Bayesian and Frequentist Inference in Partially Identified Models

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania
CEPR and NBER

September 7, 2011

---

*Correspondence: H.R. Moon: Department of Economics, KAP 300, University Park Campus, University of Southern California, Los Angeles, CA 90089. Email: hyungsikmoon@gmail.com. F. Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104. Email: schorf@ssc.upenn.edu. We thank seminar participants at the 2007 NASM, the 2008 SETA, Boston College, Duke, Johns Hopkins, Maryland, Ohio State, Rice, UC Davis, UC Irvine, UCLA, and Vanderbilt as well as Whitney Newey (co-editor) and three anonymous referees for helpful comments. We also thank Kyoungeun Kim, Minchul Shin, and Martin Weidner for research assistance. Moon gratefully acknowledges financial support from the National Science Foundation under Grant SES 0920903 and the USC Faculty Development Award. Schorfheide gratefully acknowledges financial support from the Alfred P. Sloan Foundation and the National Science Foundation under Grant SES 0617803. Gauss Programs that implement the numerical illustrations are available at *http://www.ssc.upenn.edu/˜schorf*.

Abstract

A large sample approximation of the posterior distribution of partially identified structural parameters is derived for models that can be indexed by an identifiable finite-dimensional reduced form parameter vector. It is used to analyze the differences between Bayesian credible sets and frequentist confidence sets in partially identified models. We define a plug-in estimator of the identified set and show that asymptotically Bayesian highest-posterior-density sets exclude parts of the estimated identified set, whereas it is well known that frequentist confidence sets extend beyond the boundaries of the estimated identified set. We recommend to report estimates of the identified set and information about the conditional prior along with Bayesian credible sets. A numerical illustration for a two-player entry game is provided.

JEL CLASSIFICATION: C11, C32, C33, C35

KEY WORDS: Bayesian Inference, Frequentist Confidence Sets, Partially Identified Models, Entry Games.

1 Introduction

In partially identified models one can only bound, but not point-identify the parameter vector of interest, θ. Such models arise in microeconometric as well as macroeconometric applications. Examples in the economics literature include models of censored sampling, models with missing observations or interval data (Manski and Tamer, 2002, and Manski, 2003), game-theoretic models with multiple equilibria such as entry games (Bresnahan and Reiss, 1991, Berry, 1994, Bajari, Benkard, and Levin, 2007, and Ciliberto and Tamer, 2009) and auctions (Haile and Tamer, 2003), vector autoregressions (VARs) that are identified by restricting only the sign of dynamic responses to structural shocks (Canova and De Nicolo, 2002, and Uhlig, 2005), and dynamic stochastic general equilibrium models (Lubik and Schorfheide, 2004). Given the lack of point identification researchers have focused on set estimators for the parameter of interest. While the macroeconometrics literature mostly applies Bayesian approaches, the microeconometric literature is dominated by frequentist procedures.

In regular point-identified models set estimates reported in the literature are often interpretable from both a Bayesian and a frequentist perspective. For instance, if y_i, $i = 1, \dots, n$ is a sequence of independent normal random variables with mean θ and unit variance, then the interval $[\hat{\theta}_n - 1.96/\sqrt{n},\ \hat{\theta}_n + 1.96/\sqrt{n}]$, where $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$ is both a valid 95% frequentist confidence set as well as a 95% credible set under a flat prior for θ. While the finite-sample equivalence breaks down in more complicated models, it often holds approximately as the sample size tends to infinity. This (approximate) equivalence is very convenient for the dissemination of applied research. A Bayesian reader of a frequentist paper will find the numerical results useful despite disagreements about inference methods and so will a frequentist reader of a Bayesian paper. We will show that the equivalence breaks down in the context of partially identified models.[1]

[1]Non-stationary time series models provide another prominent case in which the asymptotic equivalence does not hold, e.g. Sims and Uhlig (1991).

The contribution of this paper is to provide a formal comparison between frequentist confidence sets and Bayesian credible sets for partially identified parameters. We focus on highest-probability-density (HPD) sets, which have the smallest volume among all Bayesian credible sets and are widely used in the presentation of empirical results. For brevity, we refer to HPD sets constructed from the posterior (prior) density as posterior (prior) HPD sets. Starting point of our analysis is a likelihood function indexed by a finite-dimensional, identifiable reduced-form parameter vector ϕ. Conditional on ϕ the structural parameter of interest θ is known to lie in the identified set $\Theta(\phi)$. Using a large sample approximation of the posterior of θ we show that under some suitable regularity conditions the posterior HPD set of θ converges to the HPD set associated with the prior distribution of θ conditional on the maximum-likelihood (ML) estimate $\hat{\phi}_n$. Since this conditional prior distribution has necessarily support on $\Theta(\hat{\phi}_n)$, where $\Theta(\hat{\phi}_n)$ is the ML plug-in estimate of the identified set, we deduce that posterior HPD sets exclude portions of $\Theta(\hat{\phi}_n)$ asymptotically. Frequentist confidence sets based on $\hat{\phi}_n$ have to extend beyond the boundaries of $\Theta(\hat{\phi}_n)$ to ensure the desired coverage probabilities for all elements $\theta \in \Theta(\phi_0)$, where ϕ_0 denotes the "true" value of the reduced form parameter. Thus, in a sense that will be made precise later, credible sets for partially identified parameters tend to be smaller than confidence sets in large samples.

Our results have important implications for the interpretation of empirical findings reported in the literature. For instance, the credible bands reported in the macroeconometric literature for impulse responses in VARs identified with sign restrictions tend to be narrower than corresponding frequentist confidence bands. Vice versa, the frequentist confidence sets reported in the microeconometric literature appear overly conservative from a Bayesian perspective. While the authors of this paper have strong preferences in regard to Bayesian versus frequentist inference, the aim of the paper is not to advocate the use of one type of set estimates over another when analyzing partially identified models. Instead, the goal of the paper is to shed light on why Bayesian and frequentist set estimates are numerically different. However, the paper does have a normative message: Bayesian posteriors should be accompanied by estimates of the identified set, e.g., $\Theta(\hat{\phi}_n)$, as well as information about the prior of θ conditional on $\phi = \hat{\phi}_n$. Finally, a prior that is approximately uniform on the identified set for the parameter of interest might serve as a useful benchmark.

Earlier work on Bayesian analysis of partially identified models, most notably Poirier (1998), has focused on the characterization of finite-sample posteriors in specific econometric models.[2] We contribute to the Bayesian literature by providing a large sample approximation of the posterior distribution of a partially identified parameter θ, which is based on an insight that dates back at least to Kadane (1974): beliefs about the reduced form parameter ϕ are updated through the likelihood function, but the conditional distribution of θ given ϕ, which we denote by P^θ_ϕ, remains unchanged in view of new data. Since there exists a large literature dating back to Bernstein (1934), Le Cam (1953), and von Mieses (1965) that establishes the asymptotic normality of posterior distributions for finite-dimensional identifiable (reduced-form) parameters, we use this asymptotic normality of ϕ as the starting point for our analysis and approximate the posterior of θ as a mixture of conditional prior distributions P^θ_ϕ.

There is a rapidly growing literature on the construction of asymptotically valid frequentist confidence sets. Rather than providing a detailed review, we are highlighting some important developments. The literature distinguishes between confidence sets for the structural parameter θ and confidence sets for the identified set $\Theta(\phi)$. Our paper focuses on inference for θ because in many applications θ will be the more interesting object and it is the more natural object from a Bayesian perspective. Imbens and Manski (2004) study frequentist inference for an interval-identified parameter in a treatment model, in which the length of the identified set is estimable at a faster rate than the location. Stoye (2009) extends this analysis to a setting in which both length and location of the identified interval are estimated at the same rate. Chernozhukov, Hong, and Tamer (2007) consider inference

[2] After the first draft of this paper had been circulated Bollinger and van Hasselt (2009) derived finite-sample posteriors for a partially identified model with binary misclassification. Moreover, Liao and Jiang (2010) analyzed posterior distributions for moment inequality models based on limited-information likelihood functions. Finally, Kitagawa (2010) proposes to construct credible sets from posterior lower probabilities. The lower probability of a set is obtained by minimizing its posterior probability with respect to all priors of θ given ϕ (in our notation).

based on a generic criterion function $Q_n(\theta; Y^n) \geq 0$, where Y^n denotes the sample observations and the objective function has the property that $\text{plim}_{n\longrightarrow\infty} Q_n(\theta; Y^n) = 0$ whenever $\theta \in \Theta(\phi_0)$. Such an objective function arises, for instance, in the context of models defined by moment inequalities.

A key challenge in the frequentist literature is to find critical values $c_{\tau,n}(\theta)$ such that generalized level sets $C_F^\theta(Y^n) = \{\theta \mid Q_n(\theta; Y^n) \leq c^2_{\tau,n}(\theta)\}$ lead to asymptotically valid confidence sets. Researchers have considered fixed critical values as well as θ-specific critical values based on plug-in asymptotics, selection of binding moment conditions, sub-sampling, and bootstrap approaches, e.g., Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Bugni (2010), Romano and Shaikh (2010). To facilitate the comparison between Bayesian credible sets and frequentist confidence sets, we focus on confidence sets constructed from an objective function $Q_n(\theta; \hat{\phi}_n) \geq 0$ with the property that $Q_n(\theta; \hat{\phi}_n) = 0$ if and only if $\theta \in \Theta(\hat{\phi}_n)$. We restrict the data Y^n to enter the objective function only through the estimator $\hat{\phi}_n$, in our case the ML estimator, of the reduced form parameter. Thus, the frequentist procedures considered in this paper are based on a minimum-distance approach, which is appealing in problems with well-defined reduced-form parameters.[3]

The remainder of the paper is organized as follows. In Section 2 we derive finite-sample Bayesian credible intervals and frequentist confidence intervals for a simple moment inequality model with a scalar structural parameter. The structure of this example resembles the above-mentioned interval-identified treatment model of Imbens and Manski (2004). The large sample analysis is presented in Section 3 and Section 4 considers a two-player entry game to illustrate the theoretical results numerically. Finally, Section 5 concludes. The main proofs are collected in an Appendix to this paper. The remaining proofs as well as some derivations and computational details for the entry game estimation are provided in an Online Appendix.

Finally, a word on notation. We often use M to denote a generic finite constant and

[3]In Moon, Schorfheide, Granziera, and Lee (2009) we use the minimum-distance approach to construct frequentist error bands for a VAR identified with sign restrictions.

I to denote the identity matrix. The notation $\subseteq$ is used to denote weak inclusion and $\subset$ is used for strict inclusion. We use $A \backslash B$ to denote the set difference $\{x \mid x \in A, x \notin B\}$ and $A \ominus B$ to denote the symmetric difference $(A \backslash B) \cup (B \backslash A)$. When X is a matrix and W is a positive-definite weight matrix, let $\|X\|_W = (tr\,(WX'X))^{1/2}$. We use $\|X\|$ (without the W subscript) to denote the Euclidean norm. When P and Q are probability measures, then $\|P - Q\|$ denotes their total variation or L_1 distance. We use $N(\mu, \Sigma)$ to denote the multivariate normal distribution with mean μ and covariance matrix Σ. We let $\varphi_N(\cdot)$ and $\Phi_N(\cdot)$ denote the probability density (pdf) and cumulative density (cdf) functions of a vector of standard normal random variables. Moreover, we denote the one-sided critical value for a standard normal random variable by $z_\tau = |\Phi_N^{-1}(\tau)|$. $\mathcal{U}[a, b]$ signifies the uniform distribution on the interval $[a, b]$. We use P_b^a to denote the probability distribution of a random variable a conditional on the realization of a random variable b. $I\{x \in A\}$ is the indicator function that is equal to one if $x \in A$ and zero otherwise. "$\xrightarrow{\mathbb{P}}$" indicates convergence in probability, "$\Longrightarrow$" is convergence in distribution, and w.p.a. 1 abbreviates "with probability approaching one."

2 A Simple Example

A finite sample analysis of a simple partially identified model is conducted to illustrate that Bayesian credible sets and frequentist confidence sets are numerically different, even as the sample size tends to infinity. Consider the Gaussian location model $y_i = \phi + u_i$, where $u_i \sim iidN(0, 1)$. The log likelihood function for this model is quadratic and maximized at the sample mean $\hat{\phi}_n = \frac{1}{n}\sum_{i=1}^n y_i$. Under a flat prior $p(\phi) \propto c$ the Bayesian posterior distribution of $\sqrt{n}(\phi - \hat{\phi}_n)$ is $N(0, 1)$ and thereby identical to the sampling distribution of $\sqrt{n}(\hat{\phi}_n - \phi)$. As a consequence the Bayesian $1 - \tau$ credible interval $CS_B^\phi = \left[\hat{\phi}_n - z_{\tau/2}/\sqrt{n},\ \hat{\phi}_n + z_{\tau/2}/\sqrt{n}\right]$ is also a valid frequentist confidence interval.

Now suppose that the object of interest is a structural parameter θ that can be bounded based on ϕ as follows:

$$\phi \leq \theta \quad \text{and} \quad \theta \leq \phi + \lambda.$$

The interval $\Theta(\phi) = [\phi, \ \phi + \lambda]$ is called the identified set and in our simple example its length is known to be λ.[4] If we maintain the flat prior for ϕ and use $p(\theta|\phi)$ to denote the properly normalized prior density of θ conditional of ϕ with support $\Theta(\phi)$, then the joint distribution of data and parameters is given by

$$p(Y^n, \phi, \theta) \propto p(Y^n|\phi)p(\theta|\phi), \tag{1}$$

where $Y^n = [y_1, \dots, y_n]$. From integrating (1) with respect to θ we deduce that the marginal posterior of ϕ remains normal. It is also immediately apparent from $p(Y^n, \theta|\phi) \propto p(Y^n|\phi)p(\theta|\phi)$ that the data Y^n and the parameter vector θ are independent conditional on ϕ because θ does not enter the likelihood function. Consequently, the conditional distribution of θ given ϕ is not updated in view of the data: $p(\theta|Y^n, \phi) \propto p(\theta|\phi)$. Let $s = \sqrt{n}(\phi - \hat{\phi}_n)$. Then the marginal posterior density of θ can be expressed as the mixture

$$p(\theta|Y^n) = \int p(\theta|\hat{\phi}_n + n^{-1/2}s)\varphi_N(s)ds. \tag{2}$$

This equation suggests that as the sample size $n \longrightarrow \infty$ the posterior density of θ converges to $p(\theta|\hat{\phi}_n)$.

Now suppose that the prior for θ conditional on the reduced form parameter ϕ is uniform on the interval $\Theta(\phi)$:

$$p(\theta|\phi) = \frac{1}{\lambda} I\{\phi \leq \theta \leq \phi + \lambda\}.$$

It can be shown by direct calculation (see Online Appendix) that

$$p(\theta|Y^n) = \frac{1}{\lambda}\left[\Phi_N\left(\sqrt{n}(\theta - \hat{\phi}_n)\right) - \Phi_N\left(\sqrt{n}(\theta - \hat{\phi}_n - \lambda)\right)\right]. \tag{3}$$

[4]Our example resembles the treatment model in Imbens and Manski (2004). In their model $Y_i \in [0, 1]$ is a random outcome. Outcomes are only observed for individuals that received treatment as indicated by a binary variable $D_i \in \{0, 1\}$. The identifiable reduced form parameter can be defined as $\phi = \mathbb{E}[Y_i D_i]$ and the structural parameter of interest is $\theta = \mathbb{E}[Y_i]$. Under the assumption that the probability of treatment $p = \mathbb{E}[D_i]$ lies between zero and one and is known, the identified set is given by $\Theta(\phi) = [p\phi, \ p\phi + (1 - p)]$.

This density is depicted in Figure 1 for $\hat{\phi}_n = 0$, $n = 100$, and various choices of λ. If λ is small (relative to the sample size), the posterior density is approximately normal. For large values of λ the density resembles a step function. Theorem 1(ii) presented in Section 3 implies that for any fixed value of $\lambda > 0$

$$\int_\theta \left| p(\theta|Y^n) - \frac{1}{\lambda} I\{\hat{\phi}_n \leq \theta \leq \hat{\phi}_n + \lambda\} \right| d\theta = o_p(1). \tag{4}$$

Thus, the L_1 distance between the posterior density and the prior density $p(\theta|\hat{\phi}_n)$ on the estimated set $\Theta(\hat{\phi}_n)$ converges to zero.

The posterior density characterized by (3) is symmetric around $\hat{\phi}_n + \lambda/2$ and so are HPD sets constructed from this posterior. Since the posterior density is continuous, we can express the HPD set as

$$CS^\theta_{HPD}(Y^n) = \left[\hat{\phi}_n + \tau/2 - \eta_n,\ \hat{\phi}_n + \lambda - \tau/2 + \eta_n \right]. \tag{5}$$

Here η_n is chosen to guarantee that $P^\theta_{Y^n}\{\theta \in CS^\theta_{HPD}(Y^n)\} = 1 - \tau$. The convergence in (4) implies that

$$\left| P^\theta_{Y^n}\{\theta \in CS^\theta_{HPD}(Y^n)\} - P^\theta_{\hat{\phi}_n}\{\theta \in CS^\theta_{HPD}(Y^n)\} \right| = 2|\eta_n| \stackrel{p}{\longrightarrow} 0, \tag{6}$$

where $P^\theta_{\hat{\phi}_n}$ is the conditional prior distribution of θ given that $\phi = \hat{\phi}_n$. Thus, for n sufficiently large, the posterior HPD interval has to lie in the interior of the estimated set $\Theta(\hat{\phi}_n)$.

