likely according to the posterior for μ. (It is assumed here that the posterior is centered at the true μ_0, which will be approximately true at least in large samples under assumptions 2 and 3. More generally, the posterior will be centered at approximately $\mu_n(X)$ per assumption 3.)

Now consider the behavior of $\Pi(\cdot|X, \Theta_I \neq \emptyset)$. This is concerned with the posterior for μ, conditioning on the fact that the identified set is non-empty. In the interval identified parameter model, this is equivalent to conditioning on $\mu_L \leq \mu_U$. Therefore, $\Pi(\cdot|X, \Theta_I \neq \emptyset)$ is based on the posterior $\mu|(X, \mu_L \leq \mu_U)$. The most likely draws from $\mu|(X, \mu_L \leq \mu_U)$ will be in a neighborhood of $\tilde{\mu}_0$, which is the point in the set $\{\mu : \mu_L \leq \mu_U\}$ that has highest density according to the posterior for $\mu|X$. As a consequence, inference on the identified set conditioning on non-emptiness of the identified set will tend to be centered around the identified set evaluated at $\tilde{\mu}_0$, and in that sense this approach estimates a "pseudo-true" identified set under misspecification. The basic idea immediately generalizes to other models.

B.4. **Further Monte Carlo experiments.**

B.4.1. *A simple interval identified parameter.* This section reports the results of a Monte Carlo experiment in the context of a simple interval identified parameter, described in examples 1 and 3. The data generating process in this experiment is:

$$\begin{pmatrix} X_U \\ X_L \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \mu_{0U} \\ \mu_{0L} \end{pmatrix}, \begin{pmatrix} \Sigma_{0,UU} & \Sigma_{0,UL} \\ \Sigma_{0,UL} & \Sigma_{0,LL} \end{pmatrix} \right).$$

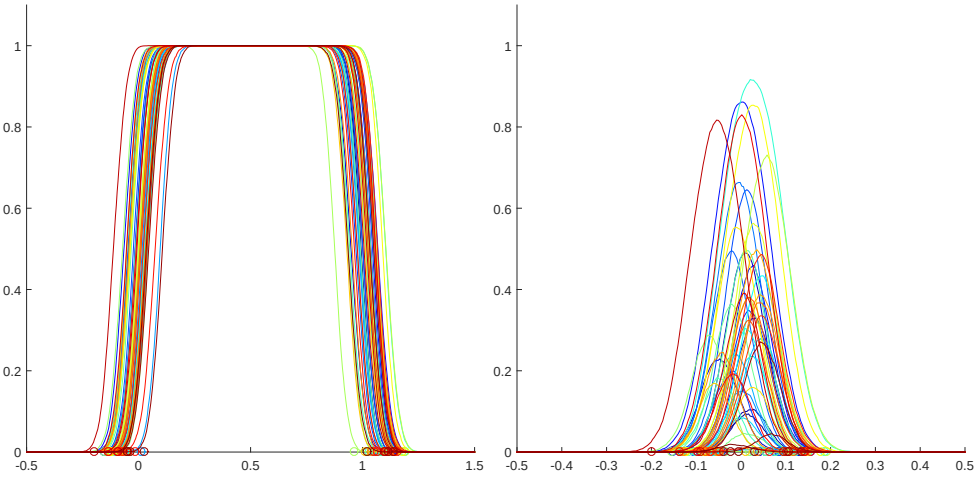
Consequently, the endpoints of the interval identified set are the first moments of the distribution of X. Suppose that the data is a random sample of $N = 500$ observations from the data generating process. Also suppose that Σ_0 is the identity matrix. The econometrician does not know that the data is normally distributed.

There are many approaches that result in a posterior distribution for μ. One possibility is to specify a normal likelihood, and specify conjugate priors for the unknown parameters. However, this entails potentially undesirable parametric distributional assumptions. Another possibility is to use the Bayesian bootstrap, which is a non-parametric approach to Bayesian inference on moments of a distribution that does not require the

specification of a parametric likelihood, and that require minimal computational investment. (See the discussion after assumption 3 for references.)

