Coverage Inducing Priors in Nonstandard Inference Problems*

Ulrich K. Müller and Andriy Norets
Princeton University and Brown University
Current version: April 2015

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.

Keywords: confidence sets, objective Bayes, conditional coverage, probability matching, nonstandard inference problems, unit roots.

*We thank the participants of the Northwestern Junior Festival on New Developments in Microeconometrics, the Cowles Foundation Summer Econometrics Conference, and the seminars at Cornell, Princeton, and Columbia for useful discussions. The second author gratefully acknowledges support from the NSF via grant SES-1260861.
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\(^1\) 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

\(^1\)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.
for highly persistent time series with an autoregressive root near unity (e.g., Phillips (1987), Chan and Wei (1987)).

In this paper 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, for instance, 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 sets for the date and magnitude of a structural break in a time series model and for the largest root in an local-to-unity autoregressive model.

2 Main Results

2.1 Definitions and Notation

We start by assuming a finite parameter space, \( \Theta = \{\theta_1, \ldots, \theta_m\} \), 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 \( \theta \), \( P(\cdot|\theta) \), has density \( p(x|\theta) \) with respect to a \( \sigma \)-finite measure \( \nu \). Without loss of generality, we assume that \( \nu \) 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, define a rejection probability function $\varphi: \Theta \times \mathcal{X} \mapsto [0,1]$, with $\varphi(\theta, x)$ the probability of not including $\theta$ conditional on having observed $X = x$. If $\varphi(\theta_j, \cdot)$ defines a level $\alpha$ test of $H_0 : \theta = \theta_j$,

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

(1)

for all $j = 1, \ldots, m$, 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 $\varphi$ 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 $j = 1, \ldots, m$.

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

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

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

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

or

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

(2)

In this definition, 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 subsection, 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 $\pi$ on $\Theta = \{\theta_1, \ldots, \theta_m\}$. Define $z_j(\pi) = \int [\varphi(\theta_j, x; \pi) - \alpha] p(x|\theta_j) d\nu(x)$ and $z(\pi) = (z_1(\pi), \ldots, z_m(\pi))'$. Assume $z(\pi)$ is continuous in $\pi$. Then there exists $\pi^*$ such that $\varphi(\theta_j, x; \pi^*)$ defines a $1 - \alpha$ confidence set, that is

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

**Proof.** The proof is analogous to the proof of the equilibrium existence in an exchange economy with $m$ goods ($\pi$ 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 $z^+_j(\pi) = \max\{0, z_j(\pi)\}$, $z^+(\pi) = (z^+_1(\pi), \ldots, z^+_m(\pi))'$, 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 \( \Delta^{m-1} \) is an \((m-1)\)-simplex. By Brouwer’s fixed point theorem there exists \( \pi^* \) such that \( \pi^* = q(\pi^*) \).

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

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

Therefore,

\[
0 = (\pi^*)'z(\pi^*) = q(\pi^*)'z(\pi^*) = z^+(\pi^*)'z(\pi^*) / \sum_{j=1}^{m} [\pi^*_j + z^+_j(\pi^*)]
\]

and \( z^+(\pi^*)'z(\pi^*) = 0 \). The latter equality implies \( z_j(\pi^*) \leq 0 \) for all \( j \in \{1, \ldots, m\} \), 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) shows 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 \( \pi \) 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 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 \( \pi^* \) puts positive mass on all \( \theta \in \Theta \) and the resulting set \( \varphi(\theta, x; \pi^*) \) is a similar confidence set of level \( 1 - \alpha \), i.e. \( \int \varphi(\theta, x; \pi^*)p(x|\theta)d\nu(x) = \alpha \) for all \( \theta \in \Theta \). (iii) There does not exist a level \( 1 - \alpha \) confidence set \( \varphi' \) that is weakly shorter than \( \varphi(\theta, x; \pi^*) \) for all \( x \in \mathcal{X} \) (\( \sum_{j=1}^m \varphi' (\theta_j, x) \geq \sum_{j=1}^m \varphi (\theta_j, x; \pi^*) \)) and strictly shorter for \( x \in \mathcal{X}_i \) with \( \nu(\mathcal{X}_i) > 0 \) (\( \sum_{j=1}^m \varphi' (\theta_j, x) > \sum_{j=1}^m \varphi (\theta_j, x; \pi^*) \)).

