NBER WORKING PAPER SERIES

BAYESIAN AND FREQUENTIST INFERENCE IN PARTIALLY IDENTIFIED MODELS

Hyungsik Roger Moon
Frank Schorfheide

Working Paper 14882
http://www.nber.org/papers/w14882

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2009

We thank seminar participants at the 2007 NASM, the 2008 SETA, Boston College, Johns Hopkins, Ohio State, Rice, UC Davis, UC Irvine, UCLA, and Vanderbilt for helpful comments. Moon gratefully acknowledges financial support from 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. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Bayesian and Frequentist Inference in Partially Identified Models
Hyungsik Roger Moon and Frank Schorfheide
NBER Working Paper No. 14882
April 2009
JEL No. C11,C32,C35

ABSTRACT

A large sample approximation of the posterior distribution of partially identified structural parameters is derived for models that can be indexed by a finite-dimensional reduced form parameter vector. It is used to analyze the differences between frequentist confidence sets and Bayesian credible sets in partially identified models. A key difference is that frequentist set estimates extend beyond the boundaries of the identified set (conditional on the estimated reduced form parameter), whereas Bayesian credible sets can asymptotically be located in the interior of the identified set. Our asymptotic approximations are illustrated in the context of simple moment inequality models and a numerical illustration for a two-player entry game is provided.

Hyungsik Roger Moon
University of Southern California
Department of Economics
KAP 300
University Park Campus
Los Angeles, CA 90089
moonr@usc.edu

Frank Schorfheide
University of Pennsylvania
Department of Economics
3718 Locust Walk
McNeil 525
Philadelphia, PA 19104-6297
and NBER
schorf@ssc.upenn.edu

1 Introduction

In partially identified models one can only bound, but not point-identify the structural parameter vector of interest, θ. Such models arise in many areas of economics. Prominent examples in macroeconomics are structural vector autoregressions (VARs) and dynamic stochastic general equilibrium (DSGE) models. In the VAR literature the identification problem has traditionally be addressed by imposing enough restrictions on the structural form such that the mapping between one-step-ahead forecast errors and structural shocks becomes one-to-one. More recently however, Canova and De Nicolo (2002) and Uhlig (2005) have developed more agnostic identification schemes that only restrict the signs of a subset of impulse responses in the initial periods after the impact of the shock, which leads to partially identified structural VAR. In DSGE models partial identification arises for instance if a subset of structural parameters guarantees the uniqueness of a rational expectations equilibrium but does not affect the equilibrium law of motion, e.g., Lubik and Schorfheide (2004).

Partially-identified models also percolate the microeconometric literature and include censored sampling models and models for interval data, surveyed at length in Manski (2003). Partial identification arises in models of industrial organization, for instance, in game-theoretic models with multiple equilibria studied by Bresnahan and Reiss (1991), Berry (1994), Halie and Tamer (2003), Pakes, Porter, Ho, and Ishi (2005), Bajari, Benkard, and Levin (2007), and Ciliberto and Tamer (2007). Given the lack of point identification researchers have rightly 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. The contribution of this paper is to compare frequentist confidence sets and Bayesian credible sets, with a special focus on the properties of Bayesian procedures.

Starting point of our analysis is a likelihood function indexed by a finite-dimensional, identifiable reduced-form parameter vector ϕ. Reduced-form and structural parameter are linked through a correspondence, which we express as $\phi = G(\theta, \alpha)$, where $\alpha \in \mathcal{A}_\theta$. The presence of the nuisance parameter α complicates the inference about θ. We present a large sample approximation of the posterior distribution of θ. The approximation is based on an insight that dates back at least to Kadane (1974) and has recently been utilized, for instance, by Poirier (1998): beliefs about the reduced form parameter ϕ are updated through the likelihood function, but the conditional distribution of θ given ϕ remains unchanged in view of new data. It is well known that under very general conditions the posterior distribution of ϕ is asymptotically normal. We construct such a normal approximation for ϕ following the analysis in Johnson (1970) and combine it with conditional prior distributions of θ given ϕ to obtain our approximation of the posterior of θ. If $H(\phi, \xi)$ is the prior probability that $\theta \in \mathcal{T}_\xi$ conditional on ϕ, then we show that under some regularity conditions an $O(n^{-1/2})$ accurate approximation of the posterior probability is given by $H(\hat{\phi}_n, \xi)$, where $\hat{\phi}_n$ is the maximum likelihood estimator of ϕ. This approximation implies there exist asymptotically valid Bayesian credible sets *inside* the identified set of θ parameters associated with $\hat{\phi}_n$, denoted by $\Theta(\hat{\phi}_n)$.

There is a rapidly growing literature on the construction of asymptotically valid frequentist confidence sets for θ, e.g. Manski and Tamer (2002), Imbens and Manski (2004), Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishi (2005), Rosen (2005), Galichon and Henry (2006), Romano and Shaikh (2006), Woutersen (2006), Andrews and Guggenberger (2007), Andrews and Soares (2007), Canay (2007), Chernozhukov, Hong, and Tamer (2007), Stoye (2007), and Beresteanu and Molinari (2008). The main challenge of this literature is to obtain large sample approximations of the sampling distribution of an estimation objective function or a test statistic that conditional on θ are uniformly valid for all $\phi = G(\theta, \alpha)$, $\alpha \in \mathcal{A}_\theta$. While we do not develop new methods to construct frequentist confidence sets, we show that frequentist sets need to extend *beyond* the boundaries of $\Theta(\hat{\phi}_n)$. Thus, we can deduce that in partially identified models, Bayesian credible sets tend to be smaller than frequentist confidence sets. This finding is in contrast with the regular point identified case, in which Bayesian and frequentist sets coincide in large samples.

The remainder of the paper is organized as follows. In Section 2 we briefly review the construction of Bayesian and frequentist set estimates for regular, point-identified models. We then generalize the setup to models in which θ is set-identified and provide a simple example of a partially identified model. The large sample approximation of the posterior of θ and the construction of asymptotically valid credible sets for partially identified models is presented in Section 3. In Section 4 we illustrate properties of the large sample approximation using simple moment inequality models. Section 5 provides a numerical illustration in the context of an entry-game model and Section 6 concludes. Proofs are collected in an Appendix.

A word on notation. We often use M to denote a generic finite constant. When X is a matrix, $\|X\| = (tr\,(X'X))^{1/2}$ denotes the Euclidean norm of X. We use $\mathcal{N}(\mu, \sigma^2)$ to denote a normal distribution with mean μ and variance σ^2 and $\phi_{\mathcal{N}}(\cdot)$ and $\Phi_{\mathcal{N}}(\cdot)$ 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 $z_\tau = \Phi_{\mathcal{N}}^{-1}(1-\tau)$. $\mathcal{U}[a,b]$ denotes the uniform distribution on the interval $[a,b]$. We use P_b^a to denote a probability distribution of a random variable a conditional on the realization of a random variable b. $I\{X \leq \xi\}$ denotes the indicator function that is equal to one if $X \leq \xi$ and zero otherwise. Finally, the notation $\subseteq$ is used to denote weak inclusion and $\subset$ is used for strict inclusion.

2 Identified and Partially Identified Models

We begin with a heuristic comparison of large sample approximations of Bayesian posterior distributions and the frequentist distribution of likelihood ratios in a point identified model in which the likelihood function is locally approximately quadratic and the maximum likelihood estimator (MLE) has a Gaussian limit distribution. It is well known that in this environment the Bayesian $1-\tau$ credible Highest Posterior Density (HPD) set is approximately a level set of the likelihood function and has a $1-\tau$ coverage probability from a frequentist perspective. A formalization and refinement of the subsequent heuristics can be found in Severini (1991), who derives asymptotic expansions for the posterior probability of confidence regions based on the likelihood ratio statistic and for the (frequentist) coverage probability of highest posterior density regions.

Suppose that a sequence of random variables $Y^n = \{Y_i\}_{i=1}^n$ is characterized by a density $p(Y^n|\phi)$ with respect to a dominating measure μ, where $\phi \in \Phi \subseteq \mathbb{R}^K$. Let $l_n(\phi) = \ln p(Y^n|\phi)$ be the log likelihood function and $\hat{\phi}_n$ denote the maximum likelihood estimator (MLE), that is $l_n(\hat{\phi}_n) \geq l_n(\phi)$ for all $\phi \in \Phi$. A large sample approximation of the Bayesian posterior density can be obtained from a second-order Taylor expansion of the log-posterior density function around the MLE $\hat{\phi}_n$. Let $-\hat{J}_n$ be the Hessian of the likelihood function evaluated at the maximum $\hat{\phi}_n$ such that

$$l_n(\phi) = l_n(\hat{\phi}_n) - \frac{1}{2}(\phi - \hat{\phi}_n)'\hat{J}_n(\phi - \hat{\phi}_n) + \mathcal{R}_l(\|\phi - \hat{\phi}_n\|^2). \tag{1}$$

Similarly, let $\pi(\phi) = \ln p(\phi)$ be the log prior density and assume that one can approximate the log prior with a first-order Taylor series expansion of the form

$$\pi(\phi) = \pi(\hat{\phi}_n) + \pi^{(1)}(\hat{\phi}_n)'(\phi - \hat{\phi}_n) + \mathcal{R}_\pi(\|\phi - \hat{\phi}_n\|).$$

Now transform ϕ according to $s = \hat{J}_n^{1/2}(\phi - \hat{\phi}_n)$, such that

$$\begin{aligned} l_n(\hat{\phi}_n + \hat{J}_n^{-1/2}s) - l_n(\hat{\phi}_n) &= -\frac{1}{2}s's + \mathcal{R}_l(\|\hat{J}_n^{-1/2}s\|^2) \\ \pi(\hat{\phi}_n + \hat{J}_n^{-1/2}s) - \pi(\hat{\phi}_n) &= \pi^{(1)}(\hat{\phi}_n)'\hat{J}_n^{-1/2}s + \mathcal{R}_\pi(\|\hat{J}_n^{-1/2}s\|^2). \end{aligned}$$

If ϕ is identifiable, then the smallest eigenvalue of J_n is positive and increasing with the sample size such that $\hat{J}_n^{-1/2}s$ tends to zero and the influence of the prior distribution on the posterior vanishes as $n \longrightarrow \infty$. Hence,

$$\ln p(\hat{\phi}_n + \hat{J}_n^{-1/2}s|Y^n) - \ln p(\hat{\phi}_n|Y^n) \approx l_n(\phi) - l_n(\hat{\phi}_n) \approx -\frac{1}{2}s's,$$

that is, the posterior distribution of ϕ is approximately normal. Under suitable regularity conditions, it can be deduced that

$$\left| P^{\phi}_{Y^n}\{2[l_n(\phi) - l_n(\hat{\phi}_n)] \geq -c_\tau\} - P\{\mathcal{Z}'\mathcal{Z} \leq c_\tau\} \right| \longrightarrow 0, \tag{2}$$

where $\mathcal{Z} \sim \mathcal{N}(0, I)$. If one chooses c_τ such that $P\{\mathcal{Z}'\mathcal{Z} \leq c_\tau\} = 1 - \tau$, then our heuristics imply that the level set

$$CS^\phi = \left\{ \phi \;\middle|\; 2[l_n(\phi) - l_n(\hat{\phi}_n)] \geq -c_\tau \right\} \tag{3}$$

provides a large sample approximation to the HPD set that is $1 - \tau$ credible.

