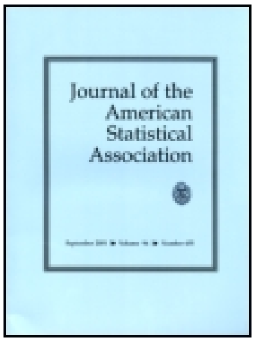

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Coverage Inducing Priors in Nonstandard Inference Problems

Ulrich K. Müller & Andriy Norets

To cite this article: Ulrich K. Müller & Andriy Norets (2016) Coverage Inducing Priors in Nonstandard Inference Problems, Journal of the American Statistical Association, 111:515, 1233-1241, DOI: 10.1080/01621459.2015.1086654

To link to this article: http://dx.doi.org/10.1080/01621459.2015.1086654

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2016, VOL. 111, NO. 515, 1233–1241, Theory and Methods
http://dx.doi.org/10.1080/01621459.2015.1086654

Taylor & Francis
Taylor & Francis Group

Coverage Inducing Priors in Nonstandard Inference Problems

Ulrich K. Müller[a] and Andriy Norets[b]

[a]Department of Economics, Princeton University, Princeton, NJ, USA; [b]Economics Department, Brown University, Providence, RI, USA

ABSTRACT
We consider the construction of set estimators that possess both Bayesian credibility and frequentist coverage properties. We show that under mild regularity conditions there exists a prior distribution that induces $(1-\alpha)$ frequentist coverage of a $(1-\alpha)$ credible set. In contrast to the previous literature, this result does not rely on asymptotic normality or invariance, so it can be applied in nonstandard inference problems. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2014
Revised June 2015

KEYWORDS
Conditional coverage; Confidence sets; Nonstandard inference problems; Objective Bayes; Probability matching; Unit roots

1. Introduction

Set estimators, such as confidence and credible sets, are standard means of describing uncertainty about model parameters and forecasts. The desirability of both Bayesian and frequentist properties for set estimators has long been understood in the literature. For instance, it is necessary (and sufficient) for a set estimator to be a credible set of level $1-\alpha$ relative to some prior to rule out the existence of relevant subsets,* and more generally to ensure "reasonable" conditional properties; see Fisher (1956), Buehler (1959), Pierce (1973), and Robinson (1977) or Section 10.4 in Lehmann (1986).

In the large class of locally asymptotically normal (LAN) models, the maximum likelihood estimator is asymptotically normally distributed, and the posterior is asymptotically Gaussian with the same variance by virtue of the Bernstein–von Mises Theorem. Frequentist and Bayesian approaches thus deliver asymptotically equivalent set estimators in LAN models, and usual classical procedures also have attractive conditional properties. In recent years, much attention has been devoted to nonstandard problems outside the LAN class, where this equivalence does not hold. For instance, nonstandard problems arise in partially identified models (e.g., Imbens and Manski 2004; Chernozhukov, Hong, and Tamer 2007), weakly identified models (e.g., Staiger and Stock 1997; Andrews, Moreira, and Stock 2006), change-point problems (e.g., Carlstein, Müller, and Siegmund 1994), and models for highly persistent time series with an autoregressive root near unity (e.g., Chan and Wei 1987; Phillips 1987).

In this article, we consider construction of set estimators that have both prespecified frequentist coverage and Bayesian credibility. For our main result, we assume a finitely discretized parameter space. We show that under a mild continuity condition, for a given type of credible set (such as the highest posterior density (HPD) set or the equal-tailed set), there always exists a prior distribution that induces frequentist coverage of level $1-\alpha$ for the $(1-\alpha)$-credible set. Previous results for Bayesian sets with frequentist coverage are either for invariant problems (see, e.g., Section 6.6.3 in Berger 1985), or for (higher order) asymptotic equivalence of coverage and credibility in LAN models under the "probability matching prior" (see Datta and Mukerjee 2004 for a survey and references). In contrast, our result is generic in the sense that it applies to any (discretized) inference problem. Our result thus generalizes the notion of "objective" probability matching priors to small sample and non-LAN contexts.

For problems without nuisance parameters, it turns out that the HPD credible set with coverage inducing prior is always similar (that is it has exact coverage for all parameter values) and length-optimal (no uniformly shorter confidence interval can exist). Thus, for such problems, this set estimator provides an attractive description of parameter uncertainty from several perspectives.

The main theoretical result can be applied and extended to different settings. First, the existence of coverage inducing priors also holds for prediction sets. Second, the result can be applied to settings with a general parameter space, hierarchical Bayesian priors, and weighted frequentist coverage. When the parameter space is continuous but bounded, this extension delivers credible sets with uniform frequentist coverage up to an arbitrarily small error. For unbounded parameter spaces, the extension can be used in a "guess and verify" method for construction of credible sets with near-uniform frequentist coverage.

The proof of the main result, which is based on a fixed-point theorem, suggests an iterative procedure for computing the coverage inducing prior. The procedure starts with an arbitrary prior distribution and then iteratively increases the relative prior weight for the parameter values at which the coverage is below the specified level. We use this method to construct confidence

CONTACT Andriy Norets *andriy_norets@brown.edu*. Economics Department, Brown University, Providence, RI 02912.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.
*A relevant subset is a subset of the sample space such that conditional on the observations falling into this set, the coverage is below the nominal level of the confidence set for all parameter values.

sets for the date and magnitude of a structural break in a time series model and for the largest root in a local-to-unity autoregressive model.

2. Main Results

2.1. Definitions and Notation

We start by assuming a finite parameter space, $\Theta = \{\theta_1, \dots, \theta_m\}$, with $\theta \in \Theta$ being the parameter of interest. In Section 2.3, we discuss implications for continuous parameter spaces.

Suppose the distribution of the data $X \in \mathcal{X}$ given parameter θ, $P(\cdot|\theta)$, has density $p(x|\theta)$ with respect to a σ-finite measure ν. Without loss of generality, we assume that ν is equivalent to $\sum_{j=1}^{m} P(\cdot|\theta_j)$. Since the parameter space is discrete, it is convenient to use randomization in defining set estimators such as confidence or credible sets. Thus, let us define a rejection probability function $\varphi : \Theta \times \mathcal{X} \mapsto [0, 1]$, with $\varphi(\theta, x)$ the probability of not including θ conditional on having observed $X = x$. If $\varphi(\theta_j, \cdot)$ defines a level α test of $H_0 : \theta = \theta_j$,

$$\int \varphi(\theta_j, x) p(x|\theta_j) d\nu(x) \leq \alpha \quad (1)$$

for all $j = 1, \dots, m$, then the set that excludes $\theta \in \Theta$ with probability $\varphi(\theta, x)$ when $X = x$ is observed forms a $1 - \alpha$ confidence set. For convenience, we will refer to φ as defining a set estimator. Also, we assume $0 < \alpha < 1$ throughout. A confidence set is called similar when (1) holds with equality for all $j = 1, \dots, m$.