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 \( \theta \) with exact probability \( 1 - \alpha \); and there does not exist a \( 1 - \alpha \) confidence set that is
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 $\varphi$ are then $\Gamma \times \mathcal{X} \mapsto [0, 1]$ functions, but priors still specify a probability distribution on $\Theta$. Theorem 1 still goes through as above 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 $\pi$, so there still exists a prior that induces coverage. The similarity of $\varphi$ in Theorem 2, however, does not necessarily hold, since coverage inducing prior of the HPD set might have zero mass for some $\theta$. The HPD confidence set might thus have coverage above $1 - \alpha$ for these $\theta$’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 $\nu_p$. 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 $\theta$ and
observed \( x \), the probability that the prediction set \( \varphi_p \) 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 \( \varphi_p \) 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 subsection, we apply Theorem 1 to settings with a general parameter space, hierarchical Bayesian priors, and weighted average coverage. In the following subsection, 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 \( \psi_j, j = 1, \ldots, m \), 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

\[
\Pi(\tilde{\theta}, \pi) = \sum_{j=1}^{m} \pi_j \cdot \psi_j(\tilde{\theta})
\]  

(3)

can be thought of as a hierarchical prior distribution on \( \tilde{\Theta} \), where \( \pi \) is a prior on a set of models \( \Theta = (\theta_1, \ldots, \theta_m) \) with common likelihood \( \tilde{p}(x|\tilde{\theta}) \) and \( \psi_j \) is a prior on \( \tilde{\theta} \) under model \( \theta_j \).
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 function of \( \pi \), with the \( \psi_j \)'s fixed. Define the marginal likelihood under model \( \theta_j \) by
\[
p(x|\theta_j) = \int \tilde{p}(x|\tilde{\theta})\psi_j(\tilde{\theta})d\tilde{\theta}.
\]
Also, let
\[
\varphi(\theta_j, x; \pi) = \frac{\int \tilde{\varphi}(\tilde{\theta}, x; \pi)\tilde{p}(x|\tilde{\theta})\psi_j(\tilde{\theta})d\tilde{\theta}}{p(x|\theta_j)} \quad (4)
\]
be the posterior non-coverage probability conditional on model \( \theta_j \). The prior weighted coverage of the set \( \tilde{\varphi}(\tilde{\theta}, x; \pi) \) in model \( \theta_j \) is then defined by
\[
\int \left[ \int [1 - \tilde{\varphi}(\tilde{\theta}, x; \pi)\tilde{p}(x|\tilde{\theta})]d\nu(x) \right] \psi_j(\tilde{\theta})d\tilde{\theta} = \int [1 - \varphi(\theta_j, x; \pi)]p(x|\theta_j)d\nu(x). \quad (5)
\]
Note that the right-hand side of (5) is the same as the coverage of \( \theta_j \) 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 \( \pi \), then Theorem 1 implies the existence of \( \pi^* \) such that (5) evaluated at \( \pi^* \) is at least \( 1 - \alpha \) for all models \( \theta_j, j = 1, \ldots, m \). In other words, Bayesian credible sets under the \( \pi^* \)-mixture prior have \( \psi_j \)-weighted average coverage of at least \( 1 - \alpha \) for all \( j = 1, \ldots, m \).

Let us note that the resulting sets are invariant to re-parameterization in the following sense. For a one-to-one transformation \( f \) on \( \tilde{\Theta} \) with differentiable inverse \( f^{-1} \), define \( \tilde{\varphi}_j(s, x; \pi^*) = \tilde{\varphi}(f^{-1}(s), x; \pi^*) \) for \( s \in f(\tilde{\Theta}) \). Then, \( \tilde{\varphi}_j \) defines a \( 1 - \alpha \)
credible set for \( f(\tilde{\theta}) \) under the prior \( \sum_j \pi_j^* \psi_j(f^{-1}(s))|df^{-1}(s)/ds| \) 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 \( \max_j \text{diam}(\tilde{\Theta}_j^m) \to 0 \) as \( m \to \infty \). For a fixed \( m \), define \( \psi_j \) to be a uniform probability density on \( \tilde{\Theta}_j^m \), \( j = 1, \ldots, m \). Define \( \tilde{\phi}_m(\tilde{\theta}, x; \pi) \) 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, \( \varphi_j(\theta, x; \pi) \), maximizes \( \sum_j \varphi_j(\theta_x) \text{vol}(\tilde{\Theta}_j^m) \) 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 \( \pi^m \) such that

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

The following theorem shows that sets \( \tilde{\phi}_m(\tilde{\theta}, x; \pi^m) \) 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 \( \tilde{\phi}_m(\tilde{\theta}, x; \pi^m) \) is within \( \epsilon \) 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 \( X_m \) with, \( \nu(X_m) > 0 \) such that

\[
\int_{\tilde{\Theta}} \tilde{\phi}(\tilde{\theta}, x)d\tilde{\theta} < (1 + \epsilon) \cdot \int_{\tilde{\Theta}} \tilde{\phi}_m(\tilde{\theta}, x; \pi^m)d\tilde{\theta} \tag{7}
\]
for all $x \in \mathcal{X}_m$.