For a frequentist analysis it is convenient to approximate the likelihood function around the probability limit ϕ_0 of the maximum likelihood estimator:

$$l_n(\phi) = l_n(\phi_0) + Z'_{n,0}(\phi - \phi_0) - \frac{1}{2}(\phi - \phi_0)'J_{n,0}(\phi - \phi_0) + \mathcal{R}(\|\phi - \phi_0\|^2). \tag{4}$$

Here, $Z_{n,0}$ and $J_{n,0}$ are the matrices of first and second derivatives of the log-likelihood function evaluated at ϕ_0. Now let $s = J_{n,0}^{1/2}(\phi - \phi_0)$ and write

$$l_n(\phi) = l_n(\phi_0) - \frac{1}{2}(s - J_{n,0}^{-1/2}Z_{n,0})'(s - J_{n,0}^{-1/2}Z_{n,0}) + \frac{1}{2}Z'_{n,0}J_{n,0}^{-1}Z_{n,0} + \mathcal{R}(\|J_{n,0}^{-1/2}s\|^2).$$

In "regular"[1] models $\frac{1}{n}J_{n,0} \stackrel{p}{\longrightarrow} J_0$ and $J_{n,0}^{-1/2}Z_{n,0} \Longrightarrow \mathcal{Z}$ uniformly in ϕ_0, where $\mathcal{Z} \sim \mathcal{N}(0, I)$. Under suitable regularity conditions one can show that

$$l_n(\phi_0) - l_n(\hat{\phi}_n) = -\frac{1}{2}(s - J_{n,0}^{-1/2}Z_{n,0})'(s - J_{n,0}^{-1/2}Z_{n,0}) + o_p(1)$$

and deduce

$$\sup_{\phi \in \Phi} \left| P^{Y^n}_{\phi}\{2[l_n(\phi) - l_n(\hat{\phi}_n)] \geq -c_\tau\} - P\{\mathcal{Z}'\mathcal{Z} \leq c_\tau\} \right| \longrightarrow 0. \tag{5}$$

[1] An important and widely studied irregular model in the econometrics literature is the autoregressive model $y_t = \phi y_{t-1} + \epsilon_t$, where $\phi \in [0, 1]$, see Sims and Uhlig (1991). Here the convergence of $J_n(\phi)^{-1/2}Z_n(\phi)$ to a limit distribution is not uniform in ϕ and Bayesian and frequentist interval estimates differ.

Thus, the set CS^{ϕ} in (3) is also a uniformly valid frequentist confidence set.

The above analysis remains essentially unchanged if one uses a smooth one-to-one function $\phi = G(\theta)$ to re-parameterize the problem in terms of a structural parameter of interest θ. The set

$$CS^{\theta} = \left\{ \theta \;\middle|\; 2[l_n(G(\theta)) - l_n(\hat{\phi}_n)] \geq -c_{\tau} \right\}$$

is (asymptotically) a $1-\tau$ credible set for a Bayesian and a $1-\tau$ confidence set for a frequentist econometrician. The point of departure in this paper is to replace the function $G(\theta)$ by a correspondence. Each value of the reduced form parameter $\phi \in \Phi$ is associated with a set of structural form parameters. This set is typically referred to as the *identified set* and will be denoted by $\Theta(\phi)$. Likewise, each structural parameter is potentially associated with multiple reduced form parameters, which we collect in the set $\Phi(\theta)$.

Example 1: Moment Inequalities. Consider the simple location model $Y_i = \phi + U_i$, where U_i is iid with some probability density function $f(u)$. Suppose that the relationship between the location parameter $\phi \in \Phi \subseteq \mathbb{R}$ and the structural parameter of interest is given by the inequalities

$$\theta - \lambda \leq \phi \leq \theta,$$

where λ is a known constant that determines the length of the identified set. The model specification is similar to the simple treatment effect model, in which observations are missing with a known probability, analyzed in Imbens and Manski (2004). In this example $\Theta(\phi) = [\phi, \phi + \lambda]$ and $\Phi(\theta) = [\theta - \lambda, \theta]$. $\square$

To study inference with respect to the partially identified parameter θ we express the correspondence $\Phi(\theta)$ in terms of a functional relationship between ϕ, θ, and an auxiliary parameter α such that

$$\phi = G(\theta, \alpha).$$

We assume that for each ϕ there exists a suitable domain $\mathcal{A}_{\theta}$ such that

$$\Phi(\theta) = \left\{ \phi \mid \phi = G(\theta, \alpha) \text{ for some } \alpha \in \mathcal{A}_{\theta} \right\}.$$

In the moment inequality example we can choose[2] $G(\theta, \alpha) = \theta - \alpha$ and $\mathcal{A}_{\theta} = [0, \lambda]$. Replacing the reduced form parameter in the likelihood function leads to the following log-likelihood ratio:

$$l_n(G(\theta, \alpha)) - l_n(\hat{\phi}_n).$$

From the perspective of inference about θ the auxiliary parameter α is a nuisance parameter.

[2]The choice of $G(\theta, \alpha)$ is not unique, because one can express α through arbitrary functions that map into the unit interval.

3 Large Sample Analysis

We now derive a large sample approximation of the posterior distribution of a parameter $\theta \in \Theta \subseteq \mathbb{R}^k$ in a partially identified model in which the identifiable reduced form parameter $\phi \in \Phi \subseteq \mathbb{R}^K$ is linked to θ through a correspondence that takes the form $\phi = G(\theta, \alpha)$, $\alpha \in \mathcal{A}_\theta$. We use the approximation to compare Bayesian credible sets and frequentist confidence intervals in partially identified models.

3.1 Bayesian Analysis

In many applications Bayesian analysis can be conveniently implemented by combining $l_n(G(\theta, \alpha))$ with a prior distribution for θ and α. One can use numerical methods such as importance sampling or Markov-Chain Monte Carlo algorithms to approximate finite-sample moments of the posterior distribution of θ. For a theoretical analysis, on the other hand, it is more convenient to work with the joint distribution of ϕ and θ, decomposed into the marginal distribution of ϕ, P^ϕ, and the conditional distribution of θ given ϕ, P^θ_ϕ.

As emphasized by Kadane (1974), the derivation of the posterior distribution can be done on the space of the reduced form parameter ϕ. Let $\mathcal{T}$ be a measurable subset of Θ. Then

$$\begin{aligned} P^\theta_{Y^n}\{\theta \in \mathcal{T}\} &= \frac{\int_\Phi \int_{\Theta(\phi)} I\{\theta \in \mathcal{T}\} \exp[l_n(\phi)] dP^\theta_\phi dP^\phi}{\int_\Phi \int_{\Theta(\phi)} \exp[l_n(\phi)] dP^\theta_\phi dP^\phi} \\ &= \int_\Phi \left[\int_{\Theta(\phi)} I\{\theta \in \mathcal{T}\} dP^\theta_\phi \right] \frac{\exp[l_n(\phi)]}{\int_\Phi \exp[l_n(\phi)] dP^\phi} dP^\phi \\ &= \int_\Phi P^\theta_\phi \{\theta \in \mathcal{T}\} dP^\phi_{Y^n}. \end{aligned} \tag{6}$$

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 has been emphasized by Poirier (1998). To obtain a large sample distribution of $P^\theta_{Y^n}$, we will replace $P^\phi_{Y^n}$ in (6) by a Gaussian approximation. There exists a long literature on normal approximations of posterior distributions in identified models, including Bernstein (1934), LeCam (1953), von Mises (1965). Our subsequent expansion of the posterior distribution of ϕ follows work by Johnson (1970). Unlike Johnson, who provides higher-order expansions of posterior distribution, we will only derive a first-order expansion. Rather starting from low-level assumptions that guarantee the existence of a maximum likelihood estimate, we begin by directly assuming the almost-sure convergence of the MLE.

Assumption 1 *(i) The MLE* $\hat{\phi}_n$ *exists and* $\hat{\phi}_n \longrightarrow \phi_0$ $[P_{\phi_0}^{Y^n}]$ *almost surely.*
(ii) For any $\delta > 0$, $\limsup_{n\to\infty} \sup_{\|\phi-\phi_0\|\geq\delta} \frac{1}{n}[l_n(\phi_0) - l_n(\phi)] > 0$ $[P_{\phi_0}^{Y^n}]$ *almost surely.*

In order to construct a normal approximation of the posterior distribution of ϕ, we need to make a few additional assumptions that guarantee the smoothness of the log likelihood function.

Assumption 2 *(i)* Φ *is a compact subset in* $\mathbb{R}^K$ *and* $\phi_0 \in int(\Phi)$.
(ii) For n *sufficiently large,* $p(Y^n|\phi)$ *is twice continuously differentiable with respect to* ϕ.
(iii) $\frac{1}{n}J_{n,0} \longrightarrow J_0$ $[P_{\phi_0}^{Y^n}]$ *almost surely and* J_n *is well-defined and negative definite.*
(iv) There exists a $\delta > 0$ *and a finite constant* M *such that* $\|\phi_1 - \phi_2\|$ *implies that* $\frac{1}{n}\|J_n(\phi_1) - J_n(\phi_2)\| \leq M\|\phi_1 - \phi_2\|$, $[P_{\phi_0}^{Y^n}]$ *almost surely.*

Under Assumption 2, the log likelihood is continuous over a compact set, therefore the MLE $\hat{\phi}_n$ is well defined. Under Assumption 2(ii), the log likelihood function is twice continuously differentiable. As in the previous Section, we use $-J_n(\phi)$ to denote the matrix of second derivatives of the log-likelihood function around ϕ. We continue to use the abbreviations $J_{n,0} = J_n(\phi_0)$ and $\hat{J}_n = J_n(\hat{\phi}_n)$. Assumptions 1 and 2 cover models with *iid* observations as well as time series models for weakly dependent data without trends. Large sample approximations of posterior distributions for non-stationary time series models can be found in Phillips and Ploberger (1996) and Kim (1998).

Assumption 3 *(i) The prior density* $p(\phi)$ *is uniformly bounded in* $\phi \in \Phi$ *and continuously differentiable in a neighborhood around* ϕ_0.
(ii) There exists a $\delta_p > 0$ *such that* $\inf_{\|\phi-\phi_0\|\leq\delta_p} p(\phi) > 0$ *and* $\sup_{\|\phi-\phi_0\|\leq\delta_p} \left\|p^{(1)}(\phi)\right\| \leq M$ *for some finite constant* M.

According to Assumption 3 the parameter ϕ_0 is drawn from the prior distribution P^ϕ whose density function is $p(\phi)$. When $p(\phi)$ is differentiable, we denote $p^{(1)}(\phi)$ to be its first derivative. We use s to denote the re-scaled parameter vector $\hat{J}_n^{1/2}(\phi - \hat{\phi}_n)$. Based on the above assumption we obtain the following approximation to the posterior distribution of ϕ.

Theorem 1 *Suppose Assumptions 1 – 3 are satisfied. Let* Y^n *be in the sure set of Assumptions 1 and 2.*

(i) There exist finite constants M and N such that whenever $n \geq N$ we have for any sequence of bounded functions $|H_n(\phi,\xi)| < M_H$

$$\left| \int_{\phi\in\Phi} H_n\left(\phi,\xi\right) dP^{\phi}_{Y^n} - \int_{\mathbb{R}^k} H_n\left(\hat{\phi}_n + \hat{J}_n^{-1/2}s, \xi\right) d\Phi_{\mathcal{N}}\left(s\right) \right| \leq \frac{M}{\sqrt{n}}.$$

(ii) There exist finite constants M and N such that whenever $n \geq N$

$$\sup_{\xi\in\mathbb{R}^K} \left| P^{\phi}_{Y^n}\{\hat{J}_n^{1/2}(\phi - \hat{\phi}_n) \leq \xi\} - \Phi_{\mathcal{N}}(\xi) \right| \leq \frac{M}{\sqrt{n}}.$$

Part (i) of Theorem 1 is proved in the Appendix. Part (ii) follows directly from Part (i) by setting $H_n(\phi,\xi) = I\{\hat{J}_n^{1/2}(\phi - \hat{\phi}_n) \leq \xi\}$ and provides a normal approximation of the posterior distribution of $\phi = \hat{\phi}_n + \hat{J}_n^{-1/2}s$. The constant M in Theorem 1 depends on the function $H(\cdot)$ only through the bound M_H.

The remainder of the paper focuses on the characterization of the posterior distribution of θ and the posterior probability of subsets of Θ. Let $\mathcal{T}_{\xi,n} \subseteq \Theta$ be a sequence of subsets of the structural parameter space, indexed my a finite-dimensional vector ξ. Moreover, define

$$H_n\left(\phi,\xi\right) = P^{\theta}_{\phi}\{\theta \in \mathcal{T}_{\xi,n}\}. \tag{7}$$

If $\mathcal{T}_{\xi,n} = \{\theta \leq \xi\}$ then $H_n\left(\phi,\xi\right)$ is the prior (and posterior) cdf of θ given ϕ and does not depend on n. If $\theta = [\theta_1', \theta_2']'$ and $\mathcal{T}_{\xi,n} = \{\theta_1 \leq \xi\}$ then $H_n\left(\phi,\xi\right)$ is the cdf of the sub-vector θ_1 conditional on ϕ. In the context of Example 1 we will be interested in the posterior probability of the sequence of sets $\mathcal{T}_{n,\xi} = [\hat{\phi}_n + \lambda\xi,\ \hat{\phi}_n + \lambda(1-\xi)]$, which can be expressed as $\int_{\Phi} H_n\left(\phi,\xi\right) dP^{\phi}_{Y^n}$. Some of our subsequent results require that $H_n(\phi,\xi)$ satisfies a Lipschitz condition.