For a prior $\pi = (\pi_1, \dots, \pi_m)'$, $\pi_j \geq 0$, $\sum_j \pi_j = 1$, the posterior probability mass function is defined as

$$p(\theta_j|x) = \frac{p(x|\theta_j)\pi_j}{\sum_{k=1}^{m} p(x|\theta_k)\pi_k}, \; j = 1, \dots, m.$$

A $1 - \alpha$ credible set is defined by any φ such that the posterior probability of excluding θ is equal to α,

$$\sum_{j=1}^{m} p(\theta_j|x)\varphi(\theta_j, x) = \alpha, \; \forall x$$

or

$$\sum_{j=1}^{m} (\alpha - \varphi(\theta_j, x)) p(x|\theta_j)\pi_j = 0, \; \forall x. \quad (2)$$

In this definition, the credibility level is evaluated prior to the realization of the randomization device if $0 < \varphi(\theta, x) < 1$ for some $\theta \in \Theta$. Similarly, the length of the set $\varphi(\cdot, x)$ is defined as its expected value $\sum_{j=1}^{m}(1 - \varphi(\theta_j, x))$, and a level $1 - \alpha$ HPD set is a shortest set of credibility level $1 - \alpha$.

2.2. Existence of Coverage Inducing Prior

In this section, we prove that for families of credible sets that satisfy a continuity restriction, there always exists a prior that turns a $1 - \alpha$ credible set into a $1 - \alpha$ confidence set. This result provides an attractive recipe for finding confidence sets: (i) choose a type of credible set suitable for the problem at hand, for example, HPD if shorter sets are desirable, (ii) determine a prior that turns this credible set into a confidence set. As we illustrate in the application section, this simple recipe is a practical and powerful approach to tackling difficult inference problems. It can also be interpreted as a way to construct default or reference priors for Bayesian inference.

Theorem 1. Suppose $\varphi(\theta, x; \pi)$ defines a $1 - \alpha$ credible set for any prior π on $\Theta = \{\theta_1, \dots, \theta_m\}$. Define $z_j(\pi) = \int [\varphi(\theta_j, x; \pi) - \alpha] p(x|\theta_j) d\nu(x)$ and $z(\pi) = (z_1(\pi), \dots, z_m(\pi))'$. Assume $z(\pi)$ is continuous in π. Then, there exists $\pi^\star$ such that $\varphi(\theta_j, x; \pi^\star)$ defines a $1 - \alpha$ confidence set, that is

$$\int \varphi(\theta, x; \pi^\star) p(x|\theta) d\nu(x) \leq \alpha, \; \forall \theta \in \Theta.$$

Proof. The proof is analogous to the proof of the equilibrium existence in an exchange economy with m goods (π corresponds to the vector of prices and $z(\pi)$ corresponds to the excess demand); see, for example, Chapter 17 in Mas-Colell, Whinston, and Green (1995).

Let $z_j^+(\pi) = \max\{0, z_j(\pi)\}$, $z^+(\pi) = (z_1^+(\pi), \dots, z_m^+(\pi))'$, and

$$q(\pi) = \frac{\pi + z^+(\pi)}{\sum_{j=1}^{m}(\pi_j + z_j^+(\pi))}.$$

Note that q is continuous (continuity of $z(\pi)$ is assumed) and $q : \Delta^{m-1} \to \Delta^{m-1}$, where Δ^{m-1} is an $(m-1)$-simplex. By Brouwer's fixed-point theorem, there exists $\pi^\star$ such that $\pi^\star = q(\pi^\star)$.

Since $\varphi(\theta, x; \pi^\star)$ defines a $1 - \alpha$ credible set,

$$(\pi^\star)' z(\pi^\star) = \int \left(\sum_{j=1}^{m} [\varphi(\theta_j, x; \pi^\star) - \alpha] p(x|\theta_j)\pi_j^\star \right) d\nu(x) = 0.$$

Therefore,

$$\begin{aligned} 0 &= (\pi^\star)' z(\pi^\star) = q(\pi^\star)' z(\pi^\star) \\ &= z^+(\pi^\star)' z(\pi^\star) \sum_{j=1}^{m} [\pi_j^\star + z_j^+(\pi^\star)] \end{aligned}$$

and $z^+(\pi^\star)' z(\pi^\star) = 0$. The latter equality implies $z_j(\pi^\star) \leq 0$ for all $j \in \{1, \dots, m\}$, and the claim of the theorem follows. □

Theorem 1 appears to be new. We are aware of the following related results. First, results on matching credible and classical sets are available for particular families of data distributions. The most well-known example is a normal likelihood with known variance and improper uniform prior for the mean. More generally, in invariant problems with continuous densities, $1 - \alpha$ Bayesian credible sets under invariant priors have $1 - \alpha$ frequentist coverage, see Section 6.6.3 in Berger (1985). Second, Joshi (1974) showed that for an unbounded parameter space the equivalence between Bayesian and classical sets cannot hold for one-sided intervals and a proper prior (there is no contradiction to the equivalence results under invariance since invariant priors are improper on unbounded spaces). Third, the Bernstein–von Mises theorem, see, for example, Section 10.2

in van der Vaart (1998), states that under weak regularity conditions implying the local asymptotic normality of the maximum likelihood estimator (MLE), standard classical and credible sets are asymptotically equivalent. There is also a literature on higher order asymptotic equivalence of coverage and credibility in LAN problems, see a monograph on the subject by Datta and Mukerjee (2004) and references therein. Note that our result in Theorem 1 does not appeal to invariance or asymptotics, and does not impose smoothness conditions on the likelihood function.

Potentially, many types of credible sets, such as HPD and equal-tailed sets, can satisfy the continuity requirement of the theorem. A sufficient condition is that $\varphi(\theta, x; \pi)$ itself is a continuous function of the prior π for almost all x. Under mild conditions, this holds for HPD, one-sided, and equal-tailed credible sets. The following theorem suggests that the use of HPD sets in Theorem 1 is particularly attractive. The proof is given in online Appendix A.

Theorem 2. Suppose that for any $\theta \in \Theta$ and $X \sim p(\cdot|\theta)$, the likelihood ratio $p(X|\theta_i)/p(X|\theta_j)$ is an absolutely continuous random variable for all $i \neq j$. Then, an HPD set $\varphi(\theta, x; \pi)$ satisfies the following: (i) The continuity conditions of Theorem 1 are satisfied. (ii) The prior $\pi^\star$ puts positive mass on all $\theta \in \Theta$ and the resulting set $\varphi(\theta, x; \pi^\star)$ is a similar confidence set of level $1-\alpha$, that is, $\int \varphi(\theta, x; \pi^\star) p(x|\theta) d\nu(x) = \alpha$ for all $\theta \in \Theta$. (iii) There does not exist a level $1-\alpha$ confidence set φ' that is weakly shorter than $\varphi(\theta, x; \pi^\star)$ for all $x \in \mathcal{X}$ ($\sum_{j=1}^m \varphi'(\theta_j, x) \geq \sum_{j=1}^m \varphi(\theta_j, x; \pi^\star)$) and strictly shorter for $x \in \mathcal{X}_l$ with $\nu(\mathcal{X}_l) > 0$ ($\sum_{j=1}^m \varphi'(\theta_j, x) > \sum_{j=1}^m \varphi(\theta_j, x; \pi^\star)$).

Thus, the HPD set with coverage inducing prior emerges as a particularly compelling description of uncertainty: it is a $1-\alpha$ credible set; it covers each θ with exact probability $1-\alpha$; and there does not exist a $1-\alpha$ confidence set that is uniformly shorter.