The frequentist analysis is markedly different. We can parameterize the correspondence as $\theta = \phi + \alpha$, where $\alpha \in [0, \lambda]$. Following Chernozhukov, Hong, and Tamer (2007) we construct a frequentist confidence set as a level set of the concentrated log inverse of the likelihood function. Define

$$\begin{aligned} Q_n(\theta; \hat{\phi}_n) &= \inf_{\alpha \in [0,\lambda]} -2[\ln p(Y^n|\theta + \alpha) - \ln p(Y^n|\hat{\phi}_n)] \qquad (7) \\ &= \begin{cases} n(\hat{\phi}_n - \theta)^2 & \text{if } \theta \leq \hat{\phi}_n \\ 0 & \text{if } \hat{\phi}_n < \theta < \hat{\phi}_n + \lambda \\ n(\hat{\phi}_n - \theta + \lambda)^2 & \text{if } \hat{\phi}_n + \lambda \leq \theta \end{cases} \end{aligned}$$

and let

$$CS_F^\theta(Y^n) = \left\{ \theta \;\middle|\; Q_n(\theta;\hat{\phi}_n) \le c_\tau^2 \right\} = \left[\hat{\phi}_n - c_\tau/\sqrt{n},\ \hat{\phi}_n + \lambda + c_\tau/\sqrt{n} \right]. \tag{8}$$

The finite-sample distribution of the maximum likelihood estimator is $\sqrt{n}(\hat{\phi}_n - \phi) \sim \mathcal{Z}$, where $\mathcal{Z} \sim N(0,1)$. It is convenient to re-scale θ according to $\vartheta = \sqrt{n}(\theta - \phi)$. In terms of the ϑ transform, the identified set $\Theta(\phi)$ is given by $0 \le \vartheta \le \sqrt{n}\lambda$. Using this reparameterization and the definition of the confidence in (8) we obtain the condition

$$\begin{aligned} \inf_{\phi\in\Phi} \inf_{\theta\in\Theta(\phi)} P_\phi^{Y^n}\{\theta \in CS_F^\theta(Y^n)\} &= \inf_{0\le\vartheta\le\sqrt{n}\lambda} P\{\vartheta - \sqrt{n}\lambda - c_\tau \le \mathcal{Z} \le \vartheta + c_\tau\} \\ &= \inf_{0\le\vartheta\le\sqrt{n}\lambda} \Phi_{\mathcal{N}}(\vartheta + c_\tau) - \Phi_{\mathcal{N}}(\vartheta - \sqrt{n}\lambda - c_\tau). \end{aligned}$$

Since the infimum is achieved for $\vartheta = 0$ and $\vartheta = \sqrt{n}\lambda$, the critical value $c_\tau > 0$ solves the following equation:

$$\Phi_N(\sqrt{n}\lambda + c_\tau) - \Phi_N(-c_\tau) = 1 - \tau. \tag{9}$$

As pointed out by Imbens and Manski (2004), if the re-scaled length of the identified set, $\sqrt{n}\lambda$, is large, then a $1-\tau$ confidence set for the parameter θ is obtained by expanding the boundaries of the interval $\Theta(\hat{\phi}_n)$ using a one sided critical value of a standard normal distribution.

A comparison of (5) and (8) indicates that the (numerical) difference between credible and confidence sets does not vanish asymptotically. The posterior HPD set $CS_{HPD}^\theta(Y^n)$ is eventually located strictly inside the estimated identified set $\Theta(\hat{\phi}_n) = [\hat{\phi}_n,\ \hat{\phi}_n + \lambda]$, while the frequentist confidence set $CS_F^\theta(Y^n)$ extends beyond the boundaries of $\Theta(\hat{\phi}_n)$. In this example we obtain the ordering

$$CS_{HPD}^\theta(Y^n) \subset \Theta(\hat{\phi}_n) \subset CS_F^\theta(Y^n) \quad \text{eventually.} \tag{10}$$

Suppose that data are generated from $P_{\phi_0}^{Y^n}$ and $\theta_0 = \phi_0$. Since (3) implies that the posterior probability $P_{Y^n}^\theta(\Theta(\hat{\phi}_n))$ converges to one, we deduce from (10) that the posterior probability of the confidence set tends to one

$$P_{Y^n}^\theta\{\theta \in CS_F^\theta(Y^n)\} \xrightarrow{\mathbb{P}} 1,$$

while the asymptotic coverage probability of the credible set is zero

$$\lim_{n\to\infty} \inf_{\phi\in\mathbb{R}} \inf_{\theta\in\Theta(\phi)} P_\phi^{Y^n}\{\theta \in CS^\theta_{HPD}(Y^n)\} = 0.$$

The frequentist result follows because the definition of $CS^\theta_{HPD}(Y^n)$ in (5) and the consistency of the estimator $\hat{\phi}_n \stackrel{\mathbb{P}}{\longrightarrow} \phi_0$, imply that values such as $\theta = \phi_0$ or $\theta = \phi_0 + \lambda$ will be excluded from the credible set w.p.a. 1. Thus, from the Bayesian perspective, the frequentist confidence set is too wide, while from the frequentist perspective, the Bayesian credible set is too narrow.

In Table 1 we compare 90% Bayesian credible sets and frequentist confidence intervals, computed based on (5) and (8), for various choices of n and the length of the identified set λ. Since in our example both the confidence interval as well as the credible interval are centered at $\hat{\phi}_n + \lambda/2$ we restrict the comparison to the length of the intervals and the posterior probability that θ falls into the frequentist confidence set. If the length of the identified set is small relative to the sample size, that is, θ is approximately point identified, then confidence and credible intervals are essentially numerically identical, that is they have the same length and and the posterior probability assigned to the interval $CS^\theta_F(Y^n)$ is approximately equal to the frequentist coverage probability of 90%. As the length of the identified interval increases or the uncertainty about ϕ decreases, the table illustrates that an 90% Bayesian credible set is shorter than an 90% frequentist confidence interval.

Now suppose the object of interest is the transformed parameter $\theta = f(\tilde{\theta})$, where $f(\cdot)$ is a strictly increasing and differentiable function. We use $f'(\cdot)$ to denote the first derivative of $f(\cdot)$ and $f^{-1}(\cdot)$ to denote its inverse. In this case the identified set becomes $\tilde{\Theta}(\phi) = [f^{-1}(\phi),\ f^{-1}(\phi+\lambda)]$ and the conditional prior density of $\tilde{\theta}$ given ϕ takes the form

$$p(\tilde{\theta}|\phi) = \frac{1}{\lambda} I\{f^{-1}(\phi) \le \tilde{\theta} \le f^{-1}(\phi+\lambda)\}|f'(\tilde{\theta})|.$$

Thus, the transformation induces a change of the prior distribution. The prior is no longer uniform if the transformation is nonlinear. If in a neighborhood of $\hat{\phi}_n$ the density $p(\tilde{\theta}|\phi)$ peaks at the boundary of the identified set $\tilde{\Theta}(\hat{\phi}_n)$, then it is possible that in any finite sample the posterior HPD set contains parameter values that lie outside of the set $\tilde{\Theta}(\hat{\phi}_n)$, albeit close to its boundary. Thus, the strict inclusion $CS^{\theta}_{B}(Y^n) \subset \Theta(\hat{\phi}_n)$ in (10) only obtains under specific prior distributions. However, we will show subsequently that the posterior probability of the symmetric difference between the finite-sample posterior HPD set as well as the prior HPD set constructed from $p(\theta|\hat{\phi}_n)$ converges to zero. The latter set is by construction a strict subset of $\Theta(\hat{\phi}_n)$.

3 Large Sample Approximations

We now generalize the analysis presented in the previous section. We assume that the joint prior distribution for the $m \times 1$ reduced form parameter vector ϕ and the $k \times 1$ structural parameter vector θ can be decomposed into a marginal distribution for $\phi \in \Phi$, denoted by P^{ϕ}, and a conditional distribution P^{θ}_{ϕ} of θ. The conditional distribution has support on the identified set $\Theta(\phi)$. Let $l_n(\phi)$ denote the log likelihood function $\ln p(Y^n|\phi)$. Kadane (1974) emphasized that the derivation of the posterior distribution can be done on the space of the reduced form parameter ϕ. For any measurable set $A \subseteq \Theta$:

$$P^{\theta}_{Y^n}(A) = \int_{\Phi} \left[\int_{\Theta(\phi)} I\{\theta \in A\} dP^{\theta}_{\phi} \right] \frac{\exp[l_n(\phi)]}{\int_{\Phi} \exp[l_n(\phi)] dP^{\phi}} dP^{\phi} = \int_{\Phi} P^{\theta}_{\phi}(A) dP^{\phi}_{Y^n}. \tag{11}$$

Since conditional on ϕ the structural parameter θ does not enter the likelihood function the prior distribution of θ given ϕ, P^{θ}_{ϕ}, is not updated in view of the data Y^n. This point also had been stressed by Poirier (1998).

The goal of this section is to provide an insightful large sample approximations of the marginal posterior distribution $P^{\theta}_{Y^n}$ and Bayesian HPD sets. Approximations of posterior distributions are stated in terms of the L_1 distance between probability measures,[5] which allows us to establish convergence results for HPD sets. In Section 3.1 $P^{\phi}_{Y^n}$ in (11) is replaced by a normal limit distribution. Under the assumption that the conditional prior distribution

[5]The L_1 distance between to probability measures can be defined as $\|P - Q\| = \sup_{|f|\leq 1} \left| \int f dP - \int f dQ \right|$, where $\sup_{|f|\leq 1}$ denotes the supremum over all functions f that are bounded by one in absolute value.

P^θ_ϕ is Lipschitz in ϕ, we show in Section 3.2 that the posterior of θ converges to $P^\theta_{\hat{\phi}_n}$ or to $P^\theta_{\phi_0}$ if $\hat{\phi}_n$ has a probability limit ϕ_0. We demonstrate in Section 3.3 that the posterior probability of the symmetric difference between finite-sample posterior HPD sets and HPD sets constructed from the distribution $P^\theta_{\hat{\phi}_n}$ converges to zero. Finally, it is shown in Section 3.4 that the posterior probability of the difference $CS^\theta_F(Y^n)\backslash CS^\theta_{HPD}(Y^n)$ is bounded below by τ, while the posterior probability of $CS^\theta_{HPD}(Y^n)\backslash CS^\theta_F(Y^n)$ tends to zero. In this sense, confidence sets in partially identified models are larger than credible sets.

3.1 First Approximation: $p(\phi|Y^n)$

There exists a large literature on the asymptotic normality of posterior distributions in identified models dating back to Bernstein (1934), Le Cam (1953), and von Mieses (1965). Since the goal of our paper is not to make an independent contribution to this literature, we will state the asymptotic normality of $P^\phi_{Y^n}$ as an assumption rather than derive it from more low-level conditions on the likelihood function.[6] Assuming that the log-likelihood function is twice continuously differentiable with respect to ϕ, define

$$\hat{J}_n = D_n^{-1}\left[-\frac{\partial^2 l_n(\phi)}{\partial\phi\partial\phi'}\right]_{\phi=\hat{\phi}_n} D_n^{-1'} \quad \text{and} \quad s = \hat{J}_n^{1/2} D_n(\phi - \hat{\phi}_n).$$

The matrix $\hat{J}_n^{-1}$ plays the role of an asymptotic posterior covariance matrix of ϕ. The matrix D_n is deterministic with elements that are diverging as $n \longrightarrow \infty$. It is chosen to ensure that $\hat{J}_n$ is convergent. The posterior distribution of the transformed reduced form parameter s will be denoted by $P^s_{Y^n}$.

Assumption 1 *Let $Y^n(\omega)$ be a sequence of random vectors defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (i) The sequence of maximum likelihood estimators is convergent: $\hat{\phi}_n \xrightarrow{\mathbb{P}} \phi_0$. The*

[6] Fundamental conditions can be found, for instance, in Johnson (1970) and in the textbook treatments of Hartigan (1983, Section 11.2), van der Vaart (1998, Theorem 10.1), or Ghosh and Ramamoorthi (2003, Theorem 1.4.2). Extensions to time series models are provided by Kim (1998) and Phillips and Ploberger (1998).

matrix $\|D_n\| \uparrow \infty$. *The likelihood function* $l_n(\phi)$ *is twice continuously differentiable w.p.a. 1, such that* $\hat{J}_n$ *is well defined. Moreover, the Hessian of the log-likelihood function has a positive definite probability limit:* $\hat{J}_n \Longrightarrow J_0 > 0$ *and* $\hat{J}_n^{-1} \Longrightarrow J_0^{-1}$. *(ii) The posterior distribution of* ϕ *is asymptotically normal:*

$$\|P^s_{Y^n} - N(0, I)\| \xrightarrow{\mathbb{P}} 0,$$

where $s = \hat{J}_n^{1/2} D_n(\phi - \hat{\phi}_n)$.

Assumption 1(ii) states the asymptotic normality in terms of L_1 convergence of probability measures. Conditions for L_1 convergence are provided in van der Vaart (1998) and Gosh and Ramamoorthi (2003). Not all of the assumptions stated in Part (i) need to be satisfied in order to obtain a posterior distribution that is asymptotically normal. For instance, it could be the case that the maximum likelihood estimator is not convergent under $\mathbb{P}$, but the posterior distribution is still asymptotically normal around $\hat{\phi}_n$. However, the low-level assumptions under which the asymptotic normality is proved, typically are consistent with the assumptions in Part (i). Assuming the convergence of $\hat{\phi}_n$ allows us to state conditions on the conditional prior P^θ_ϕ locally, in a neighborhood of ϕ_0. While in many models the sequence of Hessian matrices $\hat{J}_n$ converges in probability to a constant matrix, non-stationary time-series models provide an important class of models for which $\hat{J}_n$ converges to a stochastic limit.

In view of the large literature on extremum estimators based on non-differentiable objective functions, the assumption of a twice differentiable likelihood function might appear overly restrictive. While it rules out, for instance, likelihood functions constructed from densities with jumps, densities with singularities, or change-point problems, we note that most of the applied Bayesian analysis tends to be based on differentiable likelihood functions. When the likelihood function is only first-order stochastically differentiable, Bayesian posterior asymptotics require an alternative definition of $\hat{J}_n$ and are provided, for example, in van der Vaart (1998) and Chernozhukov and Hong (2003). Further large-sample approximations for Bayesian posteriors in irregular models can be found, for instance, in Ghosh, Ghosal, and Samanta (1994, 1995). In these models one can often establish that the posterior of a properly centered and scaled parameter converges to the limit of a suitably defined likelihood ratio process, which might lead to non-Gaussian approximations.

3.2 Second Approximation: $p(\theta|Y^n)$

The object of interest in this paper is the posterior distribution of the structural parameter θ given by (11). We begin our analysis by replacing the posterior distribution $P^{\phi}_{Y^n}$ in (11) with its asymptotic normal distribution $N(0, I)$. For any measurable set $A \subseteq \Theta$ let

$$P^{\theta}_{N,Y^n}(A) = \int_{\mathbb{R}^m} P^{\theta}_{\hat{\phi}_n + D_n^{-1}\hat{J}_n^{-1/2}s}(A)\, dN\big(0, I\big)(s). \tag{12}$$

It can be easily verified that P^{θ}_{N,Y^n} provides a large sample approximation of $P^{\theta}_{Y^n}$. For any measurable real valued function $f(\theta)$ such that $|f| \leq 1$, one can define the function $g(s) = \int f(\theta) dP^{\theta}_{\hat{\phi}_n + D_n^{-1}\hat{J}_n^{-1/2}s}$ where $|g(s)| \leq 1$. Thus,

$$\left| \int f(\theta) dP^{\theta}_{Y^n} - \int f(\theta) dP^{\theta}_{N,Y^n} \right| = \left| \int g(s) dP^{s}_{Y^n} - \int g(s) dN\big(0, I\big)(s) \right| \tag{13}$$

According to Assumption 1(ii) and the definition of L_1 convergence, the right-hand-side of (13) converges in probability to zero for every function $|g(s)| \leq 1$ and we can deduce that

$$\|P^{\theta}_{Y^n} - P^{\theta}_{N,Y^n}\| \xrightarrow{\mathbb{P}} 0. \tag{14}$$

While the distribution P^{θ}_{N,Y^n} provides a valid large sample approximation of the posterior, we shall construct a second approximation that is more insightful in regard to comparisons between Bayesian credible sets and frequentist confidence sets. This second approximation requires an additional assumption, which we will discuss in more detail after stating the main result.

Assumption 2 *Let $N_\delta(\phi_0) = \{\phi \in \Phi \mid \|\phi - \phi_0\| < \delta\}$. There exists a $\delta > 0$ and a constant $M(\phi_0, \delta)$ such that $\|P^{\theta}_{\phi_1} - P^{\theta}_{\phi_2}\| \leq M(\phi_0, \delta)\|\phi_1 - \phi_2\|$ for $\phi_1, \phi_2 \in N_\delta(\phi_0)$.*

Theorem 1 *Let $s = \hat{J}_n^{1/2} D_n(\phi - \hat{\phi}_n)$. (i) If Assumption 1 is satisfied*

$$\|P^\theta_{Y^n} - P^\theta_{N,Y^n}\| \stackrel{\mathbb{P}}{\longrightarrow} 0.$$

(ii) If Assumptions 1 and 2 are satisfied, then

$$\|P^\theta_{Y^n} - P^\theta_{\hat{\phi}_n}\| \stackrel{\mathbb{P}}{\longrightarrow} 0 \quad \text{and} \quad \|P^\theta_{Y^n} - P^\theta_{\phi_0}\| \stackrel{\mathbb{P}}{\longrightarrow} 0,$$

where $P^\theta_{\hat{\phi}_n}$ ($P^\theta_{\phi_0}$) denotes the conditional prior P^θ_ϕ evaluated at $\phi = \hat{\phi}_n$ ($\phi = \phi_0$).

The proof of Part (i) was provided above. Part (ii) states that the posterior distribution of θ can be approximated either by the conditional prior distribution P^θ_ϕ evaluated at the maximum likelihood estimate $\hat{\phi}_n$ or at its probability limit ϕ_0. A proof is presented in the Appendix.

