Supplement to "Bayesian inference in a class of partially identified models"

BRENDAN KLINE
Department of Economics, University of Texas at Austin

ELIE TAMER
Department of Economics, Harvard University

This Supplement contains additional material. Section S1 provides further examples of the model framework, Section S2 provides results on measurability, and Section S3 provides further Monte Carlo experiments.

S1. FURTHER EXAMPLES OF MODEL FRAMEWORK

EXAMPLE 5 (Moment Inequalities). Suppose that θ is known to satisfy the moment inequality conditions $E_{\mathbb{P}_0}(m(X, \theta)) \geq 0$, where $m(X, \theta)$ is a known vector-valued "moment function" of the data, X, and the parameter, θ. The expectation is taken with respect to the true unknown data generating process $\mathbb{P}_0$ for X. Suppose that (perhaps just as an approximation) the random vector X has a discrete distribution with J support points $(x_1, \ldots, x_J)$, such that $\mathbb{P}(X = x_j) = p_j$. In this model, the parameter μ is equal to a specification of $(p_1, \ldots, p_J)$, and so the identified set at μ is $\Theta(\mu) = \{\theta \in \Theta : \sum_{j=1}^{J} m(x_j, \theta) p_j \geq 0\}$.

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

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

Brendan Kline: brendan.kline@austin.utexas.edu
Elie Tamer: elietamer@fas.harvard.edu

[1] See for example Section S3.2 concerning regression with interval data.

EXAMPLE 6 (Posterior Probabilities for the Simple Interval Identified Parameter, Continued). This example continues the discussion in Example 4.

Consider $\Pi(\Theta_I \subseteq \Theta^*|X)$. This is the posterior probability that all values in the identified set are contained in Θ^* or, equivalently, the posterior probability that all values of the parameter that could have generated the data are contained in Θ^*. Note that $\bigcap_{\theta \in (\Theta^*)^C} \mu_I(\theta)^C = \bigcap_{\theta \in (-\infty, a) \cup (b, \infty)} \{\mu : \mu_L \leq \theta \leq \mu_U\}^C = \bigcap_{\theta \in (-\infty, a) \cup (b, \infty)} \{\mu : \mu_L > \theta \text{ or } \mu_U < \theta\} = \{\mu : \mu_L > \mu_U \text{ or } a \leq \mu_L \leq \mu_U \leq b\} \supset \{\mu : a \leq \mu_L, \mu_U \leq b\}.$[2] So $\Pi(\Theta_I \subseteq \Theta^*|X) \equiv \Pi(\{\mu : \mu_L > \mu_U \text{ or } a \leq \mu_L \leq \mu_U \leq b\}|X)$. In other words, this posterior probability is the posterior probability of the set of μ such that the identified set evaluated at μ is either empty or nonempty but contained within Θ^*.

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

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

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

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

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

---

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

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

$\Pi(\mu_{0L} \subseteq \Theta_I | X) \to \text{Uniform}[0, 1]$ and $\Pi(\mu_{0U} \subseteq \Theta_I | X) \to \text{Uniform}[0, 1]$. So the boundary points of the identified set are "covered" with the same distribution as a p-value in repeated large samples, providing frequentist coverage properties. (See also Section 5.)

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

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

Case 5. Conversely, suppose that $\Theta_I \nsubseteq [a, b]$ and $\text{int}(\Theta_I) \neq \emptyset$. This implies that $\mu_{0L} < \mu_{0U}$ and either $\mu_{0L} < a$ or $\mu_{0U} > b$. Consider the case that $\mu_{0L} < a$; the case that $\mu_{0U} > b$ is similar. Let θ^* be some point that is in $(\mu_{0L}, \min\{a, \mu_{0U}\})$. Note that if $a \leq \mu_{0U}$, then this interval is nonempty since $\mu_{0L} < a$. Alternatively, if $a > \mu_{0U}$, then this interval is nonempty since $\mu_{0L} < \mu_{0U}$. Then $\mu_0 \in \text{int}(\mu_I(\theta^*)) = \text{int}(\{\mu : \mu_L \leq \theta^* \leq \mu_U\}) = \{\mu : \mu_L < \theta^* < \mu_U\}$, since $\mu_{0L} < \theta^* < \mu_{0U}$ by choice of θ^*. So by part (ii) of Theorem 3, $\Pi(\Theta_I \subseteq \Theta^* | X) \to 0$.