Assumption 4 *The sequence of functions $H_n(\phi,\xi)$ defined in (7) is Lipschitz in ϕ, that is, $|H_n(\phi_1,\xi) - H_n(\phi_2,\xi)| \leq M(\xi)\|\phi_1 - \phi_2\|$, where $M(\xi)$ is a constant that depends on ξ.*

Since we are interested in using large sample approximations of posterior distributions to characterize asymptotically valid credible sets, we provide the following formal definition.

Definition 1 *A sequence of sets $CS^{\theta}_{B,\tau}(Y^n)$ is asymptotically $1-\tau$ credible if $P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{B,\tau}(Y^n)\} \longrightarrow 1-\tau$.*

Combining (6) and Theorem 1 we obtain the following approximation to the posterior probability that $\{\theta \in \mathcal{T}_{\xi,n}\}$:

Corollary 1 *Suppose Assumptions 1 – 3 are satisfied. Let Y^n be in the sure set of Assumptions 1 and 2. The function $H_n(\phi, \xi)$ is defined in (7).*
(i) There exist finite constants M and N such that whenever $n \geq N$

$$\left| P^{\theta}_{Y^n}\{\theta \in \mathcal{T}_{\xi,n}\} - \int_{\mathbb{R}^K} H_n(\hat{\phi}_n + \hat{J}_n^{-1/2}s, \xi)\phi_{\mathcal{N}}(s)ds \right| \leq \frac{M}{\sqrt{n}}.$$

(ii) If the sequence of functions $H_n(\phi, \xi)$ satisfies Assumption 4, then there exist finite constants $M(\xi)$ and N such that whenever $n \geq N$

$$\left| P^{\theta}_{Y^n}\{\theta \in \mathcal{T}_{\xi,n}\} - H_n(\hat{\phi}_n, \xi) \right| \leq \frac{M(\xi)}{\sqrt{n}}.$$

(iii) If the sequence of functions $H_n(\phi, \xi)$ satisfies Assumption 4 and for every $\phi \in \Phi$ and $\tau \geq \xi > 0$ there is a set $\mathcal{T}_\xi(\phi) \subset \Theta(\phi)$ such that $P^{\theta}_{\phi}\{\theta \in \mathcal{T}_\xi(\phi)\} \geq 1 - \xi$, then there exists a sequence of sets $CS^{\theta}_{B,\tau}(Y^n) \subseteq \mathcal{T}_\xi(\hat{\phi}_n) \subset \Theta(\hat{\phi}_n)$ that is asymptotically $1 - \tau$ credible.

Part (i) of Corollary 1 is a direct consequence of Theorem 1. Part (ii) is proved in the Appendix and implies that an $O(n^{-1/2})$ accurate approximation of the posterior distribution of θ can be calculated from the conditional prior distribution $P^{\theta}_{\hat{\phi}_n}$, provided that $H_n(\phi, \xi)$ satisfies the Lipschitz condition. For instance, the Lipschitz condition is satisfied in Example 1, if P^{θ}_{ϕ} is $\mathcal{U}[\phi, \phi + \lambda]$ and $\mathcal{T}_{\xi,n} = \{\theta \leq \xi\}$. According to Corollary 1(iii) one can construct asymptotically valid credible sets as subsets of $\mathcal{T}_\xi(\hat{\phi}_n)$. By construction, these sets lie strictly inside of the identified set $\Theta(\hat{\phi}_n)$.

3.2 Frequentist Analysis

Starting point for our frequentist analysis is the log likelihood ratio $l_n(G(\theta, \alpha)) - l_n(\hat{\phi}_n)$. We begin by concentrating out the nuisance parameter α. Let

$$\hat{\alpha}(\theta) = \text{argmax}_{\alpha \in \mathcal{A}_\theta} l_n(G(\theta, \alpha))$$

and define the profile objective function

$$Q_n(\theta) = 2[l_n(G(\theta, \hat{\alpha}(\theta))) - l_n(\hat{\phi}_n)]. \tag{8}$$

We consider consider confidence intervals that are of the form

$$CS^{\theta}_{F,\tau} = \{Q_n(\theta) \geq -c_\tau(\theta)\}. \tag{9}$$

If the critical value function $c_\tau(\theta)$ is constant, then the confidence interval is a level set of the profile objective function $Q_n(\theta)$. For $CS^{\theta}_{F,\tau}$ to be a confidence set that is uniformly valid asymptotically the following condition has to be satisfied:

$$\lim_{n \longrightarrow \infty} \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} P_{\phi}^{Y^n}\{Q_n(\theta) \geq -c_{\tau}(\theta)\} \geq 1 - \tau. \tag{10}$$

Constructing a critical value function such that (10) holds with equality can be challenging and is the subject of a number of recent papers, including Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2005), Andrews and Guggenberger (2007), and Andrews and Soares (2007). For the remainder of this section we will assume that such a critical value function is available and we will conduct a comparison between Bayesian and frequentist confidence sets.

Since the objective function $Q_n(\theta) = 0$ if $\theta \in \Theta(\hat{\phi}_n)$ we can deduce immediately that the frequentist confidence interval contains $\Theta(\hat{\phi}_n)$: $\Theta(\hat{\phi}_n) \subseteq CS^{\theta}_{F,\tau}$. We now proceed by providing some conditions under which $\Theta(\hat{\phi}_n) \subset CS^{\theta}_{F,\tau}$. Suppose to the contrary that $\Theta(\hat{\phi}_n) = CS^{\theta}_{F,\tau}$. As long as the likelihood function has a unique maximum $\hat{\phi}_n$, $Q_n(\theta) = 0$ if and only if $\theta \in \Theta(\hat{\phi}_n)$. Using our definition of $\Phi(\theta)$ notice that

$$P_{\phi}^{Y^n}\{Q_n(\theta) = 0\} = P_{\phi}^{Y^n}\{\hat{\phi}_n \in \Phi(\theta)\}.$$

Now consider

$$\inf_{\theta \in \Theta(\phi)} P_{\phi}^{Y^n}\{\hat{\phi}_n \in \Phi(\theta)\} = \inf_{\theta \in \Theta(\phi)} P_{\phi}^{Y^n}\{\sqrt{n}(\hat{\phi}_n - \phi) \in \sqrt{n}(\Phi(\theta) - \phi)\}.$$

Let $\tilde{\theta}$ be such that ϕ is on the boundary of $\Phi(\tilde{\theta})$. Moreover, assume that $\sqrt{n}(\Phi(\tilde{\theta}) - \phi)$ can be covered with a convex cone $\mathcal{C}$ that is centered at ϕ. Then we obtain

$$\inf_{\theta \in \Theta(\phi)} P_{\phi}^{Y^n}\{\sqrt{n}(\hat{\phi}_n - \phi) \in \sqrt{n}(\Phi(\theta) - \phi)\} \leq P_{\phi}^{Y^n}\{\sqrt{n}(\hat{\phi}_n - \phi) \in \mathcal{C}\}.$$

This argument proves the following theorem.

Theorem 2 *Suppose there exists a pair $\tilde{\theta}$ and $\tilde{\phi}$ such that (i) $\tilde{\phi}$ is a boundary point of $\Phi(\tilde{\theta})$, (ii) $\sqrt{n}(\Phi(\tilde{\theta}) - \tilde{\phi})$ can be covered with a convex cone $\mathcal{C}$, and (iii) $P\{\mathcal{Z} \in \mathcal{C}\} \leq 1 - \tau$. Then $\Theta(\hat{\phi}_n) \subset CS^{\theta}_{F,\tau}$.*

The set $CS^{\theta}_{F,\tau}$ is a confidence set for the entire parameter vector θ. To conduct inference for a subset of parameters θ_1, one could project $CS^{\theta}_{F,\tau}$ onto the relevant subspace of Θ. In this case, it is still true that the projection of $\Theta(\hat{\phi}_n)$ is a strict subset of the projection of $CS^{\theta}_{F,\tau}$.

3.3 Bayesian versus Frequentist Sets

According to Theorem 2 an asymptotically valid frequentist confidence set for θ extends beyond $\Theta(\hat{\phi}_n)$, whereas Corollary 1(iii) implies that asymptotically valid Bayesian credible sets can be constructed as subsets of $\Theta(\hat{\phi}_n)$. Thus, unlike in the identified case discussed in Section 2, frequentist and Bayesian set estimates are numerically different in large samples if a model is partially identified. In particular, one can obtain Bayesian credible sets that are strict subsets frequentist confidence sets.

The frequentist literature on partially identified models is also concerned about estimates of the identified set $\Theta(\phi)$. Imbens and Manski (2004) highlight that confidence sets for the set $\Theta(\phi)$ tend to be larger than confidence sets for an element $\theta \in \Theta(\phi)$. The existing literature is not very clear about the instances in which an empirical researcher might prefer an a confidence set for $\Theta(\phi)$ over a confidence set for $\theta \in \Theta(\phi)$. A loose argument for reporting an estimate of $\Theta(\phi)$ is that the econometrician's audience might be interested in solving a minimax decision problem of the form[3]

$$\min_{\delta(Y^n)\in\mathcal{D}} \; \max_{\phi\in\Phi,\, \theta\in\Theta(\phi)} \; P_\phi^{Y^n}[\mathcal{L}(\delta(Y^n),\theta)] \tag{11}$$

by replacing $\Theta(\phi)$ in (11) with a confidence set that covers $\Theta(\phi)$. Here $\delta(Y^n)$ is a decision function and $\mathcal{L}(\delta,\theta)$ a loss function. In a Bayesian framework, the natural approach for the econometrician would be to compute the posterior distribution for θ and solve the decision problem by minimizing posterior expected loss

$$\min_{\delta(Y^n)\in\mathcal{D}} \; P^\theta_{Y^n}[\mathcal{L}(\delta(Y^n),\theta)],$$

which does not require a credible set for $\Theta(\phi)$. Nonetheless, a posterior credible set could be obtained, for instance, by taking unions of $\Theta(\phi)$ for values of ϕ in a set CS_τ^ϕ:

$$CS_\tau^* = \bigcup_{\phi\in CS_\tau^\phi} \Theta(\phi).$$

By construction, $I\{\Theta(\phi) \subseteq CS_\tau^*\} \geq I\{\phi \in CS_\tau^\phi\}$. If CS_τ^ϕ is both a valid frequentist confidence set as well as a valid Bayesian credible set (see Section 2) we can deduce that CS_τ^* is a valid set estimate for $\Theta(\phi)$ from both the Bayesian and the frequentist perspective.

[3] As an alternative to the expected loss one could consider the regret $\mathcal{L}(\delta(Y^n),\theta) - \mathcal{L}(\delta^{opt},\theta)$.

4 Illustrations

The large sample approximations obtained in the previous section are now applied to several specific examples, beginning with the moment inequality example presented in Section 2. With our extensions of Example 1, we illustrate that Bayesian credible sets are asymptotically located inside $\Theta(\hat{\phi}_n)$, whereas frequentist confidence sets extend beyond the boundaries of $\Theta(\hat{\phi}_n)$ (Section 4.1). This result also holds if frequentist inference is based on an integrated instead of a profile likelihood function (Section 4.2, Example 1 continued). Our large sample Bayesian inference can be extended to cover the models in which the volume of the identified set depends on an estimable parameters and is potentially zero. We show how to modify the approximation of the posterior to allow for reduced form parameters that lie on the boundary of Φ (Section 4.2, Example 2). Finally, we consider a model in which the reduced form parameter is uniquely determined by the structural parameter, but not vice versa. If inference is conducted for the entire parameter vector θ (instead of a subset of θ), then certain Bayesian $1-\tau$ credible sets are in fact valid $1-\tau$ frequentist confidence sets (Section 4.2, Example 3).