As a consequence of considering posterior probabilities over the identified set (rather than a posterior over the partially identified parameter), the theoretical results show that both priors will result in the same large sample approximations to the posterior probabilities over the identified set. Consequently, this Monte Carlo experiment works directly with the large sample approximations based on the Bayesian bootstrap so that $\mu|X \sim Normal(\mu_n, \frac{\Sigma_n}{n})$ where μ_n is the sample average of the moments corresponding to μ, and Σ_n is the sample covariance of the moments corresponding to μ. By the logic of those approximations, those approximations are still functions of the sample size n: these results can be viewed as using a *numerical approximation* to the posterior (which is still a function of sample size n). As expected from the theoretical results, results not reported here show that almost exactly the same results obtain from the "exact" posteriors under reasonable prior specifications. The Bayesian bootstrap does not entail parametric distributional assumptions, so it does not assume that the data generating process is normal.

The experiment involves multiple different specifications of μ_0.



(A) $\Pi(\theta \in \Theta_I|X)$ when $\mu_{0L} = 0$ and $\mu_{0U} = 1$

(B) $\Pi(\theta \in \Theta_I|X)$ when $\mu_{0L} = 0$ and $\mu_{0U} = 0$

FIGURE 4. Posterior probabilities that various parameter values belong to the identified set

First, suppose that $\mu_{0L} = 0$ and $\mu_{0U} = 1$, so that there is a non-singleton identified set. Figure 4a displays the values of $\Pi(\theta \in \Theta_I|X)$ for various values of θ, and various draws from the data generating process. Each "curve" corresponds to $\Pi(\theta \in \Theta_I|X)$ for a particular value of X drawn from the data generating process, treating θ as the argument that is plotted along the horizontal axis. Consequently, the distribution over $\Pi(\theta \in \Theta_I|X)$ (i.e., the existence of multiple curves in the figure) is the distribution induced by the data generating process.

As discussed in example 3, for essentially all draws of X, $\Pi(\theta \in \Theta_I|X) \approx 1$ for values in approximately $[0.1, 0.9]$. So, $\Pi(\theta \in \Theta_I|X) \approx 1$ on a large subset of the interior of the identified set. (In larger samples, per the discussion in example 3, the interval on which $\Pi(\theta \in \Theta_I|X) \approx 1$ would be wider.) Also, for essentially all draws of X, $\Pi(\theta \in \Theta_I|X) \approx 0$ for values outside approximately $[-0.1, 1.1]$. (In larger samples, per the discussion in example 3, the interval on which $\Pi(\theta \in \Theta_I|X) \not\approx 0$ would be narrower.) And finally, per the discussion in example 3, in the neighborhoods of the two points on the boundary of the identified set, the values of $\Pi(\theta \in \Theta_I|X)$ vary depending on the particular draw of X. Note that, as discussed throughout this paper, this figure should not be interpreted to mean that there is a "posterior for" θ that is uniform on (most of) the identified set, $[0, 1]$. Indeed, if $\mu_{0U} = 2$ instead, then the analogous figure would have values of 1 on about $[0.1, 1.9]$, which would obviously not be a "uniform" posterior on the identified set. Instead, the interpretation is that there is essentially posterior certainty that all such points are in the identified set, or equivalently, there is essentially posterior certainty that all such points could have generated the data.

The circles along the horizontal axis of figure 4a are the endpoints of the 95% credible set for the identified set, for each draw from the data generating process. The credible set of a given color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 95.8% of the draws from the data generating process, the 95% credible set indeed does contain the true identified set, so the credible set is also a valid frequentist confidence set.

Now, second, suppose that $\mu_{0L} = 0$ and $\mu_{0U} = 0$, so that there is point identification; however, this is not known a priori by the econometrician. Figure 4b similarly displays the values of $\Pi(\theta \in \Theta_I|X)$ for various values of θ, and various draws from the data generating process. $\Pi(\theta \in \Theta_I|X)$ tends to be largest for values around 0, the singleton value of the identified set. However, unlike in the above case of a non-singleton identified set, in general $\Pi(\theta \in \Theta_I|X)$ is bounded away from 1. This is consistent with the discussion in example 3 that concludes that in large samples, $\Pi(0 \in \Theta_I|X) \approx P_{N(0,\Sigma_0)}(\mu_L \leq -\sqrt{n}\mu_{nL}(X), \mu_U \geq -\sqrt{n}\mu_{nU}(X))$. A "typical" value of $(\sqrt{n}\mu_{nL}(X), \sqrt{n}\mu_{nU}(X))$ in large samples is $(0, 0)$, which would imply that $\Pi(0 \in \Theta_I|X) \approx P_{N(0,\Sigma_0)}(\mu_L \leq 0, \mu_U \geq 0)$. Since Σ_0 is the identity matrix, $\Pi(0 \in \Theta_I|X) \approx \frac{1}{4}$.