Case 6. Additionally, in large samples, by part (iii) of Theorem 3, the posterior probability is approximately $\Pi(\Theta_I \subseteq \Theta^* | X) \approx P_{N(0, \Sigma_0)}(\sqrt{n}(\bigcap_{\theta \in (\Theta^*)^C} \mu_I(\theta)^C - \mu_n(X)))$.

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

Case 7. Suppose that $\mu_{0L} < \mu_{0U}$, so that the identified set is nonempty and not a singleton. Then $\mu_0 \in \text{int}(\bigcup_{\theta \in \Theta} \mu_I(\theta))$. So by part (iv) of Theorem 3, $\Pi(\Theta_I \neq \emptyset | X) \to 1$.

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

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

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

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

S2. MEASURABILITY

To establish the measurability of the events corresponding to the posterior probability statements, the following definitions are introduced relative to the measurable sets introduced in Assumption 1.

DEFINITION 5 (Measurable Inverse Included in Sets). The term $\mathcal{M}_1$ is a collection of subsets of $\mathbb{R}^{d_\delta}$ such that for all $\Delta^* \in \mathcal{M}_1$, $\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is a measurable subset of M, that is, $\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) \in \mathcal{B}(M)$.

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

DEFINITION 6 (Measurable Inverse Included Sets). The term $\mathcal{M}_2$ is a collection of sub-sets of $\mathbb{R}^{d_\delta}$ such that for all $\Delta^* \in \mathcal{M}_2$, $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C$ is a measurable subset of M, that is, $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C \in \mathcal{B}(M)$.

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

DEFINITION 7 (Measurable Inverse Intersection Sets). The term $\mathcal{M}_3$ is a collection of subsets of $\mathbb{R}^{d_\delta}$ such that for all $\Delta^* \in \mathcal{M}_3$, $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is a measurable subset of M, that is, $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) \in \mathcal{B}(M)$.

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

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

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

   (i)   *The set Δ_I is compact and $\Delta(\Theta)$ is compact.*

   (ii) *For any $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is closed.*

   (iii) *For any open $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C$ is open.*

   (iv) *For any closed $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is closed.*

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

   (v)   *It holds that $\mathcal{M}_1 = \mathcal{P}(\mathbb{R}^{d_\delta})$.*

   (vi) *It holds that $\mathcal{B}(\mathbb{R}^{d_\delta}) \subseteq \mathcal{M}_2$.*

   (vii) *It holds that $\mathcal{B}(\mathbb{R}^{d_\delta}) \subseteq \mathcal{M}_3$.*

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

For part (ii), suppose that $\{\mu_n\}_n$ is a sequence in $\bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ that converges to some point $\mu^* \in M$. Since $\mu_n \in \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ there must be θ_n such that $\Delta(\theta_n) = \delta$ and $\mu_n \in \mu_I(\theta_n)$. Since Δ is a continuous function and Θ is compact, $\{\theta : \Delta(\theta) = \delta\}$ is compact. Therefore there is a convergent subsequence $\theta_{n_k} \to \theta^* \in \{\theta : \Delta(\theta) = \delta\}$. Since Q is continuous and $Q(\theta_{n_k}, \mu_{n_k}) = 0$ along this subsequence, also $Q(\theta^*, \mu^*) = 0$. So $\mu^* \in \mu_I(\theta^*)$, and thus $\mu^* \in \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$. Therefore $\bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is closed.

And since an arbitrary intersection of closed sets is a closed set, for any $\Delta^* \subseteq \mathbb{R}^{d_\delta}$, $\bigcap_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is closed.