Remark 1: Consider an interval-identified location with $m = 1$, $k = 1$, $\Theta(\phi) = [\phi, \phi + \lambda]$ and a prior density of the form

$$p(\theta|\phi) = \frac{1}{\lambda} f\left(\frac{\theta - \phi}{\lambda}\right), \tag{15}$$

where $\int f(x)dx = 1$ and $f(x) = 0$ outside of the unit interval. Here $\lambda > 0$ is the known length of the identified set. Moreover, assume that $f(x)$ is differentiable with uniformly bounded first derivative $|df/dx| < M_1$ for $0 \leq x \leq 1$. Our assumptions imply that the function $f(x)$ can be bounded by a constant M_0. This model provides a slight generalization of the example studied in Section 2. We begin by verifying Assumption 2. Without loss of generality, let $\phi_1 \leq \phi_2$. For $\phi_2 \geq \phi_1 + \lambda$ we obtain the bound[7]

$$\int |p(\theta|\phi_1) - p(\theta|\phi_2)| d\theta = 2 \leq 2\frac{\|\phi_1 - \phi_2\|}{\lambda}. \tag{16}$$

For $\phi_2 < \phi_1 + \lambda$ we obtain

$$|p(\theta|\phi_1) - p(\theta|\phi_2)| = \begin{cases} p(\theta|\phi_1) & \text{if } \phi_1 \leq \theta \leq \phi_2 \\ |p(\theta|\phi_1) - p(\theta|\phi_2)| & \text{if } \phi_2 \leq \theta \leq \phi_1 + \lambda \\ p(\theta|\phi_2) & \text{if } \phi_1 + \lambda \leq \theta \leq \phi_2 + \lambda \\ 0 & \text{otherwise} \end{cases}$$

[7]If P and Q have densities $p(\omega)$ and $q(\omega)$ with respect to Lebesgue measure, then $\|P - Q\| = \int |p(\omega) - q(\omega)| d\omega$.

Both $p(\theta|\phi_1)$ and $p(\theta|\phi_2)$ are bounded by M_0. On the set $\phi_2 \leq \theta \leq \phi_1 + \lambda$ a first-order Taylor expansion of the prior density implies that

$$|p(\theta|\phi_1) - p(\theta|\phi_2)| \leq \frac{M_1}{\lambda^2}\|\phi_1 - \phi_2\|.$$

Thus, integrating over θ we deduce

$$\int |p(\theta|\phi_1) - p(\theta|\phi_2)| d\theta \leq 2M_0\|\phi_1 - \phi_2\| + \frac{M_1}{\lambda}\|\phi_1 - \phi_2\|. \quad (17)$$

To obtain a bound for the integral over the range $\phi_2 \leq \theta \leq \phi_1 + \lambda$ we used the fact that $\phi_1 \leq \phi_2$, which implies that $(\phi_1 + \lambda - \phi_2)/\lambda \leq 1$. Combining (16) and (17) by setting $M = \max\{2/\lambda,\ 2M_0 + M_1/\lambda\}$ provides a verification of Assumption 2. ■

Remark 2: If the length λ of the identified set is zero or the continuous prior distribution on the identified set with length $\lambda > 0$ is replaced by a point mass, the Lipschitz condition will fail. Let P^θ_ϕ be the probability measure that assigns mass one to $\theta = \phi + \tau$, where $0 \leq \tau \leq \lambda$. Then for every $\phi_1 \neq \phi_2$

$$\|P^\theta_{\phi_1} - P^\theta_{\phi_2}\| \geq |P^\theta_{\phi_1}(\{\phi_1 + \tau\}) - P^\theta_{\phi_2}(\{\phi_1 + \tau\})| = 1.$$

Alternatively, if one considers a sequence of models with $\Theta_n(\phi) = [\phi,\ \phi + \lambda_n]$ and priors

$$p_n(\theta|\phi) = \frac{1}{\lambda_n} f\left(\frac{\theta - \phi}{\lambda_n}\right)$$

where the length of the identified $\lambda_n \longrightarrow 0$ fast enough such that $D_n\lambda_n \longrightarrow 0$, it can be verified by direct calculation (see Online Appendix) that

$$\|P^\theta_{Y^n} - N(\hat{\phi}_n, D_n^{-2}\hat{J}_n^{-1})\| \overset{\mathbb{P}}{\longrightarrow} 0. \quad (18)$$

The approximation in Theorem 1(ii) implies that the uncertainty about the reduced-form parameter ϕ is asymptotically negligible and the shape of the posterior is determined by the conditional prior $P^\theta_{\hat{\phi}_n}$. However, this conclusion only holds, if the identified set is large relative to the uncertainty about its location. If the size of the identified set is small, the uncertainty about ϕ dominates the posterior and the shape of P^θ_ϕ is essentially irrelevant as in (18). This result is consistent with the simulations presented in Section 2. According to Table 1 for $\lambda = 0.01$ the length of the credible interval for $n = 100$ ($n = 500$) is 0.33 (0.15), which corresponds to $2z_{0.05}/\sqrt{n}$, that is an interval constructed from a $N(\hat{\phi}_n, 1/n)$ distribution. For $\lambda = 10$, on the other hand, the posterior credible interval has length 9.0 and corresponds to an interval with 90% probability under a $\mathcal{U}[\hat{\phi}_n, \hat{\phi}_n + 10]$ distribution, which is $P^{\theta}_{\hat{\phi}_n}$.

A similar phenomenon arises from a frequentist perspective. If the identified set is small in the sense that $D_n \lambda_n$ converges to zero fast enough, then the estimated set, which collapses to $\hat{\phi}_n$ asymptotically, needs to be expanded by two-sided critical values to obtain a valid confidence set. In this case Bayesian credible sets and frequentist confidence sets are approximately the same in large enough samples. This can also be seen from the third column (with header "0.01") in Table 1. If, on the other hand, the size of the identified set is asymptotically large, e.g. $\lambda_n = \lambda > 0$, then it suffices to expand the estimated identified set with one-sided critical values to obtain a valid confidence set, as shown in Imbens and Manski (2004). Stoye (2004) studies frequentist confidence intervals for the case $D_n \lambda_n \longrightarrow \lambda$. According to Theorem 1(i) an asymptotic approximation of the posterior for Stoye's case is given by the more complicated mixture $P^{\theta}_{N,Y}$. ■

Remark 3: In many applications, including the entry game discussed in Section 4, the identified set $\Theta(\phi)$ lies in a lower dimensional subspace of Θ. This automatically leads to a violation of Assumption 2 with respect to the joint distribution of the elements of θ. Consider, for instance, the following modification of the interval-identified location model in Section 2. We introduce a second structural parameter θ_2, defined as $\theta_2 = \theta_1 - \phi$. The identified set for $\theta = [\theta_1, \theta_2]'$ is $\Theta(\phi) = \big\{(\theta_1, \theta_2) \in \mathbb{R}^2 \mid \phi \leq \theta_1 \leq \phi + \lambda,\ \theta_2 = \theta_1 - \phi\big\}$. Moreover, we equip $\Theta(\phi)$ with the following prior distribution: $P^{\theta_1}_{\phi}$ is characterized by the density function (15) and $P^{\theta_2}_{(\phi,\theta_1)}$ is a point mass at $\theta_2 = \theta_1 - \phi$. By construction the intersection of $\Theta(\phi_1)$ and $\Theta(\phi_2)$ for $\phi_1 \neq \phi_2$ is empty. Thus, for our choice of P^{θ}_{ϕ} as well as any other prior of θ given ϕ the L_1 distance between measures $P^{\theta}_{\phi_1}$ and $P^{\theta}_{\phi_2}$ is 2 if $\phi_1 \neq \phi_2$ and the Lipschitz condition is violated.

Now consider inference for the element θ_1 of the parameter vector θ. Given our choice of $P_\phi^{\theta_1}$, the calculations in Remark 1 imply that Assumption 2 is satisfied if θ is replaced by (the subvector) θ_1. The change of variables $\theta_2 = \theta_1 - \phi$ can be used to verify that the Lipschitz condition is also satisfied for $P_\phi^{\theta_2}$, which has support on $\Theta_2(\phi) = [0, \lambda]$. Thus, in the context of this example the approximations in Theorem 1 are valid for the marginal posterior distributions of θ_1 and θ_2. More generally, if the object of interest is the marginal posterior of a subvector θ_i of θ, then Assumption 2 only needs to be verified for $P_\phi^{\theta_i}$.[8] ■

3.3 Posterior HPD Sets for θ

We now examine the large sample behavior of posterior HPD sets. It is assumed that the prior distributions P_ϕ^θ, the posterior distribution $P_{Y^n}^\theta$, and its large sample approximations P_{N,Y^n}^θ and $P_{\hat{\phi}_n}^\theta$ have densities $p(\theta|\phi)$, $p(\theta|Y^n)$, $p_N(\theta|Y^n)$, and $p(\theta|\hat{\phi}_n)$ with respect to Lebesgue measure in $\mathbb{R}^k$. If the object of inference is a subvector θ_i of θ, then the joint densities $p(\theta|Y^n)$ and $p(\theta|\phi)$ can simply be replaced by marginal densities $p(\theta_i|Y^n)$ and $p(\theta_i|\phi)$, respectively. Define the finite-sample HPD set as

$$CS_{HPD}^\theta(Y^n) = \{\theta \mid p(\theta|Y^n) \geq \kappa_{Y^n}\},$$

where the cut-off κ_{Y^n} is chosen to ensure that $P_{Y^n}^\theta(CS_{HPD}^\theta(Y^n)) = 1 - \tau$. We begin with the case in which for each ϕ the prior density allows the construction of a unique prior HPD set, defined as

$$CS_{HPD}^\theta(\phi) = \{\theta \mid p(\theta|\phi) \geq \kappa_\phi\}$$

with $P_\phi^\theta(CS_{HPD}^\theta(\phi)) = 1 - \tau$. At the end of this subsection we discuss the special case of conditional prior distributions that are uniform on $\Theta(\phi)$.

[8] If Assumption 2 is satisfied for the entire vector θ, it is also satisfied for any subvector. If the Lipschitz condition is not satisfied for the entire vector, the condition is *not automatically* satisfied for subvectors with a dimension that is equal to or smaller than the dimension of the subspace that contains $\Theta(\phi)$. For instance, the projection of $\Theta(\phi)$ onto the domain of a subvector θ_i could have Lebesgue measure zero.

The following theorem states that the posterior probability of the symmetric differences $CS^{\theta}_{HPD}(Y^n) \ominus CS^{\theta}_{HPD}(\phi_0)$ and $CS^{\theta}_{HPD}(Y^n) \ominus CS^{\theta}_{HPD}(\hat{\phi}_n)$ converges to zero. We adopt this notion of convergence of a sequence of finite sample credible sets from Severini (1991, Page 613).

Assumption 3 *At $\phi = \phi_0$ the prior density $p(\theta|\phi)$ leads to a unique $1-\tau$ HPD set with a threshold $\kappa_\phi < \infty$ and $\int I\{\, p(\theta|\phi) = \kappa_\phi\} p(\theta|\phi) d\theta = 0$.*

Theorem 2 *Suppose Assumptions 1, 2, and 3 are satisfied. Then (i):*

$$P^{\theta}_{Y^n}\left(CS^{\theta}_{HPD}(Y^n) \ominus CS^{\theta}_{HPD}(\phi_0) \right) \xrightarrow{\mathbb{P}} 0$$

and (ii)

$$P^{\theta}_{Y^n}\left(CS^{\theta}_{HPD}(Y^n) \ominus CS^{\theta}_{HPD}(\hat{\phi}_n) \right) \xrightarrow{\mathbb{P}} 0.$$

A proof of the theorem can be found in the appendix. In view of Theorem 1 the distribution $P^{\theta}_{Y^n}$ can also be replaced by $P^{\theta}_{\hat{\phi}_n}$ or $P^{\theta}_{\phi_0}$. Theorem 2 generalizes the results obtained in the context of the simple example in Section 2. Since by construction $CS^{\theta}_{HPD}(\hat{\phi}) \subset \Theta(\phi)$, the posterior mass of the subsets of the posterior HPD set that are not located inside of $\Theta(\phi_0)$ and $\Theta(\hat{\phi}_n)$, respectively, tends to zero. Moreover, there exist subsets of $\Theta(\phi_0)$ and $\Theta(\hat{\phi}_n)$ that are asymptotically excluded from the Bayesian credible set.

If the prior P^{θ}_{ϕ} is uniform, then there does not exist a unique prior HPD set. However, in any finite sample, the posterior HPD set is uniquely defined as long as the posterior density

$$p(\theta|Y^n) = \int \frac{I\{\theta \in \Theta(\phi)\}}{\int I\{\theta \in \Theta(\phi)\} d\theta} p(\phi|Y^n) d\phi$$

is not flat near the cutoff κ_{Y^n}. This is the case in the example of Section 2. We can deduce from Theorem 1 that if $CS^{\theta}_{B}(\hat{\phi}_n)$ is a $1-\tau$ credible set under the conditional prior distribution $P^{\theta}_{\hat{\phi}_n}$, the posterior probability $P^{\theta}_{Y^n}(CS^{\theta}_{B}(\hat{\phi}_n)) \xrightarrow{\mathbb{P}} 1-\tau$. Moreover, since the posterior distribution asymptotically concentrates on $\Theta(\hat{\phi}_n)$, it follows directly from Theorem 1 (see also Corollary 1 below) that the probability of the subset of the posterior HPD set $CS^{\theta}_{HPD}(Y^n)$ that is not located inside of the estimated identified set, $P^{\theta}_{Y^n}(CS^{\theta}_{HPD}(Y^n) \backslash \Theta(\hat{\phi}_n))$, tends to zero.

3.4 Comparison to Frequentist Confidence Sets

Since we model partial identification through the identified set correspondence $\Theta(\phi)$, we can relate the structural parameter of interest θ and the identified reduced form parameter ϕ via nuisance parameters, say α, $\phi = G(\theta, \alpha)$, where $\alpha \in \mathcal{A}_\theta$:

$$\theta \in \Theta(\phi) \quad \text{iff} \quad \exists\, \alpha \in \mathcal{A}_\theta \text{ such that } \phi = G(\theta, \alpha). \tag{19}$$

Suppose that $\hat{\phi}_n$ is a consistent and asymptotically normal estimator of ϕ. In this case a natural objective function for frequentist inference is the minimum distance criterion function

$$Q_n(\theta; \hat{\phi}_n) = \min_{\alpha \in \mathcal{A}_\theta} n \big\| \hat{\phi}_n - G(\theta, \alpha) \big\|_{W_n} \geq 0, \tag{20}$$

where $\{W_n\}$ is a sequence of positive definite weight matrices. Now suppose that we consider frequentist confidence sets of the form

$$CS_F^\theta(Y^n) = \left\{ \theta \;\middle|\; Q_n(\theta; \hat{\phi}_n) \leq c_n^2(\theta) \right\}. \tag{21}$$

For CS_F^θ to be a confidence set that is asymptotically valid the following condition has to be satisfied

$$\lim_{n \longrightarrow \infty} \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} P_\phi^{Y^n} \{ Q_n(\theta; \hat{\phi}_n) \leq c_n^2(\theta) \} \geq 1 - \tau. \tag{22}$$

Constructing sequences of critical value functions such that (22) holds is the subject of a rapidly growing literature, e.g., Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Rosen (2008), Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), and Bugni (2010). The goal of this literature is to sharpen the confidence set to achieve the desired coverage probability in (22). However, regardless of the construction of $c_n(\theta)$, a key property of the frequentist objective function is that $Q_n(\theta; \hat{\phi}_n) = 0$ if $\theta \in \Theta(\hat{\phi}_n)$, which follows from (19). This implies $\Theta(\hat{\phi}_n) \subseteq CS_F^\theta(Y^n)$ and leads to the following corollary.

Corollary 1 *Suppose Assumptions 1 and 2 are satisfied and the frequentist confidence set takes the form (21). Then (i),*

$$P_{Y^n}^\theta \Big(CS_{HPD}^\theta(Y^n) \backslash CS_F^\theta(Y^n) \Big) \xrightarrow{\mathbb{P}} 0$$

and (ii)

$$P^{\theta}_{Y^n}\left(CS^{\theta}_{HPD}(Y^n)\backslash CS^{\theta}_F(Y^n)\right) \geq P^{\theta}_{Y^n}\left(\Theta(\hat{\phi}_n)\right) - (1-\tau) \stackrel{\mathbb{P}}{\longrightarrow} \tau.$$

Corollary 1 is a consequence of Theorem 1 and formalizes the sense in which the Bayesian credible set is asymptotically "contained" in the frequentist confidence set. In view of the corollary, we conclude that the frequentist confidence set $CS^{\theta}_F(Y^n)$ is too large from a Bayesian perspective.

4 Numerical Illustration: A Two-Player Entry Game

To provide a second numerical illustration of Bayesian credible sets and frequentist confidence sets, we consider an example that has received a lot of attention in the microeconometric literature on partially identified models: a two-player entry game, see for instance Bresnahan and Reiss (1991), Berry (1994), Tamer (2003), and Ciliberto and Tamer (2009). Depending on the entry decision of the competing firm, Firm j either does not enter market i, operates as monopolist, or operates as duopolist. Potential monopoly (M) and duopoly (D) profits are given by

$$\pi^M_{i,j} = \beta_j + \epsilon_{i,j}, \quad \pi^D_{i,j} = \beta_j - \gamma_j + \epsilon_{i,j}, \quad j = 1,2, \quad i = 1,\ldots,n. \tag{23}$$

The $\epsilon_{i,j}$'s capture latent profit components that are known to the two firms but unobserved by the econometrician. We restrict our attention to pure strategy Nash equilibria. Non-uniqueness arises if both firms are profitable as monopolists but not as duopolists. The econometrician observes which firm(s) enter each of the n markets and we use n_{11}, n_{00}, n_{10}, n_{01} to denote the frequency with which the four possible market configurations are observed.