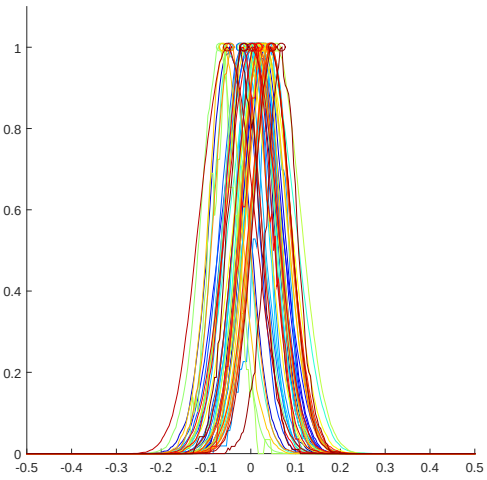
The reason for this is that $\Pi(0 \in \Theta_I|X) = \Pi(\mu_L \leq 0 \leq \mu_U|X)$. If $\mu_{0L} = 0 = \mu_{0U}$, then the posterior for μ does not necessarily satisfy $\mu_L \leq 0 \leq \mu_U$ with high probability, since consistency of the posterior allows that $\mu_L > \mu_U$ with high probability, and also that $0 < \mu_L \leq \mu_U$ or $\mu_L \leq \mu_U < 0$ with high probability. This is roughly analogous to the "boundary" problem that would arise in existing frequentist approaches to this model. But note that, unlike in existing frequentist approaches, it is not necessary to use an ad hoc rule like a "tolerance parameter." This is because the Bayesian approach, including for data such that $\mu_{nL} > \mu_{nU}$, results in a non-degenerate posterior distribution over the

identified set. In particular, if $\mu_{nL} > \mu_{nU}$, then some of the draws of the identified set will be the "empty" identified set (for draws such that $\mu_L > \mu_U$), while others will be a narrow identified set (for draws such that $\mu_L \leq \mu_U$ and $\mu_L \approx \mu_U$). Therefore, the Bayesian approach "automatically" accounts for the fact that $\mu_{nL} > \mu_{nU}$ does not necessarily mean that the true identified set is empty, whereas existing frequentist approaches have to impose this fact using an ad hoc rule.

Nevertheless, even in large samples there will not be a large amount of posterior evidence that $0 \in \Theta_I$. And, since consistency allows that $0 < \mu_L \leq \mu_U$ or $\mu_L \leq \mu_U < 0$ with high probability, this is true even for $\Pi(0 \in \Theta_I | X, \Theta_I \neq \emptyset)$, since $\Pi(0 \in \Theta_I | X, \Theta_I \neq \emptyset) = \Pi(0 \in \Theta_I | X, \mu_L \leq \mu_U)$. However, this is not a deficiency of this approach. Rather, it is the logical Bayesian inference based on the structure of the model, as just discussed.

More similarly to before with partial identification, for essentially all draws of X, $\Pi(\theta \in \Theta_I | X) \approx 0$ for values outside approximately $[-0.2, 0.2]$.

As above, the circles along the horizontal axis of of figure 4b are the endpoints of the 95% credible set for the identified set, for each draw from the data generating process. The credible set of a given color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 90.6% of the draws from the data generating process, the 95% credible set indeed does contain the true identified set, so the credible set is not quite (but is almost) a valid frequentist confidence set. The lack of exact frequentist coverage is not surprising because when $\mu_{0L} = \mu_{0U}$, the discussion in remark 5 does not hold.



(A) $\frac{\Pi(\theta \in \Theta_I | X)}{\max_\theta \Pi(\theta \in \Theta_I | X)}$