For part (iv), note by Lemma 2 that $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) = \bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$. Suppose that $\{\mu_n\}_n$ is a sequence in $\bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ that converges to some point $\mu^* \in M$. Since $\mu_n \in \bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$, there must be $\delta_n \in \Delta^* \cap \Delta(\Theta)$ and $\theta_n \in \Theta$ such that $\Delta(\theta_n) = \delta_n$ and $\mu_n \in \mu_I(\theta_n)$. Since Θ is compact and since $\Delta^* \cap \Delta(\Theta)$ is compact (since it is the intersection of a closed set and a compact set), there is a convergent subsequence $\delta_{n_k} \to \delta^* \in \Delta^* \cap \Delta(\Theta)$ and $\theta_{n_k} \to \theta^* \in \Theta$. Since $\delta_{n_k} = \Delta(\theta_{n_k})$, $\delta^* = \Delta(\theta^*)$. Since Q is continuous and $Q(\theta_{n_k}, \mu_{n_k}) = 0$ along this subsequence, also $Q(\theta^*, \mu^*) = 0$. So $\mu^* \in \mu_I(\theta^*)$, and thus $\mu^* \in \bigcup_{\{\theta : \Delta(\theta) = \delta^*\}} \mu_I(\theta)$. And so $\mu^* \in \bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$. Therefore $\bigcup_{\delta \in \Delta^* \cap \Delta(\Theta)} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is closed.

For part (iii), note by Lemma 2 that $(\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C)^C = \bigcup_{\delta \in (\Delta^*)^C} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) = \bigcup_{\delta \in (\Delta^*)^C \cap \Delta(\Theta)} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$. By part (iv), this is closed because $(\Delta^*)^C$ is closed. So $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)^C$ is open.

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

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

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

For part (vii), let Θ_m be defined as above and also let $\Delta_m = \mathbb{R}^{d_\delta} \cap \overline{B}(m)$.[4] Since $\mathbb{R}^{d_\delta}$ is closed and $\overline{B}(m)$ is compact, Δ_m is compact. Also note that $\bigcup_{m \geq 1} \bigcup_{\delta \in \Delta^* \cap \Delta_m} \bigcup_{\{\theta \in \Theta_m : \Delta(\theta) = \delta\}} \mu_I(\theta) = \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta \in \Theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$. If $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta \in \Theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$, then there is $\delta \in \Delta^*$ and θ such that $\Delta(\theta) = \delta$ and $\mu \in \mu_I(\theta)$. Consequently, for large enough m, $\delta \in \Delta^* \cap \Delta_m$ and $\theta \in \Theta_m$, so $\mu \in \bigcup_{m \geq 1} \bigcup_{\delta \in \Delta^* \cap \Delta_m} \bigcup_{\{\theta \in \Theta_m : \Delta(\theta) = \delta\}} \mu_I(\theta)$. Conversely, if $\mu \in \bigcup_{m \geq 1} \bigcup_{\delta \in \Delta^* \cap \Delta_m} \bigcup_{\{\theta \in \Theta_m : \Delta(\theta) = \delta\}} \mu_I(\theta)$, then it is immediate that $\mu \in \bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta \in \Theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$.

The proof of part (iv) also establishes that if Δ^* is closed, then $\bigcup_{\delta \in \Delta^* \cap \Delta_m} \bigcup_{\{\theta \in \Theta_m : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is closed. Consequently, $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta \in \Theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is the countable union of closed sets, so is a Borel set. Suppose that Δ^* is either a countable union or a countable intersection of sets Δ_n^* such that $\bigcup_{\delta \in \Delta_n^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is a Borel set for each Δ_n^*. In the case that Δ^* is a countable union, $\Delta^* = \bigcup_{n \geq 1} \Delta_n^*$. Therefore, $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta) = \bigcup_{n \geq 1} \bigcup_{\delta \in \Delta_n^*} \bigcup_{\{\theta : \Delta(\theta) = \delta\}} \mu_I(\theta)$ is a countable union of Borel

---

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

sets, so is a Borel set. In the case that Δ^* is a countable intersection, $\Delta^* = \bigcap_{n\geq 1} \Delta_n^*$. There-fore, $\bigcap_{\delta \in \Delta^*} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C = \bigcap_{n\geq 1} \bigcap_{\delta \in \Delta_n^*} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is the countable intersec-tion of Borel sets, so is a Borel set. This is because $\bigcup_{\delta \in \Delta_n^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ are Borel sets, so also $\bigcap_{\delta \in \Delta_n^*} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ are Borel sets. Consequently, $\bigcup_{\delta \in \Delta^*} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a Borel set for any Borel set Δ^*. This is because the Borel sets of a metrizable space are contained in any collection of sets that has the property that all closed sets are elements of the collection, and the collection is closed under countable unions and countable intersections. See for example Aliprantis and Border (2006, Corollary 4.18).