Assume that $\epsilon_{i,j} \sim iidN(0,1)$ and define the vector of structural parameters $\theta = [\beta_1, \gamma_1, \beta_2, \gamma_2]'$, where $\gamma_1, \gamma_2 \geq 0$. It follows from (23) that the probabilities that firm j in market i is profitable as monopolist and duopolist, respectively, are given by $m_j(\theta) = \Phi_N(\beta_j)$ and $d_j(\theta) = \Phi_N(\beta_j - \gamma_j)$. Moreover, we use $\phi = [\phi_{11},\ \phi_{00},\ \phi_{10}]'$ to denote the non-redundant reduced form probabilities of observing a monopoly, no entry, or the entry of Firm 1. The relationship between ϕ and θ implied by the pure strategy Nash equilibrium concept is given by the following set of equalities and inequalities which defines the identified set $\Theta(\phi)$:

$$\phi_{11} = d_1(\theta)d_2(\theta), \quad \phi_{00} = (1 - m_1(\theta))(1 - m_2(\theta)) \tag{24}$$

$$m_1(\theta)(1 - m_2(\theta)) + d_1(\theta)(m_2(\theta) - d_2(\theta)) \leq \phi_{10} \leq m_1(\theta)(1 - d_2(\theta)). \tag{25}$$

The relationships between the reduced form probabilities ϕ_{11} and ϕ_{00} and the structural parameter θ imply that β_2 and γ_2 are uniquely determined conditional on ϕ, β_1, and γ_1. Thus, $\Theta(\phi)$ lies in a 2-dimensional subspace of $\mathbb{R}^4$ (see Remark 3 in Section 3.2). In Section 4.1 we characterize the identified set in terms of a functional relationship between ϕ, θ, and an auxiliary parameter. Section 4.2 discusses the priors for Bayesian inference as well as the objective function for the frequentist confidence sets. In Section 4.3 we generate two samples, one of size $n = 100$ and one of size $n = 1,000$ and compute 60% and 90% credible and confidence sets as well as estimates of the identified set. Computational details are relegated to the Online Appendix.

4.1 Identified Set Correspondence and Data Generation

The correspondence between θ and ϕ can be represented by the relationship $\phi = \tilde{G}(\theta, \psi)$ where $\psi \in [0, 1]$ and the two inequalities (25) are replaced by (omitting the θ-arguments):

$$\phi_{10} = m_1(1 - m_2) + d_1(m_2 - d_2) + \psi(m_1 - d_1)(m_2 - d_2). \tag{26}$$

The last term, which is pre-multiplied by ψ, is the probability that both firms are profitable as monopolists but not as duopolists. Consequentially, the auxiliary parameter ψ can be interpreted as the probability of a sunspot shock that selects Firm 1 if the Nash equilibrium is not unique. This representation is convenient for data generation. Parameter values of our data generating process for θ and ψ are reported in the second column of Table 2. We assume that Firm 1 is slightly more profitable than Firm 2, that is $\beta_1 > \beta_2$, and more likely to enter a market that can only sustain a single monopolist, $\psi > 0.5$. Using (24) and (26) we calculate the associated reduced form probabilities ϕ, which are also reported in the table. The probabilities of a Firm 1 monopoly, and Firm 2 monopoly, and a duopoly are 48%, 33%, and 12%, respectively.

In order to implement frequentist inference the following characterization of the relationship between ϕ and θ in (26) is more convenient: $\phi_{10} = m_1(1-d_2) - \alpha$, where $0 \leq \alpha \leq \bar{\alpha}(\theta)$ and $\bar{\alpha}(\theta) = (m_1 - d_1)(m_2 - d_2)$. In turn, we can define the non-negative function

$$Q(\theta;\phi) = \min_{0 \leq \alpha \leq \bar{\alpha}(\theta)} \big\| \phi - G(\theta,\alpha) \big\|, \tag{27}$$

where

$$G(\theta,\alpha) = \big[d_1 d_2,\ (1-m_1)(1-m_2),\ m_1(1-d_2) \big]' - \big[0,\ 0,\ \alpha \big]'.$$

It is straightforward to verify that $\theta \in \Theta(\phi)$ if and only if $Q(\theta;\phi) = 0$. Subsequently, we will restrict inference to the sub-vector $\theta_1 = [\beta_1, \gamma_1]'$. The identified set of θ_1 is given by the projection $\Theta_1(\phi)$ defined as $\Theta_1(\phi) = \big\{ \theta_1 \bigm| \exists\, \theta_2 \text{ s.t. } Q([\theta_1', \theta_2']'; \phi) = 0 \big\}$.

4.2 Bayesian and Frequentist Inference

For the Bayesian analysis we consider two priors specified on the (θ, ψ) space as well as one prior directly specified on the (θ_1, ϕ) space. The third column of Table 2 provides information on Priors 1 and 2. Both priors are based on the same distribution for θ but differ with respect to the distribution of ψ. The priors are constructed as products of marginal distributions:

$$\text{Priors 1, 2:} \qquad p_{(\theta,\psi)}(\theta,\psi) = p(\beta_1)p(\gamma_1)p(\beta_2)p(\gamma_2)p(\psi). \tag{28}$$

The distributions of γ_1 and γ_2 are truncated at zero to ensure that duopoly profits are less than monopoly profits. Under Prior 1 the mean of ψ, which can be interpreted as an equilibrium selection probability, is 0.5. Under Prior 2 the mean of ψ is set to 0.8, which reflects the belief that it is more likely that Firm 1 enters if a monopoly of either firm is profitable. By evaluating the function $\tilde{G}(\theta,\psi)$ at random draws from the prior distribution of (θ,ψ) we obtain draws from the prior distribution of ϕ. Means and standard deviation are reported in the last four rows of Table 2 for Prior 1.

The theoretical analysis in Section 3 highlighted that a crucial ingredient of Bayesian inference in partially identified models is the conditional prior distribution of $p(\theta|\phi)$. Since Priors 1 and 2 are specified on the (θ, ψ) space, the shape of this conditional distribution is not directly evident. We can use (24), and (26) to construct a one-to-one mapping f between (θ_1, ϕ) and (θ, ψ) such that

$$p_{(\theta_1)}(\theta_1|\phi) \propto p_{(\theta_1,\phi)}(\theta_1, \phi) = p_{(\theta,\psi)}(f(\theta_1, \phi)) \cdot |f^{(1)}((\theta_1, \phi))|. \tag{29}$$

The change of variables introduces the Jacobian term $|f^{(1)}((\theta_1, \phi))|$. In addition to Priors 1 and 2 we construct a Prior 3 that is flat with respect to ϕ and uniform on the identified set $\Theta_1(\phi)$ conditional on ϕ. This prior is directly specified on (θ_1, ϕ) space:

$$\text{Prior 3:} \qquad p_{(\theta_1,\phi)}(\theta_1, \phi) \propto \frac{I\{\theta_1 \in \Theta_1(\phi)\}}{\int I\{\theta_1 \in \Theta_1(\phi)\} d\theta_1}. \tag{30}$$

To obtain the posterior distributions, the three prior distributions are combined with the likelihood function $p(Y^n|\phi) = p(Y^n|\tilde{G}(\theta, \psi))$. Draws from the posterior associated with Priors 1 and 2 are generated with a Random Walk Metropolis Algorithm, e.g. Geweke (2005). Posterior draws under Prior 3 are obtained through direct sampling.

Frequentist confidence sets are constructed based on the minimum-distance approach described in Section 3.4. Let $\hat{\phi}_n$ be the maximum likelihood estimator of ϕ, which has the property that $\sqrt{n}(\hat{\phi}_n - \phi) \Longrightarrow N(0, \Lambda)$. Let $\hat{\Lambda}_n$ be a consistent estimator of Λ and define a sample analogue of $Q(\theta; \phi)$ in (27) as

$$Q_n(\theta; \hat{\phi}_n) = \min_{0 \le \alpha \le \bar{\alpha}(\theta)} n \big\| \hat{\phi}_n - G(\theta, \alpha) \big\|_{\hat{\Lambda}_n^{-1}}. \tag{31}$$

For the numerical illustration we use for simplicity a fixed critical value and the projection method to obtain a confidence set for θ_1

$$CS_F^{\theta_1}(Y^n) = \left\{ \theta_1 \, \middle| \, \exists\, \theta_2 \text{ s.t } Q_n([\theta_1', \theta_2']'; \hat{\phi}_n) \le c_\tau^2 \right\}. \tag{32}$$

It can be shown that a fixed critical value that delivers uniformly valid confidence sets over the entire parameter space is given by the $1 - \tau$ quantile of a $\chi^2(df = 3)$ distribution with three degrees of freedom. Three degrees of freedom arise because as (one of) the interaction parameters γ_1 and γ_2 approach zero, both inequalities (25) become binding. Since the equalities (24) always need to hold, it can also be shown that any θ-dependent critical value $c^2_\tau(\theta)$ never falls below the $1-\tau$ quantile of a $\chi^2(df=2)$ distribution. We will use this insight to obtain a lower bound on the frequentist confidence set.

4.3 Numerical Results

The maximum likelihood estimator $\hat{\phi}_n$ combined with the sample size n provides a sufficient statistic. To compute the credible and confidence sets we set $\hat{\phi}_n = \phi_0$, where ϕ_0 is given in the second column of Table 2. We consider the sample sizes $n = 100$ and $n = 1,000$ and set $\tau = 0.4$ and 0.1, respectively. Results for $n = 100$ are depicted in Figure 2. The solid ellipsoid-like contour in the six panels indicates the identified set $\Theta_1(\phi)$ conditional on $\phi = \hat{\phi}_n = \phi_0$. The dashed contours in the top panels depict credible sets conditional on $\hat{\phi}_n = \phi_0$ under Prior 1 and Prior 2. The plots highlight that this conditional distribution is not uniform and that it is sensitive to beliefs about the structural parameters θ and ψ. In case of Prior 2 most of the mass concentrates near the upper edge of the identified set. The posterior confidence sets obtained under Priors 1 and 2 are dominated by the uncertainty about the reduced form parameter ϕ, though the shape of $p(\theta_1|\phi)$ has some noticeable effect on inference as well. The bottom left panel shows posterior credible sets under Prior 3, which is uniform on $\Theta_1(\phi)$ conditional on ϕ.

Finally, the frequentist confidence sets are shown in the bottom right panel. Most strikingly, the confidence sets are substantially larger than the credible sets. As discussed above, the frequentist confidence sets are conservative because they are constructed with critical values from a $\chi^2(df=3)$. These critical values are $c^2_{\tau=0.4} = 2.95$ and $c^2_{\tau=0.1} = 6.25$. The 10% critical values for a $\chi^2(df=2)$ is 4.61. This implies that a 90% confidence set constructed from the same objective function with θ-dependent critical values cannot be smaller than the 60% confidence set depicted in the figure.

In Figure 3 we show the same credible and confidence sets but now computed based on a sample of $n = 1,000$ observations. As predicted by the large sample theory presented in Section 3, the posterior credible sets closely resemble the credible sets associated with the conditional prior distribution $p(\theta_1|\phi)$. Under Prior 3, the remaining uncertainty about ϕ induces some curvature into the posterior density of θ_1 and the credible sets mimic the contours of the identified set. The increased sample size tightens the frequentist confidence set, but by construction it extends beyond the boundaries of $\Theta_1(\hat{\phi}_n)$ for any sample size.

Remark 1: The entry game is incompletely specified in the sense that it is silent about the equilibrium selection in case both a Firm 1 and Firm 2 monopoly are profitable, but a duopoly is not. To characterize the relationship between structural and reduced form parameters we introduced the parameter ψ, which can be interpreted as an equilibrium selection probability. Since it is an inherent feature of Bayesian analysis to place probability distributions over unknown parameters, the specification of a prior entails placing a probability distribution on the likelihood ψ that Firm 1 instead of Firm 2 enters the market. According to Priors 1 and 2, the distribution of ψ is independent of the profit function parameters θ. However, it does not have to be independent. Bajari, Hong, and Ryan (2010), for instance, consider more sophisticated, albeit still restrictive, probabilistic equilibrium selection mechanisms. In general, the equilibrium selection could depend on all primitives of the game as is the case in Beresteanu, Molchanov, and Molinari (2011). Prior 3 masks the distribution of the equilibrium selection probability because it is specified on (θ_1, ϕ) space. However, a change of variables can easily recover the implicit beliefs about ψ that lead to a prior that is uniform on $\Theta_1(\phi)$ conditional on ϕ. ■

Remark 2: We excluded covariates from the firms' profit function. The introduction of covariates X leads to covariate-specific reduced form and auxiliary parameters, that is, ϕ would have to be replaced by $\phi(X)$ and ψ by $\psi(X)$. If the regressor space is discretized, as it is in many applications in industrial organization, the reduced-form parameter vector remains finite-dimensional and the large sample analysis in Section 3 remains applicable. ■

Remark 3: The extension of the Bayesian analysis to mixed strategy equilibria is conceptually straightforward but computationally more involved. Suppose that $\beta_1 = \beta_2 = 0$.

Whenever $[\epsilon_{i,1}, \epsilon_{i,2}] \in [0, \gamma_1] \otimes [0, \gamma_2]$ one has to consider three cases: a Firm 1 monopoly, a Firm 2 monopoly, or a mixed strategy equilibrium in which the firms enter with probabilities $\epsilon_{i,2}/\gamma_2$ and $\epsilon_{i,1}/\gamma_1$. Linking reduced form parameters and structural parameters now requires a two-dimensional auxiliary parameter ψ that summarizes the selection probabilities for the three possible Nash equilibria. Under the assumption that $\epsilon_{i,j} \sim iidN(0,1)$ computation of $\phi = \tilde{G}(\theta, \psi)$ involves the evaluation of normal cdfs (as in the case of pure strategy equilibria) and the calculation of means of truncated normal random variates. Methods to compute the identified set are discussed in Beresteanu, Molchanov, and Molinari (2011). ■

5 Conclusion

In regular identified models credible sets and confidence sets are numerically approximately identical if the sample size is large, despite different probabilistic underpinnings. We derived a large sample approximation for the posterior distribution of a structural parameter vector in a partially identified model that is characterized by a finite-dimensional vector of reduced form parameters to compare Bayesian credible sets and frequentist confidence sets. Frequentist confidence sets extend beyond the boundaries of the estimated identified set, whereas Bayesian credible sets asymptotically exclude part of the estimated identified set. Thus, in large samples frequentist sets are too large from a Bayesian perspective and Bayesian sets are too small from a frequentist perspective. The Bayesian approach induces a probability distribution on the identified set $\Theta(\phi)$ conditional on ϕ. This distribution is not updated through the likelihood function and creates a challenge for the reporting of Bayesian inference. In this regard it is important to report estimates of the identified set $\Theta(\hat{\phi}_n)$ as well as the conditional prior $P^{\theta}_{\hat{\phi}_n}$ along with Bayesian posteriors so that the audience can assess whether due to the choice of prior the posterior concentrates in a small subset of the identified set. In the context of a two-player entry game we were able to conduct inference based on a reference prior that is conditionally uniformly distributed on the (compact) identified set $\Theta(\phi)$. Such a prior might have appeal to a broad audience.

References

Andrews, Donald and Patrik Guggenberger (2009): "Validity of Subsampling and 'Plug-in Asymptotics' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, **25**, 669-709.

Andrews, Donald and Gustavo Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, **78**, 119-158.

Bajari, Patrick, Lanier Benkard, and Jonathan Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, **75**, 1331-1370.

Bajari, Patrick, Han Hong, and Stephen Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, **78**, 1529-1568.

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, forthcoming.

Bernstein, S.N. (1934): *Theory of Probability*, Gostekhizdat, Moscow.

Berry, Steven (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, **25**, 242-262.

Bresnahan, Timothy and Peter Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, **48**, 57-81.

Bollinger, Christopher and Martijn van Hasselt (2009): "A Bayesian Analysis of Binary Misclassification: Inference in Partially Identified Models," *Manuscript*, University of Kentucky and University of Western Ontario.

Bugni, Federico (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, **78**, 735-754.

Canova, Fabio and Gianni De Nicolo (2002): "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7," *Journal of Monetary Economics*, **49**, 1131-59.

Chernozhukov, Victor and Han Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics* **115**, 293-346.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, **75**, 1243-1284.

Ciliberto, Federico and Elie Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, **77**, 1791-1828.

Geweke, John (2005): *Contemporary Bayesian Econometrics and Statistics*, John Wiley & Sons, Hoboken.

Ghosh, Jayanta K., Subhashia Ghosal, and Tapas Samanta (1994): "Stability and Convergence of the Posterior in Non-Regular Problems," in S.S. Gupta and J.O. Berger (eds.) *Statistical Decision Theory and Related Topics V*, 183-199, Springer Verlag, New York.

Ghosh, Jayanta K., Subhashia Ghosal, and Tapas Samanta (1995): "On the Convergence of Posterior Distributions," *Annals of Statistics*, **23**(6), 2145-2152.

Ghosh, Jayanta K and R.V. Ramamoorthi (2003): "Bayesian Nonparametrics," *Springer Verlag*, New York.

Haile, Phil and Elie Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, **111**, 1-51.

Hartigan, John (1983): *Bayes Theory*, Springer Verlag, New York.

Imbens, Guido and Charles Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, **72**, 1845-57.

Johnson, Richard, (1970): "Asymptotic Expansions of Associated with Posterior Distributions," *Annals of Mathematical Statistics*, **41**, 851-864.

Kadane, Joseph B. (1974): "The Role of Identification in Bayesian Theory," in S.E. Fienberg and A. Zellner (eds.) *Studies in Bayesian Econometrics and Statistics*, 175-191. North Holland, Amsterdam.

Kim, Jae-Young (1998): "Large Sample Properties of Posterior Densities, Bayesian Information Criterion and the Likelihood Principle in Nonstationary Time Series Models," *Econometrica*, **66**, 359-380.

Kitagawa, Toru (2010): "Inference and Decision for Set Identified Parameters Using Posterior Lower and Upper Probabilities," Manuscript, Cemmap and UCL.

LeCam, Lucien (1953): "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes Estimates," *University of California Publications in Statistics*, **1**, 277-330.