Now suppose that the parameter of interest is $\gamma = g(\theta) \in \Gamma$, where g is not one-to-one. This arises for set estimation in the presence of nuisance parameters, for instance. Set estimators φ are then $\Gamma \times \mathcal{X} \mapsto [0, 1]$ functions, but priors still specify a probability distribution on Θ. Theorem 1 still goes through as long as $z(\pi)$ with $z_j(\pi) = \int [\varphi(g(\theta_j), x; \pi) - \alpha] p(x|\theta_j) d\nu(x)$ is continuous in π, so there still exists a prior that induces coverage. The similarity of φ in Theorem 2, however, does not necessarily hold, since the coverage inducing prior of the HPD set might have zero mass for some θ. The HPD confidence set might thus have coverage above $1-\alpha$ for these θ's.

2.3. Extensions

2.3.1. Predictive Sets

It is particularly intuitive to insist on compelling conditional properties when describing uncertainty in a forecasting setting. The extension of our framework and theoretical results to prediction sets is straightforward: Suppose we seek to describe the uncertainty about a yet unobserved random variable $Y \sim p_p(\cdot|\theta, x)$ after observing $X = x$, where $X \sim p(\cdot|\theta)$, $Y \in \mathcal{Y}$ and p_p is a conditional density on $\mathcal{Y}$ with respect to a generic measure ν_p. Let $\varphi_p(y, x)$ denote the probability that y is not included in a prediction set when x is observed (typically, $y \mapsto 1 - \varphi_p(y, x)$ is the characteristic function of the prediction set). Then, for a given parameter θ and observed x, the probability that the prediction set φ_p will not cover Y is given by

$$\int \varphi_p(y, x) p_p(y|\theta, x) d\nu_p(y).$$

If we denote this probability by $\varphi(\theta, x)$, then the definition of frequentist coverage and Bayesian credibility for φ_p are exactly given by (1) and (2), respectively. Therefore, the existence of a prior that guarantees frequentist nominal coverage for credible sets (Theorem 1) also holds for prediction sets. Note that a prior that induces coverage for predictive sets can be different from a prior that induces coverage for the sets on the parameter space.

2.3.2. General Parameter Space and Weighted Coverage

In this section, we apply Theorem 1 to settings with a general parameter space, hierarchical Bayesian priors, and weighted average coverage. In the following section, we show that this extension implies existence of credible sets that have approximate frequentist coverage for bounded parameter spaces.

Consider a parameter space $\tilde{\Theta} \subseteq \mathbb{R}^k$ and a likelihood function $\tilde{p}(x|\tilde{\theta})$ defined for every $\tilde{\theta} \in \tilde{\Theta}$. Suppose ψ_j, $j = 1, \ldots, m$, are probability densities with respect to Lebesgue measure on $\tilde{\Theta}$ and $\pi = (\pi_1, \ldots, \pi_m)$, $\pi_j \geq 0$, $\sum_j \pi_j = 1$. The mixture density

$$\Pi(\tilde{\theta}, \pi) = \sum_{j=1}^{m} \pi_j \cdot \psi_j(\tilde{\theta}) \tag{3}$$

can be thought of as a hierarchical prior distribution on $\tilde{\Theta}$, where π is a prior on a set of models $\Theta = (\theta_1, \ldots, \theta_m)$ with common likelihood $\tilde{p}(x|\tilde{\theta})$ and ψ_j is a prior on $\tilde{\theta}$ under model θ_j.

As in the setup of Theorem 1, fix a type of credible set such as the HPD set and let $\tilde{\varphi}(\tilde{\theta}, x; \pi)$ denote the rejection probability characterizing a $1-\alpha$ credible set on $\tilde{\Theta}$ of this type under prior $\Pi(\tilde{\theta}, \pi)$. The set $\tilde{\varphi}$ here is viewed as a function of π, with the ψ_j's fixed. Define the marginal likelihood under model θ_j by

$$p(x|\theta_j) = \int \tilde{p}(x|\tilde{\theta}) \psi_j(\tilde{\theta}) d\tilde{\theta}.$$

Also, let

$$\varphi(\theta_j, x; \pi) = \frac{\int \tilde{\varphi}(\tilde{\theta}, x; \pi) \tilde{p}(x|\tilde{\theta}) \psi_j(\tilde{\theta}) d\tilde{\theta}}{p(x|\theta_j)} \tag{4}$$

be the posterior noncoverage probability conditional on model θ_j. The prior weighted coverage of the set $\tilde{\varphi}(\tilde{\theta}, x; \pi)$ in model θ_j is then defined by

$$\begin{aligned} &\int \left[\int [1 - \tilde{\varphi}(\tilde{\theta}, x; \pi) \tilde{p}(x|\tilde{\theta})] d\nu(x) \right] \psi_j(\tilde{\theta}) d\tilde{\theta} \\ &= \int [1 - \varphi(\theta_j, x; \pi)] p(x|\theta_j) d\nu(x). \end{aligned} \tag{5}$$

Note that the right-hand side of (5) is the same as the coverage of θ_j in the discrete model of Section 2.1. Also, Bayesian credibility of the set $\tilde{\varphi}(\tilde{\theta}, x; \pi)$ at level $1-\alpha$ implies credibility (2) in the discrete model of Section 2.1 of the set $\varphi(\theta, x; \pi)$ defined in (4). Thus, if the weighted coverage rate in (5) is a continuous function of π, then Theorem 1 implies the existence of $\pi^\star$ such that (5) evaluated at $\pi^\star$ is at least $1-\alpha$ for all models θ_j, $j = 1, \ldots, m$. In other words, Bayesian credible sets under the $\pi^\star$-mixture prior have ψ_j-weighted average coverage of at least $1-\alpha$ for all $j = 1, \ldots, m$.

Let us note that the resulting sets are invariant to reparameterization in the following sense. For a one-to-one transformation f on $\tilde{\Theta}$ with differentiable inverse f^{-1}, define $\tilde{\varphi}_f(s, x; \pi^\star) = \tilde{\varphi}(f^{-1}(s), x; \pi^\star)$ for $s \in f(\tilde{\Theta})$. Then, $\tilde{\varphi}_f$ defines a $1-\alpha$ credible set for $f(\tilde{\theta})$ under the prior $\sum_j \pi_j^\star \psi_j(f^{-1}(s))|df^{-1}(s)/ds|$ on $f(\tilde{\Theta})$, and this set has $\psi_j(f^{-1}(s))|df^{-1}(s)/ds|$-weighted coverage for $j = 1, \ldots, m$.