(B) $\Pi(\Theta_I \neq \emptyset | X)$

FIGURE 5. Other posterior probabilities when $\mu_{0L} = 0$ and $\mu_{0U} = 0$

In some applications, it may be of interest to know the value(s) of θ that are "most likely" to be in the identified set, and/or to compare the relative odds that various values of θ are in the identified set. Figure 5a displays the values of the posterior odds $\frac{\Pi(\theta\in\Theta_I|X)}{\max_\theta \Pi(\theta\in\Theta_I|X)}$. The relative odds of θ_1^* and θ_2^* is the ratio of the displayed posterior odds, since the denominator cancels. $\frac{\Pi(\theta\in\Theta_I|X)}{\max_\theta \Pi(\theta\in\Theta_I|X)}$ behaves more like $\Pi(\theta \in \Theta_I|X)$ behaved before in the case of a non-singleton identified set. The posterior odds for values outside approximately $[-0.2, 0.2]$ are approximately zero. And the posterior odds for essentially a single value of θ in a neighborhood of the true identified set is 1. The value of θ that has maximal posterior odds depends on the draw of X; it tends to be approximately $\frac{\mu_{nL}+\mu_{nU}}{2}$, which is indicated for each draw from the data generating process by a circle that is the same color as the corresponding posterior "curve." If there is a non-singleton identified set, then $\max_\theta \Pi(\theta \in \Theta_I|X) \approx 1$ so $\frac{\Pi(\theta\in\Theta_I|X)}{\max_\theta \Pi(\theta\in\Theta_I|X)} \approx \Pi(\theta \in \Theta_I|X)$, so the posterior odds are essentially the same as $\Pi(\theta \in \Theta_I|X)$ in the case of a non-singleton identified set.

Particularly in the case that $\mu_{0L} = 0 = \mu_{0U}$, it may also be of interest to know the posterior probability that the identified set is non-empty. Figure 5b displays the posterior probability that the identified set is non-empty, for various draws from the data generating process. The posterior probability in this figure of a given color corresponds to the same draw of X as the posterior "curves" displayed above of the same color. As expected from example 3, these posterior probabilities are distributed approximately according to Uniform$[0, 1]$ in repeated samples. If there is a non-singleton identified set, then the posterior probability that the identified set is non-empty is essentially 1 for all draws from the data generating process, so those posterior probabilities are not displayed.

B.4.2. *Regression with interval data.* This section reports the results of a Monte Carlo experiment in the context of interval data on the outcome in a linear regression model. The data generating process in this experiment is:

$$Y = Z\beta + U = -1 + 1Z_1 + 2Z_2 + 3Z_3 + U,$$

where $\beta = (-1, 1, 2, 3)$ is the true parameter and

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ U \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 & 0.3 & 0 \\ 0.3 & 1 & 0.3 & 0 \\ 0.3 & 0.3 & 1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}\right)$$

The observed outcome is the interval $[floor(Y), ceil(Y)]$. The data is therefore $X = (floor(Y), ceil(Y), Z)$. The data is a random sample of $N = 2000$ observations from the data generating process. See also for example Manski and Tamer (2003).

This model implies the conditional moment inequality conditions $E(floor(Y)|Z) \leq Z\beta \leq E(ceil(Y)|Z)$ for all Z, and therefore $E(f(Z)(ceil(Y)-Z\beta)) \geq 0$ and $E(f(Z)(Z\beta-$

$floor(Y))) \geq 0$ for any non-negative vector-valued function $f(\cdot)$. The Monte Carlo experiment uses the four "instruments" in $f(Z) = (1, Z_1^2, Z_2^2, Z_3^2)$. Therefore, there are eight moment inequality conditions, and the point identified parameter μ is a 19×1 vector of non-redundant moments of various products of $floor(Y)$, $ceil(Y)$, and the components of Z. The posterior for μ comes from the large sample approximation to the Bayesian bootstrap, so $\mu|X \sim Normal(\mu_n, \frac{\Sigma_n}{n})$ where μ_n is the sample average of the moments corresponding to μ, and Σ_n is the sample covariance of the moments corresponding to μ. The partially identified parameter of interest is β_3, the coefficient on Z_3. Consequently, $\Delta(\beta) = \beta_3$. By numerical approximation, the true identified set for β_3 corresponding to these eight moment inequality conditions is $\Delta_I \approx [1.84, 4.16]$. The identified set $\Theta_I(\mu)$ is a convex polytope (i.e., the set of solutions of the moment inequality conditions). Therefore, computation of $\Delta_I(\mu)$ is a linear programming problem.