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

$$\int_{\tilde{\Theta}} [1 - \tilde{\phi}(\tilde{\theta}, x)] d\tilde{\theta} > \int_{\tilde{\Theta}} [1 - \tilde{\varphi}^m(\bar{\theta}, x; \pi^{m*})] d\bar{\theta} - \varepsilon \cdot \text{vol}(\Theta).$$

Thus, part (ii) of Theorem B 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 $\nu$ or its volume is strictly larger on a set of positive measure $\nu$.

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 improper prior can induce coverage of one-sided confidence sets under an unbounded parameter space.

At the same time, a non-standard 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 $\zeta(\tilde{\theta}) \to \infty$ for some function $\zeta: \tilde{\Theta} \mapsto \mathbb{R}$. Pick $\kappa_S$ large enough so that $\tilde{\Theta}_S = \{\tilde{\theta} : \zeta(\tilde{\theta}) > \kappa_S\}$ 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 $\tilde{\Theta}_{NS} = \tilde{\Theta} \setminus \tilde{\Theta}_S$ 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 $\tilde{\Theta}_m = \{\tilde{\theta} : \kappa_S < \zeta(\tilde{\theta}) < \kappa_m\}$, where $\kappa_m \to \infty$. Proceed similarly to Section 2.3.3 and define $\Pi^m(\tilde{\theta}, \pi)$ with $\psi_j$ 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 $\tilde{\Theta}_j^m$ for $j = 1, \ldots, m - 1$, but without any restrictions on $\tilde{\Theta}_m^m$. The results of the previous subsections then imply that there exists a large enough $m$ and prior $\pi_{m*}$ such that $\tilde{\phi}^m(\tilde{\theta}, x; \pi_{m*})$ has near uniform coverage for parameters in $\tilde{\Theta}_{NS}$. Also, the prior $\Pi^m(\tilde{\theta}, \pi_{m*})$, 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 $\kappa_S \ll \zeta(\tilde{\theta}) \ll \kappa_m$. Further, continuously extend $\Pi^m(\tilde{\theta}, \pi_{m*})$ to an improper prior $\Pi^m(\tilde{\theta}, \pi_{m*})$ that is flat on $\tilde{\Theta}_S$, and define $\bar{\phi}^m(\tilde{\theta}, x; \pi_{m*})$ to be the shortest $1 - \alpha$ credible set relative to $\Pi^m(\tilde{\theta}, \pi_{m*})$ that is constant on $\tilde{\Theta}_j^m$, $j = 1, \ldots, m - 1$. Then, one would expect $\bar{\phi}^m(\tilde{\theta}, x; \pi_{m*})$ to have approximately $1 - \alpha$ coverage for all parameter values with $\zeta(\tilde{\theta}) \gg \kappa_S$, and also for
\( \tilde{\theta} \in \tilde{\Theta}_{NS}, \) since \( \tilde{\phi}^m(\tilde{\theta}, x; \pi^{m*}) \) is presumably close to \( \tilde{\phi}^m(\bar{\theta}, x; \pi^{m*}) \) for all \( \bar{\theta} \in \bar{\Theta}_{NS} \).

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

Nevertheless, it is possible to numerically check whether \( \tilde{\phi}^m(\tilde{\theta}, x; \pi^{m*}) \) has approximately nominal coverage for \( \tilde{\theta} \) with \( \zeta(\tilde{\theta}) \) close to \( \kappa_S \). 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) provide 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 \( \pi \) and iteratively increase its coordinates \( \pi_j \) in proportion to undercoverage of \( \varphi(\theta, x; \pi) \) at \( \theta_j \). 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 for additional implementation details). This procedure is not guaranteed to converge in general; some counterexamples are given in 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 \( \theta_i \) for all \( i \neq j \) (see, for example, 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 \( \theta_i \)'s that are close to \( \theta_j \). A weaker sufficient condition, \((\pi^*)'z(\pi) > 0\) for any \( \pi \neq \pi^* \) 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 subsections illustrate the approach for the construction of 95% nominal level confidence sets in two non-standard inference problems. Additional computational details may be found in Appendix C.

4.1 Break Date and Magnitude

In this subsection, 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

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

where \( \varepsilon_t \sim iid \mathcal{N}(0,1) \). As argued by Elliott and Müller (2007), a moderate break magnitude is usefully modelled 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 \( \mu \), we find that the partial sum process of the demeaned observations \( \tilde{y}_t = y_t - T^{-1} \sum_{s=1}^T y_s \) satisfies

\[
T^{-1/2} \sum_{t=1}^{sT} \tilde{y}_t \sim X(s) = W(s) - sW(1) - \delta(\min(\lambda, s) - \lambda s)
\]
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 show 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 $\beta$ undergoes a break of magnitude $\delta/\sqrt{T I_{\beta\beta}}$, with $I_{\beta\beta}$ the Fisher information about $\beta$ in the stable model. The same limiting problem may also be motivated using the framework in Müller (2011) without reference to a parametric model.