2.3.3. Bounded Parameter Space and Approximations

Suppose $\tilde{\Theta}$ is a bounded subset of $\mathbb{R}^k$. Let $\{\tilde{\Theta}_j^m, j = 1, \ldots, m\}$ be a sequence of partitions of $\tilde{\Theta}$ such that $\max_j \operatorname{diam}(\tilde{\Theta}_j^m) \to 0$ as $m \to \infty$. For a fixed m, define ψ_j to be a uniform probability density on $\tilde{\Theta}_j^m$, $j = 1, \ldots, m$. Define $\tilde{\varphi}^m(\tilde{\theta}, x; \pi)$ to be the shortest $(1-\alpha)$-credible set with respect to the prior $\Pi^m(\tilde{\theta}, \pi)$ in (3) that is constant on each $\tilde{\Theta}_j^m$. This implies that the corresponding set on the space of models, $\varphi^m(\theta, x; \pi)$, maximizes $\sum_j \varphi(\theta_j, x)\mathrm{vol}(\tilde{\Theta}_j^m)$ subject to the credibility constraint (2). Let us further assume that for any $\tilde{\theta} \in \tilde{\Theta}$ and $X \sim \tilde{p}(\cdot|\tilde{\theta})$, the marginal likelihood ratio $p(X|\theta_i)/p(X|\theta_j)$ is an absolutely continuous random variable for any $i \neq j$. Then, Theorem 2 and the discussion above imply that for every m there exists $\pi^{m\star}$ such that

$$\alpha = \int \int \tilde{\varphi}^m(\tilde{\theta}, x; \pi^{m\star})\tilde{p}(x|\tilde{\theta})]d\nu(x)\psi_j(\tilde{\theta})d\tilde{\theta},\ j = 1, \ldots, m. \quad (6)$$

The following theorem shows that sets $\tilde{\varphi}^m(\tilde{\theta}, x; \pi^{m\star})$ have approximately uniform coverage and an approximate volume optimality property.

Theorem 3. Assume that $\tilde{p}(\cdot|\tilde{\theta})$ is uniformly continuous in $\tilde{\theta}$ under the $L_1(\nu)$ distance. Then, for every $\epsilon > 0$, there exists M such that $\forall m \geq M$: (i) the coverage of $\tilde{\varphi}^m(\tilde{\theta}, x; \pi^{m\star})$ is within ϵ of $1-\alpha$ for all $\tilde{\theta} \in \tilde{\Theta}$; and (ii) for any $(1-\alpha)$-confidence set $\tilde{\phi}(\tilde{\theta}, x)$, there exists $\mathcal{X}_m$ with, $\nu(\mathcal{X}_m) > 0$ such that

$$\int_{\tilde{\Theta}} \tilde{\phi}(\tilde{\theta}, x)d\tilde{\theta} < (1+\epsilon) \cdot \int_{\tilde{\Theta}} \tilde{\varphi}^m(\tilde{\theta}, x; \pi^{m\star})d\tilde{\theta} \quad (7)$$

for all $x \in \mathcal{X}_m$.

The theorem is proved in Appendix B (online). Inequality (7) for the volumes of set complements implies the following inequality for the volumes of the sets

$$\int_{\tilde{\Theta}} [1 - \tilde{\phi}(\tilde{\theta}, x)]d\tilde{\theta} > \int_{\tilde{\Theta}} [1 - \tilde{\varphi}^m(\tilde{\theta}, x; \pi^{m\star})]d\tilde{\theta} - \epsilon \cdot \mathrm{vol}(\tilde{\Theta}).$$

Thus, part (ii) of Theorem 3 delivers an approximate version of the following equivalent restatement of Part (iii) of Theorem 2: any competing confidence set either has the same volume almost surely ν or its volume is strictly larger on a set of positive measure ν.

The assumptions of the theorem seem to be weak. The uniform continuity assumption holds, for example, when $\tilde{p}(x|\tilde{\theta})$ is continuous in $\tilde{\theta}$ for every x and $\tilde{\Theta}$ is compact (pointwise convergence for densities implies convergence in the total variation distance and continuity implies uniform continuity on compacts).

2.3.4. Unbounded Parameter Space

The mixture approach of Section 2.3.2 is also applicable to unbounded parameter spaces. However, there might well not exist a finite mixture that induces uniform coverage: Joshi's (1974) results imply that only an improper prior can induce coverage of one-sided confidence sets under an unbounded parameter space.

At the same time, a nonstandard problem with unbounded parameter space can sometimes be well approximated by a standard Gaussian shift experiment outside a sufficiently large but bounded subset of the parameter space. For example, the problem of inference for an autoregressive root near unity, which we consider below, has this property (see Section 4.1 of Elliott, Müller, and Watson (2015) for a formal statement and additional examples). In this case, one can use the hierarchical mixture approach to implement a "guess and verify" method for the construction of a credible set with approximate frequentist coverage that employs a flat prior on the nearly standard part of the parameter space.

For concreteness, suppose the problem converges to the Gaussian shift experiment as $\varsigma(\tilde{\theta}) \to \infty$ for some function $\varsigma : \tilde{\Theta} \mapsto \mathbb{R}$. Pick κ_S large enough so that $\tilde{\Theta}_S = \{\tilde{\theta} : \varsigma(\tilde{\theta}) > \kappa_S\}$ is the approximately standard part of the parameter space, where a flat prior induces an HPD set with near-nominal coverage, and let the bounded set $\tilde{\Theta}_{NS} = \tilde{\Theta} \backslash \tilde{\Theta}_S$ be the remaining nonstandard part. For $m \to \infty$, let $\{\tilde{\Theta}_1^m, \ldots, \tilde{\Theta}_{m-1}^m\}$ be a finer and finer partition of $\tilde{\Theta}_{NS}$, and let $\tilde{\Theta}_m^m = \{\tilde{\theta} : \kappa_S < \varsigma(\tilde{\theta}) < \kappa_m\}$, where $\kappa_m \to \infty$. Proceed similarly to Section 2.3.3 and define $\Pi^m(\tilde{\theta}, \pi)$ with ψ_j uniform on $\tilde{\Theta}_j^m$, $j = 1, \ldots, m$, and let $\tilde{\phi}^m(\tilde{\theta}, x; \pi)$ be the shortest $(1-\alpha)$-credible set relative to $\Pi^m(\tilde{\theta}, \pi)$ that is constant on $\tilde{\Theta}_j$ for $j = 1, \ldots, m-1$, but without any restrictions on $\tilde{\Theta}_m^m$. The results of the previous sections then imply that there exists a large enough m and prior $\pi^{m\star}$ such that $\tilde{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ has near-uniform coverage for parameters in $\tilde{\Theta}_{NS}$. Also, the prior $\Pi^m(\tilde{\theta}, \pi^{m\star})$, which is flat on $\tilde{\Theta}_m^m$, is expected to induce coverage close to the nominal level for parameters $\tilde{\theta} \in \tilde{\Theta}_S$ with $\kappa_S \ll \varsigma(\tilde{\theta}) \ll \kappa_m$. Further, continuously extend $\Pi^m(\tilde{\theta}, \pi^{m\star})$ to an improper prior $\bar{\Pi}^m(\tilde{\theta}, \pi^{m\star})$ that is flat on $\tilde{\Theta}_S$, and define $\bar{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ to be the shortest $1-\alpha$ credible set relative to $\bar{\Pi}^m(\tilde{\theta}, \pi^{m\star})$ that is constant on $\tilde{\Theta}_j^m$, $j = 1, \ldots, m-1$. Then, one would expect $\bar{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ to have approximately $1-\alpha$ coverage for all parameter values with $\varsigma(\tilde{\theta}) \gg \kappa_S$, and also for $\tilde{\theta} \in \tilde{\Theta}_{NS}$, since $\bar{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ is presumably close to $\tilde{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ for all $\tilde{\theta} \in \tilde{\Theta}_{NS}$.