4.1 Moment Inequalities, Part I

Consider Example 1 of Section 2: $Y_i = \phi + U_i$, U_i is *iid* with pdf $f(u)$, and $\theta - \lambda \leq \phi \leq \theta$, where λ is known. Assume that the density function $f(u)$ satisfies Assumptions 1 and 2 with $J_0 = 1$. Moreover, assume that the prior density $p(\phi, \alpha) = p(\phi)p(\alpha)$ where $p(\phi)$ satisfies Assumption 3 and the prior on the auxiliary parameter $\alpha = \theta - \phi$ is uniform on the interval $[0, \lambda]$. To obtain a large sample approximation of the posterior cdf of θ, let $\mathcal{T}_{\xi,n} = \{\theta \leq \xi\}$. Thus, the function $H_n(\phi, \xi)$ is the cdf of a $\mathcal{U}[\phi, \phi+\lambda]$ random variable and of the form

$$H_n(\phi,\xi) = P^{\theta}_{\phi}\{\theta \leq \xi\} = \begin{cases} 0 & \text{if } \xi < \phi \\ (\phi - \xi)/\lambda & \text{if } \phi \leq \xi \leq \phi + \lambda \\ 1 & \text{otherwise} \end{cases} . \tag{12}$$

If $\lambda > 0$ the function $H_n(\phi, \xi)$ in (12) satisfies the Lipschitz condition in Assumption 4 and we obtain the following two approximations of the posterior probability $\theta \leq \xi$:

$$\begin{aligned} \hat{P}^{\theta}_{Y^n(i)}\{\theta \leq \xi\} &= \frac{1}{\lambda}(\xi - \hat{\phi}_n)\Phi_{\mathcal{N}}(\sqrt{n}(\xi - \hat{\phi}_n)) \\ &\quad -\frac{1}{\lambda}(\xi - (\hat{\phi}_n + \lambda))\Phi_{\mathcal{N}}(\sqrt{n}(\xi - (\hat{\phi}_n + \lambda))) \\ &\quad +\frac{1}{\lambda\sqrt{n}}\Big[\phi_{\mathcal{N}}(\sqrt{n}(\xi - \hat{\phi}_n)) - \phi_{\mathcal{N}}(\sqrt{n}(\xi - (\hat{\phi}_n + \lambda)))\Big], \end{aligned} \tag{13}$$

$$\hat{P}^{\theta}_{Y^n(ii)}\{\theta \leq \xi\} = H_n(\hat{\phi}_n, \xi). \tag{14}$$

The posterior density associated with approximation (i) can be obtained by differentiating the cdf $\hat{P}^{\theta}_{Y^n(i)}\{\theta \le \xi\}$ with respect to θ:

$$\hat{p}_{(i)}(\theta|Y^n) = \frac{1}{\lambda}\bigg[\Phi_{\mathcal{N}}(\sqrt{n}(\theta - \hat{\phi}_n)) - \Phi_{\mathcal{N}}(\sqrt{n}(\theta - (\hat{\phi}_n + \lambda)))\bigg]. \tag{15}$$

Since $\Phi_{\mathcal{N}}(x) = 1 - \Phi_{\mathcal{N}}(-x)$, it is straightforward to verify that the approximate posterior density $\hat{p}_{(i)}(\theta|Y^n)$ is symmetric around the mode $\hat{\theta}_n = \hat{\phi}_n + \lambda/2$. For large values of n $\hat{p}_{(i)}(\theta|Y^n)$ approaches the density function associated with $\hat{P}^{\theta}_{Y^n(ii)}$: it jumps from 0 to $1/\lambda$ at $\theta = \hat{\phi}_n$, stays constant, and drops back to zero around $\theta = \hat{\phi}_n + \lambda$.

According to $\hat{P}^{\theta}_{Y^n(ii)}$ the posterior distribution of θ is uniform on the $\Theta(\hat{\phi}_n)$ asymptotically. This suggests that the set

$$CS^{\theta}_{B,\xi}(Y^n) = [\hat{\phi}_n + \xi\lambda/2,\ \hat{\phi}_n + \lambda - \xi\lambda/2] \subset \Theta(\hat{\phi}_n)$$

is an asymptotically valid $1-\xi$ credible interval for θ. To verify this claim, let $\mathcal{T}_{\xi,n} = CS^{\theta}_{B,\xi}(Y^n)$ and define

$$H_n(\phi, \xi) = P^{\theta}_{\phi}\{\theta \in \mathcal{T}_{\xi,n}\}. \tag{16}$$

This function is piecewise linear with a Lipschitz constant of $1/\lambda$ (see Assumption 4). Thus, provided that $\lambda > 0$, the posterior probability $P^{\theta}_{Y^n}\{\theta \in CS^{\theta}_{B,\xi}(Y^n)\}$ can according to Corollary 1 be approximated by $H_n(\hat{\phi}_n, \xi) = 1 - \xi$ in (16), which verifies the claim. If $\lambda = 0$ and θ is point identified, the Lipschitz condition is violated and the asymptotic approximation of the posterior cdf is of the form $\hat{P}^{\theta}_{Y^n(i)}\{\theta \le \xi\} = \Phi_{\mathcal{N}}(\xi)$, which leads to the "standard" interval $\hat{\phi}_n \pm z_{\tau/2}/\sqrt{n}$.

To illustrate the frequentist analysis we assume that $f(u) = \phi_{\mathcal{N}}(u)$. Hence the profile objective function is given by

$$-Q_n(\theta) = \begin{cases} n(\hat{\phi}_n - \theta)^2 & \text{if } \theta \le \hat{\phi}_n \\ 0 & \text{if } \hat{\phi}_n < \theta < \hat{\phi}_n + \lambda \\ n(\hat{\phi}_n - \theta + \lambda)^2 & \text{if } \hat{\phi}_n + \lambda \le \theta \end{cases}. \tag{17}$$

The finite-sample distribution of the maximum likelihood estimator is $\sqrt{n}(\hat{\phi}_n - \phi) \sim \mathcal{Z}$, where $\mathcal{Z}$ is a standard normal random variable. It is convenient to re-scale θ according to $s_\theta = \sqrt{n}(\theta - \phi)$. In terms of the s_θ transform, the identified set $\Theta(\phi)$ is given by $0 \le s_\theta \le \sqrt{n}\lambda$. We can now characterize the distribution of the profile objective function as

$$-Q_n(\phi + n^{-1/2}s_\theta) \sim \begin{cases} (\mathcal{Z} - s_\theta)^2 & \text{if } s_\theta \le \mathcal{Z} \\ 0 & \text{if } \mathcal{Z} < s_\theta < \mathcal{Z} + \sqrt{n}\lambda \\ (\mathcal{Z} - s_\theta + \sqrt{n}\lambda)^2 & \text{if } \mathcal{Z} + \sqrt{n}\lambda \le s_\theta \end{cases}. \tag{18}$$

Now suppose that the critical value c_τ solves the following equation:

$$\Phi_{\mathcal{N}}(\sqrt{n}\lambda + \sqrt{c_\tau}) - \Phi_{\mathcal{N}}(-\sqrt{c_\tau}) = 1 - \tau. \tag{19}$$

In view of (18), we deduce

$$\begin{aligned} & \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} P_\phi^{Y^n}\{Q_n(\theta) \geq -c_\tau\} \\ = & \inf_{0 \leq s_\theta \leq \sqrt{n}\lambda} P\{s_\theta - \sqrt{n}\lambda - \sqrt{c_\tau} \leq \mathcal{Z} \leq s_\theta + \sqrt{c_\tau}\} \\ = & \inf_{0 \leq s_\theta \leq \sqrt{n}\lambda} \Phi_{\mathcal{N}}(s_\theta + \sqrt{c_\tau}) - \Phi_{\mathcal{N}}(s_\theta - \sqrt{n}\lambda - \sqrt{c_\tau}) \\ = & \; 1 - \tau, \end{aligned}$$

where the last line follows since the infimum is achieved at $s_\theta = 0$ or $s_\theta = \sqrt{n}\lambda$ and by the definition of c_τ in (19) . Therefore, the resulting confidence interval is of the form

$$CS^\theta_{F,\tau}(Y^n) = \left[\hat{\phi}_n - \sqrt{c_\tau/n},\ \hat{\phi}_n + \lambda + \sqrt{c_\tau/n}\right]. \tag{20}$$

As pointed out by Imbens and Manski (2004), if the re-scaled length of the identified set 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 $\sqrt{c_\tau} \approx z_\tau$. If, on the other hand, the length of the identified set is zero (exact identification) or $\sqrt{n}\lambda$ is close to zero, then the boundaries of $\Theta(\hat{\phi}_n)$ have to be expanded by a two-sided critical value $\sqrt{c_\tau} \approx z_{\tau/2}$.

A comparison of the frequentist and the Bayesian interval leads to the relationship $CS^\theta_{B,\tau} \subset \Theta(\hat{\phi}_n) \subset CS^\theta_{F,\tau}$, as postulated in Corollary 1(iii) and Theorem 2. To see that the conditions of Theorem 2 are satisfied, choose, for instance, $\tilde{\theta} = 0$, $\tilde{\phi} = -\lambda$. Hence $\Phi(\tilde{\theta}) - \tilde{\phi} = [0, \lambda]$, which expands to $\mathcal{C} = \mathbb{R}^+$ if scaled by $\sqrt{n}$. Since $P\{\mathcal{Z} \in \mathbb{R}^+\} = 1/2$, a confidence set with coverage probability greater than 50% has to extend beyond $\Theta(\hat{\phi}_n)$.

A graphical comparison of the frequentist confidence intervals and Bayesian credible intervals is provided in Figure 1, assuming that $f(u) = \phi_{\mathcal{N}}(u)$. The two panels of the Figure are drawn for a data set in which $\hat{\phi}_n = 0$. We overlay sample sizes $n = 5$ and $n = 500$. The top panel depicts posterior densities $p(\theta|Y^n) = \hat{p}_{(i)}(\theta|Y^n)$ given in (15) and exact 90% HPD intervals, calculated numerically. The bottom panel depicts the standardized frequentist objective function $\frac{1}{n}Q_n(\theta)$ from Equation (18), the critical values $-c_\tau/n$ that solve (19), and 90% frequentist confidence intervals.

4.2 Moment Inequalities, Part II

Example 1 (continued): Frequentist Analysis with Integrated Likelihood. Previously, our frequentist analysis was based on a profile likelihood function, whereas the

Bayesian inference was based on an integrated likelihood function in which the nuisance parameter α was integrated out with respect to the prior distribution. We will now construct a frequentist confidence interval for θ based on the integrated objective function obtained with a prior (or weight function) $\alpha \sim \mathcal{U}[0, \lambda]$:

$$l_{n,int}(\theta) = \ln \int_{\alpha \in \mathcal{A}_\theta} \exp[l_n(G(\theta, \alpha))] d\alpha.$$

If $f(u) = \phi_{\mathcal{N}}(u)$ then $l_{n,int}(\theta) = \ln \hat{p}_{(i)}(\theta | Y^n)$ defined in (15). Now consider the distribution of $\exp[l_{n,int}(\theta)]$ near the boundaries of the identified set $\Theta(\phi) = [\phi, \phi + \lambda]$. If $\lambda > 0$ and the sample size is sufficiently large, then

$$P\left\{ l_{n,int}(\theta) \leq \ln t \right\} \approx \begin{cases} \Phi_{\mathcal{N}}\left(\Phi_{\mathcal{N}}^{-1}(\lambda t) - \sqrt{n}(\theta - \phi_0) \right) & \text{for small } |\theta - \phi_0| \\ \Phi_{\mathcal{N}}\left(\Phi_{\mathcal{N}}^{-1}(\lambda t) - \sqrt{n}(\theta - (\phi_0 + \lambda)) \right) & \text{for small } |\theta - (\phi_0 + \lambda)| \end{cases}.$$

Thus, we obtain the following approximation for level sets:

$$CS^{\theta}_{F,\tau} = \left\{ \theta \;\middle|\; l_{n,int}(\theta) \leq \ln(\tau/\lambda) \right\} = \left[\hat{\phi}_n - z_\tau/\sqrt{n} \;,\; \hat{\phi}_n + \lambda + z_\tau/\sqrt{n} \right], \tag{21}$$

where $z_\tau = \Phi_{\mathcal{N}}^{-1}(1 - \tau)$. From a comparison of (20) and (21) we deduce that the frequentist intervals constructed from the profile and the integrated likelihood function are approximately the same. In particular, the interval obtained from the profile likelihood function also has the property that it extends beyond $\Theta(\hat{\phi}_n)$. Thus, it is not the absence of a distribution over the identified set $\Theta(\phi)$, but rather the requirement that the $1 - \tau$ coverage probability is guaranteed for all $\theta \in \Theta(\phi)$ that leads to frequentist set to be larger than the Bayesian set.