Figure 1: Break Application

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

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

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 $\tilde{\Theta}_m^j$ of size 0.0025 $\times$ 0.05. Panel (a) in Figure 1 displays the mixture prior $\Pi^m(\tilde{\theta}, \pi^{m*})$ that induces the 95% constrained HPD set $\tilde{\varphi}^m$ to have average coverage of 95% on $\tilde{\Theta}_m^j$. 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 $\times$ 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 in Jorgenson, Ho, and 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 this 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 seems 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 to 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

$$y_t = \mu + u_t, \ t = 1, \ldots, T \tag{10}$$

$$(1 - \rho L)\phi(L)u_t = \varepsilon_t$$

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^{p-1}$, $\varepsilon_t \sim iid(0, \sigma^2)$ and $u_0 = O_p(1)$. The largest root $\rho$ is close to unity, while the roots of $\phi$ are all bounded away from the complex unit circle. As in Phillips (1987), 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 $\rho$ is close to one.

The appropriate limit experiment under $\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2)$ 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 $W(r)$ a standard Wiener process. The density of $X$ relative to the measure of a standard Wiener process is given by

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

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) discuss 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 \( \rho \) 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 application of inference about \( \rho \) arises in international economics: the theory of purchasing power parity (PPP) implies that real exchange rates are mean reverting, and the value of \( \rho \) 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 \( \tilde{\Pi}_m(\tilde{\theta}, \pi^m) \) normalized to one on \( \tilde{\Theta}_S \). Figure 2 displays \( \tilde{\Pi}_m(\tilde{\theta}, \pi^m) \) for \( m \in \{11, 51, 201, 501\} \). The global shape of \( \tilde{\Pi}_m(\tilde{\theta}, \pi^m) \) 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 \( \tilde{\Pi}_m(\tilde{\theta}, \pi^m) \) 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 $\rho$. Unreported results show that if one casts the small sample problem of inference about $\rho$ in the asymptotic form \((11)\), and applies the prior of Figure 2 with $m = 500$, then one obtains a HPD set with frequentist coverage close to the nominal level also in small samples.

![Graphs showing HPD sets for $\rho$ with different $m$ values](image)

Figure 2: Autoregressive Root: Priors that Induce Average Frequentist Coverage of Constrained HPD Set

Table 1 reports the empirical results of Lopez, Murray, and Papell (2013) for the largest autoregressive root of 9 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 this data,
their description of uncertainty about $\rho$ also has reasonable conditional properties. Interestingly, even though the prior peaks at $\tilde{\theta} = 0$, the HPD set does not include the value of unity for any of 9 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 zero. 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.

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

<table>
<thead>
<tr>
<th>Country</th>
<th>Sample</th>
<th>Lopez et al.</th>
<th>ES</th>
<th>Hansen</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1870-1998</td>
<td>[0.85, 1.00]</td>
<td>[0.84, 1.00]</td>
<td>[0.89, 1.00]</td>
<td>[0.85, 0.99]</td>
</tr>
<tr>
<td>Belgium</td>
<td>1880-1998</td>
<td>[0.77, 0.98]</td>
<td>[0.80, 0.99]</td>
<td>[0.92, 1.00]</td>
<td>[0.77, 0.98]</td>
</tr>
<tr>
<td>Finland</td>
<td>1881-1998</td>
<td>[0.58, 0.85]</td>
<td>[0.84, 0.98]</td>
<td>[0.90, 1.00]</td>
<td>[0.58, 0.84]</td>
</tr>
<tr>
<td>Germany</td>
<td>1880-1998</td>
<td>[0.89, 1.00]</td>
<td>[0.88, 1.00]</td>
<td>[0.88, 1.00]</td>
<td>[0.85, 0.99]</td>
</tr>
<tr>
<td>Italy</td>
<td>1880-1998</td>
<td>[0.73, 0.95]</td>
<td>[0.41, 0.85]</td>
<td>[0.89, 1.00]</td>
<td>[0.73, 0.94]</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1870-1998</td>
<td>[0.88, 1.00]</td>
<td>[0.84, 1.00]</td>
<td>[0.88, 1.00]</td>
<td>[0.84, 0.99]</td>
</tr>
<tr>
<td>Spain</td>
<td>1880-1998</td>
<td>[0.85, 1.00]</td>
<td>[0.86, 1.00]</td>
<td>[0.92, 1.00]</td>
<td>[0.85, 0.99]</td>
</tr>
<tr>
<td>Sweden</td>
<td>1880-1998</td>
<td>[0.83, 0.99]</td>
<td>[0.81, 0.99]</td>
<td>[0.88, 1.00]</td>
<td>[0.82, 0.98]</td>
</tr>
<tr>
<td>UK</td>
<td>1870-1998</td>
<td>[0.78, 0.99]</td>
<td>[0.82, 1.00]</td>
<td>[0.83, 1.00]</td>
<td>[0.82, 0.98]</td>
</tr>
</tbody>
</table>
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).