It seems possible to make the above discussion technically precise under some conditions. We do not pursue this, though, because we cannot show that $\bar{\Pi}^m(\tilde{\theta}, \pi^{m\star})$ necessarily becomes approximately flat for $\tilde{\theta} \in \tilde{\Theta}$ with $\varsigma(\tilde{\theta})$ close to κ_S, even for very large κ_S. Thus, for $\tilde{\theta} \in \tilde{\Theta}_S$ with $\varsigma(\tilde{\theta})$ close to κ_S, the coverage of $\bar{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ might fall substantially short of the nominal level.

Nevertheless, it is possible to numerically check whether $\bar{\phi}^m(\tilde{\theta}, x; \pi^{m\star})$ has approximately nominal coverage for $\tilde{\theta}$ with $\varsigma(\tilde{\theta})$ close to κ_S. In several examples we considered, including the autoregression example below, we find that it does. We conjecture that this result holds for a wide range of problems, but leave a proof to future research.

3. Computing Coverage Inducing Priors

The proof of Theorem 1 is based on the Brouwer fixed-point theorem with the coverage inducing prior being a fixed point. There is large applied and theoretical literature on computing fixed points. Scarf (1967) introduced the first algorithm that is guaranteed to get an approximate fixed point in a finite number of steps. Eaves (1972) proposed a homotopy based algorithm for computing fixed points. These two algorithms can be used to solve our fixed-point problem. A presentation of more recent theoretical literature that extends these early contributions can be found in a monograph by Yang (1999). It appears that homotopy based methods are favored in the theoretical literature. Dixon and Parmenter (1996) provided an applied perspective on methods for computing fixed points in the context of solving for equilibria in economic models. It seems that applied researchers prefer Newton-type methods even though they are theoretically guaranteed to converge only when they start from a point sufficiently close to the solution.

We do not intend to contribute to the literature on computing fixed points. In our applications, the following simple algorithm works well. We start with some prior π and iteratively increase its coordinates π_j in proportion to undercoverage of $\varphi(\theta, x; \pi)$ at θ_j. Thus, the algorithm essentially iterates on the mapping q defined in the proof of Theorem 1, for which the coverage inducing prior is a fixed point (see Appendix C online for additional implementation details). This procedure is not guaranteed to converge in general; some counterexamples were given by Scarf and Hansen (1973). It is guaranteed to converge when parameters have the following substitution property: changing the prior $\pi = (\pi_1, \ldots, \pi_m)$ to $\pi' = (\pi_1, \ldots, \pi_{i-1}, \pi_i + \epsilon, \pi_{i+1}, \ldots \pi_m)/(1+\epsilon)$ for any j and $\epsilon > 0$ leads to a strict decrease in coverage at θ_i for all $i \neq j$ (see, e.g., Section 17.H of Mas-Colell, Whinston, and Green 1995). Unfortunately, this sufficient condition seems implausible in our settings as the substitution property is likely to be violated for θ_i's that are close to θ_j. A weaker sufficient condition, $(\pi^\star)'z(\pi) > 0$ for any $\pi \neq \pi^\star$ and $z(\pi)$ defined in the proof of Theorem 1 also appears difficult to verify. Nevertheless, the algorithm converges very quickly in our applications. Thus, we do not pursue homotopy algorithms described in Yang (1999) or Newton-type methods that would require approximations to derivatives.

4. Applications

The following two sections illustrate the approach for the construction of 95% nominal-level confidence sets in two nonstandard inference problems. Additional computational details may be found in Appendix C (online).

4.1. Break Date and Magnitude

In this section, we construct an (approximate) joint confidence set for the date and magnitude of a parameter shift in a time-series model. A large literature considers tests for parameter instability, as well as inference about the date of a potential break; see Carlstein, Müller, and Siegmund (1994), Stock (1994), and Perron (2006) for surveys and references. We are not aware, however, of a previous construction of a joint confidence set for the date and magnitude.

Consider first the simple case where the mean of the Gaussian time series y_t undergoes a single shift at time $t = \tau$ of magnitude d, that is

$$y_t = \mu + d\mathbf{1}[t > \tau] + \varepsilon_t, t = 1, \ldots, T, \tag{8}$$

where $\varepsilon_t \sim \text{iid}\mathcal{N}(0, 1)$. As argued by Elliott and Müller (2007), a moderate break magnitude is usefully modeled via asymptotics where $d = d_T = \delta/\sqrt{T}$, so that the parameter change is of the same magnitude as the sampling uncertainty. Imposing translation invariance to deal with the nuisance parameter μ, we find that the partial sum process of the demeaned observations $\bar{y}_t = y_t - T^{-1}\sum_{s=1}^{T} y_s$ satisfies

$$T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} \bar{y}_t \sim X(s) = W(s) - sW(1) - \delta(\min(\lambda, s) - \lambda s) \tag{9}$$

for any $s = t/T$, where W is a standard Wiener process, and $\lambda = \tau/T$ is the break date measured in the fraction of the sample size. This suggests that the relevant observation in the limit experiment is the continuous time process X. Elliott and Müller (2014) formally showed that this is indeed the limit experiment in the sense of LeCam in model (8), and also in well-behaved parametric time-series models where a single parameter β undergoes a break of magnitude $\delta/\sqrt{T\mathcal{I}_{\beta\beta}}$, with $\mathcal{I}_{\beta\beta}$ the Fisher information about β in the stable model. The same limiting problem may also be motivated using the framework in Müller (2011) without reference to a parametric model.

The density of X under $\tilde{\theta} = (\lambda, \delta)$ relative to the measure of a standard Brownian bridge $W(s) - sW(1)$ is given by

$$p(x|\tilde{\theta}) = \exp[-\delta x(\lambda) - \tfrac{1}{2}\delta^2 \lambda(1-\lambda)].$$

As in much of the literature, we rule out break dates arbitrarily close to the beginning and end of the sample, and assume $\lambda \in L = [0.15, 0.85]$. Also, we restrict the break magnitude to $\delta \in D = [-15, 15]$, which should cover the empirically relevant part of the parameter space for most applications, so that $\tilde{\theta} \in \tilde{\Theta} = L \times D$.

We implement the approach discussed in Section 2.3.3 with $\tilde{\Theta}$ partitioned into $m = 168,000$ equal-sized rectangles $\tilde{\Theta}_j^m$ of size 0.0025×0.05. Panel (a) in Figure 1 displays the mixture prior $\Pi^m(\tilde{\theta}, \pi^{m\star})$ that induces the 95% constrained HPD set $\tilde{\varphi}^m$ to have average coverage of 95% on $\tilde{\Theta}_j^m$. Pointwise coverage of $\tilde{\varphi}^m$ is between 94.8% and 95.2% for all $\tilde{\theta} \in \tilde{\Theta}$. A coarser partition with rectangles of size 0.005×0.1 yields a numerically similar average coverage inducing prior, with pointwise coverage between 94.5% and 95.4%.