Example 2: Unknown Length of Identified Set. We previously assumed that the length of the identified set is known and made a distinction between identified intervals of length zero and length greater than zero. Now suppose that the length itself depends on an unknown but estimable reduced form parameter $\phi_2 \geq 0$: $\theta - \phi_2 \leq \phi_1 \leq \theta$ and $\Theta(\phi) = [\phi_1, \phi_1 + \phi_2]$. This modified version of the inequality moment example is a stylized representation of the treatment effect model studied by Imbens and Manski (2004). The problem has been recently analyzed from a frequentist perspective in Stoye (2007). Let $\phi = [\phi_1, \phi_2]$ and assume that the prior distribution for θ given ϕ is uniform on $\Theta(\phi)$.

We previously derived the approximation of the posterior distribution under the assumption that the "true" reduced form parameter lies in the interior of the domain Φ. This assumption guaranteed that $\hat{\phi}_n$ is also in the interior and the score $\hat{Z}_n = l_n^{(1)}(\hat{\phi}_n) = 0$ eventually. To accommodate reduced form parameters on the boundary of Φ, that is $\phi_2 = 0$ in this example, the approximation in Theorem 1 can be modified as follows:

$$\left| \int_{\hat{J}_n^{1/2}(\Phi - \hat{\phi}_n)} H\left(\hat{\phi}_n + \hat{J}_n^{-1/2} s, \xi\right) dP^s_{Y^n} \right. \tag{22}$$

$$\left. - \int_{\hat{J}_n^{1/2}(\Phi - \hat{\phi}_n)} H\left(\hat{\phi}_n + \hat{J}_n^{-1/2} s, \xi\right) \frac{\phi_{\mathcal{N}}(s - \hat{J}_n^{-1/2}\hat{Z}_n)}{\int_{\hat{J}_n^{1/2}(\Phi - \hat{\phi}_n)} \phi_{\mathcal{N}}(s - \hat{J}_n^{-1/2}\hat{Z}_n) ds} ds \right| = o_p(n^{-1/2}).$$

If the score $\hat{Z}_n$ is non-zero and the Hessian $-\hat{J}_n$ is negative definite, the normal approximation of s has to be centered at $\hat{J}_n^{-1/2}\hat{Z}_n$ instead of zero. Moreover, the distribution of s is restricted to the set $\hat{J}_n^{-1/2}(\Phi - \hat{\phi}_n)$, which guarantees that the resulting posterior of ϕ has support on the domain Φ. The above approximation requires an additional assumption that guarantees that $\hat{J}_n^{-1/2}\hat{Z}_n$ is stochastically bounded.[4] Due to the behavior of the score $\hat{Z}_n$, the bound is only valid with probability approaching one, rather than almost surely. If we let $H(\phi, \xi) = P^{\theta}_{\phi}\{\theta \leq \xi\}$, which is given by

$$H(\phi, \xi) = \begin{cases} H_0(\phi, \xi) & \text{if } \phi_2 = 0 \\ H_+(\phi, \xi) & \text{if } \phi_2 > 0 \end{cases},$$

where

$$H_0(\phi, \xi) = \begin{cases} 0 & \text{if } \xi < \phi_1 \\ 1 & \text{if } \phi_1 \leq \xi \end{cases}, \quad H_+(\phi, \xi) = \begin{cases} 0 & \text{if } \xi < \phi_1 \\ \frac{\phi_1 - \xi}{\phi_2} & \phi_1 \leq \xi \leq \phi_1 + \phi_2 \\ 1 & \text{otherwise} \end{cases},$$

then (22) provides a large sample approximation to the posterior of θ that is valid regardless of whether the identified set has zero or non-zero length.

Example 3: Singleton $\Phi(\theta)$. In some models the set of structural parameters uniquely determines the reduced form parameters, that is $\Phi(\theta)$ is a singleton, while $\Theta(\phi)$ is set-valued. An example of such a model is a structural vector autoregression, in which ϕ corresponds to the regression coefficients and the non-redundant elements of the variance-covariance matrix of the one-step-ahead forecast errors. The vector θ corresponds to a collection of structural impulse response functions. These impulse responses depend in addition to ϕ on a non-identifiable orthonormal matrix that rotates orthogonalized one-step-ahead forecast errors into a vector of structural shocks.

Consider the following modification of Example 1: $\theta = [\theta_1, \theta_2]'$, $\theta_1 - \lambda \leq \phi \leq \theta_1$, and $\theta_2 = \theta_1 - \phi$. Notice that the slackness parameter α that arose in Example 1 is now called

---

[4] Classical analysis is typically based on the assumption that $J_{n,0}^{-1/2} Z_{n,0} = O_p(1)$. Thus, stochastic equicontinuity of the standardized score process $J_n^{-1/2}(\phi) l_n^{(1)}(\phi)$ would suffice to ensure that $\hat{J}_n^{-1/2}\hat{Z}_n = O_p(1)$.

θ_2 and part of the structural parameter vector θ. $\Theta(\phi)$ is located in a one-dimensional subspace of Θ and remains set-valued, while $\Phi(\theta) = \theta_1 - \theta_2$ is a singleton. The projections of the identified set $\Theta(\phi)$ on the domains of θ_1 and θ_2 are given by $\Theta_1(\phi) = [\phi, \phi + \lambda]$ and $\Theta_2(\phi) = [0, \lambda]$ We maintain that conditional on ϕ the prior for θ_1 is $\mathcal{U}[\phi, \phi + \lambda]$, which implies the prior on $\Theta_2(\phi)$ is also uniform.

Using the same arguments as for Example 1, replacing α by θ_2, one can deduce that the set

$$CS^{\theta_1}_{B,\tau} = [\hat{\phi}_n + \tau\lambda/2, \hat{\phi}_n + \lambda - \tau\lambda/2]$$

is an asymptotically valid $1 - \tau$ credible set for θ_1. Similarly, $CS^{\theta_2}_{B,\tau} = [\tau\lambda/2,\ \lambda - \tau\lambda/2]$ is a $1 - \tau$ credible set for θ_2. Likewise, the set characterized in (20) is a valid $1 - \tau$ frequentist confidence set for θ_1. Thus, the lessons learned from Example 1 still apply to inference about the θ_1 element of the θ vector.

More interestingly, we will now consider inference for the vector θ. Consider the following subsets of Θ:

$$\mathcal{T}_{\xi,n} = \left\{ \theta_1, \theta_2 \;\middle|\; \hat{\phi}_n - \xi_1/\sqrt{n} \leq \theta_1 - \theta_2 \leq \hat{\phi}_n + \xi_1/\sqrt{n},\ \lambda\xi_2/2 \leq \theta_2 \leq \lambda(1 - \xi_2) \right\} \tag{23}$$

and define

$$H_n(\phi, \xi) = P^{\theta}_{\phi}\{\theta \in \mathcal{T}_{\xi,n}\} = \begin{cases} 1 - \xi_2 & \text{if } \hat{\phi}_n - \xi_1/\sqrt{n} \leq \phi \leq \hat{\phi}_n + \xi_1/\sqrt{n} \\ 0 & \text{otherwise} \end{cases}$$

Since the sequence of functions $H_n(\phi, \xi)$ does not satisfy the Lipschitz condition in Assumption 4 we use the following approximation

$$\hat{P}^{\theta}_{Y^n(i)}\{\theta \in \mathcal{T}_{\xi,n}\} = (1 - \xi_2) \int_{-\xi_1/\sqrt{n}}^{\xi_1/\sqrt{n}} \phi_{\mathcal{N}}(s) ds = (1 - \xi_2)(1 - 2\Phi_{\mathcal{N}}(\xi_1)).$$

Thus, an asymptotically valid credible set can be obtained by choosing $\xi_1 \geq 0$ and $0 \leq \xi_2 \leq 1$ such that $(1 - \xi_2)(1 - 2\Phi_{\mathcal{N}}(\xi_1)) = 1 - \tau$. It turns out that the volume of the $1 - \tau$ credible set is minimized by setting $\xi_2 = 0$ and $\xi_1 = z_{\tau/2}$. The set $\Theta(\hat{\phi}_n)$, which is obtained from $\mathcal{T}_{\xi,n}$ in (23) by setting $\xi_1 = 0$ and $\xi_2 = 0$, is not an asymptotically valid credible set – it is too small. However, since one can construct valid credible sets with $\xi_2 > 0$, it is not the case that $\Theta(\hat{\phi}_n)$ is nested in every asymptotically valid credible set. In this example $\Theta(\hat{\phi}_n)$ happens to be nested in the $1 - \tau$ credible set with the smallest volume among the $\mathcal{T}_{\xi,n}$ sets.

Now consider the following construction of a frequentist confidence interval for θ. Since $\Phi(\theta)$ is a singleton, we can express $\phi = G(\theta)$, without having to introduce an α. Moreover, the natural relationship between the domains Φ and Θ is: $\Phi = \{\phi \mid \phi = G(\theta), \theta \in \Theta\}$. Thus,

$$\begin{aligned} & \inf_{\phi \in \Phi} \inf_{\theta \in \Theta(\phi)} P_{\phi}^{Y^n}\{2[l_n(G(\theta)) - l_n(\hat{\phi}_n)] \geq -c_\tau\} \\ &= \inf_{\theta \in \Theta} \inf_{\phi \in \Phi(\theta)} P_{\phi}^{Y^n}\{2[l_n(G(\theta)) - l_n(\hat{\phi}_n)] \geq -c_\tau\} \\ &= \inf_{\theta \in \Theta} P_{G(\theta)}^{Y^n}\{2[l_n(G(\theta)) - l_n(\hat{\phi}_n)] \geq -c_\tau\} \\ &= \inf_{\phi \in \Phi} P_{\phi}^{Y^n}\{2[l_n(\phi) - l_n(\hat{\phi}_n)] \geq -c_\tau\} \end{aligned} \tag{24}$$

Using the large sample approximation described in Section 2 one can obtain an asymptotically valid confidence set CS^ϕ that takes the form of the level set (3). In our example:

$$CS^\phi = [\hat{\phi}_n - z_{\tau/2}/\sqrt{n},\ \hat{\phi}_n + z_{\tau/2}/\sqrt{n}].$$

According to (24) the corresponding confidence set for θ is given by $CS^\theta_{F,\tau} = \bigcup_{\phi \in CS^\phi} \Theta(\phi)$. This set equals $\mathcal{T}_\xi$ in (23) for $\xi_1 = z_{\tau/2}$ and $\xi_2 = 0$. Thus, the frequentist confidence set is identical to Bayesian credible set that has the smallest volume among the $\mathcal{T}_\xi$ sets.

5 A Numerical Example: Bayesian Analysis of a Two-Player Entry Game

At last, 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 (2007). Rather than directly working with the asymptotic approximation derived in Section 3.1, we will use Markov-Chain Monte Carlo techniques to generate posterior draws of θ for a small ($n = 50$) and a large ($n = 1,000$) sample. We will focus on a fairly simple version of the entry game without firm-specific regressors. Depending on the entry decision of the second firm, Firm l either does not enter market i, operates as monopolist, or operates as duopolist. Potential monopoly (M) and duopoly (D) profits are given by

$$\pi^M_{i,l} = x_i'\beta_l + \epsilon_{i,l}, \quad \pi^D_{i,l} = x_i'\beta_l - \gamma_l + \epsilon_{i,l}, \quad l = 1, 2 \quad i = 1, \dots, n \tag{25}$$

The $\epsilon_{i,l}'s$ capture latent profit components that are known to the two firms but unobserved by the econometrician and x^i is a vector of observable market characteristics. We assume that the outcome of the entry game in each market is a pure strategy Nash equilibrium. It is straightforward to verify that the Nash equilibrium is unique, except if both firms are profitable as monopolist but not as duopolist. In the latter case, the model is silent about which firm actually enters the market. As a consequence, the model only delivers bounds for the probability of observing a particular monopoly.