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 paper 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.
A Proof of Theorem 2

(i) If the likelihood ratio \( p(X|\theta_i)/p(X|\theta_j) \) is an absolutely continuous random variable for any \( i \neq j \), then \( p(X|\theta_j), j = 1, \ldots, m \), have the same support and the posterior distribution is well defined for any prior \( \pi \) and \( \nu \)-almost all \( x \in \mathcal{X} \). Moreover, ties in the posterior probabilities \( (p(\theta_i|X) = p(\theta_j|X), i \neq j) \) happen with probability zero under any \( \theta \in \Theta \). An HPD credible set \( \varphi(\cdot, \cdot; \pi) \) is uniquely defined and continuous in \( \pi \) whenever there are no ties in the posterior probabilities. The function \( z(\pi) \) defined in Theorem 1 is therefore continuous in \( \pi \) and Theorem 1 implies that there exists a prior \( \pi^* \) for which \( \varphi(\cdot, \cdot; \pi^*) \) has coverage of at least \( 1 - \alpha \).

(ii) Next, let us show that \( \pi^*_j > 0 \) for any \( j \) and \( \varphi(\cdot, \cdot; \pi^*) \) is a similar \( 1 - \alpha \) confidence set. If \( \pi^*_j = 0 \) for some \( j \) then \( \theta_j \) is not contained in the \( 1 - \alpha \) HPD credible set for any \( x \) and \( \varphi(\cdot, \cdot; \pi^*) \) has zero coverage at \( \theta_j \). Thus, \( \pi^*_j > 0 \) for all \( j \).

Since \( \varphi(\cdot, \cdot; \pi^*) \) is a \( 1 - \alpha \) credible set
\[
\sum_{j=1}^{m} \varphi(\theta_j, x; \pi^*) p(x|\theta_j) \pi^*_j = \alpha \sum_{j=1}^{m} p(x|\theta_j) \pi^*_j.
\]
Integration of the last display implies
\[
\sum_{j=1}^{m} \left[ \int \varphi(\theta_j, x; \pi^*) p(x|\theta_j) d\nu(x) \right] \pi^*_j = \alpha. \tag{12}
\]
Since the coverage of \( \varphi(\cdot, \cdot; \pi^*) \) is at least \( 1 - \alpha \), \( \int \varphi(\theta_j, x; \pi^*) p(x|\theta_j) d\nu(x) \leq \alpha \). Because \( \pi^*_j > 0 \) for all \( j \) the equality in (12) can hold only if \( \int \varphi(\theta_j, x; \pi^*) p(x|\theta_j) d\nu(x) = \alpha \) for all \( j \) or, in other words, \( \varphi(\cdot, \cdot; \pi^*) \) is similar.

(iii) It suffices to show that if \( \sum_{j=1}^{m} \varphi'(\theta_j, x) \geq \sum_{j=1}^{m} \varphi(\theta_j, x; \pi^*) \) for all \( x \in \mathcal{X} \) and \( \sum_{j=1}^{m} \varphi'(\theta_j, x) > \sum_{j=1}^{m} \varphi(\theta_j, x; \pi^*) \) for all \( x \in \mathcal{X}' \) with \( \nu(\mathcal{X}') > 0 \), then
\[ \int \varphi'(\theta_j, x)p(x|\theta_j)\,d\nu(x) > \alpha \text{ for some } j. \]

The HPD set \( \varphi(\theta, x; \pi^*) \) can be defined for \( \nu \)-almost all \( x \) by the minimum length property that for all \( \varphi'' \) with \( \sum_{j=1}^{m} \varphi''(\theta_j, x) = \sum_{j=1}^{m} \varphi(\theta_j, x; \pi^*) \), \( \sum_{j=1}^{m} \varphi(\theta_j, x; \pi^*)p(\theta_j|x) \leq \sum_{j=1}^{m} \varphi''(\theta_j, x)p(\theta_j|x) \). Thus, for \( \nu \)-almost all \( x \in X \),

\[ \sum_{j=1}^{m} \varphi(\theta_j, x; \pi^*)p(x|\theta_j)\pi^*_j < \sum_{j=1}^{m} \varphi'(\theta_j, x)p(x|\theta_j)\pi^*_j \]

and for all \( x \in X \) the inequality holds weakly. Integrating this inequality with respect to \( \nu \) yields \( \sum_{j=1}^{m} \pi^*_j \int (\varphi(\theta_j, x; \pi^*) - \varphi'(\theta_j, x))p(x|\theta_j)\,d\nu(x) < 0 \). Since \( \sum_{j=1}^{m} \pi^*_j \int \varphi(\theta_j, x; \pi^*)p(x|\theta_j)\,d\nu(x) = \alpha \) by part (ii), this implies that there exists \( j \) such that \( \int \varphi'(\theta_j, x)p(x|\theta_j)\,d\nu(x) > \alpha \).