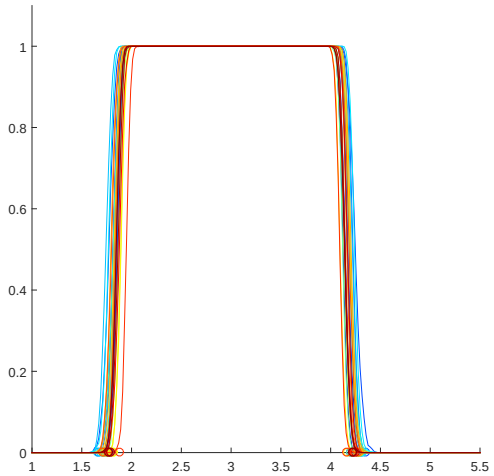


FIGURE 6. $\Pi(\beta_3 \in \Delta_I|X)$

Figure 6 displays the values of $\Pi(\beta_3 \in \Delta_I|X)$ for various values of β_3, and various draws from the data generating process. As before, each "curve" corresponds to $\Pi(\beta_3 \in \Delta_I|X)$ for a particular value of X drawn from the data generating process. For essentially all draws of X, $\Pi(\beta_3 \in \Delta_I|X) \approx 1$ for values in approximately $[1.9, 4.1]$. So, $\Pi(\beta_3 \in \Delta_I|X) \approx 1$ on essentially the entirety of the identified set. Also, for essentially all draws of X, $\Pi(\beta_3 \in \Delta_I|X) \approx 0$ for values outside approximately $[1.5, 4.5]$. In the neighborhoods of the two points on the boundary of the identified set, the values of $\Pi(\beta_3 \in \Delta_I|X)$ vary depending on the particular draw of X.

The circles along the horizontal axis of figure 6 are the endpoints of the 95% credible set for the identified set, for each draw from the data generating process. The credible set of a given color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 95.6% of the draws from the data generating process, the 95% credible set indeed does contain the true identified set, so the credible set is also a valid frequentist confidence set. Note that this concerns just part of the

partially identified parameter, but is nevertheless consistent-in-level. Other frequentist approaches might require conservative projection methods.

B.5. **Further empirical results.**

B.5.1. *Model with market presence.* One specification uses only the firm- and market-specific binary explanatory variable: *market presence*. In this specification, the payoff of firm LCC if it enters market m is

$$\beta_{LCC}^{cons} + \beta_{LCC}^{pres} X_{LCCm,pres} + \Delta_{LCC} y_{OAm} + \epsilon_{LCCm}$$

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

$$\beta_{OA}^{cons} + \beta_{OA}^{pres} X_{OAm,pres} + \Delta_{OA} y_{LCCm} + \epsilon_{OAm}.$$

In this specification, θ is a 7-dimensional vector. The point identified parameter μ is a 16-dimensional vector of conditional choice probabilities: there are four types of markets (because there are two binary explanatory variables per market) and each type of market is summarized by four choice probabilities (because there are two firms each of which has a binary decision to enter or not).