As an empirical illustration, we revisit the question of a potential break in the mean of U.S. labor productivity after the second oil price shock. As reviewed by Jorgenson, Ho, and

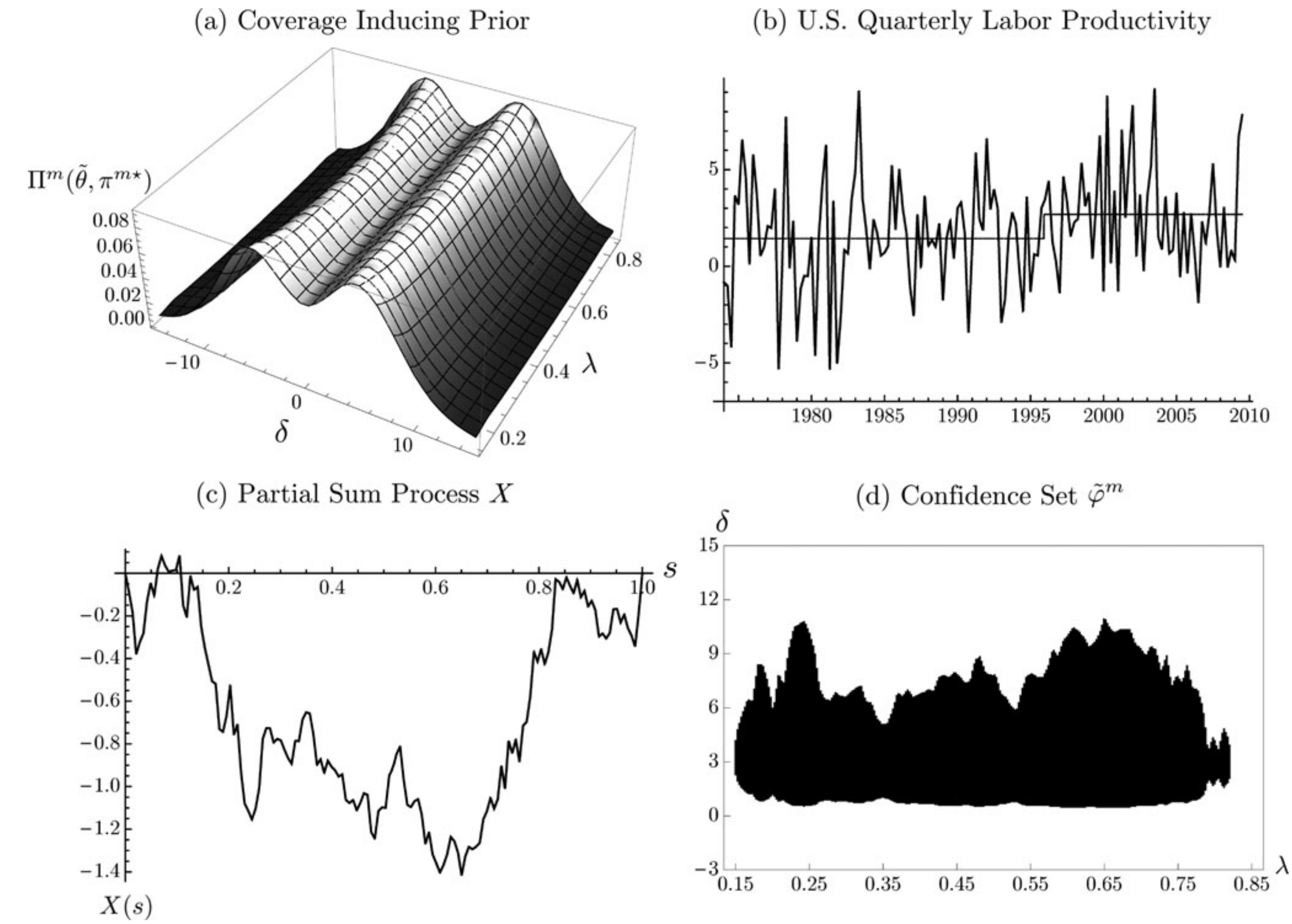

Figure 1. Break application.

Stiroh (2008), a number of authors argue for an upward shift in the mid 90's. Panel (b) in Figure 1 plots the series, taken from Elliott and Müller (2014), with the sample means pre- and post-1995Q4. Panel (c) plots the linearly interpolated partial sum process X for these data (normalized by the square root of the long-run variance to induce the limit in (9)), and panel (d) reports $\tilde{\varphi}^m$ for this X. The set $\tilde{\varphi}^m$ excludes all parameter values with $\delta \leq 0$, corroborating previous evidence on the existence of an upward shift in mean productivity over this period. The data seem compatible with a wide range of break dates and break magnitudes of approximately $\delta \in [0.5, 10]$, corresponding to a mean shift in quarterly productivity of about 0.1–2.2 percentage points.

4.2. *Autoregressive Root Near Unity*

As a second illustration, consider inference about the largest autoregressive root in a univariate autoregressive process

$$\begin{aligned} y_t &= \mu + u_t,\ t = 1, \ldots, T \\ (1 - \rho L)\phi(L)u_t &= \varepsilon_t, \end{aligned} \tag{10}$$

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_{p-1} z^{p-1}$, $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ and $u_0 = O_p(1)$. The largest root ρ is close to unity, while the roots of ϕ are all bounded away from the complex unit circle. As in Phillips (1987) and Chan and Wei (1987) and a large subsequent literature, consider asymptotics with $\rho = \rho_T = 1 - \tilde{\theta}/T$ for some fixed $\tilde{\theta}$, which yield accurate approximations to the small sampling distribution of statistics computed from (10) when ρ is close to 1.

The appropriate limit experiment under $\varepsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$ in the sense of LeCam involves observing the Ornstein–Uhlenbeck process X on the unit interval, where $X(s) = \int_0^s \exp[-\tilde{\theta}(s - r)]dW(r)$ with $W(r)$ being a standard Wiener process. The density of X relative to the measure of a standard Wiener process is given by

$$p(x|\tilde{\theta}) = \exp\left[-\tilde{\theta}\int_0^1 x(s)dx(s) - \tfrac{1}{2}\tilde{\theta}^2\int_0^1 x(s)^2 ds\right]. \tag{11}$$

While the exponent in (11) is a quadratic function of $\tilde{\theta}$, so that the likelihood has a Gaussian shape, inference about $\tilde{\theta}$ based on X is nevertheless a nonstandard problem, since the information $\int_0^1 X(s)^2 ds$ is random (and correlated with the score $\int_0^1 X(s)dW(s)$). Sims (1988) and Sims and Uhlig (1991) discussed how these features affect conditional properties of inference, and argue for the desirability of a Bayesian perspective in the near unit root model.

A large number of procedures have been derived to construct confidence sets for ρ near unity in (10), which correspond to a particular confidence set construction for $\tilde{\theta}$ in the limiting experiment of observing X; see, for instance, Stock (1991), Andrews (1993), Andrews and Chen (1994), Hansen (1999), Elliott and Stock (2001), and Mikusheva (2007). An important