Liao, Yuan and Wenxin Jiang (2010): "Bayesian Analysis in Moment Inequality Models," *Annals of Statistics*, **38**, 275-316.

Lubik, Thomas and Frank Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, **94**, 190-217.

Manski, Charles and Elie Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, **70**, 519-547.

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

von Mises, Richard (1965): *Mathematical Theory of Probability and Statistics,* Academic Press, New York.

Moon, Hyungsik Roger, Frank Schorfheide, Eleonora Granziera, and Mihye Lee (2009): "Inference for VARs Identified with Sign Restrictions," *Manuscript*, University of Southern California and University of Pennsylvania.

Phillips, Peter C.B. and Werner Ploberger (1996): "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica*, **64**, 381-412.

Poirier, Dale (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, **14**, 483-509.

Romano, Joseph P. and Azeem M. Shaikh (2010): "Inference for Partially Identified Set in Partially Identified Econometric Models," *Econometrica*, **78**, 169-212.

Rosen, Adam (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, **146**, 107-117.

Severini, Thomas (1991): "On the Relationship between Bayesian and Non-Bayesian Interval Estimates," *Journal of the Royal Statistical Society. Series B*, **53** 611-618.

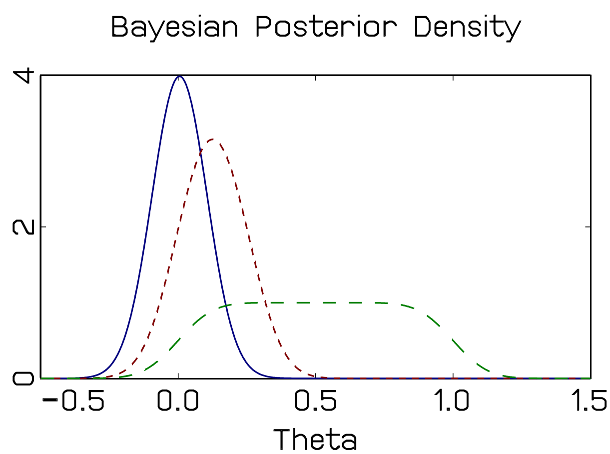
Sims, Christopher and Harald Uhlig (1991): "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, **59**, 1591-1599.

Stoye, Jörg (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, **77**, 1299-1315.

Uhlig, Harald (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, Aad (1998): *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Figure 1: Inference in the Inequality Condition Model, Known Length



Notes: The figure depicts posterior pdfs of θ for λ equal to 0.01, 0.25, and 1.00. The sample size is $n = 100$ and $\hat{\phi}_n = 0$.

Table 1: 90% Frequentist Confidence Intervals and Bayesian Credible Sets

		Length of $\Theta(\phi)$			
n		0.01	0.25	1.00	10
10	$P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{F}\}$	0.90	0.92	0.97	1.00
	Length of CS^{θ}_{F}	1.04	1.12	1.81	10.8
	Length of CS^{θ}_{HPD}	1.04	1.07	1.40	9.02
50	$P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{F}\}$	0.90	0.95	0.99	1.00
	Length of CS^{θ}_{F}	0.47	0.61	1.36	10.4
	Length of CS^{θ}_{HPD}	0.47	0.52	1.03	9.00
100	$P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{F}\}$	0.90	0.96	0.99	1.00
	Length of CS^{θ}_{F}	0.33	0.51	1.26	10.3
	Length of CS^{θ}_{HPD}	0.33	0.41	0.96	9.00
500	$P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{F}\}$	0.90	0.98	1.00	1.00
	Length of CS^{θ}_{F}	0.15	0.37	1.11	10.1
	Length of CS^{θ}_{HPD}	0.15	0.27	0.91	9.00

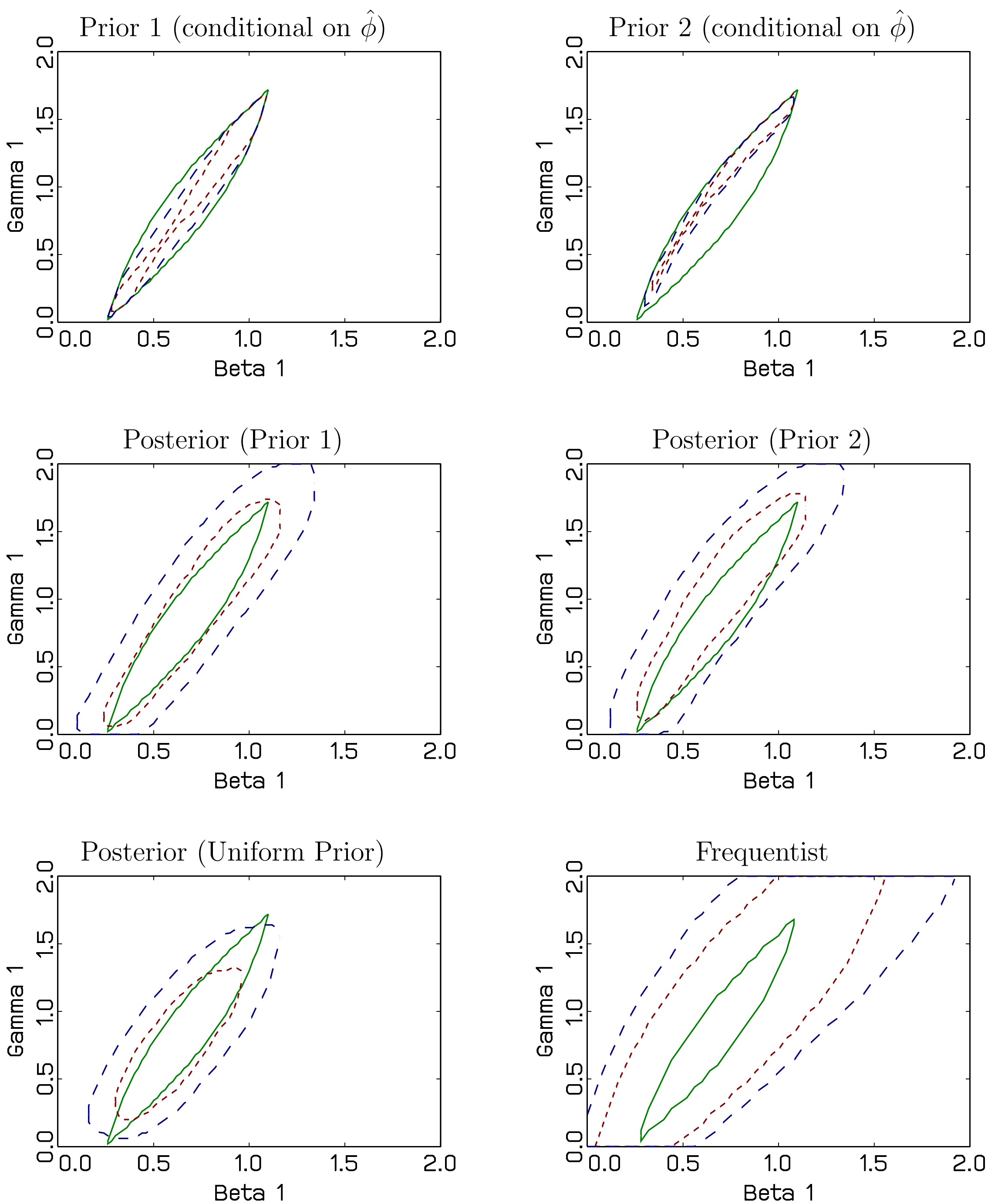
Notes: CS^{θ}_{F} is the frequentist confidence interval, CS^{θ}_{HPD} is the Bayesian credible interval, and $P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{F}\}$ is the posterior probability that θ lies in the frequentist confidence interval.

Table 2: Entry Game: "True" Parameters and Prior

Parameter	True Value	Prior Distribution
Structural Parameters θ		
β_1	0.7	$N(0, 4^2)$
γ_1	1.0	$N_+(0, 2^2)$
β_2	0.5	$N(0, 4^2)$
γ_2	1.0	$N_+(0, 2^2)$
Auxiliary Parameter ψ		
ψ	0.7	Prior 1: $\mathcal{B}(0.5, 0.2^2)$
	0.7	Prior 2: $\mathcal{B}(0.8, 0.1^2)$
Implied Reduced Form Parameters ϕ (Prior 1)		
ϕ_{11}	0.12	$\mu_{11} = 0.13$, $\sigma_{11} = 0.28$
ϕ_{00}	0.07	$\mu_{00} = 0.25$, $\sigma_{00} = 0.37$
ϕ_{10}	0.48	$\mu_{10} = 0.31$, $\sigma_{10} = 0.40$

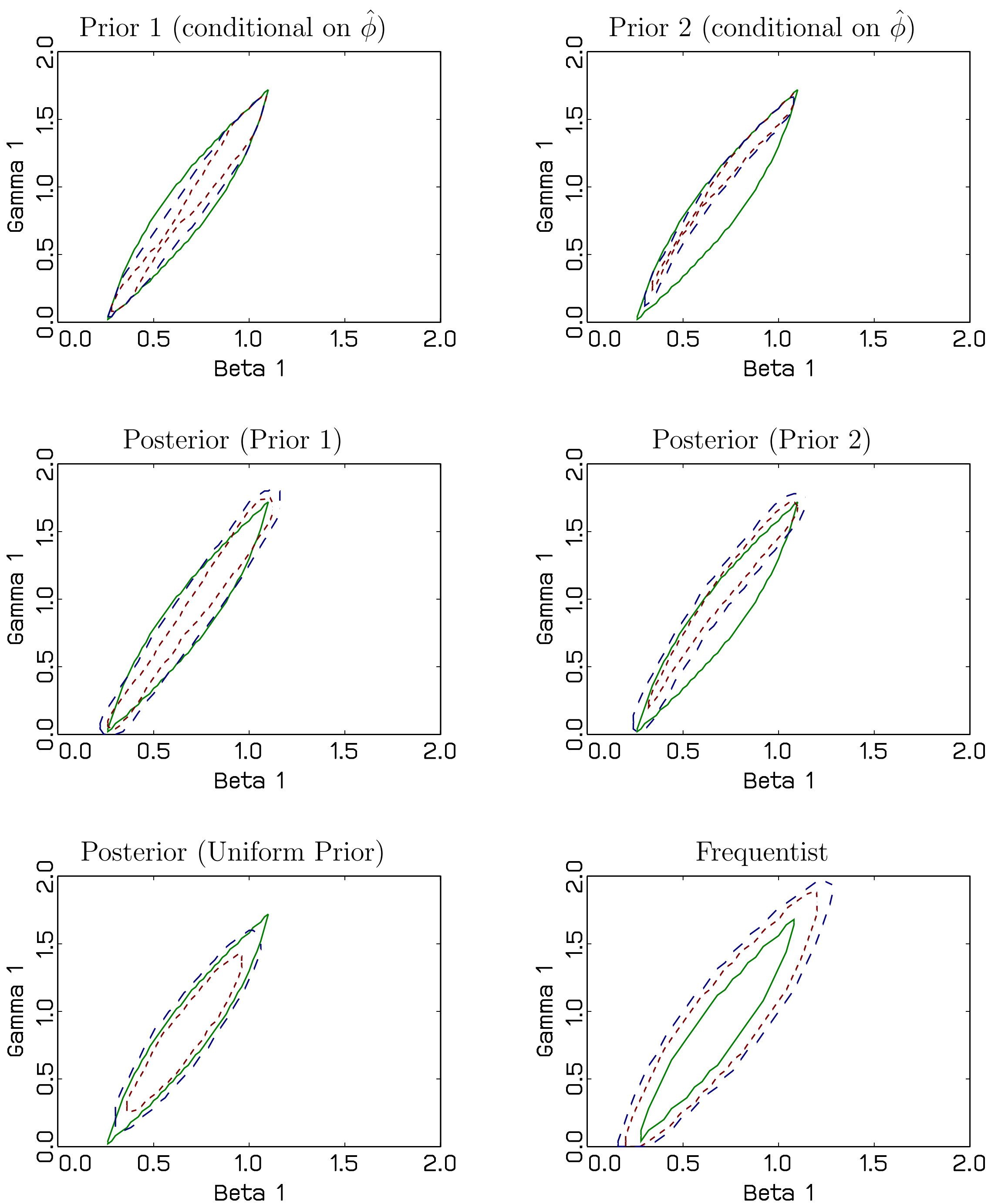
Notes: For the prior distribution of the reduced form parameters we report means μ and standard deviations σ under $\alpha \sim \mathcal{B}(0.5, 0.2^2)$. $N(\mu, \sigma^2)$ and $\mathcal{B}(\mu, \sigma^2)$ refer to normal and Beta distributions with mean μ and variance σ^2. $N_+(\nu, \sigma^2) = I\{X \geq 0\}N(\nu, \sigma^2)$ denotes a truncated normal distribution.

Figure 2: Posterior Credible Sets and Frequentist Confidence Sets, $n = 100$



Notes: Figure depicts identified sets (solid), 90% credible (confidence) sets (long dashes), and 60% credible (confidence) sets (short dashes).

Figure 3: Posterior Credible Sets and Frequentist Confidence Sets, $n = 1,000$



Notes: Figure depicts identified sets (solid), 90% credible (confidence) sets (long dashes), and 60% credible (confidence) sets (short dashes).

Bayesian and Frequentist Inference in Partially Identified Models Supplementary Material: Proofs and Derivations

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide[1]
University of Pennsylvania
CEPR and NBER

This supplement contains proofs and derivations for results presented in the paper "Bayesian and Frequentist Inference in Partially Identified Models" by H. Moon and F. Schorfheide.

[1]Correspondence: H.R. Moon: Department of Economics, KAP 300, University Park Campus, University of Southern California, Los Angeles, CA 90089. Email: hyungsikmoon@gmail.com. F. Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104. Email: schorf@ssc.upenn.edu.

A Proofs of Theorems and Corollary in Main Text

This section contains proofs for Theorems 1 and 2 as well as Corollary 1. The proof of Theorem 1 requires Lemma A.1 stated below.

Proof of Theorem 1(ii): Since the L_1 distance satisfies the triangle inequality

$$\|P^\theta_{Y^n} - P^\theta_{\hat{\phi}_n}\| \leq \|P^\theta_{Y^n} - P^\theta_{N,Y^n}\| + \|P^\theta_{N,Y^n} - P^\theta_{\hat{\phi}_n}\|$$

it suffices to show that $\|P^\theta_{N,Y^n} - P^\theta_{\hat{\phi}_n}\| \stackrel{\mathbb{P}}{\longrightarrow} 0$.

$$\begin{aligned}
&\left\|P^\theta_{N,Y^n} - P^\theta_{\hat{\phi}_n}\right\| \\
&\quad \leq \int_{\mathbb{R}^m} \left\|P^\theta_{\hat{\phi}_n+\hat{J}_n^{-1/2}D_n^{-1}s} - P^\theta_{\hat{\phi}_n}\right\| dN\big(0,I\big)(s) \\
&\quad \leq \int_{\mathbb{R}^m} I\{\|\hat{\phi}_n-\phi_0\|<\delta\} I\{\|\hat{\phi}_n-\phi_0+\hat{J}_n^{-1/2}D_n^{-1}s\|<\delta\} \left\|P^\theta_{\hat{\phi}_n+\hat{J}_n^{-1/2}D_n^{-1}s} - P^\theta_{\hat{\phi}_n}\right\| dN\big(0,I\big)(s) \\
&\qquad +2I\{\|\hat{\phi}_n-\phi_0\|\geq\delta\} + 2\int_{\mathbb{R}^m} I\{\|\hat{\phi}_n-\phi_0+\hat{J}_n^{-1/2}D_n^{-1}s\|\geq\delta\} dN\big(0,I\big)(s) \\
&\quad \leq \int_{\mathbb{R}^m} M(\phi_0,\delta)\|\hat{J}_n^{-1/2}D_n^{-1}s\| dN\big(0,I\big)(s) + 2I\{\|\hat{\phi}_n-\phi_0\|\geq\delta\} \\
&\qquad +2I\{\|\hat{\phi}_n-\phi_0\|\geq\delta/2\} + 2\int_{\mathbb{R}^m} I\{\|\hat{J}_n^{-1/2}D_n^{-1}s\|\geq\delta/2\} dN\big(0,I\big)(s) \\
&\quad \leq M(\phi_0,\delta)\|\hat{J}_n^{-1/2}\|\|D_n^{-1}\| \int_{\mathbb{R}^m} \|s\| dN\big(0,I\big)(s) + o_p(1) \stackrel{\mathbb{P}}{\longrightarrow} 0.
\end{aligned}$$

For the second inequality we bound the L_1 distance $\|P^\theta_{\hat{\phi}_n+\hat{J}_n^{-1/2}D_n^{-1}s} - P^\theta_{\hat{\phi}_n}\|$ by 2 if either $\hat{\phi}_n$ or $\hat{\phi}_n + \hat{J}_n^{-1/2}D_n^{-1}s$ lie outside of the $N_\delta(\phi_0)$ neighborhood. For the third inequality we use the Lipschitz bound of Assumption 2 and the inequality $I\{\|x+y\|\geq\delta\} \leq I\{\|x\|\geq\delta/2\} + I\{\|y\|\geq\delta/2\}$. The last line follows from Assumption 1 that $\hat{\phi}_n$ converges in probability to ϕ_0 , $\|D_n\| \uparrow \infty$, and $\hat{J}_n^{-1/2} = O_p(1)$. A similar argument can be used to establish the convergence of $P^\theta_{Y^n}$ to $P^\theta_{\phi_0}$. ■

The following Lemma is needed for the subsequent proof of Theorem 2. To simplify the notation let $p_Y(\theta) = p(\theta|Y^n)$, and $p_0(\theta) = p(\theta|\phi_0)$. Similarly, we abbreviate the thresholds κ_{Y^n} and κ_{ϕ_0} by κ_Y and κ_0. A proof is provided in the Online Appendix.