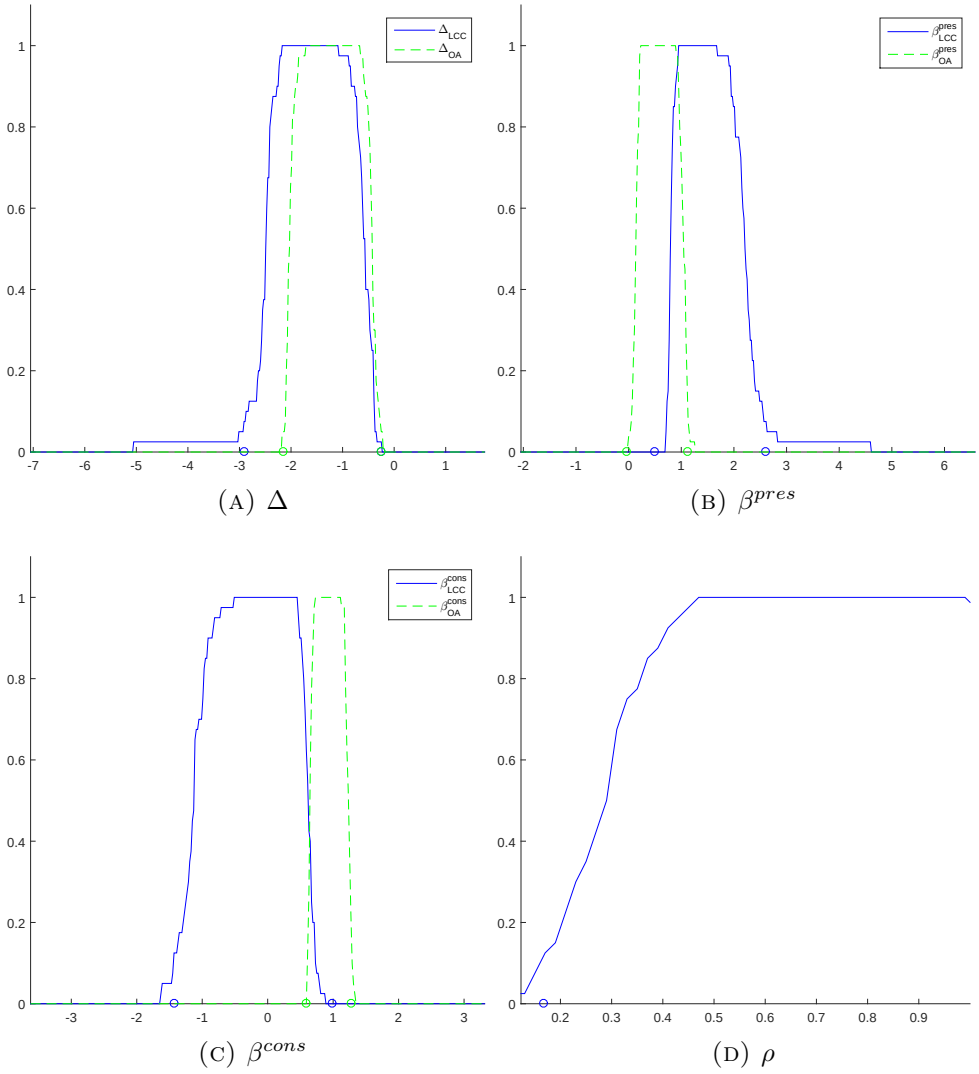
Figure 7 reports the posterior probabilities that various parameter values belong to the identified set. The posterior probabilities for the identified sets for the Δ parameters seem essentially the same across the two types of firms. The effect of market presence on the payoff of both firms is almost certainly positive, in the sense that the posterior "curve" for both firms is close to zero on negative values, but the effect of having a greater market presence seems higher for LCC firms, in the sense that the posterior "curve" for the LCC firms is relatively greater on larger values of the parameter space. The monopoly profits associated with below-median market presence (i.e., the constant term) seems lower for LCC firms. Finally, the "curve" of posterior probabilities associated with ρ is basically flat and equal to one for values of ρ greater than approximately 0.5, implying that any sufficiently high correlation almost certainly could have generated the data. The circles along the horizontal axes in figure 7 are the endpoints of the 95% credible sets for the identified set for the corresponding parameter. Figure 8 displays the posterior probabilities over the identified sets for a few pairs of parameters.

B.5.2. *Model with market size.* Another specification also has only one binary explanatory variable: *market size*. In this specification, the payoff of firm LCC if it enters market m is

$$\beta_{LCC}^{cons} + \beta_{LCC}^{size} X_{m,size} + \Delta_{LCC} y_{OAm} + \epsilon_{LCCm}$$

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

$$\beta_{OA}^{cons} + \beta_{OA}^{size} X_{m,size} + \Delta_{OA} y_{LCCm} + \epsilon_{OAm}.$$



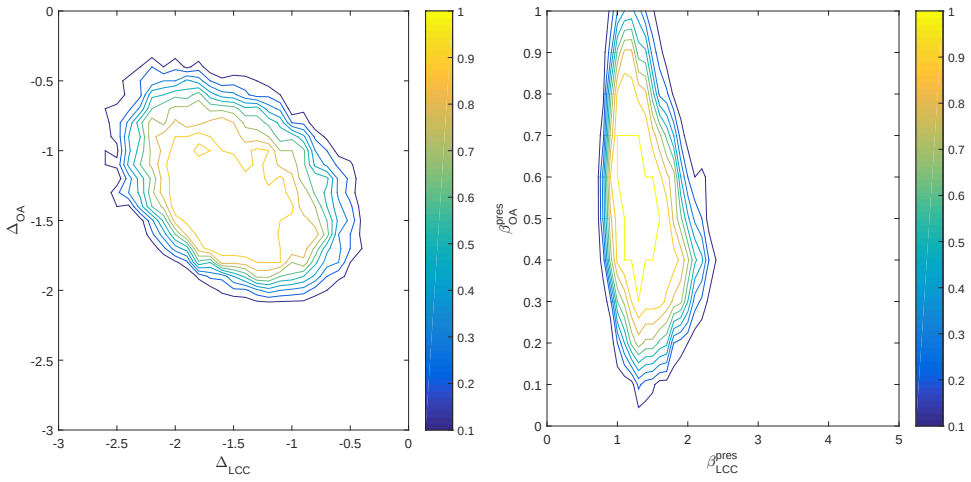
(A) Δ (B) β^{pres}

(C) β^{cons} (D) ρ

FIGURE 7. Posterior probabilities that various parameter values belong to the identified set in model with market presence only