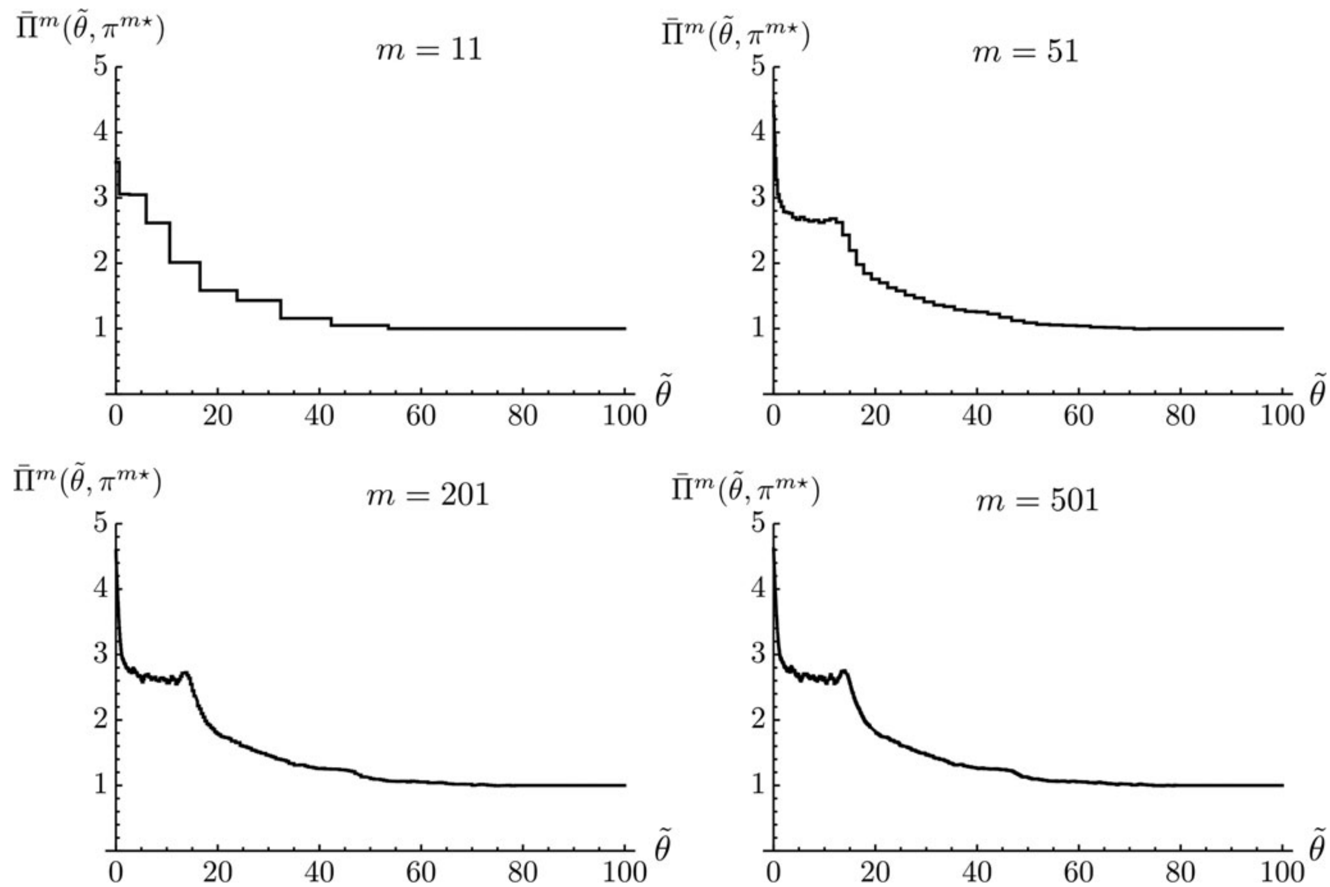

Figure 2. Autoregressive root: Priors that induce average frequentist coverage of constrained HPD set.

application of inference about ρ arises in international economics: the theory of purchasing power parity (PPP) implies that real exchange rates are mean reverting, and the value of ρ governs the half-life of deviations of this long-run equilibrium. See Murray and Papell (2002), Gospodinov (2004), Rossi (2005), and Lopez, Murray, and Papell (2013) for additional methods and empirical evidence.

We follow Lopez, Murray, and Papell (2013) and impose $\rho \leq 1$, so that the parameter space in the limiting problem becomes $\tilde{\Theta} = [0, \infty)$. We apply the "guess and verify" construction of Section 2.3.4 above, with $\tilde{\Theta}_j^m = [80(\frac{j-1}{m-1})^2, 80(\frac{j}{m-1})^2)$ for $j = 1, \ldots, m-1$, $\tilde{\Theta}_S = [80, \infty)$, and the prior $\bar{\Pi}^m(\tilde{\theta}, \pi^m)$ normalized to 1 on $\tilde{\Theta}_S$. Figure 2 displays $\bar{\Pi}^m(\tilde{\theta}, \pi^{m\star})$ for $m \in \{11, 51, 201, 501\}$. The global shape of $\bar{\Pi}^m(\tilde{\theta}, \pi^{m\star})$ is seen to be quite similar for all considered values of m. Pointwise coverage for $\tilde{\theta} \in \tilde{\Theta}$ is between 0.897–0.963 for $m = 11$, 0.939–0.955 for $m = 51$, 0.946–0.951 for $m = 201$, and 0.948–0.951 for $m = 501$. For all m, the priors $\bar{\Pi}^m(\tilde{\theta}, \pi^{m\star})$ put much more mass on small values of $\tilde{\theta}$, counteracting the well-known positive bias of the peak of the likelihood, that is the downward bias of the MLE of ρ. Unreported results show that if one casts the small sample problem of inference about ρ in the asymptotic form (11), and applies the prior of Figure 2 with $m = 501$, then one obtains an HPD set with frequentist coverage close to the nominal level also in small samples.

Table 1 reports the empirical results of Lopez, Murray, and Papell (2013) for the largest autoregressive root of nine real exchange rates relative to the U.S., along with our HPD set for $m = 501$. The HPD set is numerically very close to what is obtained by Lopez, Murray, and Papell (2013), suggesting that at least for these data, their description of uncertainty about ρ also has reasonable conditional properties. Interestingly, even

Table 1. 95% confidence sets for largest autoregressive root in real exchange rates.

Country	Sample	Lopez et al.	ES	Hansen	HPD
Australia	1870–1998	[0.85, 1.00]	[0.84, 1.00]	[0.89, 1.00]	[0.85, 0.99]
Belgium	1880–1998	[0.77, 0.98]	[0.80, 0.99]	[0.92, 1.00]	[0.77, 0.98]
Finland	1881–1998	[0.58, 0.85]	[0.84, 0.98]	[0.90, 1.00]	[0.58, 0.84]
Germany	1880–1998	[0.89, 1.00]	[0.88, 1.00]	[0.88, 1.00]	[0.85, 0.99]
Italy	1880–1998	[0.73, 0.95]	[0.41, 0.85]	[0.89, 1.00]	[0.73, 0.94]
Netherlands	1870–1998	[0.88, 1.00]	[0.84, 1.00]	[0.88, 1.00]	[0.84, 0.99]
Spain	1880–1998	[0.85, 1.00]	[0.86, 1.00]	[0.92, 1.00]	[0.85, 0.99]
Sweden	1880–1998	[0.83, 0.99]	[0.81, 0.99]	[0.88, 1.00]	[0.82, 0.98]
UK	1870–1998	[0.78, 0.99]	[0.82, 1.00]	[0.83, 1.00]	[0.82, 0.98]

Notes: The data and the Lopez et al. (2013), Elliott and Stock (2001) (ES), and Hansen (1999) intervals are from Tables 3 and 5 in Lopez et al. (2013).

though the prior peaks at $\tilde{\theta} = 0$, the HPD set does not include the value of unity for any of nine series (the upper point of the interval for Germany is 0.998), reflecting some evidence for (weak) mean reversion. The three methods considered by Lopez, Murray, and Papell (2013) all amount to inverting 5% level two-sided hypothesis tests with substantial null rejection probability on both sides for all hypothesized values of $\tilde{\theta}$. In contrast, the 5% level hypothesis tests corresponding to the HPD set become essentially one-sided for $\tilde{\theta}$ close to 0. The 5% level DF-GLS unit root test of Elliott, Rothenberg, and Stock (1996) does reject the unit root null hypothesis $\tilde{\theta} = 0$ for all series, so our empirical results in that regard do not contradict what is obtained by standard frequentist reasoning.