Suppose that $\epsilon_{i,l} \sim iid\mathcal{N}(0,1)$ and let $\theta = [\beta_1', \gamma_1, \beta_2', \gamma_2]'$. Using (25) it is straightforward to calculate probabilities that firm l is profitable as monopolist (duopolist) in market i. For $x_i = x$ we denote these probabilities by $\mu_l(\theta, x)$ and $\delta_l(\theta, x)$, respectively. Moreover, we use

$$\phi(x) = [\phi_{00}(x),\ \phi_{01}(x),\ \phi_{10}(x),\ \phi_{11}(x)]'$$

to denote the reduced form probabilities of observing no entry, entry of Firm 1, entry of Firm 2, or entry of both firms in a market with characteristics x. We observe no entry if neither firm is profitable as monopolist, we observe a duopoly if both firms are profitable as duopolists. An upper bound on the probability that Firm 1 operates as monopolist is given by the probability that Firm 1 is profitable as monopolist and Firm 2 is not profitable as duopolist. The lower bound is given by the sum of the probability that Firm 1 is profitable as monopolist and Firm 2 is not profitable as monopolist and of the probability that Firm 1 would be profitable as duopolist, but Firm 2 would only be profitable as monopolist. Formally,

$$\phi_{00}(x) = (1 - \mu_1(x))(1 - \mu_2(x)) \tag{26}$$

$$\phi_{11}(x) = \delta_1(x)\delta_2(x) \tag{27}$$

$$\phi_{10}(x) \leq \mu_1(x)(1 - \delta_2(x)) \tag{28}$$

$$\phi_{10}(x) \geq \mu_1(x)(1 - \mu_2(x)) + \delta_1(x)(\mu_2(x) - \delta_2(x)). \tag{29}$$

It can be verified that the Nash equilibrium restriction for a Firm 2 monopoly does not add any further restrictions on the reduced form probabilities.

In order to be able to uniquely determine the reduced form parameters as a function of the probabilities μ_i and δ_i, we introduce for each x an auxiliary parameter $\alpha(x) \in [0,1]$ that captures the slackness in the inequality restrictions for $\phi_{10}(x)$:

$$\phi_{10}(x) = \mu_1(x)(1-\mu_2(x)) + \delta_1(x)(\mu_2(x)-\delta_2(x)) + \alpha(x)(\mu_1(x)-\delta_1(x))(\mu_2(x)-\delta_2(x)). \tag{30}$$

The second term, which is pre-multiplied by α, can be interpreted as the probability that both firms are profitable as monopolists but not as duopolists. Consequentially, the slackness can be viewed as the probability of a sunspot shock that selects Firm 1 if the Nash equilibrium is not unique. Equations (26), (27), and (30) define the function $G(\theta, \alpha)$.

For the large sample theory presented in Section 3 to be applicable to the entry game we need to assume that the regressor x is discretized. The discretization ensures that the reduced-form parameter vector ϕ is finite dimensional and is not uncommon in the empirical literature. These regressors are assumed to take only finitely many values. In the subsequent numerical illustration we only use an intercept as regressor.

We proceed in several steps: (i) we specify a data generating process by choosing "true" values of θ and α, which imply a "true" ϕ. (ii) Instead of specifying a prior distributions P^{ϕ} and P^{θ}_{ϕ}, we start from a prior on θ and α and generate draws from the implied distributions P^{ϕ} and P^{θ}_{ϕ}. (iii) Finally, will generate two samples of size $n = 50$ and $n = 1,000$ and compare the posterior distributions.

The parameterization of the data generating process is summarized in the second column of Table 1. The probabilities of a a Firm 1 monopoly, and Firm 2 monopoly, and a duopoly are 48%, 33%, and 12%, respectively. The third column of Table 1 specifies the prior distributions. We use fairly diffuse Gaussian priors for the elements of the θ vector. The distributions of γ_1 and γ_2 are truncated at zero to ensure that duopoly profits are less than monopoly profits. The auxiliary parameter α has support on the unit interval. We consider three different priors, centered at 0.2 (low α), 0.5 (Benchmark), and 0.8 (high α), respectively. By evaluating the function $G(\theta, \alpha)$ at random draws from the prior distribution of θ and α we obtain draws from the prior distribution of ϕ. Means and standard deviation are reported in the last four rows of Table 1 under the Benchmark prior for α.

According to our previous analysis the prior distribution of θ given ϕ plays an important role in Bayesian inference for partially identified models. We depict unconditional prior densities as well as prior densities conditional on the "true" value of ϕ in Figure 2. Except for α, the unconditional prior densities are essentially invisible because they are very diffuse compared to the conditional priors. While in a fully identified model the prior P^{θ}_{ϕ} should be a pointmass at the singleton $\theta(\phi)$, the entry game model is partially identified and leads to a non-degenerate P^{θ}_{ϕ}. The prior distribution on α induces a prior distribution for the profit function parameters given the reduced form entry probabilities. Figure 2 illustrates how P^{θ}_{ϕ} shifts as one changes the prior for α. While the prior for α could in principle be correlated with the prior for θ, for instance to reflect the belief that the firm with higher expected monopoly profits is more likely to enter the market if the equilibrium is not unique, we will treat α and θ as independent.

We now generate samples of $n = 50$ and $n = 1,000$ observations from the data generating process and us a random-walk Metropolis Algorithm to generate draws from the posterior of θ and α. Using the relationship $\phi = G(\theta, \alpha)$ we convert the θ-α draws into ϕ draws. Figure 3 indicates that after 50 observations there is still substantial uncertainty about the reduced form parameters. Since we specified the prior distribution for ϕ implicitly through a prior for θ and α, changes in the prior for α can in principle affect the prior and posterior of ϕ. However, according to Figure 3 this effect is negligible in our illustration. Figure 4 depicts posterior densities for the profit function parameters β and δ. While the prior distribution of α and hence P_ϕ^θ has some effect on the posterior, overall the posterior distribution is dominated by the uncertainty about the reduced form parameter. Finally, the two panels of Figure 5 show scatter plots of draws from the posterior distribution of β_1 and γ_1. Moreover, we outline the projection of the identified set $\Theta(\hat{\phi}_n)$ onto the domain of β_1 and γ_1. Here $\hat{\phi}_n$ is the posterior mean of the reduced form parameter vector ϕ. According to our asymptotic theory, the posterior distribution concentrates near $\Theta(\hat{\phi}_n)$, which is evident from the posterior draws obtained with $n = 1,000$.

6 Conclusion

We derived a large sample approximation for the posterior distribution of a structural parameter vector in a partially identified model to compare Bayesian credible sets and frequentist confidence sets. Unlike in regular models, Bayesian and frequentist set estimates differ not just with respect to their philosophical underpinnings. Frequentist confidence intervals have to extend beyond the boundaries of the identified set (conditional on the estimated reduced form parameter), whereas Bayesian credible sets can be be located in the interior of the identified set asymptotically. The main challenge to frequentist inference is to establish the uniform validity of the set estimate. The main challenge to Bayesian inference is to control the shape of the prior distribution on the identified set conditional on the reduced form parameter to avoid highly informative priors on the identified set induced by nonlinearities of parameter transformations and to document the sensitivity of posterior inference to the choice of prior even in large samples.

Appendix

The proof of Theorem 1 will closely follow the arguments in Johnson (1970). We rewrite the posterior density of ϕ as

$$
\begin{aligned}
q^{\phi}(\phi|Y^n) &= \begin{cases} \frac{p(\phi)\exp[l_n(\phi)]}{p(\hat{\phi}_n)\exp[l_n(\hat{\phi}_n)]} & \text{if } \phi \in \Phi \\ 0 & \text{otherwise} \end{cases} \\
p^{\phi}(\phi|Y^n) &= \frac{q^{\phi}(\phi|Y^n)}{\int_{\mathbb{R}^K} q^{\phi}(\phi|Y^n)d\phi}.
\end{aligned}
$$

Define $\hat{\Sigma}_n = n^{1/2}\hat{J}_n^{-1/2}$ and $z = \hat{\Sigma}_n^{-1}(\phi - \hat{\phi}_n) \in \Phi_z$. Moreover, let $q^z(z|Y^n) = q^{\phi}(\hat{\phi}_n + \hat{\Sigma}_n z)$. Then the posterior density of z can be expressed as

$$
p^z(z|Y^n) = \frac{q^z(z|Y^n)}{\int_{\mathbb{R}^K} q^z(z|Y^n)dz}.
$$

Let Ω_0 and Ω_1 denote the sure sets for which Assumptions 1 and 2 hold, respectively. We begin by introducing several Lemmas that are useful for the proof of the theorem.

Lemma 1 *Suppose that Assumptions 1 – 2 hold. Fix a constant κ_1 with $0 < \kappa_1 < 1$. Then, one can choose a constant $\delta_1 > 0$ and, for each $\omega \in \Omega_1$ a constant $N_{1\omega}$ such that the following statements hold: (a) If $n \geq N_{1\omega}$ and $\|\phi - \phi_0\| \leq \delta_1$, then there exists finite constants $M_{\min}$ and $M_{\max}$ such that*

$$
0 < M_{\min} \leq \lambda_{\min}(n^{-1}J_n(\phi)) \leq \lambda_{\max}(n^{-1}J_n(\phi)) \leq M_{\max} < \infty.
$$

(b) If $n \geq N_{1\omega}$, then

$$
0 < M_{\min} \leq \lambda_{\min}(n^{-1}\hat{J}_n) \leq \lambda_{\max}(n^{-1}\hat{J}_n) \leq M_{\max} < \infty
$$

and (c)

$$
0 < M_{\max}^{-1} \leq \lambda_{\min}(n^{-1}\hat{J}_n) \leq \lambda_{\max}(n^{-1}\hat{J}_n) \leq M_{\min}^{-1} < \infty.
$$

Lemma 2 (Lemma 2.2 in Johnson) *Suppose that Assumptions 1 – 2 hold. Then, we can choose a constant δ_2 $(0 < \delta_2 < 1)$, a constant $\kappa_2 < \frac{1}{2}$, and, for each $\omega \in \Omega_1$, a constant $N_{2\omega}$ $(\geq N_{1\omega})$ such that if $n \geq N_{2\omega}$ and $\|z\| \leq \delta_2$, then*

$$
\frac{1}{n}l_n\left(\hat{\phi}_n + \hat{\Sigma}_n z\right) - \frac{1}{n}l_n\left(\hat{\phi}_n\right) \leq -\kappa_2 \|z\|^2.
$$

Lemma 3 (Lemma 2.3 in Johnson) *Suppose that Assumptions 1 – 2 hold. Suppose that $\delta > 0$ is given. Then, we can choose a constant $\kappa_3 > 0$ and, for each $\omega \in \Omega_1$, a constant $N_{3\omega}$ $(\geq N_{2\omega})$ such that whenever $n \geq N_{3\omega}$ and $\|z\| \geq \delta$, we have*

$$
\frac{1}{n}l_n\left(\hat{\phi}_n + \hat{\Sigma}_n z\right) - \frac{1}{n}l_n\left(\hat{\phi}_n\right) \leq -\kappa_3.
$$

Now define (our definition differs from Johnson's)

$$p_1(\phi - \hat{\phi}_n; \hat{\phi}_n) = 1 + \frac{p^{(1)}(\hat{\phi}_n)'}{p(\hat{\phi}_n)}(\phi - \hat{\phi}_n).$$

Since $\hat{\phi}_n \to \phi_0$ *a.s.* and $p(\phi_0) > 0$, $p(\hat{\phi}_n) > 0$ near ϕ_0.