Lemma A.1 *Suppose that* $\int |p_Y(\theta) - p_0(\theta)| d\theta = o_p(1)$ *and* $\int I\{p_0(\theta) = \kappa_0\} p_0(\theta) d\theta = 0$, *where* $\kappa_0 < \infty$. *Then*

$$\int \Big| I\{p_Y(\theta) \geq \kappa_0\} - I\{p_0(\theta) \geq \kappa_0\} \Big| p_Y(\theta) d\theta = o_p(1).$$

Proof of Lemma A.1: (This Lemma is used to prove Theorem 2. Write

$$\begin{aligned}
&\int \Big| I\{p_Y(\theta) \geq \kappa_0\} - I\{p_0(\theta) \geq \kappa_0\} \Big| p_Y(\theta) d\theta \\
&= \int I\{\theta \mid p_Y(\theta) \geq \kappa_0,\ p_0(\theta) < \kappa_0\} p_Y(\theta) d\theta + \int I\{\theta \mid p_Y(\theta) < \kappa_0,\ p_0(\theta) \geq \kappa_0\} p_Y(\theta) d\theta \\
&= \int_{\theta \in A_n} p_Y(\theta) d\theta + \int_{\theta \in B_n} p_Y(\theta) d\theta = (I) + (II),
\end{aligned}$$

say. We will subsequently construct $o_p(1)$ bounds for terms (I) and (II).

Bound for (I): We deduce from the L_1 convergence assumption of $p_Y(\theta)$ to $p_0(\theta)$ that

$$(I) = \int_{\theta \in A_n} p_Y(\theta) d\theta = \int_{\theta \in A_n} p_0(\theta) d\theta + o_p(1) = (Ia) + o_p(1).$$

Thus, it suffices construct an $o_p(1)$ bound for Ia. Define the function

$$f_n(\theta) = p_Y(\theta) - p_0(\theta)$$

and notice that $f_n(\theta) > 0$ for $\theta \in A_n$. With this definition,

$$\begin{aligned}
\int_{A_n} f_n(\theta) p_0(\theta) d\theta &= \int_{A_n} |p_Y(\theta) - p_0(\theta)| p_0(\theta) d\theta \\
&\leq \kappa_0 \int_{A_n} |p_Y(\theta) - p_0(\theta)| d\theta = o_p(1).
\end{aligned} \tag{A.1}$$

The inequality follows from $p_0(\theta) < \kappa_0$ on the set A_n. The $o_p(1)$ statement is a consequence of the assumptions that $p_Y(\theta)$ converges to $p_0(\theta)$ in L_1 and that κ_0 is finite.

Now notice that

$$I\{\theta \in A_n\} = I\Big\{ I\{\theta \in A_n\} f_n(\theta) > 0 \Big\}. \tag{A.2}$$

If $\theta \in A_n$ then $f_n(\theta) > 0$, which means that $I\{\theta \in A_n\} f_n(\theta) > 0$. Moreover, for any $\eta > 0$ we obtain the inequality

$$I\Big\{ I\{\theta \in A_n\} f_n(\theta) > \eta \Big\} \leq \frac{1}{\eta} I\{\theta \in A_n\} f_n(\theta). \tag{A.3}$$

Thus,

$$\begin{aligned}
(Ia) &= \int I\Big\{I\{\theta \in A_n\} f_n(\theta) > 0\Big\} p_0(\theta) d\theta \\
&\leq \int I\Big\{I\{\theta \in A_n\} f_n(\theta) > 0\Big\} p_0(\theta) d\theta - \int I\Big\{I\{\theta \in A_n\} f_n(\theta) > \eta\Big\} p_0(\theta) d\theta \\
&\quad + \frac{1}{\eta} \int_{A_n} f_n(\theta) p_0(\theta) d\theta \\
&= \int I\Big\{0 < I\{\theta \in A_n\} f_n(\theta) \leq \eta\Big\} p_0(\theta) d\theta + \frac{1}{\eta} \int_{A_n} f_n(\theta) p_0(\theta) d\theta \\
&= (Ib) + (Ic),
\end{aligned}$$

say. The first equality follows from (A.2). The inequality is a consequence of (A.3).

To bound (Ib) notice that

$$I\Big\{0 < I\{\theta \in A_n\} f_n(\theta) \leq \eta\Big\} \leq I\Big\{\kappa_0 - \eta \leq p_0(\theta) \leq \kappa_0 + \eta\Big\}.$$

For the indicator function on the left-hand-side to be one, it has to be the case that $\theta \in A_n$ and $f_n(\theta) \leq \eta$. On the set A_n $p_Y(\theta) \geq \kappa_0$ which leads to

$$\kappa_0 \leq p_Y(\theta) = p_0(\theta) + f_n(\theta) \leq p_0(\theta) + \eta,$$

that is,

$$\kappa_0 - \eta \leq p_0(\theta).$$

Moreover, $p_0(\theta) < \kappa_0 \leq \kappa_0 + \eta$ and therefore the following inequality is satisfied:

$$\kappa_0 - \eta \leq p_0(\theta) \leq \kappa_0 + \eta.$$

Thus,

$$(Ib) \leq \int I\Big\{\kappa_0 - \eta \leq p_0(\theta) \leq \kappa_0 + \eta\Big\} p_0(\theta) d\theta.$$

Based on the Dominated Convergence Theorem and the assumption $\int I\{p_0(\theta) = \kappa_0\} p_0(\theta) = 0$ we deduce that

$$\lim_{\eta \longrightarrow 0} \int I\Big\{\kappa_0 - \eta \leq p_0(\theta) \leq \kappa_0 + \eta\Big\} p_0(\theta) d\theta = \int I\Big\{p_0(\theta) = \kappa_0\Big\} p_0(\theta) = 0. \tag{A.4}$$

Notice that our bound for (Ib) is deterministic.

To establish that $(Ia) \xrightarrow{\mathbb{P}} 0$ it suffices to show that for every $\epsilon > 0$ and $\delta > 0$ there exists an $N(\epsilon, \delta)$ such that for $n \geq N(\epsilon, \delta)$

$$\mathbb{P}\{Ia > \epsilon\} \leq \mathbb{P}\{Ib > \epsilon/2\} + \mathbb{P}\{Ic > \epsilon/2\} < \delta.$$

Based on (A.4) we can find an $\eta(\epsilon) > 0$ such that $\mathbb{P}\{Ib > \epsilon/2\} = 0$. To obtain a bound for (Ic) define $Z_n = \int_{A_n} f_n(\theta)p_0(\theta)d\theta$ such that $(Ic) = Z_n/\eta$. According to (A.1), $Z_n = o_p(1)$. Thus, we can find an $N(\epsilon, \delta)$ such that

$$\mathbb{P}\left\{|Z_n| > \eta(\epsilon)\frac{\epsilon}{2}\right\} < \delta$$

whenever $n \geq N(\epsilon, \delta)$, which shows that $(Ia) = o_p(1)$.

Bound for (II): This bound can be obtained following the same steps. Change the definition of $f_n(\theta)$ to

$$f_n(\theta) = p_0(\theta) - p_Y(\theta).$$

Using this definition we obtain that

$$\begin{aligned}
\int_{\theta \in B_n} f_n(\theta)p_Y(\theta)d\theta &= \int_{\theta \in B_n} (p_0(\theta) - p_Y(\theta))p_Y(\theta)d\theta \\
&\leq \kappa_0 \int_{\theta \in B_n} \big|p_0(\theta) - p_Y(\theta)\big|d\theta = o_p(1)
\end{aligned}$$

because on the set B_n the density $p_Y(\theta)$ is bounded by κ_0. Now consider

$$\begin{aligned}
(II) &= \int_{B_n} p_Y(\theta)d\theta = \int I\Big\{I\{\theta \in B_n\}f_n(\theta) > 0\Big\}p_Y(\theta)d\theta \\
&\leq \int I\Big\{I\{\theta \in B_n\}f_n(\theta) > 0\Big\}p_Y(\theta)d\theta - \int I\Big\{I\{\theta \in B_n\}f_n(\theta) > \eta\Big\}p_Y(\theta)d\theta \\
&\quad + \frac{1}{\eta}\int_{B_n} f_n(\theta)p_Y(\theta)d\theta \\
&= \int I\Big\{0 < I\{\theta \in B_n\}f_n(\theta) \leq \eta\Big\}p_0(\theta)d\theta + \frac{1}{\eta}\int_{B_n} f_n(\theta)p_Y(\theta)d\theta + o_p(1) \\
&= (IIb) + (IIc) + o_p(1).
\end{aligned}$$

In the last line we used the L_1 convergence to replace $p_Y(\theta)$ by $p_0(\theta)$ in the definition of term (IIb) which introduces an additional $o_p(1)$ term.

To bound (IIb) notice that

$$I\Big\{0 < I\{\theta \in B_n\} f_n(\theta) \leq \eta\Big\} \leq I\Big\{\kappa_0 - \eta \leq p_n(\theta) \leq \kappa_0 + \eta\Big\}.$$

For the indicator function on the left-hand-side to be one, it has to be the case that $\theta \in B_n$ and $f_n(\theta) \leq \eta$. On the set B_n $p_Y(\theta) < \kappa_0$ which leads to

$$\kappa_0 > p_Y(\theta) = p_0(\theta) - f_n(\theta) \geq p_0(\theta) - \eta.$$

that is,

$$\kappa_0 + \eta \geq p_0(\theta).$$

Moreover, $p_0(\theta) \geq \kappa_0 \geq \kappa_0 - \eta$ and therefore the following inequality is satisfied:

$$\kappa_0 - \eta \leq p_0(\theta) \leq \kappa_0 + \eta.$$

Thus,

$$(IIb) \leq \int I\Big\{\kappa_0 \leq p_0(\theta) < \kappa_0 + \eta\Big\} p_0(\theta) d\theta.$$

Dominated convergence implies that the bound converges to zero as $\eta \longrightarrow 0$. The remaining steps needed to establish that $(II) = o_p(1)$ are identical to the steps followed for term (I). ■

Proof of Theorem 2: Throughout the proof we express the symmetric difference between two sets in terms of indicator functions: $A \ominus B = |I\{x \in A\} - I\{x \in B\}|$. Part (i): To simplify the notation let $p_Y(\theta) = p(\theta|Y^n)$ and $p_0(\theta) = p(\theta|\phi_0)$. Similarly, we abbreviate the thresholds κ_{Y^n} and κ_{ϕ_0} by κ_Y and κ_0. Write

$$\begin{aligned}
&\int \Big| I\{p_Y(\theta) \geq \kappa_Y\} - I\{p_0(\theta) \geq \kappa_0\} \Big| p_Y(\theta) d\theta \\
&= \int \Big| I\{p_Y(\theta) \geq \kappa_Y\} - I\{p_Y(\theta) \geq \kappa_0\} \Big| p_Y(\theta) d\theta \\
&\quad + \int \Big| I\{p_Y(\theta) \geq \kappa_0\} - I\{p_0(\theta) \geq \kappa_0\} \Big| p_Y(\theta) d\theta \\
&= (I) + (II),
\end{aligned}$$

say. In view of our assumptions Lemma A.1 provides an $o_p(1)$ bound for term (II). Now consider term (I). Since by construction

$$\int I\{p_Y(\theta) \geq \kappa_Y\} p_Y(\theta) d\theta = 1 - \tau,$$

we can write term I as

$$\begin{aligned}
(I) &= \int I\Big\{p_Y(\theta) \geq \min\{\kappa_0, \kappa_Y\}\Big\} p_Y(\theta) d\theta - \int I\Big\{p_Y(\theta) \geq \max\{\kappa_0, \kappa_Y\}\Big\} p_Y(\theta) d\theta \\
&= I\{\kappa_0 \geq \kappa_Y\} \left[(1-\tau) - \int I\{p_Y(\theta) \geq \kappa_0\} p_Y(\theta) d\theta\right] \\
&\quad + I\{\kappa_0 < \kappa_Y\} \left[\int I\{p_Y(\theta) \geq \kappa_0\} p_Y(\theta) d\theta - (1-\tau)\right] \\
&= \left| \int I\{p_Y(\theta) \geq \kappa_0\} p_Y(\theta) d\theta - (1-\tau) \right|.
\end{aligned}$$

In order to show that $I = o_p(1)$ we add and subtract $\int I\{p_0(\theta) \geq \kappa_0\} p_Y(\theta) d\theta$ and using the triangle inequality:

$$\begin{aligned}
(I) &\leq \left| \int I\{p_Y(\theta) \geq \kappa_0\} p_Y(\theta) d\theta - \int I\{p_0(\theta) \geq \kappa_0\} p_Y(\theta) d\theta \right| \\
&\quad + \left| \int I\{p_0(\theta) \geq \kappa_0\} p_Y(\theta) d\theta - (1-\tau) \right| \\
&= \left| \int I\{p_Y(\theta) \geq \kappa_0\} p_Y(\theta) d\theta - \int I\{p_0(\theta) \geq \kappa_0\} p_Y(\theta) d\theta \right| \\
&\quad + \left| \int I\{p_0(\theta) \geq \kappa_0\} p_Y(\theta) d\theta - \int I\{p_0(\theta) \geq \kappa_0\} p_0(\theta) d\theta \right| \\
&\leq \int \Big| I\{p_Y(\theta) \geq \kappa_0\} - I\{p_0(\theta) \geq \kappa_0\} \Big| p_Y(\theta) d\theta \\
&\quad + \int I\{p_0(\theta) \geq \kappa_0\} \big| p_Y(\theta) - p_0(\theta) \big| d\theta = o_p(1).
\end{aligned}$$

The first equality holds because $\int I\{p_0(\theta) \geq \kappa_0\} p_0(\theta) d\theta = 1 - \tau$. The final $o_p(1)$ result follows from Lemma A.1 and the L_1 convergence of the posterior densities established in Theorem 1.

Part (ii): The triangle inequality implies that

$$\|P^\theta_{\hat{\phi}_n} - P^\theta_{\phi_0}\| \leq \|P^\theta_{Y^n} - P^\theta_{\hat{\phi}_n}\| + \|P^\theta_{Y^n} - P^\theta_{\phi_0}\| \stackrel{\mathbb{P}}{\longrightarrow} 0$$

by Theorem 1(ii). Let $p_n(\theta) = p(\theta|\hat{\phi}_n)$ and $\kappa_n = \kappa_{\hat{\phi}_n}$. Then using the same argument as for Part (i), replacing $p_Y(\theta)$ by $p_n(\theta)$ and κ_Y by κ_n we can easily establish that

$$\int \Big| I\{\theta \in CS^\theta_{HPD}(\hat{\phi}_n)\} - I\{\theta \in CS^\theta_{HPD}(\phi_0)\} \Big| dP^\theta_{Y^n} \xrightarrow{\mathbb{P}} 0. \tag{A.5}$$

Now consider the following inequality

$$\begin{aligned} \big| I\{\theta \in A\} - I\{\theta \in B\} \big| &\leq \big| I\{\theta \in A\} - I\{\theta \in C\} \big| + \big| I\{\theta \in B\} - I\{\theta \in C\} \big| \qquad \text{(A.6)} \\ &= (I) + (II). \end{aligned}$$

If the left-hand side of (A.6) is zero, then the inequality is trivially satisfied. The left-hand side of (A.6) is one if $\theta \in A$ and $\theta \notin B$ or if $\theta \notin A$ and $\theta \in B$. Since the statement of the inequality is symmetric in A and B we focus on the first case. If $\theta \in A$, $\theta \notin B$, and $\theta \in C$, then $(I) = |1-1| = 0$ and $(II) = |0-1| = 1$. If $\theta \in A$, $\theta \notin B$, and $\theta \notin C$, then $(I) = |1-0| = 1$ and $(II) = |0+0| = 0$. We deduce that whenever the left-hand side of (A.6) is equal to one, the right-hand side is equal to one as well, which confirms the inequality. Now let

$$A = CS^\theta_{HPD}(Y^n), \quad B = CS^\theta_{HPD}(\hat{\phi}_n), \quad \text{and} \quad C = CS^\theta_{HPD}(\phi_0).$$

Integrating both sides of (A.6) yields

$$\begin{aligned} &\int \big| I\{\theta \in A\} - I\{\theta \in B\} \big| p_Y(\theta) d\theta \\ &\leq \int \big| I\{\theta \in A\} - I\{\theta \in C\} \big| p_Y(\theta) d\theta + \int \big| I\{\theta \in B\} - I\{\theta \in C\} \big| p_Y(\theta) d\theta = o_p(1). \end{aligned}$$

The $o_p(1)$ statement follows from Part (i) and (A.5). ■

Proof of Corollary 1: Recall that $\Theta(\hat{\phi}_n) \subset CS^\theta_F(Y^n)$ and $CS^\theta_{HPD}(Y^n) \subset \Theta$. Part (i) follows from the inequalities

$$\begin{aligned} & P^\theta_{Y^n}(CS^\theta_{HPD}(Y^n) \backslash CS^\theta_F(Y^n)) \\ &\leq P^\theta_{Y^n}(\Theta \backslash \Theta(\hat{\phi}_n)) \\ &= 1 - P^\theta_{Y^n}(\Theta(\hat{\phi}_n)) \\ &\leq 1 - P^\theta_{\hat{\phi}_n}(\Theta(\hat{\phi}_n)) + \Big| P^\theta_{\hat{\phi}_n}(\Theta(\hat{\phi}_n)) - P^\theta_{Y^n}(\Theta(\hat{\phi}_n)) \Big| \\ &\xrightarrow{\mathbb{P}} 0. \end{aligned}$$

The probability limit is obtained from $P^{\theta}_{\hat{\phi}_n}(\Theta(\hat{\phi}_n)) = 1$ and Theorem 1(ii).