As reviewed above, the literature has developed a large number of confidence sets for the largest autoregressive root. Traditional frequentist objectives, such as unbiasedness, short expected length, or maximizing the probability of excluding wrong parameter values are not easily implemented, and they do not pin down a unique set. What is more, as stressed in Sims (1988) and Sims and Uhlig (1991), an exclusive concern for frequentist properties might lead to unappealing conditional properties. Researchers famously expressed opposing frequentist and Bayesian views on near-unit root econometrics in the 1991 Oct.-Dec. issue of the *Journal of Applied Econometrics*. In this context, the approach to set estimation developed here seems particularly attractive: the determination of the coverage inducing prior is a fairly straightforward numerical problem, and the resulting set estimator has appealing frequentist and conditional properties by construction.

5. Conclusion

In non-LAN inference problems, frequentist and Bayesian set estimators do not generally coincide, even in large samples. Such problems thus bring to the forefront the different underlying rationales of classical and Bayesian inference. We prove in this article that it is always possible to reconcile these two perspectives for set estimation: there generically exists a prior that turns a $1 - \alpha$ credible set into a $1 - \alpha$ confidence set, so it is possible "to have it both ways." What is more, if the set estimation concerns the entire parameter vector, then the HPD set with coverage inducing prior is always similar and length optimal. This suggests a practical and potentially powerful approach to set estimation in nonstandard inference problems.

Supplementary Materials

The online supplement contains the appendices for the article.

Acknowledgment

We thank the anonymous referees whose comments and suggestions helped us to improve the manuscript.

Funding

The second author gratefully acknowledges support from the NSF via grant SES-1449346.

References

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856. [1238]

Andrews, D. W. K., and Chen, H. (1994), "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics*, 12, 187–204. [1238]

Andrews, D. W. K., Moreira, M. J., and Stock, J. H. (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752. [1233]

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, New York: Springer. [1233,1234]

Buehler, R. J. (1959), "Some Validity Criteria for Statistical Inferences," *The Annals of Mathematical Statistics*, 30, 845–863. [1233]

Carlstein, E., Müller, H.-G., and Siegmund, D. (eds.) (1994), *Change-Point Problems* (vol. 23 of *Lecture Notes—Monograph Series*), Hayward, CA: IMS. [1233,1237]

Chan, N. H., and Wei, C. Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063. [1233,1238]

Chernozhukov, V., Hong, H., and Tamer, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [1233]

Datta, G. S., and Mukerjee, R. (2004), *Probability Matching Priors: Higher Order Asymptotics*, New York: Springer. [1233,1235]

Dixon, P., and Parmenter, B. (1996), "Computable General Equilibrium Modelling for Policy Analysis and Forecasting," in *Handbook of Computational Economics* (vol. 1, 1 ed.), eds. M. A. H., A. K. D., and R. J., Elsevier, chap. 01, pp. 3–85. [1237]

Eaves, B. (1972), "Homotopies for Computation of Fixed Points," *Mathematical Programming*, 3, 1–22. [1237]

Elliott, G., and Müller, U. K. (2007), "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, 141, 1196–1218. [1237]

——— (2014), "Pre and Post Break Parameter Inference," *Journal of Econometrics*, 180, 141–157. [1237]

Elliott, G., Müller, U. K., and Watson, M. W. (2015), "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811. [1236]

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836. [1240]

Elliott, G., and Stock, J. H. (2001), "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics*, 103,155–181. [1238,1239]

Fisher, S. R. A. (1956), *Statistical Methods and Scientific Inference* (1st ed.), London: Oliver & Boyd. [1233]

Gospodinov, N. (2004), "Asymptotic Confidence Intervals for Impulse Responses of Near-Integrated Processes," *Econometrics Journal*, 7, 505–527. [1239]

Hansen, B. E. (1999), "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607. [1238,1239]

Imbens, G., and Manski, C. F. (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [1233]

Jorgenson, D., Ho, M. S., and Stiroh, K. J. (2008), "A Retrospective Look at the US Productivity Growth Resurgence," *Journal of Economic Perspectives*, 22, 3–24. [1238]

Joshi, V. M. (1974), "A Note on the Incompatibility of One-Sided Bayes and Frequency Confidence Intervals," *Journal of the Royal Statistical Society*, Series B, 36, 237–242. [1234,1236]

Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: Wiley. [1233]

Lopez, C., Murray, C. J., and Papell, D. (2013), "Median-Unbiased Estimation in DF-GLS Regressions and the PPP Puzzle," *Applied Economics*, 45, 455–464. [1239]

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995), *Microeconomic Theory*, Oxford, UK: Oxford University Press. [1234,1237]

Mikusheva, A. (2007), "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452. [1238]

Müller, U. K. (2011), "Efficient Tests Under a Weak Convergence Assumption," *Econometrica*, 79, 395–435. [1237]

Murray, C. J., and Papell, D. (2002), "The Purchasing Power Parity Persistence Paradigm," *Journal of International Economics*, 56, 1–19. [1239]

Perron, P. (2006), "Dealing With Structural Breaks," in *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, ed. K. Patterson, and T. C. Mills, Basingstoke, UK: Palgrave Macmillan, pp. 278–352. [1237]

Phillips, P. C. B. (1987), "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547. [1233,1238]

Pierce, D. A. (1973), "On Some Difficulties in a Frequency Theory of Inference," *The Annals of Statistics*, 1, 241–250. [1233]

Robinson, G. K. (1977), "Conservative Statistical Inference," *Journal of the Royal Statistical Society*, Series B, 39, 381–386. [1233]

Rossi, B. (2005), "Confidence Intervals for Half-Life Deviations from Purchasing Power Parity," *Journal of Business and Economic Statistics*, 23, 432–442. [1239]

Scarf, H. (1967), "The Approximation of Fixed Points of a Continuous Mapping," *SIAM Journal on Applied Mathematics*, 15, 1328–1343. [1237]

Scarf, H., and Hansen, T. (1973), *The Computation of Economic Equilibria*, Cowles Foundation Monograph Series, New Haven, CT: Yale University Press. [1237]

Sims, C. A. (1988), "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control*, 12, 463–474. [1238,1240]

Sims, C. A., and Uhlig, H. (1991), "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, 59, 1591–1599. [1238,1240]

Staiger, D., and Stock, J. H. (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [1233]

Stock, J. H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459. [1238]

——— (1994), "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics* (vol. 4), eds. R. F. Engle, and D. McFadden, New York: North Holland, pp. 2740–2841. [1237]

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge, UK: Cambridge University Press. [1235]

Yang, Z. (1999), *Computing Equilibria and Fixed Points : The Solution of Nonlinear Inequalities*, Boston, MA: Kluwer Academic. [1237]