For part (vi), note that $(\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C)^C = \bigcup_{\delta \in (\Delta^*)^C} \bigcup_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)$ is a Borel set for any Borel set $(\Delta^*)^C$, by part (vii). So since the Borel sets are closed under complements, $\bigcap_{\delta \in (\Delta^*)^C} \bigcap_{\{\theta:\Delta(\theta)=\delta\}} \mu_I(\theta)^C$ is a Borel set for any Borel set Δ^*. $\square$

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

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

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

REMARK 10 (Measurability of $Q(\theta, \mu)$). Because of the connection (by definition) of the measurability of a function and the measurability of pre-images of measurable sets, it is tempting to ask for an analogue of Lemma 3 that assumes only measurability of $Q(\theta, \mu)$. Unfortunately, such a result (in general) is not available. Suppose that $M = [0, 1]$ and $\Theta = [0, 1]$, and let A be any set in $\Theta \times M$ with the property that A is Borel measurable, but the projection of A onto M is not Borel measurable. That such sets exist is the same as the existence of analytic but not Borel measurable sets. Let $Q(\theta, \mu)$ be 1 minus the characteristic function for A, which is measurable (by definition). Then consider the set of μ such that $\Theta_I(\mu) \cap \Theta \neq \emptyset$, that is, the set of μ corresponding to the posterior proba-bility of a nonempty identified set or, equivalently, quantity (ii) in Definition 3. That set of μ is the projection of A onto M, which is not Borel measurable by construction. That suggests an inability to assign a posterior probability to nonemptiness of the identified set with this $Q(\theta, \mu)$, despite measurability of $Q(\theta, \mu)$.

S3. Further Monte Carlo experiments

S3.1 *A simple interval identified parameter*

This section reports the results of a Monte Carlo experiment in the context of a simple interval identified parameter, described in Examples 1, 3, 4, and 6. The data generating process in this experiment is

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

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

There are many approaches that result in a posterior distribution for μ. One possibility is to specify a normal likelihood, and specify conjugate priors for the unknown parameters. However, this entails potentially undesirable parametric distributional assumptions. Another possibility is to use the Bayesian bootstrap, which is a nonparametric approach to Bayesian inference on moments of a distribution that does not require the specification of a parametric likelihood, and that requires minimal computational investment. See the discussion after Assumption 3 for references.

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

The experiment involves multiple different specifications of μ_0.

First, suppose that $\mu_{0L} = 0$ and $\mu_{0U} = 1$, so that there is a nonsingleton identified set. Figure 3(a) displays the values of $\Pi(\theta \in \Theta_I|X)$ for various values of θ and various draws from the data generating process. Each "curve" corresponds to $\Pi(\theta \in \Theta_I|X)$ for a particular value of X drawn from the data generating process, treating θ as the argument that is plotted along the horizontal axis. Consequently, the distribution over $\Pi(\theta \in \Theta_I|X)$ (i.e., the existence of multiple curves in the figure) is the distribution induced by the data generating process.

As discussed in Example 4, for essentially all draws of X, $\Pi(\theta \in \Theta_I|X) \approx 1$ for values in approximately $[0.1, 0.9]$. So $\Pi(\theta \in \Theta_I|X) \approx 1$ on a large subset of the interior of the