Part (ii) can be deduced from the following inequalities:

$$\begin{aligned} & P^{\theta}_{Y^n}(CS^{\theta}_F(Y^n) \backslash CS^{\theta}_{HPD}(Y^n)) \\ \geq \quad & P^{\theta}_{Y^n}(\Theta(\hat{\phi}_n) \backslash CS^{\theta}_{HPD}(Y^n)) \\ \geq \quad & P^{\theta}_{Y^n}(\Theta(\hat{\phi}_n)) - P^{\theta}_{Y^n}(CS^{\theta}_{HPD}(Y^n)) \\ \geq \quad & P^{\theta}_{\hat{\phi}_n}(\Theta(\hat{\phi}_n)) - P^{\theta}_{Y^n}(CS^{\theta}_{HPD}(Y^n)) - \left| P^{\theta}_{Y^n}(\Theta(\hat{\phi}_n)) - P^{\theta}_{\hat{\phi}_n}(\Theta(\hat{\phi}_n)) \right| \\ \xrightarrow{\mathbb{P}} \quad & 1 - (1-\tau) = \tau. \end{aligned}$$

The probability limit is obtained from $P^{\theta}_{\hat{\phi}_n}(\Theta(\hat{\phi}_n)) = 1$, $P^{\theta}_{Y^n}(CS^{\theta}_{HPD}(Y^n)) = 1 - \tau$, and Theorem 1(ii). ■

B Derivations of Results Presented in Main Text

This section contains derivations for Section 2, derivations for Remark 2 in Section 3, as well as detailed derivations for the entry game illustration in Section 4.

Derivations for Section 2

Direct calculation of the posterior density of θ:

$$\begin{aligned} p(\theta|Y^n) &= \frac{1}{\sqrt{2\pi/n}} \int_{-\infty}^{\infty} \frac{1}{\lambda} I\{\phi \leq \theta \leq \phi + \lambda\} \exp\left\{-\frac{n}{2}(\phi - \hat{\phi}_n)^2\right\} d\phi \\ &= \frac{1}{\lambda}\frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}(\theta-\hat{\phi}_n-\lambda)}^{\sqrt{n}(\theta-\hat{\phi}_n)} \exp\left\{-\frac{s^2}{2}\right\} ds \\ &= \frac{1}{\lambda}\Big[\Phi_N(\sqrt{n}(\theta-\hat{\phi}_n)) - \Phi_N(\sqrt{n}(\theta-\hat{\phi}_n-\lambda))\Big]. \end{aligned}$$

The second equality follows from re-arranging the inequalities in the indicator function and the change of variables $s = \sqrt{n}(\phi - \hat{\phi}_n)$. It is straightforward to verify that $p(\theta|Y^n)$ has a single mode at $\theta = \hat{\phi}_n + \lambda/2$ and is symmetric around the mode. ■

Derivations for Section 3

Direct Calculations to Verify Equation (18): We begin with the change of variable $s = \hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n + \tilde{s})$, which leads to

$$\begin{aligned} p(\theta|Y^n) &= p_N(\theta|Y^n) \\ &= \frac{1}{\lambda_n} \int f\left(\frac{\theta - \hat{\phi}_n - \hat{J}_n^{-1/2} D_n^{-1} s}{\lambda_n}\right) \varphi_N(s) ds \\ &= \frac{1}{\lambda_n} |\hat{J}_n^{1/2} D_n| \int_{\tilde{s}=-\lambda_n}^{0} f(-\lambda_n^{-1}\tilde{s}) \varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n + \tilde{s})) d\tilde{s}. \end{aligned}$$

The second equality makes use of the assumption that $f(x) = 0$ outside of the unit interval. The L_1 distance can be bounded as follows:

$$
\begin{aligned}
&\int_\theta \left| p_N(\theta|Y^n) - |\hat{J}_n^{1/2} D_n| \varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n)) \right| d\theta \\
&= |\hat{J}_n^{1/2} D_n| \int_\theta \left| \int_{\tilde{s}=-\lambda_n}^0 \frac{1}{\lambda_n} f(-\lambda_n^{-1}\tilde{s}) \left[\varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n + \tilde{s})) - \varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n)) \right] d\tilde{s} \right| d\theta \\
&\leq |\hat{J}_n^{1/2} D_n| \int_{\tilde{s}=-\lambda_n}^0 \int_\theta \frac{1}{\lambda_n} f(-\lambda_n^{-1}\tilde{s}) \left| \varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n + \tilde{s})) - \varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n)) \right| d\theta d\tilde{s} \\
&\leq \int_{\tilde{s}=-\lambda_n}^0 \frac{1}{\lambda_n} f(-\lambda_n^{-1}\tilde{s}) \int_{\tilde{\theta}} \left| \varphi_N(\tilde{\theta} + \hat{J}_n^{1/2} D_n \tilde{s}) - \varphi_N(\tilde{\theta}) \right| d\tilde{\theta} d\tilde{s}. \qquad \text{(B.1)}
\end{aligned}
$$

The first equality follows because $\int_0^1 f(x)dx = 1$ and $\varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n))$ does not depend on $\tilde{s}$. The last inequality is based on the change of variables $\tilde{\theta} = \hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n)$.

Now consider the difference $\varphi_N(\tilde{\theta} + h) - \varphi_N(\tilde{\theta})$ for $-\bar{h} \leq h \leq 0$. By direct calculation we obtain

$$
\begin{aligned}
|\varphi_N(\tilde{\theta} + h) - \varphi_N(\tilde{\theta})| &= \left| (2\pi)^{-1/2} \exp\left\{ -\frac{1}{2}(\tilde{\theta} + h)^2 \right\} - \varphi_N(\tilde{\theta}) \right| \\
&= \left| \exp\left\{ -\frac{1}{2}(2\tilde{\theta}h + h^2) \right\} - 1 \right| \varphi_N(\tilde{\theta}).
\end{aligned}
$$

A first-order Taylor series expansion around $h = 0$ yields

$$
\exp\left\{ -\frac{1}{2}(2\tilde{\theta}h + h^2) \right\} - 1 = -(\tilde{\theta} + h_*(\tilde{\theta})) \exp\{-\tilde{\theta}h_*(\tilde{\theta})\} \exp\{-h_*^2(\tilde{\theta})/2\} h,
$$

where $-\bar{h} \leq h_*(\tilde{\theta}) \leq 0$. Thus, on the interval $-\bar{h} \leq h \leq 0$ we obtain the bound

$$
\left| \exp\left\{ -\frac{1}{2}(2\tilde{\theta}h + h^2) \right\} - 1 \right| \varphi_N(\tilde{\theta}) \leq (|\tilde{\theta}| + \bar{h}) \exp\{-\tilde{\theta}\bar{h} I\{\tilde{\theta} \leq 0\}\} \bar{h} \varphi_N(\tilde{\theta}). \qquad \text{(B.2)}
$$

Replacing $\bar{h}$ by $\hat{J}_n^{1/2} D_n \lambda_n$ in (B.2) and combining (B.1) with (B.2) leads to

$$
\begin{aligned}
&\int_\theta \left| p_N(\theta|Y^n) - |\hat{J}_n^{1/2} D_n| \varphi_N(\hat{J}_n^{1/2} D_n(\theta - \hat{\phi}_n)) \right| d\theta \\
&\leq \hat{J}_n^{1/2} D_n \lambda_n \int_{\tilde{\theta}} (|\tilde{\theta}| + \hat{J}_n^{1/2} D_n \lambda_n) \exp\{-\tilde{\theta} \hat{J}_n^{1/2} D_n \lambda_n I\{\tilde{\theta} \leq 0\}\} \varphi_N(\tilde{\theta}) d\tilde{\theta} = o_p(1).
\end{aligned}
$$

The $o_p(1)$ statement follows because $D_n\lambda_n \longrightarrow 0$ and we can find a finite constant M and an N_M such that for $n > N_M$

$$\int_{\tilde{\theta}} (|\tilde{\theta}| + \hat{J}_n^{1/2} D_n\lambda_n) \exp\{-\tilde{\theta}\hat{J}_n^{1/2} D_n\lambda_n I\{\tilde{\theta} \leq 0\}\}\varphi_N(\tilde{\theta})d\tilde{\theta} \leq M$$

with probability approaching one. ■

Derivations for Section 4

The probabilities that firm i is profitable as monopolist and duopolist are

$$m_i = \Phi_N(\beta_i) \quad \text{and} \quad d_i = \Phi_N(\beta_i - \gamma_i). \tag{B.3}$$

The relationship between the reduced-form entry probabilities and m_i and d_i, $i = 1, 2$ is given by

$$\begin{aligned} \phi_{11} &= d_1 d_2 && \text{(B.4)} \\ \phi_{00} &= (1 - m_1)(1 - m_2) && \text{(B.5)} \\ \phi_{10} &= m_1(1 - m_2) + d_1(m_2 - d_2) + \psi(m_1 - d_1)(m_2 - d_2) && \text{(B.6)} \\ &= m_1(1 - d_2) - (1 - \psi)(m_1 - d_1)(m_2 - d_2), \end{aligned}$$

where $\psi \in [0, 1]$. The vector of non-redundant reduced form parameters is given by $\phi = [\phi_{11}, \phi_{00}, \phi_{10}]'$ and the structural parameters are $\theta = [\beta_1, \gamma_1, \beta_2, \gamma_2]'$. In addition, there is an auxiliary parameter ψ.

Identified Set

We will now provide a characterization of the identified set $\Theta(\phi)$. Define

$$G(\theta, \alpha) = \begin{bmatrix} G_1(\theta) \\ G_2(\theta) \end{bmatrix} - \begin{bmatrix} 0_{2\times 1} \\ \alpha \end{bmatrix}, \tag{B.7}$$

where

$$G_1(\theta) = \begin{bmatrix} d_1 d_2 \\ (1-m_1)(1-m_2) \end{bmatrix}, \quad G_2(\theta) = m_1(1-d_2).$$

and

$$\alpha = (1-\psi)(m_1-d_1)(m_2-d_2).$$

Moreover, let

$$\bar{\alpha}(\theta) = (m_1-d_1)(m_2-d_2) \tag{B.8}$$

and

$$Q(\theta;\phi) = \min_{0\le\alpha\le\bar{\alpha}(\theta)} \Big\| \phi - G(\theta,\alpha) \Big\|. \tag{B.9}$$

Notice that by construction $Q(\theta;\phi) \ge 0$. In view of (B.4) to (B.6) and (B.7) it is straightforward to verify that the identified set can be characterized as follows:

$$\theta \in \Theta(\phi) \quad \text{iff} \quad Q(\theta;\phi) = 0.$$

Suppose we partition θ into $\theta = [\theta_1', \theta_2']'$. (B.4) and (B.5) imply that conditional on ϕ and θ_1 the subvector θ_2 is uniquely determined. Thus, the dimension of the identified set $\Theta(\phi)$ is 2. Since the entry game is symmetric with respect to Firm 1 and Firm 2, our illustration focuses on inference for θ_1. We denote the identified set for this subvector by $\Theta_1(\phi)$ and it can be characterized by the projection

$$\Theta_1(\phi) = \left\{ \theta_1 \,\middle|\, \exists\, \theta_2 \text{ s.t. } Q([\theta_1', \theta_2']';\phi) = 0 \right\}.$$

Frequentist Inference

Starting point of the frequentist inference is a large sample approximation of the sampling distribution of $\hat{\phi}_n$, defined as

$$\hat{\phi}_n = \left[\frac{n_{11}}{n}, \frac{n_{00}}{n}, \frac{n_{10}}{n} \right]', \tag{B.10}$$

where n_{11} is the number of markets with a duopoly, n_{00} is the number of markets without entry, and n_{10} is the number of markets with a Firm 1 monopoly. We assume that

$$\sqrt{n}(\hat{\phi}_n - \phi) \Longrightarrow N(0, \Lambda(\phi)) \tag{B.11}$$

uniformly in ϕ, where $\Lambda(\phi)$ can be consistently estimated by $\hat{\Lambda}$. Now define

$$Q_n(\theta;\hat{\phi}_n) = \min_{0\leq\alpha\leq\bar{\alpha}(\theta)} n\left\|\hat{\phi}_n - G(\theta,\alpha)\right\|_{\hat{\Lambda}^{-1}}. \tag{B.12}$$

We shall construct a confidence set for θ as a level set of $Q_n(\theta;\hat{\phi}_n)$. To do so, we examine the sampling distribution of $Q_n(\theta;\hat{\phi}_n)$ for $\theta \in \Theta(\phi)$.

We partition $\hat{\phi}_n$ into $\hat{\phi}_{1,n}$ and $\hat{\phi}_{2,n}$ where the partitions conform with $G_1(\theta)$ and $G_2(\theta)$. Moreover, define

$$\hat{H}_1(\theta) = \hat{\phi}_{1,n} - G_1(\theta), \quad \hat{H}_2(\theta) = \hat{\phi}_{2,n} - G_2(\theta),$$

and partition $\hat{\Lambda}$ accordingly. In addition, let

$$\hat{H}_{2.11}(\theta) = \hat{H}_2(\theta) - \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}\hat{H}_1(\theta), \quad \hat{\Lambda}_{2.11} = \hat{\Lambda}_{22} - \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}\hat{\Lambda}_{12}.$$

Using the formula for factorizing a joint normal density into a marginal and a conditional density we can re-write the objective function as

$$Q_n(\theta;\hat{\phi}_n) = \min_{0\leq\alpha\leq\bar{\alpha}(\theta)} n\left(\left\|\hat{H}_1(\theta)\right\|_{\hat{\Lambda}_{11}^{-1}} + \left\|\hat{H}_{2.11}(\theta)+\alpha\right\|_{\hat{\Lambda}_{2.11}^{-1}}\right). \tag{B.13}$$

The minimizing value of α which we denote by $\hat{\alpha}(\theta)$ is given by

$$\hat{\alpha}(\theta) = \begin{cases} 0 & \text{if } 0 \leq \hat{H}_{2.11}(\theta) \\ -\hat{H}_{2.11}(\theta) & \text{if } -\bar{\alpha}(\theta) \leq \hat{H}_{2.11}(\theta) < 0 \\ \bar{\alpha}(\theta) & \text{otherwise} \end{cases}. \tag{B.14}$$

In turn, the objective function becomes

$$Q_n(\theta;\hat{\phi}_n) = \begin{cases} n\left\|\hat{H}_1(\theta)\right\|_{\hat{\Lambda}_{11}^{-1}} + n\left\|\hat{H}_{2.11}(\theta)\right\|_{\hat{\Lambda}_{2.11}^{-1}} & \text{if } 0 \leq \hat{H}_{2.11}(\theta) \\ n\left\|\hat{H}_1(\theta)\right\|_{\hat{\Lambda}_{11}^{-1}} & \text{if } -\bar{\alpha}(\theta) \leq \hat{H}_{2.11}(\theta) < 0 \\ n\left\|\hat{H}_1(\theta)\right\|_{\hat{\Lambda}_{11}^{-1}} + n\left\|\hat{H}_{2.11}(\theta)+\bar{\alpha}(\theta)\right\|_{\hat{\Lambda}_{2.11}^{-1}} & \text{otherwise} \end{cases}. \tag{B.15}$$

As shown in Andrews and Guggenberger (2009), critical values for the construction of uniformly valid confidence sets can be obtained by considering the behavior of the objective function $Q_n(\cdot)$ under sequences of parameters. To do so, suppose data are generated based on $\phi_n = G(\theta_n, \alpha_n)$. To approximate the distribution of $Q_n(\theta_n; \hat{\phi}_n)$, notice that

$$\begin{aligned}
\hat{H}_1(\theta_n) &= \hat{\phi}_{1,n} - G_1(\theta_n) \\
&= \hat{\phi}_{1,n} - \phi_{1,n} \\
\hat{H}_{2.11}(\theta_n) &= \hat{\phi}_{2,n} - G_2(\theta_n) - \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}[\hat{\phi}_{1,n} - G_1(\theta_n)] \\
&= \hat{\phi}_{2,n} - \phi_{2,n} - \alpha_n - \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}(\hat{\phi}_{1,n} - \phi_{1,n}).
\end{aligned}$$

Let

$$Z_{1,n} = \sqrt{n}\hat{\Lambda}_{11}^{-1/2}(\hat{\phi}_{1,n} - \phi_{1,n}), \quad Z_{2.11,n} = \sqrt{n}\hat{\Lambda}_{2.11}^{-1/2}[\hat{\phi}_{2,n} - \phi_{2,n} - \hat{\Lambda}_{21}\hat{\Lambda}_{11}^{-1}(\hat{\phi}_{1,n} - \phi_{1,n})].$$

Using this notation, we can rewrite the objective function as

$$Q_n(\theta_n; \hat{\phi}_n) = \begin{cases} \|Z_{1,n}\| + \|Z_{2.11,n} - \sqrt{n}\hat{\Lambda}_{2.11}^{-1/2}\alpha_n\| & \text{if } \sqrt{n}\hat{\Lambda}_{2.11}^{-1/2}\alpha_n \leq Z_{2.11,n} \\ \|Z_{1,n}\| + \|Z_{2.11,n} + \sqrt{n}\hat{\Lambda}_{2.11}^{-1/2}(\bar{\alpha}(\theta_n) - \alpha_n)\| & \text{if } Z_{2.11,n} < -\sqrt{n}\hat{\Lambda}_{2.11}^{-1/2}(\bar{\alpha}(\theta_n) - \alpha_n) \\ \|Z_{1,n}\| & \text{otherwise} \end{cases}. \tag{B.16}$$

Now suppose that $\sqrt{n}\Lambda_{2.11}^{-1/2}\alpha_n \longrightarrow a$, $\sqrt{n}\Lambda_{2.11}^{-1/2}(\bar{\alpha}(\theta_n) - \alpha_n) \longrightarrow \bar{a}$, where $a \in \mathbb{R}^+ \cup \infty$ and $\bar{a} \in \mathbb{R}^+ \cup \infty$. Thus,

$$Q_n(\theta_n; \hat{\phi}_n) \Longrightarrow \begin{cases} \|Z_1\| + \|Z_{2.11} - a\| & \text{if } a \leq Z_{2.11} \\ \|Z_1\| + \|Z_{2.11} + \bar{a}\| & \text{if } Z_{2.11} < -\bar{a} \\ \|Z_1\| & \text{otherwise} \end{cases}, \tag{B.17}$$

where $Z_1 \sim N(0, I_2)$ and $Z_{2.11} \sim N(0, 1)$ and Z_1 and $Z_{2.11}$ are independent. We have to distinguish three cases. First,

$$Q_n(\theta_n; \hat{\phi}_n) \Longrightarrow \|Z_1\| \leq \|Z_1\| + \|Z_{2.11}\| I\{Z_{2.11} \geq 0\} \quad \text{if} \quad a = \infty, \ \bar{a} = \infty.$$

Second,

$$Q_n(\theta_n; \hat{\phi}_n) \Longrightarrow \|Z_1\| + \|Z_{2.11} - a\| I\{Z_{2.11} \geq a\} \leq \|Z_1\| + \|Z_{2.11}\| I\{Z_{2.11} \geq 0\} \quad \text{if} \quad a < \infty, \ \bar{a} = \infty.$$

Third,

$$\begin{aligned} Q_n(\theta_n;\hat{\phi}_n) &\Longrightarrow \|Z_1\| + \|Z_{2.11} - a\| I\{Z_{2.11} \geq a\} + \|Z_{2.11} + \bar{a}\| I\{Z_{2.11} < -\bar{a}\} \quad \text{if} \quad a < \infty,\ \bar{a} < \infty \\ &\leq \|Z_1\| + \|Z_{2.11}\|. \end{aligned}$$

The bound for this last case is weaker than the bounds for the first two cases. The case $\bar{a} < 0$ arises only if $\bar{\alpha}(\theta_n) \longrightarrow 0$ sufficiently fast, meaning that θ_n approaches an area of the parameter space in which the model is point identified. From the definition of $\bar{\alpha}(\theta)$ in (B.8) it follows that the third case arises if one of the interaction parameters is close to zero. In our numerical illustration we use a conservative fixed critical value obtained from the $1-\tau$ quantile of a $\chi^2(df = 3)$.