Therefore, this specification of the model does not contain any excluded regressors. In this specification, the point identified parameter μ is an 8-dimensional vector of conditional choice probabilities, while the partially identified parameter θ remains a 7-dimensional vector.

Figure 9 reports the posterior probabilities that various parameter values belong to the identified set. As compared to the figures for the model with market presence in section B.5.1, this specification with market size seems to contain less information about θ, as expected from the literature on identification that suggests the key role of excluded regressors. In particular, this specification of the model implies 8 moment equality conditions (2 of which are redundant) compared to the 16 moment equality conditions (4 of which are redundant) in the specification with market presence.

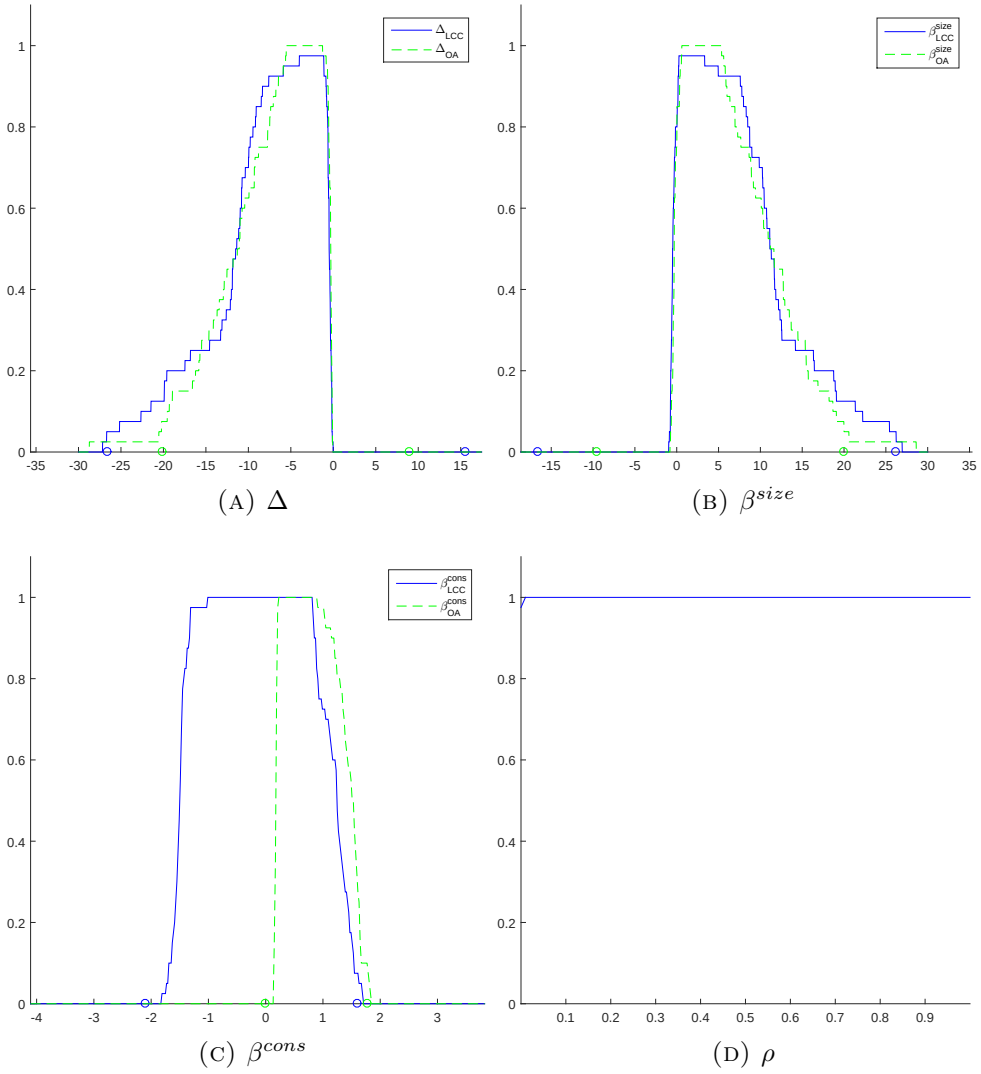


(A) $(\Delta_{LCC}, \Delta_{OA})$ (B) $(\beta^{pres}_{LCC}, \beta^{pres}_{OA})$

(C) $(\beta^{cons}_{LCC}, \Delta_{LCC})$ (D) $(\beta^{cons}_{OA}, \Delta_{OA})$

FIGURE 8. Posterior joint probabilities that various pairs of parameter values belong to the identified set in model with market presence only

The posterior "curve" for the identified sets for the Δ parameters suggests that the data in this specification is much less informative about the magnitude of the interaction effect than it was in the specification in section B.5.1. The same is true for the posterior "curve" for the identified set for the β^{size} parameters. The posterior "curve" for monopoly profits associated with below-median market size (i.e., the constant terms) seems about the same compared to the specification in section B.5.1, but the identified sets are slightly larger with less evidence of a difference across types of firms. Finally, it is not surprising that evidently nothing is learned about ρ, since the literature on identification suggests that it is difficult to distinguish between correlated unobservables and other factors without an excluded regressor. The circles along the horizontal axes in figure 9 are the endpoints of the 95% credible sets for the identified set for the corresponding parameter.



(A) Δ (B) β^{size}

(C) β^{cons} (D) ρ

FIGURE 9. Posterior probabilities that various parameter values belong to the identified set in the model with market size only

References

Aliprantis, C. and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer: Berlin, 3 ed.

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

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

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

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile regression under misspecification, with an application to the US wage structure," *Econometrica*, 74, 539–563.

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

——— (2012): "Partial identification using random set theory," *Journal of Econometrics*, 166, 17–32.

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

Bernardo, J. M. and A. F. Smith (2009): *Bayesian theory*, vol. 405, John Wiley & Sons.

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

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

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

Bickel, P. and P. Millar (1992): "Uniform convergence of probability measures on classes of functions," *Statistica Sinica*, 2, 15.

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

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

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