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

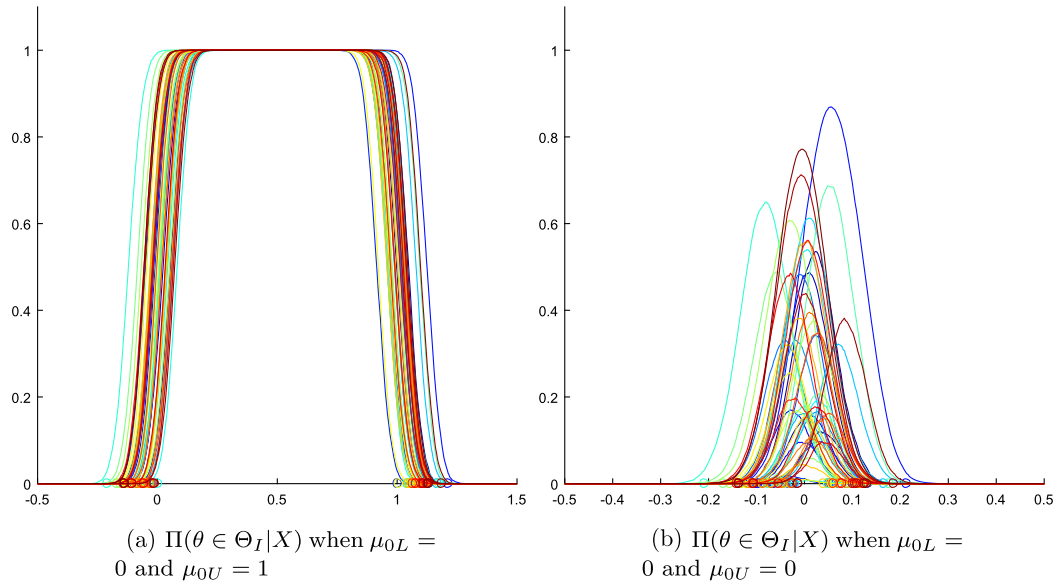(b) $\Pi(\theta \in \Theta_I|X)$ when $\mu_{0L} = 0$ and $\mu_{0U} = 0$

FIGURE 3. Posterior probabilities that various parameter values belong to the identified set.

identified set. (In larger samples, per the discussion in Example 4, the interval on which $\Pi(\theta \in \Theta_I|X) \approx 1$ would be wider.) Also, for essentially all draws of X, $\Pi(\theta \in \Theta_I|X) \approx 0$ for values outside approximately $[-0.1, 1.1]$. (In larger samples, per the discussion in Example 4, the interval on which $\Pi(\theta \in \Theta_I|X) \napprox 0$ would be narrower.) And finally, per the discussion in Example 4, in the neighborhoods of the two points on the boundary of the identified set, the values of $\Pi(\theta \in \Theta_I|X)$ vary depending on the particular draw of X. Note that, as discussed throughout this paper, this figure should not be interpreted to mean that there is a "posterior for" θ that is uniform on (most of) the identified set, $[0, 1]$. Indeed, if $\mu_{0U} = 2$ instead, then the analogous figure would have values of 1 on about $[0.1, 1.9]$, which would obviously not be a "uniform" posterior on the identified set. Instead, the interpretation is that there is essentially posterior certainty that all such points are in the identified set or, equivalently, there is essentially posterior certainty that all such points could have generated the data.

The circles along the horizontal axis of Figure 3(a) are the endpoints of the 95% credible set for the identified set, for each draw from the data generating process. The credible set of a given color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 94.6% of the draws from the data generating process, the 95% credible set indeed does contain the true identified set, so the credible set is also a valid frequentist confidence set.

Now, second, suppose that $\mu_{0L} = 0$ and $\mu_{0U} = 0$, so that there is point identification; however, this is not known a priori by the econometrician. Figure 3(b) similarly displays the values of $\Pi(\theta \in \Theta_I|X)$ for various values of θ and various draws from the data generating process. The posterior $\Pi(\theta \in \Theta_I|X)$ tends to be largest for values around 0, the singleton value of the identified set. However, unlike in the above case of a nonsingleton identified set, in general $\Pi(\theta \in \Theta_I|X)$ is bounded away from 1.

This is consistent with the discussion in Example 4 that concludes that in large samples, $\Pi(0 \in \Theta_I | X) \approx P_{N(0, \Sigma_0)}(\mu_L \leq -\sqrt{n}\mu_{nL}(X), \mu_U \geq -\sqrt{n}\mu_{nU}(X))$. A "typical" value of $(\sqrt{n}\mu_{nL}(X), \sqrt{n}\mu_{nU}(X))$ in large samples is $(0, 0)$, which would imply that $\Pi(0 \in \Theta_I | X) \approx P_{N(0, \Sigma_0)}(\mu_L \leq 0, \mu_U \geq 0)$. Since Σ_0 is the identity matrix, $\Pi(0 \in \Theta_I | X) \approx \frac{1}{4}$.