\section{Proof of Theorem 3}

(i) By the assumed uniform continuity and \( \max_j \text{diam}(\tilde{\Theta}^m_j) \to 0 \), there exists \( M_\epsilon \) such that for any \( m \geq M_\epsilon \) and \( \tilde{\theta}_1, \tilde{\theta}_2 \in \tilde{\Theta}^m_j, j = 1, \ldots, m, \)

\[ \int |\tilde{p}(x|\tilde{\theta}_1) - \tilde{p}(x|\tilde{\theta}_2)|\,d\nu(x) < \epsilon. \] (13)

In order to obtain a contradiction, assume there exists \( j^* \) and \( \tilde{\theta}^* \in \tilde{\Theta}^m_j \) such that

\[ \int \tilde{\varphi}^m(\tilde{\theta}^*, x; \pi^{m*})\tilde{p}(x|\tilde{\theta}^*)\,d\nu(x) < \alpha - \epsilon. \] (14)

For any \( \tilde{\theta}_1 \in \tilde{\Theta}^m_j, \tilde{\varphi}^m(\tilde{\theta}_1, x; \pi^{m*}) = \tilde{\varphi}^m(\tilde{\theta}_1, x; \pi^{m*}) \) as \( \tilde{\varphi}^m \) is constant on \( \tilde{\Theta}^m_j \), by definition. Therefore, by (13) and (14), \( \int \tilde{\varphi}^m(\tilde{\theta}_1, x; \pi^{m*})\tilde{p}(x|\tilde{\theta}_1)\,d\nu(x) < \alpha, \forall \tilde{\theta}_1 \in \tilde{\Theta}^m_j \), which would make the equality in (6) impossible. A contradiction for \( \int \tilde{\varphi}^m(\tilde{\theta}^*, x; \pi^{m*})\tilde{p}(x|\tilde{\theta}^*)\,d\nu(x) > \alpha + \epsilon \) can be obtained in the same way.
(ii) Suppose the claim does not hold. Then, there exists a subsequence \( \{m_k\} \) with
\[
\int \tilde{\phi}(\theta, x) d\tilde{\theta} \geq (1 + \epsilon) \cdot \int \tilde{\varphi}^{m_k}(\theta, x; \pi^{m_k*}) d\tilde{\theta}
\]
for \( \nu \)-almost all \( x \). Pick \( m_k > M_\epsilon \), with \( M_\epsilon \) defined in part (i) of this proof. For this \( m_k \), and any \( \tilde{\theta} \in \tilde{\Theta}^{m_k}_j \), define
\[
\phi'(\theta_j, x) = \tilde{\phi}'(\theta, x) = \int_{\tilde{\Theta}_j} \tilde{\phi}(\theta_1, x) d\tilde{\theta}_1/V_j \text{ and } \phi''(\theta_j, x) = \tilde{\phi}''(\tilde{\theta}, x) = \tilde{\phi}'(\tilde{\theta}, x)/(1 + \epsilon),
\]
where \( V_j = \text{vol}(\tilde{\Theta}^{m_k}_j) \), \( j = 1, \ldots, m_k \). Note that
\[
\int \phi(\tilde{\theta}, x) d\tilde{\theta} = \int \tilde{\phi}'(\tilde{\theta}, x) d\tilde{\theta} = (1 + \epsilon) \int \tilde{\phi}''(\tilde{\theta}, x) d\tilde{\theta} = (1 + \epsilon) \sum_j \phi''(\theta_j, x)V_j.
\]
Thus, \( \sum_j \phi''(\theta_j, x)V_j \geq \int \tilde{\varphi}^{m_k}(\theta, x; \pi^{m_k*}) d\tilde{\theta} \geq \sum_j \varphi^{m_k}(\theta_j, x; \pi^{m_k*})V_j \), and since \( \varphi^{m_k}(\theta, x; \pi^{m_k*}) \) maximizes \( \sum_j \varphi(\theta_j, x)V_j \) subject to \( \sum_j |\alpha - \varphi(\theta_j, x)|p(x|\theta_j)\pi^{m_k*}_j \leq 0 \),
\[
\sum_j \phi''(\theta_j, x)p(x|\theta_j)\pi^{m_k*}_j \geq \sum_j \varphi^{m_k}(\theta_j, x; \pi^{m_k*})p(x|\theta_j)\pi^{m_k*}_j.
\]
Integration of this inequality with respect to \( \nu \) gives
\[
\sum_j \pi^{m_k*}_j \int \phi''(\theta_j, x)p(x|\theta_j) d\nu(x) \geq \alpha.
\]
Thus, there exists \( j^* \) such that \( \int \phi''(\theta_{j^*}, x)p(x|\theta_{j^*}) d\nu(x) \geq \alpha \) and
\[
\int \phi'(\theta_{j^*}, x)p(x|\theta_{j^*}) d\nu(x) \geq (1 + \epsilon)\alpha. \tag{15}
\]
At the same time,
\[
\int \phi'(\theta_{j^*}, x)p(x|\theta_{j^*}) d\nu(x) = \int_{\tilde{\Theta}^{m_k*}_j} \int_{\tilde{\Theta}^{m_k*}_j} \int_{\tilde{\Theta}^{m_k*}_j} \phi(\tilde{\theta}_1, x)p(x|\tilde{\theta}_2) d\nu(x) d\tilde{\theta}_1 d\tilde{\theta}_2/V^2_{j^*}.
\]
Because \( \int \phi(\tilde{\theta}_1, x)p(x|\tilde{\theta}_1) d\nu(x) \leq \alpha \) for all \( \tilde{\theta}_1 \in \tilde{\Theta} \) by assumption on \( \phi \),
\[
\int \phi'(\theta_{j^*}, x)p(x|\theta_{j^*}) d\nu(x) \leq \alpha + \sup_{\tilde{\theta}_1, \tilde{\theta}_2 \in \tilde{\Theta}^{m_k*}_j} \int |\tilde{\nu}(x|\tilde{\theta}_1) - \tilde{\nu}(x|\tilde{\theta}_2)| d\nu(x) < (1 + \epsilon)\alpha. \tag{16}
\]
Combining (15) and (16) yields the desired contradiction.
C Computational Details for Applications