A frequentist confidence set for the 4-dimensional parameter vector θ can then be defined as the level set

$$CS_F^\theta(Y^n) = \{\theta \mid Q_n(\theta;\hat{\phi}_n) \leq c_\tau^2\}. \tag{B.18}$$

We are restricting our attention to confidence sets constructed from fixed (rather than sample-size and θ dependent) critical values. In principle, one can construct the set $CS_F^\theta(Y^n)$ by evaluating the objective function $Q_n(\theta\ \hat{\phi}_n)$ on a 4-dimensional grid. However, since the identified set $\Theta(\phi)$ lies in a 2-dimensional subspace the specification of a suitable grid is difficult. Moreover, our goal is to construct a confidence set for the subvector θ_1. Thus, we let

$$\underline{Q}_n(\theta_1;\hat{\phi}_n) = \min_{\theta_2}\ Q_n([\theta_1', \theta_2]';\hat{\phi}_n)$$

and define

$$CS_F^{\theta_1}(Y^n) = \{\theta \mid \underline{Q}_n(\theta_1;\hat{\phi}_n) \leq c_\tau^2\}. \tag{B.19}$$

The confidence set $CS_F^{\theta_1}(Y^n)$ is the projection of $CS_F^\theta(Y^n)$ onto the domain of θ_1. To compute the projection-based confidence set we specify a 2-dimensional grid for θ_1 and evaluate the objective function $\underline{Q}_n(\theta_1;\hat{\phi}_n)$ for each grid point. A parameter value is included in the confidence set if $\underline{Q}_n(\theta_1;\hat{\phi}_n) \leq c_\tau^2$.

Bayesian Inference – Draws from Conditional Prior

Prior 1 and Prior 2 are specified on the θ - ψ space through densities $p(\theta,\psi)$. These priors induce a prior distribution on the reduced form parameters ϕ. As explained in the main text, the conditional prior of θ given ϕ will not get updated through the likelihood function and the posterior will converge to $p(\theta|\hat{\phi}_n)$. In order to characterize the conditional prior $p(\theta_1|\phi)$ we conduct the following change of variables. Let

$$Z = [\beta_1, \gamma_1, \beta_2, \gamma_2, \psi]' \tag{B.20}$$

and

$$X = [\beta_1, \gamma_1, \phi_{11}, \phi_{00}, \phi_{10}]'. \tag{B.21}$$

To convert a prior density for $Z = f(X)$ into a prior for X, we can use

$$p_X(X) = p_Z(f(X))|f'(X)|. \tag{B.22}$$

Once we have derived $p_X(X)$ we can proceed as follows. Notice that

$$p(\theta_1|\phi) \propto p(\theta_1, \phi). \tag{B.23}$$

We use a Random-Walk Metropolis Algorithm to generate draws from $p(\theta_1|\phi)$. For this algorithm it is sufficient to be able evaluate the joint density $p(\theta_1,\phi)$ numerically. Descriptions of the algorithm can be found in many textbooks, e.g., Geweke (2005). Our proposal density is multivariate Gaussian with a covariance matrix that equals a suitably scaled identity matrix.

We shall proceed by characterizing the function $f(X)$, component by component and then derive the Jacobian $f'(X)$. The following functional relationships will be useful:

$$m_1 = \Phi_N(\beta_1), \quad m_2 = \Phi_N(\beta_2), \quad d_1 = \Phi_N(\beta_1 - \gamma_1), \quad d_2 = \Phi_N(\beta_2 - \gamma_2).$$

Since we will have to solve for β_2 and γ_2, notice that

$$\beta_2 = \Phi_N^{-1}(m_2), \quad \gamma_2 = \Phi_N^{-1}(m_2) - \Phi_N^{-1}(d_2).$$

The Nash equilibrium conditions imply that

$$\begin{aligned}\phi_{00} &= (1-m_1)(1-m_2)\\ \phi_{11} &= d_1 d_2\\ \phi_{10} &= m_1(1-m_2)+d_1(m_2-d_2)+\psi(m_1-d_1)(m_2-d_2).\end{aligned}$$

We can use these conditions to solve for m_2, d_2, and ψ:

$$\begin{aligned}m_2 &= 1-\frac{\phi_{00}}{1-m_1}\\ d_2 &= \frac{\phi_{11}}{d_1}\\ \psi &= \frac{\phi_{10}-m_1(1-m_2)-d_1(m_2-d_2)}{(m_1-d_1)(m_2-d_2)}.\end{aligned}$$

The expression for ψ can be simplified by replacing m_2 and d_2:

$$\begin{aligned}\psi &= \frac{\phi_{10}-m_1(1-m_2)-d_1(m_2-d_2)}{(m_1-d_1)(m_2-d_2)}\\ &= \frac{\phi_{10}-\phi_{00}\frac{m_1}{1-m_1}-d_1\left(1-\frac{\phi_{00}}{1-m_1}-\frac{\phi_{11}}{d_1}\right)}{(m_1-d_1)\left(1-\frac{\phi_{00}}{1-m_1}-\frac{\phi_{11}}{d_1}\right)}\\ &= \frac{\phi_{10}(1-m_1)-\phi_{00}m_1-d_1\left(1-m_1-\phi_{00}-\frac{\phi_{11}(1-m_1)}{d_1}\right)}{(m_1-d_1)\left(1-m_1-\phi_{00}-\frac{\phi_{11}(1-m_1)}{d_1}\right)}\\ &= \frac{\phi_{10}(1-m_1)-\phi_{00}m_1-d_1 g(X)}{(m_1-d_1)g(X)},\end{aligned}$$

where

$$g(X)=\left(1-m_1-\phi_{00}-\frac{\phi_{11}(1-m_1)}{d_1}\right).$$

Combining terms, we obtain the following expressions for the components of $f(X)$:

$$\begin{aligned}f_1(X) &= \beta_1\\ f_2(X) &= \gamma_1\\ f_3(X) &= \Phi_N^{-1}\left(1-\frac{\phi_{00}}{1-\Phi_N(\beta_1)}\right)\\ f_4(X) &= f_3(X)-\Phi_N^{-1}\left(\frac{\phi_{11}}{\Phi_N(\beta_1-\gamma_1)}\right)\\ f_5(X) &= \frac{A_5(X)}{B_5(X)}=\frac{\phi_{10}(1-\Phi_N(\beta_1))-\phi_{00}\Phi_N(\beta_1)-\Phi_N(\beta_1-\gamma_1)g(X)}{(\Phi_N(\beta_1)-\Phi_N(\beta_1-\gamma_1))g(X)}\end{aligned}$$

where

$$g(X) = \left(1 - \Phi_N(\beta_1) - \phi_{00} - \frac{\phi_{11}(1 - \Phi_N(\beta_1))}{\Phi_N(\beta_1 - \gamma_1)}\right).$$

Now we can calculate the derivatives for the jacobian matrix. For this define

$$\psi(z) = \frac{\partial \Phi_N^{-1}(z)}{\partial z} = \frac{1}{\phi_N(\Phi_N^{-1}(z))}.$$

Term $f_1(X)$:

$$\frac{\partial f_1(X)}{\partial \beta_1} = 1.$$

Term $f_2(X)$:

$$\frac{\partial f_2(X)}{\partial \gamma_1} = 1.$$

Term $f_3(X)$:

$$\begin{aligned}
\frac{\partial f_3(X)}{\partial \beta_1} &= -\psi\left(1 - \frac{\phi_{00}}{1 - \Phi_N(\beta_1)}\right) \frac{\phi_{00}}{[1 - \Phi_N(\beta_1)]^2} \phi_N(\beta_1) \\
\frac{\partial f_3(X)}{\partial \phi_{00}} &= -\psi\left(1 - \frac{\phi_{00}}{1 - \Phi_N(\beta_1)}\right) \frac{1}{1 - \Phi_N(\beta_1)}.
\end{aligned}$$

Term $f_4(X)$:

$$\begin{aligned}
\frac{\partial f_4(X)}{\partial \beta_1} &= \frac{\partial f_3(X)}{\partial \beta_1} + \psi\left(\frac{\phi_{11}}{\Phi_N(\beta_1 - \gamma_1)}\right) \frac{\phi_{11}\phi_N(\beta_1 - \gamma_1)}{\Phi_N^2(\beta_1 - \gamma_1)} \\
\frac{\partial f_4(X)}{\partial \gamma_1} &= -\psi\left(\frac{\phi_{11}}{\Phi_N(\beta_1 - \gamma_1)}\right) \frac{\phi_{11}\phi_N(\beta_1 - \gamma_1)}{\Phi_N^2(\beta_1 - \gamma_1)} \\
\frac{\partial f_4(X)}{\partial \phi_{11}} &= -\psi\left(\frac{\phi_{11}}{\Phi_N(\beta_1 - \gamma_1)}\right) \frac{1}{\Phi_N(\beta_1 - \gamma_1)} \\
\frac{\partial f_4(X)}{\partial \phi_{00}} &= \frac{\partial f_3(X)}{\partial \phi_{00}}.
\end{aligned}$$

Term $f_5(X)$:

$$\frac{\partial f_5(X)}{\partial x} = \frac{\frac{\partial A(X)}{\partial x} B(X) - A(X) \frac{\partial B(X)}{\partial x}}{B(X)^2}.$$

Term $A(X)$:

$$\begin{aligned}
\frac{\partial A(X)}{\partial \beta_1} &= -(\phi_{10}+\phi_{00})\phi_N(\beta_1) - \phi_N(\beta_1-\gamma_1)g(X) - \Phi_N(\beta_1-\gamma_1)\frac{\partial g(X)}{\partial \beta_1} \\
\frac{\partial A(X)}{\partial \gamma_1} &= \phi_N(\beta_1-\gamma_1)g(X) - \Phi_N(\beta_1-\gamma_1)\frac{\partial g(X)}{\partial \gamma_1} \\
\frac{\partial A(X)}{\partial \phi_{11}} &= -\Phi_N(\beta_1-\gamma_1)\frac{\partial g(X)}{\partial \phi_{11}} \\
\frac{\partial A(X)}{\partial \phi_{00}} &= -\Phi_N(\beta_1) - \Phi_N(\beta_1-\gamma_1)\frac{\partial g(X)}{\partial \phi_{00}} \\
\frac{\partial A(X)}{\partial \phi_{10}} &= (1-\Phi_N(\beta_1)) - \Phi_N(\beta_1-\gamma_1)\frac{\partial g(X)}{\partial \phi_{10}}.
\end{aligned}$$

Term $B(X)$:

$$\begin{aligned}
\frac{\partial B(X)}{\partial \beta_1} &= (\phi_N(\beta_1) - \phi_N(\beta_1-\gamma_1))g(X) + (\Phi_N(\beta_1) - \Phi_N(\beta_1-\gamma_1))\frac{\partial g(X)}{\partial \beta_1} \\
\frac{\partial B(X)}{\partial \gamma_1} &= \phi_N(\beta_1-\gamma_1)g(X) + (\Phi_N(\beta_1) - \Phi_N(\beta_1-\gamma_1))\frac{\partial g(X)}{\partial \gamma_1} \\
\frac{\partial B(X)}{\partial \phi_{11}} &= (\Phi_N(\beta_1) - \Phi_N(\beta_1-\gamma_1))\frac{\partial g(X)}{\partial \phi_{11}} \\
\frac{\partial B(X)}{\partial \phi_{00}} &= (\Phi_N(\beta_1) - \Phi_N(\beta_1-\gamma_1))\frac{\partial g(X)}{\partial \phi_{00}}.
\end{aligned}$$

Term $g(X)$:

$$\begin{aligned}
\frac{\partial g(X)}{\partial \beta_1} &= -\phi_N(\beta_1) + \frac{\phi_{11}\phi_N(\beta_1)}{\Phi_N(\beta_1-\gamma_1)} + \frac{\phi_{11}(1-\Phi_N(\beta_1))\phi_N(\beta_1-\gamma_1)}{\Phi_N^2(\beta_1-\gamma_1)} \\
\frac{\partial g(X)}{\partial \gamma_1} &= -\frac{\phi_{11}(1-\Phi_N(\beta_1))\phi_N(\beta_1-\gamma_1)}{\Phi_N^2(\beta_1-\gamma_1)} \\
\frac{\partial g(X)}{\partial \phi_{11}} &= -\frac{1-\Phi_N(\beta_1)}{\Phi_N(\beta_1-\gamma_1)} \\
\frac{\partial g(X)}{\partial \phi_{00}} &= -1.
\end{aligned}$$

Bayesian Inference – Draws from Posterior

According to Equations (B.3) to (B.6) we can express the reduced form probabilities as functions of θ and ψ. Thus, the likelihood function is given by

$$p(Y^n|\theta,\psi) = \phi_{11}^{n_{11}}(\theta,\psi)\phi_{00}^{n_{00}}(\theta,\psi)\phi_{10}^{n_{10}}(\theta,\psi)\phi_{01}^{n_{01}}(\theta,\psi). \tag{B.24}$$

If this prior distribution is combined with a prior specified on the θ - ψ space, then the posterior is given by

$$p(\theta, \psi|Y^n) \propto p(Y^n|\theta, \psi)p(\theta, \psi) \tag{B.25}$$

and draws can be generated with a Random Walk Metropolis Algorithm.

In addition to Priors 1 and 2 we consider a prior that is flat with respect to the reduced form parameters. Conditional on ϕ, the prior for θ_1 is uniform on the identified set $\Theta_1(\phi)$. In order to obtain draws from the posterior distribution of θ_1 we sample from (i) $p(\phi|Y^n)$ and (ii) from $p(\theta_1|\phi)$. For Step (i) notice that under the flat prior for ϕ, the posterior distribution $P^{\phi}_{Y^n}$ takes the form of a Dirichlet distribution

$$[\phi_{11}, \phi_{00}, \phi_{10}, \phi_{01}]' \sim \text{Dirichlet } (n_{11} + 1, n_{00} + 1, n_{10} + 1, n_{01}).$$

A draw from this Dirichlet distribution can be generated as follows: Let $a_j \sim \mathcal{G}(n_j + 1, 1)$, where $j \in \{11, 00, 10, 01\}$ and $\mathcal{G}(\alpha, 1)$ denotes a Gamma distribution with shape parameter α and scale parameter 1. Then set

$$\phi = [a_{11}, a_{00}, a_{10}, a_{01}]'/(a_{11} + a_{00} + a_{10} + a_{01}).$$

For Step (ii) we specify a two-dimensional grid for θ_1 in order construct projections of the identified set $\Theta_1(\phi)$ onto the β_1 and γ_1 ordinates. Let these projections be delimited by $\underline{\beta}_1$, $\bar{\beta}_1$, $\underline{\gamma}_1$, and $\bar{\gamma}_1$. We then use an acceptance sampler with a proposal density that is uniform on $[\underline{\beta}_1, \bar{\beta}_1] \otimes [\underline{\gamma}_1, \bar{\gamma}_1]$ to obtain a draw of θ_1 conditional on ϕ.

Bayesian Inference – Credible Sets

Credible sets are computed according to the following steps:

1. Construct two independent sequences $\{\theta^{(1)}_{1,s}\}^S_{s=1}$ and $\{\theta^{(2)}_{1,s}\}^S_{s=1}$ of draws from the distribution of θ_1.

2. Use the $\{\theta^{(1)}_{1,s}\}^S_{s=1}$ draws to construct Kernel density estimates $\hat{p}(\theta^{(2)}_{1,s})$ for each $\theta^{(2)}_{1,s}$, $s = 1, \ldots, S$.

3. Find a cutoff κ such that $(1-\tau)S$ of the density estimates $\hat{p}(\theta_{1,s}^{(2)})$ are greater or equal than κ.

4. Use the $\{\theta_{1,s}^{(1)}\}_{s=1}^{S}$ draws to construct Kernel density estimates $\hat{p}(\theta_1)$ for values of θ_1 on a 2-dimensional grid. Include a particular grid point into the credible set if $\hat{p}(\theta_1) \geq \kappa$.