The reason for this is that $\Pi(0 \in \Theta_I | X) = \Pi(\mu_L \leq 0 \leq \mu_U | X)$. If $\mu_{0L} = 0 = \mu_{0U}$, then the posterior for μ does not necessarily satisfy $\mu_L \leq 0 \leq \mu_U$ with high probability, since consistency of the posterior allows that $\mu_L > \mu_U$ with high probability, and also that $0 < \mu_L \leq \mu_U$ or $\mu_L \leq \mu_U < 0$ with high probability. This is roughly analogous to the "boundary" problem that would arise in existing frequentist approaches to this model. But note that, unlike in existing frequentist approaches, it is not necessary to use an ad hoc rule like a "tolerance parameter." This is because the Bayesian approach, including for data such that $\mu_{nL} > \mu_{nU}$, results in a nondegenerate posterior distribution over the identified set. In particular, if $\mu_{nL} > \mu_{nU}$, then some of the draws of the identified set will be the "empty" identified set (for draws such that $\mu_L > \mu_U$), while others will be a narrow identified set (for draws such that $\mu_L \leq \mu_U$ and $\mu_L \approx \mu_U$). Therefore, the Bayesian approach "automatically" accounts for the fact that $\mu_{nL} > \mu_{nU}$ does not necessarily mean that the true identified set is empty, whereas existing frequentist approaches have to impose this fact using an ad hoc rule.

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

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

As above, the circles along the horizontal axis of Figure 3(b) are the endpoints of the 95% credible set for the identified set, for each draw from the data generating process. The credible set of a given color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 89.2% of the draws from the data generating process, the 95% credible set indeed does contain the true identified set, so the credible set is not quite (but is almost) a valid frequentist confidence set. The lack of exact frequentist coverage is not surprising because when $\mu_{0L} = \mu_{0U}$, the discussion in Remark 5 does not hold.

In some applications, it may be of interest to know the value(s) of θ that is (are) "most likely" to be in the identified set, and/or to compare the relative odds that various values of θ are in the identified set. Figure 4(a) displays the values of the posterior odds $\frac{\Pi(\theta \in \Theta_I | X)}{\max_\theta \Pi(\theta \in \Theta_I | X)}$. The relative odds of θ_1^* and θ_2^* is the ratio of the displayed posterior odds, since the denominator cancels. The term $\frac{\Pi(\theta \in \Theta_I | X)}{\max_\theta \Pi(\theta \in \Theta_I | X)}$ behaves more like $\Pi(\theta \in \Theta_I | X)$ behaved before in the case of a nonsingleton identified set. The posterior odds for values outside approximately $[-0.2, 0.2]$ are approximately 0. And the posterior odds for essentially a single value of θ in a neighborhood of the true identified set is 1. The value of θ that has maximal posterior odds depends on the draw of X; it tends to be approximately $\frac{\mu_{nL} + \mu_{nU}}{2}$, which is indicated for each draw from the data generating process by

(a) $\dfrac{\Pi(\theta \in \Theta_I | X)}{\max_\theta \Pi(\theta \in \Theta_I | X)}$

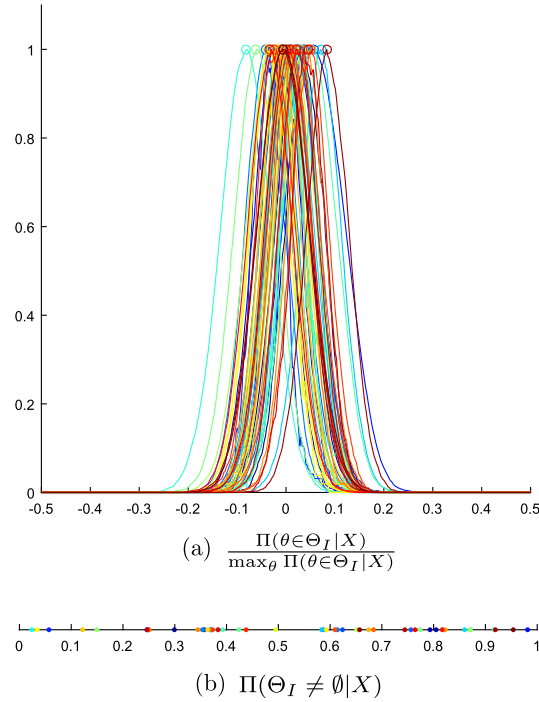(b) $\Pi(\Theta_I \neq \emptyset | X)$

FIGURE 4. Other posterior probabilities when $\mu_{0L} = 0$ and $\mu_{0U} = 0$.