C.1 Algorithm

The fixed point iterations described in Section 3 require repeated evaluation of coverage probabilities. These may be computed using an importance sampling approach: Let $\bar{p}$ be a proposal density such that $p(\theta_j|x)$ is absolutely continuous with respect to $\bar{p}$, and let $X_i, i = 1, \ldots, N$ be $N$ i.i.d. draws from $\bar{p}$. Then non-coverage probability of a set $\varphi$ at $\theta_j$ can then be written as

$$RP_j = \int \varphi(\theta_j, x)p(\theta_j|x)\nu(x) = \int \varphi(\theta_j, x)\frac{p(\theta_j|x)}{\bar{p}(x)}\bar{p}(x)\nu(x),$$

yielding the approximation

$$\hat{R}P_j(\varphi) = N^{-1} \sum_{i=1}^{N} \varphi(\bar{\theta}, X_i) \frac{p(\theta_j|X_i)}{\bar{p}(X_i)}.$$

Write $\varphi_\pi$ for the set $\varphi(\theta_j, x; \pi)$ of Theorem 1. We employ the following algorithm to obtain an approximate $\pi^*$ such that the HPD set $\varphi_\pi^*$ has nearly coverage close to the nominal level:

1. Compute and store $\frac{p(\theta_j|X_i)}{\bar{p}(X_i)}$, $i = 1, \ldots, N, j = 1, \ldots, m$.

2. Initialize $\pi^{(0)}$ at $\pi_j^{(0)} = 1/m, j = 1, \ldots, m$.

3. For $l = 0, 1, \ldots$
   
   (a) Compute $z_j = \hat{R}P_j(\varphi_\pi^{(l)}) - \alpha, j = 1, \ldots, m$.

   (b) If $\max_j z_j - \min_j z_j < \varepsilon$, set $\pi^* = \pi^{(l)}$ and end.

   (c) Otherwise, set $\pi_j^{(l+1)} = \exp(\omega z_j)\pi_j^{(l)} / \sum_{k=1}^{m} \exp(\omega z_k)\pi_k^{(l)}, j = 1, \ldots, m$, and go to step 3a.
We set $\varepsilon = 0.0003$, and found $\omega = 1.5$ to yield reliable results as long as $N$ is chosen large enough.

In the context of obtaining a credible set with approximately uniform coverage in a bounded but continuous set $\tilde{\Theta}$, we employ the above algorithm for a given partition $m$ with $\varphi = \varphi^m(\theta_j, x; \pi)$ now defined as described in Section 2.3.3. In addition, we evaluate the uniform coverage properties of the resulting set estimator on $\tilde{\Theta}$, $\tilde{\varphi}^m(\tilde{\theta}, x; \pi^{m*})$, by computing the (approximate) non-coverage probabilities

$$\hat{\text{RP}}(\tilde{\theta}) = N^{-1} \sum_{i=1}^{N} \hat{\varphi}^m(\tilde{\theta}, X_i; \pi^{m*}) \frac{p(\tilde{\theta}|X_i)}{\bar{p}(X_i)}$$

over a fine grid of values of $\tilde{\theta}$. If these uniform properties are unsatisfactory, then the algorithm is repeated using a finer partition.