Lemma 4 (Lemma 2.4 in Johnson) *Suppose that Assumptions 1 – 3 hold. Then, there exists a constant δ_4,a constant M, and, for each $\omega \in \Omega_0$, a constant $N_{4\omega}(> N_{3\omega})$ such that if $n \geq N_{4\omega}$, then*

$$\int_{\|z\|\leq\delta_4} \left| q^z\left(z|Y^n\right) - p_1\left(\hat{\Sigma}_n z; \hat{\phi}_n\right)\exp\left(-\frac{1}{2}nz'z\right)\right| dz \leq \frac{M}{n}.$$

Lemma 5 *Suppose Assumptions 1 – 3 are satisfied. Let Y^n be in the sure set of Assumptions 1 and 2. Then, there exist a finite constant M and a finite constant N such that whenever $n \geq N$ we have*

$$\left|\int_{\mathbb{R}^K}\left[q^z(z|Y^n) - \exp\left(-\frac{1}{2}nz'z\right)\right]dz\right| \leq \frac{M}{n}.$$

Proof of Lemma 5 For a given $\omega \in \Omega_0$, choose $N_\omega \geq N_{4\omega}$ in Lemma 4 such that when $n \geq N_{4\omega}$, the statements of Lemmas 1, 2, 3, and 4 hold, and for the δ_p in Assumption 3, $\|\hat{\phi}_n - \phi_0\| \leq \delta_p$ by Lemma 1 in Wu (1981). We bound

$$\begin{aligned}
&\int_{\mathbb{R}^K}\left|q^z\left(z|Y^n\right) - \exp\left(-\frac{1}{2}nz'z\right)\right|dz \\
\leq\; & \int_{\mathbb{R}^K}\left|q^z\left(z|Y^n\right) - p_1\left(\hat{\Sigma}_n z; \hat{\phi}_n\right)\exp\left(-\frac{1}{2}nz'z\right)\right|dz \\
&+ \int_{\mathbb{R}^K}\left|1 - p_1\left(\hat{\Sigma}_n z; \hat{\phi}_n\right)\right|\exp\left(-\frac{1}{2}nz'z\right)dz \\
=\; & I + II
\end{aligned}$$

Term I can be bounded by

$$\begin{aligned}
III + IV + IV \;=\; & \int_{\|z\|\leq\delta_4}\left|q^z\left(z|X_n\right) - p_1\left(\hat{\Sigma}_n z; \hat{\phi}_n\right)\exp\left(-\frac{1}{2}nz'z\right)\right|dz \\
+\; & \int_{\|z\|>\delta_4} q^z\left(z|Y^n\right)dz + \int_{\|z\|>\delta_4}\left|p_1\left(\hat{\Sigma}_n z; \hat{\phi}_n\right)\right|\exp\left(-\frac{1}{2}nz'z\right)dz,
\end{aligned}$$

where δ_4 is defined in Lemma 4. The $O(n^{-1})$ bound for III follows directly from Lemma 4. Now consider term IV:

$$\begin{aligned}
IV \;=\; & \int_{\|z\|>\delta_4}\frac{p(\hat{\phi}_n + \hat{\Sigma}_n z)}{p(\hat{\phi}_n)}\exp\left(l_n(\hat{\phi}_n + \hat{\Sigma}_n z) - l_n(\hat{\phi}_n)\right)dz \\
\leq\; & M\exp(-\kappa_3 n) \leq \frac{M}{n}.
\end{aligned}$$

The bound follows from Assumption 3 and Lemma 3. Finally, to obtain a bound for term V, by Assumption 3 with $\|\hat{\phi}_n - \phi_0\| \leq \delta_p$ and by Lemma 1(c), we can choose M such that

$$\left|p_1(\hat{\Sigma}_n z;\hat{\phi}_n)\right| \leq 1 + \left\|\frac{p^{(1)}(\hat{\phi}_n)}{p(\hat{\phi}_n)}\right\| \left\|\hat{\Sigma}_n z\right\| \leq 1 + M\left\|z\right\|.$$

Then,

$$V \leq \int_{\|z\|>\delta_4} \exp\left(-\frac{1}{2}n\left\|z\right\|^2\right) dz + M\int_{\|z\|>\delta_4} \left\|z\right\| \exp\left(-\frac{1}{2}n\left\|z\right\|^2\right) dz \leq \frac{M}{n}.$$

Combining the bounds for terms III, IV, and V, we have

$$I \leq \frac{M}{n}.$$

For the term II, from the definition $p_1\left(\hat{\Sigma}_n z;\hat{\phi}_n\right)$ and (6) and by change of variable $v = \sqrt{n}\left\|z\right\|$, we have

$$II \leq M\int_{\mathbb{R}^K} \left\|z\right\| \exp\left(-\frac{1}{2}n\left\|z\right\|^2\right) dz \leq \frac{M}{n}\int_0^\infty \exp\left(-\frac{1}{2}v^2\right) v^K dv \leq \frac{M}{n},$$

as required for the lemma. ■

Proof of Theorem 1(i): For $s = \sqrt{n}z \in \Phi_s$ the posterior is $p^s(s|Y^n) = \sqrt{n}p^z(\sqrt{n}z|Y^n)$. We now abbreviate $H(\hat{\phi}_n + \hat{J}_n^{-1/2}s,\xi) = H(s,\xi)$. Then,

$$\int_{\Phi_s} H(s,\xi)dP^s_{Y^n} - \int_{\mathbb{R}^K} H(s,\xi)d\Phi_{\mathcal{N}}(s) = \int_{\mathbb{R}^K} H(n^{1/2}z,\xi)[p^z(z|Y^n) - n^{1/2}\phi_{\mathcal{N}}(n^{1/2}z)]dz.$$

To prove the theorem it suffices to show that

$$\sqrt{\frac{2\pi}{n}}\left|\int_{\mathbb{R}^K} H(n^{1/2}z,\xi)[p^z(z|Y^n) - n^{1/2}\phi_{\mathcal{N}}(n^{1/2}z)]dz\right| \leq \frac{M}{n}.$$

Consider the following bound

$$\begin{aligned}
&\sqrt{\frac{2\pi}{n}}\left|\int_{\mathbb{R}^K} H(n^{1/2}z,\xi)\left(p^z(z|Y^n) - n^{1/2}\phi_{\mathcal{N}}\left(n^{1/2}z\right)\right)dz\right| \\
&= \left|\int_{\mathbb{R}^K} H(n^{1/2}z,\xi)\left(\frac{\sqrt{2\pi/n}\ \ q^z(z|Y^n)}{\int_{\mathbb{R}^K} q^z\left(z|Y^n\right)dz} - \exp\left(-\frac{1}{2}nz'z\right)\right)dz\right| \\
&\leq \left|\int_{\mathbb{R}^K} H(n^{1/2}z,\xi)q^z\left(z|Y^n\right)\left(1 - \frac{\sqrt{2\pi/n}}{\int_{\mathbb{R}^K} q^z\left(z|Y^n\right)dz}\right)dz\right| \\
&\quad + \left|\int_{\mathbb{R}^K} H(n^{1/2}z,\xi)\left(q^z\left(z|Y^n\right) - \exp\left(-\frac{1}{2}nz'z\right)\right)dz\right| \\
&= I + II, \text{ say.}
\end{aligned}$$

Since $|H(n^{1/2}z,\xi)| < M_H$, the first term can be bounded by

$$\begin{aligned} I &\leq M_H \left| \int_{\mathbb{R}^K} q^z(z|Y^n)\,dz - \sqrt{\frac{2\pi}{n}} \right| \\ &= M_H \left| \int_{\mathbb{R}^K} q^z(z|Y^n)\,dz - \int_{\mathbb{R}^K} \exp\left(-\frac{1}{2}nz'z\right) dz \right| \\ &\leq M_H \int_{\mathbb{R}^K} \left| q^z(z|Y^n) - \exp\left(-\frac{1}{2}nz'z\right) \right| dz \\ &\leq \frac{M}{n}. \end{aligned}$$

The third inequality follows from Lemma 5. The bound for term II can be obtained in a similar manner. ■

Proof of Corollary 1(ii): Consider the following bound:

$$\begin{aligned} &\left| P^{\theta}_{Y^n}\{\theta \in \mathcal{T}_{\xi,n}\} - H_n(\hat{\phi}_n,\xi) \right| \\ &\leq \left| P^{\theta}_{Y^n}\{\theta \in \mathcal{T}_{\xi,n}\} - \int_{\mathbb{R}^K} H_n(\hat{\phi}_n + \hat{J}_n^{-1/2}s,\xi)\phi_{\mathcal{N}}(s)ds \right| \\ &\quad + \left| \int_{\mathbb{R}^K} \left[H_n(\hat{\phi}_n + \hat{J}_n^{-1/2}s,\xi) - H_n(\hat{\phi}_n,\xi) \right] \phi_{\mathcal{N}}(s)ds \right| \\ &= I + II, \text{ say.} \end{aligned}$$

Theorem 1 provides a bound for I. Using the Lipschitz assumption we deduce

$$II \leq M^*(\xi) \left| \int_{\mathbb{R}^K} \|\hat{J}_n^{-1/2}s\| \phi_{\mathcal{N}}(s)ds \right| \leq n^{-1/2} M^*(\xi) \left| \int_{\mathbb{R}^K} \|\hat{\Sigma}_n\| \|s\| \phi_{\mathcal{N}}(s)ds \right| \leq \frac{M(\xi)}{\sqrt{n}}.$$

The last inequality is a consequence of Lemma 1. ■

References

Andrews, Donald, Steven Berry, and Panle Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," *Manuscript*, Department of Economics, Yale University.

Andrews, Donald and Patrik Guggenberger (2007): "Validity of Subsampling and 'Plug-in Asymptotics' Inference for Parameters Defined by Moment Inequalities," *Manuscript*, Department of Economics, Yale University.

Andrews, Donald and Gustavo Soares (2007): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Manuscript*, Department of Economics, Yale University.

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

Beresteanu, Arie and Francesca Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, **76**, 763-814.

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.

Canay, Ivan A. (2007): "EL Inference for Partially Identified Models: Large Deviation Optimality and Bootstrap Validity," *Manuscript*, Department of Economics, University of Wisconsin.

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, 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 (2007): "Market Structure and Multiple Equilibria in Airline Markets," *Manuscript*, Department of Economics, University of Virginia and Northwestern University.

Galichon, Alfred and Marc Henry (2006): "Inference in Incomplete Models," *Manuscript*, Columbia University.

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

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.

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

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

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

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

Pakes, Ariel, Jack Porter, Kate Ho and Joy Ishi (2005): "Moment Inequalities and Their Application," *Manuscript*, Department of Economics, Harvard University.

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 (2006): "Inference for Partially Identified Econometric Models," *Manuscript*, Stanford University.

Rosen, Adam (2005): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Manuscript*, Department of Economics, University College London.

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

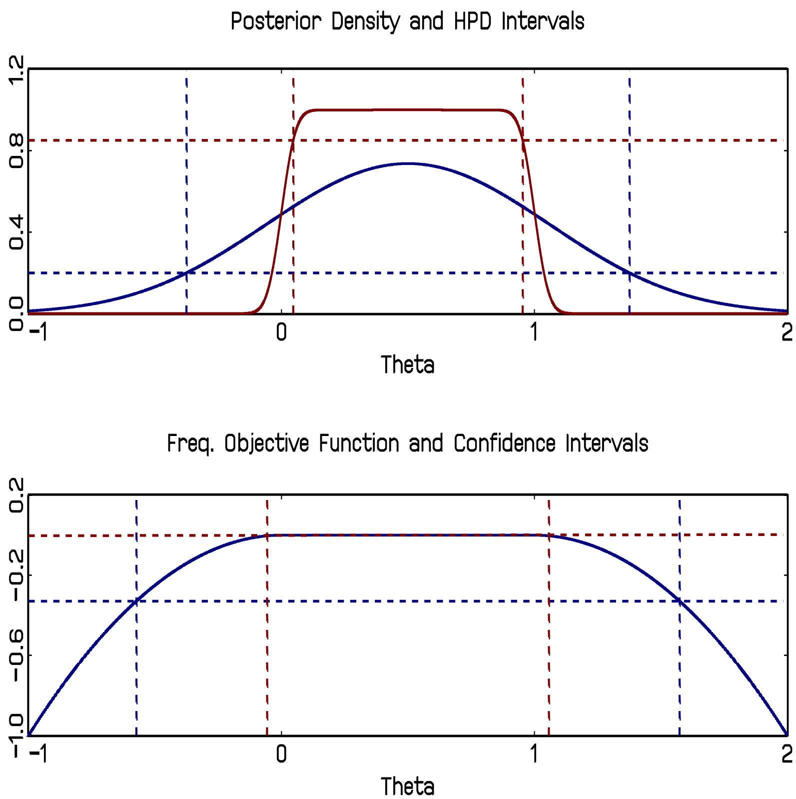
Stoye, Jörg (2007): "More on Confidence Intervals for Partially Identified Parameters," *Manuscript,* Department of Economics, New York University.