a circle that is the same color as the corresponding posterior "curve." If there is a nonsingleton identified set, then $\max_\theta \Pi(\theta \in \Theta_I | X) \approx 1$, so $\frac{\Pi(\theta \in \Theta_I | X)}{\max_\theta \Pi(\theta \in \Theta_I | X)} \approx \Pi(\theta \in \Theta_I | X)$, so the posterior odds are essentially the same as $\Pi(\theta \in \Theta_I | X)$ in the case of a nonsingleton identified set.

Particularly in the case that $\mu_{0L} = 0 = \mu_{0U}$, it may also be of interest to know the posterior probability that the identified set is nonempty. Figure 4(b) displays the posterior probability that the identified set is nonempty, for various draws from the data generating process. The posterior probability in this figure of a given color corresponds to the same draw of X as the posterior "curves" displayed above of the same color. As expected from Example 6, these posterior probabilities are distributed approximately according to Uniform[0, 1] in repeated samples. If there is a nonsingleton identified set, then the posterior probability that the identified set is nonempty is essentially 1 for all draws from the data generating process, so those posterior probabilities are not displayed.

S3.2 *Regression with interval data*

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

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

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

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

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

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

Figure 5 displays the values of $\Pi(\beta_3 \in \Delta_I | X)$ for various values of β_3, and various draws from the data generating process. As before, each "curve" corresponds to $\Pi(\beta_3 \in \Delta_I | X)$ for a particular value of X drawn from the data generating process. For
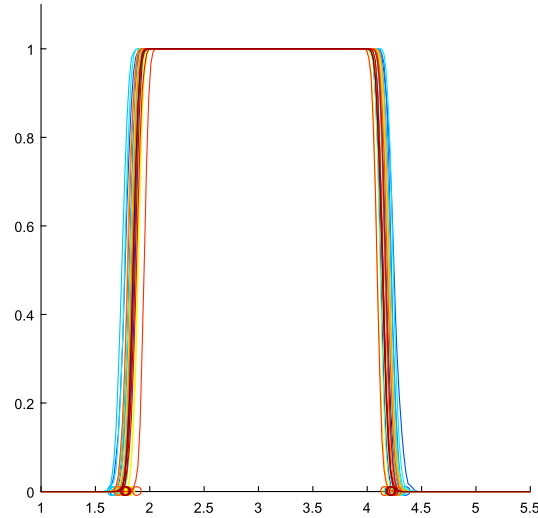

FIGURE 5. The values of $\Pi(\beta_3 \in \Delta_I | X)$.

essentially all draws of X, $\Pi(\beta_3 \in \Delta_I|X) \approx 1$ for values in approximately $[1.9, 4.1]$. So $\Pi(\beta_3 \in \Delta_I|X) \approx 1$ on essentially the entirety of the identified set. Also, for essentially all draws of X, $\Pi(\beta_3 \in \Delta_I|X) \approx 0$ for values outside approximately $[1.5, 4.5]$. In the neighborhoods of the two points on the boundary of the identified set, the values of $\Pi(\beta_3 \in \Delta_I|X)$ vary depending on the particular draw of X.

The circles along the horizontal axis of Figure 5 are the endpoints of the 95% credible set for the identified set, for each draw from the data generating process. The credible set of a given color corresponds to the same draw of X as the posterior "curve" displayed in the same color. In approximately 95.6% of the draws from the data generating process, the 95% credible set indeed does contain the true identified set, so the credible set is also a valid frequentist confidence set. Note that this concerns just part of the partially identified parameter, but nevertheless avoids conservative coverage. Other frequentist approaches might require conservative projection methods.

References

Aliprantis, C. D. and K. C. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third edition. Springer, Berlin. [6]

Kitagawa, T. (2012), "Estimation and inference for set-identified parameters using posterior lower probability." Working paper. [6]

Manski, C. F. and E. Tamer (2003), "Inference on regressions with interval data on a regressor or outcome." *Econometrica*, 70 (2), 519–546. [11]

Molchanov, I. (2006), *Theory of Random Sets*. Springer, London. [6]