For an unbounded parameter space $\tilde{\Theta}$, we implement the guess and verify approach described in Section 2.3.4. We first choose an appropriate $\kappa_S$ by computing the coverage of the HPD set relative to a flat prior on $\tilde{\Theta}$, and select $\kappa_S$ to be just large enough for coverage to be sufficiently close to $1 - \alpha$ for all $\tilde{\theta}$ with $\varsigma(\tilde{\theta}) > \kappa_S$. We then partition $\tilde{\Theta}_{NS}$ into $m - 1$ subsets, and set $\tilde{\Theta}_m = \tilde{\Theta}_S$. As discussed in Section 2.3.4, it makes sense to rule out a large discontinuity of $\bar{\Pi}^m(\tilde{\theta}, \pi^{m*})$ at the boundary between $\tilde{\Theta}_{NS}$ and $\tilde{\Theta}_S$. Thus, in the above algorithm, we directly adjust the $m - 1$ values of $\bar{\Pi}^m(\tilde{\theta}, \pi)$ on $\tilde{\Theta}_j$, $j = 1, \ldots, m - 1$ without any scale normalization, and simply set $\bar{\Pi}^m(\tilde{\theta}, \pi)$ on $\tilde{\Theta}_S$ equal to the value of $\bar{\Pi}^m(\tilde{\theta}, \pi)$ of a subset $\tilde{\Theta}_{j}^m$ neighboring $\tilde{\Theta}_S$. This has the additional advantage of avoiding computation of the potentially ill-defined $\text{RP}_m$. After the iterations have concluded, we evaluate the coverage properties of the resulting set estimator $\bar{\varphi}^m(\tilde{\theta}, x; \pi^{m*})$ on a fine grid on $\tilde{\Theta}_{NS}$, and on a fine grid in the part of $\tilde{\Theta}_S$ where $\bar{\varphi}^m(\tilde{\theta}, x; \pi^{m*})$ is affected by the shape of the prior.
If these are unsatisfactory, we increase $m$ and/or $\kappa_S$. 

C.2 Details for Break Date and Magnitude

The continuous process $X$ is approximated with 800 steps. Uniform coverage is evaluated on the Cartesian grid with $\lambda \in \{0.15, 0.15125, 0.1525, \ldots, 0.85\}$ and $\delta \in \{-15.0, -14.99, -14.98, \ldots, 15\}$. We set $N = 3 \cdot 10^6$, and $\bar{p}$ to be uniform on $\{ (\lambda, \delta) : 0.13 \leq \lambda \leq 0.87, -16 \leq \delta \leq 16 \}$, which yields Monte Carlo standard deviations of coverage probabilities uniformly smaller than 0.001. We impose symmetry with respect to the sign of $\delta$, and around $\lambda = 0.5$, in the computation of $\pi^{m*}$. Computations for the finer partition take about 6 hours on a modern PC.

In the application, we use Elliott and Müller’s (2014) estimate of 2.6 for the long-run standard deviation of the quarterly data $y_t$.

C.3 Details for Autoregressive Root Near Unity

The continuous process $X$ is approximated with 800 steps. Uniform coverage is evaluated on 5001 values $\tilde{\theta} \in \{120(\frac{j}{5000})^2\}_{j=0}^{5000}$. We set $N = 1.5 \cdot 10^6$, and set $\bar{p}$ to be uniform on the 101 values $\tilde{\theta} \in \{160(\frac{j}{100})^2\}_{j=0}^{100}$, which yields Monte Carlo standard deviations of coverage probabilities uniformly smaller than 0.001.

For the application, we rely on output of the DF-GLS regression also employed in Lopez, Murray, and Papell (2013) to obtain small sample analogues to $\int_0^1 X(s)dX(s)$ and $\int_0^1 X(s)^2ds$. Specifically, let $\hat{\rho}$, $\hat{\sigma}_\rho$ and $\hat{\phi}$ be the usual OLS estimate of $\rho$, its standard error and the additional coefficients in an aug-
mented Dickey-Fuller regression using GLS demeaned data (with lag length as determined by Lopez, Murray, and Papell (2013)). We then employ the analogue

\[ \left( T^{-1} \hat{\phi}(1)(\hat{\rho} - 1)/\hat{\sigma}_\rho^2, T^{-2} \hat{\phi}(1)^2/\hat{\sigma}_\rho^2 \right) \Rightarrow \left( \int_0^1 X(s) dX(s), \int_0^1 X(s)^2 ds \right) \]

for the empirical results in Table 1.
References