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

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

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

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

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

Ferguson, T. (1973): "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, 209–230.

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

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

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

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

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

——— (1971): "46656 varieties of Bayesians," *The American Statistician*, 25, 62–63.

Hwang, C.-R. (1980): "Laplace's method revisited: weak convergence of probability measures," *The Annals of Probability*, 1177–1182.

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

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

Kitagawa, T. (2012): "Estimation and inference for set-identified parameters using posterior lower probability," Working paper.

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, 3131–3141.

——— (2015a): "The empirical content of games with bounded regressors," Working paper.

——— (2015b): "Identification of complete information games," Working paper.

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

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

Lo, A. (1987): "A large sample study of the Bayesian bootstrap," *The Annals of Statistics*, 1, 360–375.

Manski, C. (2003): *Partial Identification of Probability Distributions*, Springer: New York.

——— (2007): *Identification for Prediction and Decision*, Harvard University Press.

Manski, C. and E. Tamer (2003): "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70, 519–546.

Molchanov, I. (2006): *Theory of Random Sets*, Springer.

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

——— (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80, 755–782.

Neal, R. (2003): "Slice sampling," *The Annals of Statistics*, 31, 705–741.

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

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

Ponomareva, M. and E. Tamer (2011): "Misspecification in moment inequality models: back to moment equalities?" *The Econometrics Journal*, 14, 186–203.

Rao, R. R. (1962): "Relations between weak and uniform convergence of measures with applications," *The Annals of Mathematical Statistics*, 659–680.

Rockafellar, R. and R.-B. Wets (2009): *Variational Analysis*, vol. 317 of *Grundlehren der mathematischen Wissenschaften*, Springer: Dordrecht.

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

——— (2010): "Inference for the identified set in partially identified econometric models," *Econometrica*, 78, 169–211.

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

Rubin, D. (1981): "The Bayesian bootstrap," *The Annals of Statistics*, 9, 130–134.

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

Shi, X. and M. Shum (2012): "Simple two-stage inference for a class of partially identified models," Working paper.

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

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

——— (2010): "Partial identification in econometrics," *Annual Review of Economics*, 2, 167–195.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

Van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press: Cambrige.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

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