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.

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

Wu, Chien-Fu, (1981): "Asymptotic Theory of Nonlinear Least Squares Estimation," *Annals of Statistics*, **9**, 501-513.

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



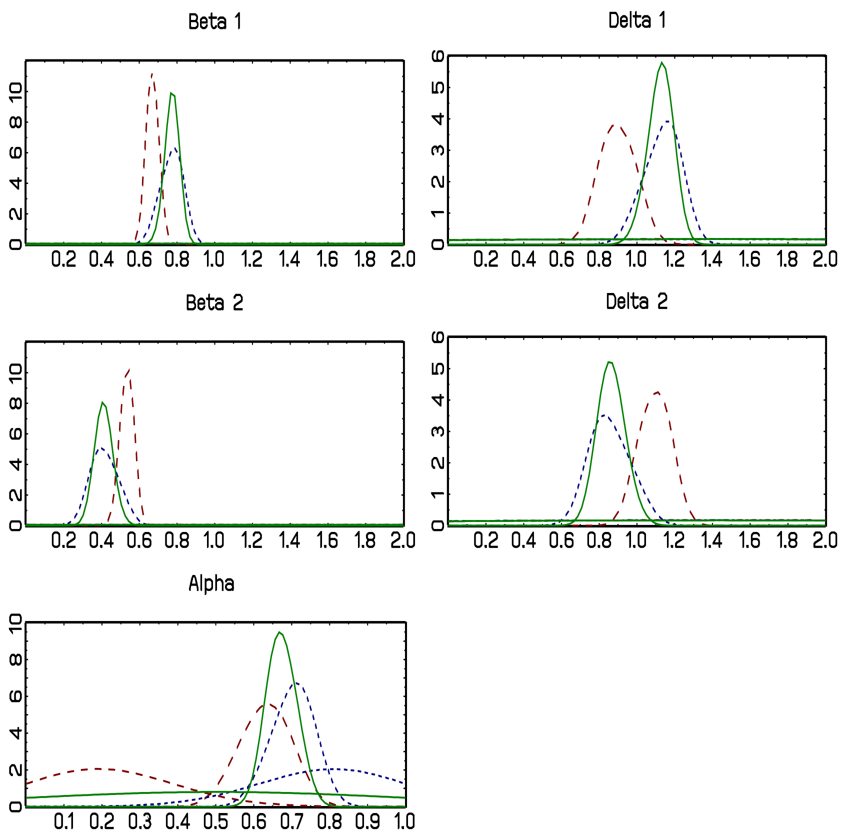
Notes: The figures are drawn for $\hat{\phi}_n = 0$ and overlay $n = 5$ and $n = 500$. The top panel depicts posterior densities $p(\theta|Y^n)$ and 90% credible intervals. The bottom panel depicts the standardized frequentist objective function $\frac{1}{n}Q_n(\theta)$, the cut-off value c_τ/n for $\tau = 0.1$, and 90% frequentist confidence intervals.

Table 1: Entry Game: "True" Parameters and Prior

Parameter	True Value	Prior Distribution
Structural Parameters θ		
β_1	0.7	$\mathcal{N}(0, 4^2)$
γ_1	1.0	$\mathcal{N}_+(0, 4^2)$
β_2	0.5	$\mathcal{N}(0, 4^2)$
γ_2	1.0	$\mathcal{N}_+(0, 4^2)$
Auxiliary Parameter α		
α	0.7	Benchmark: $\mathcal{B}(0.5, 0.2^2)$
	0.7	Low α: $\mathcal{B}(0.2, 0.1^2)$
	0.7	High α: $\mathcal{B}(0.8, 0.1^2)$
Implied Reduced Form Parameters ϕ		
ϕ_{00}	0.07	$\mu_{00} = 0.25$, $\sigma_{00} = 0.37$
ϕ_{10}	0.48	$\mu_{10} = 0.31$, $\sigma_{10} = 0.40$
ϕ_{01}	0.33	$\mu_{01} = 0.31$, $\sigma_{01} = 0.40$
ϕ_{11}	0.12	$\mu_{11} = 0.13$, $\sigma_{11} = 0.28$

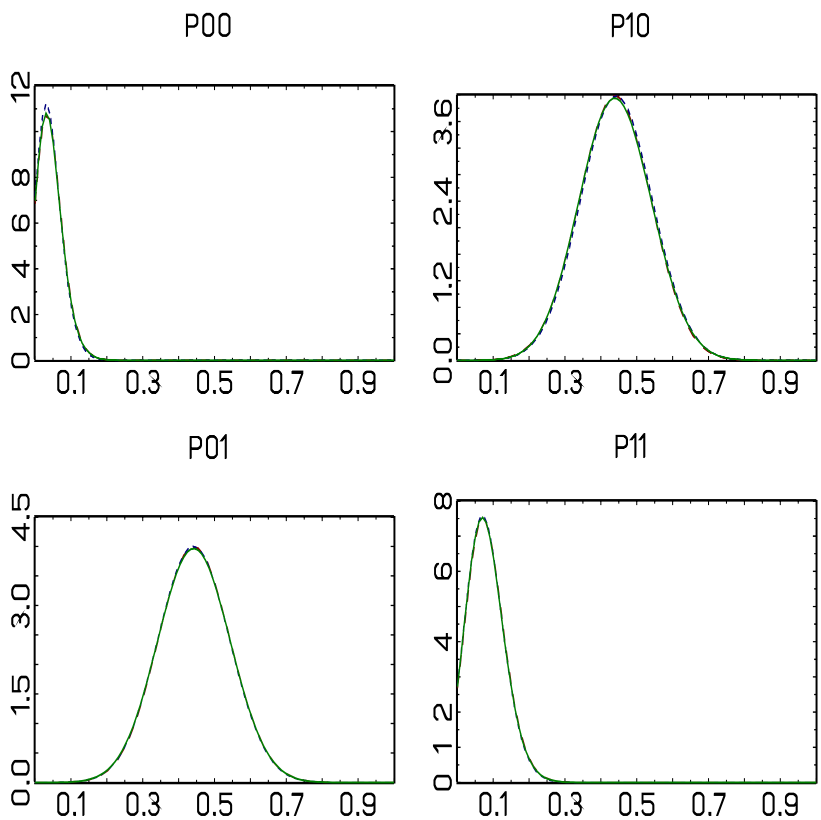
Notes: for the prior distribution of the reduced form parameters we report means μ and standard deviations σ under $\alpha \sim \mathcal{B}(0.5, 0.2^2)$. $\mathcal{N}(\nu, \sigma^2)$ and $\mathcal{B}(\mu, \sigma^2)$ refer to Normal and Beta distributions with mean μ and variance σ^2.

Figure 2: Conditional Distribution of θ Given ϕ



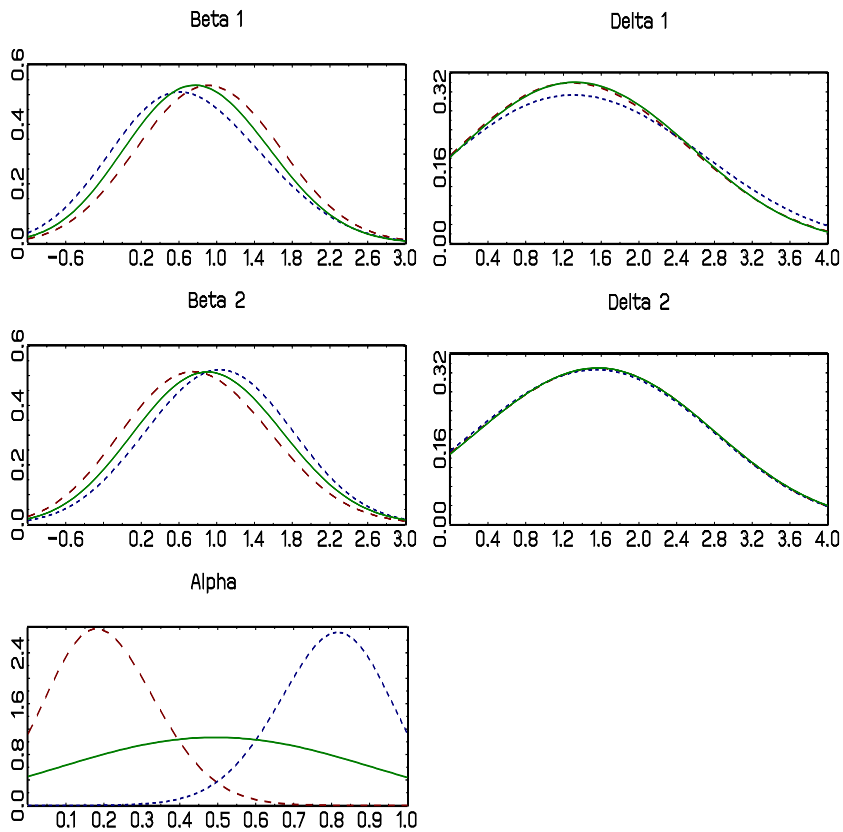
Notes: Benchmark Prior (solid, green), Low α Prior (long dashes, red), High α Prior (short dashes, blue). Each panel depicts 3 unconditional prior densities and 3 densities conditional on the "true" ϕ. Except for α the unconditional prior densities appear invisible because they are very diffuse compared to the conditional densities.

Figure 3: Posterior Distribution of ϕ, $n = 50$



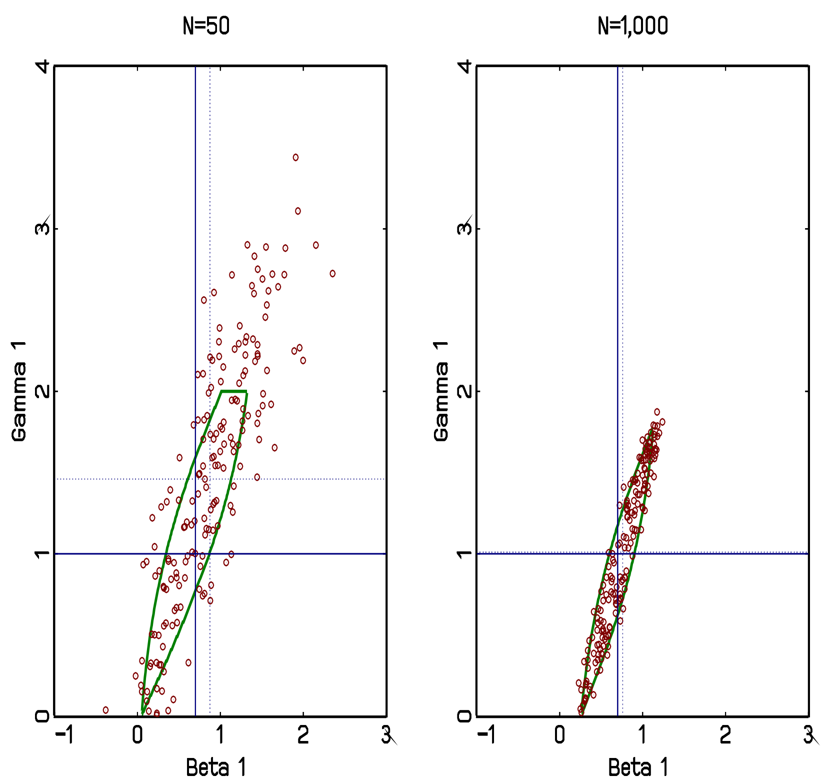
Notes: Benchmark Prior (solid, green), Low α Prior (long dashes, red), High α Prior (short dashes, blue). Since the posterior of ϕ is insensitive to the prior on α the three densities appear on top of each other.

Figure 4: Posterior Distribution of θ, $n = 50$



Notes: Benchmark Prior (solid, green), Low α Prior (long dashes, red), High α Prior (short dashes, blue).

Figure 5: Posterior Distribution of β_1 and γ_1



Notes: The panels depict draws from the posterior distribution and an outline of the projection of $\Theta(\hat{\phi}_n)$ onto the β_1-γ_1 space.