Credibility of Confidence Sets in Nonstandard Econometric Problems*

Ulrich K. Müller and Andriy Norets
Princeton University

September 2012
Revised July 2013

Abstract

Confidence intervals are commonly used to describe parameter uncertainty. In nonstandard problems, however, their frequentist coverage property does not guarantee that they do so in a reasonable fashion. For instance, confidence intervals may be empty or extremely short with positive probability, even if they are based on inverting powerful tests. We apply a betting framework to formalize the "reasonableness" of confidence intervals as descriptions of parameter uncertainty, and use it for two purposes. First, we quantify the degree of unreasonableness of previously suggested confidence intervals in nonstandard problems. Second, we derive alternative confidence sets that are reasonable by construction. We apply our framework to inference about a parameter near a boundary and a local-to-unity autoregressive root. We find that previously suggested confidence intervals are not reasonable, and numerically determine alternative confidence sets that satisfy our criteria.

JEL classification: C18

Keywords: confidence sets, betting, Bayes, conditional coverage, recognizable subsets, probability matching, nonstandard econometric problems, unit roots, parameters near boundary.

*We are grateful to the co-editor and two anonymous referees for comments and suggestions. We also thank Bruce Hansen and 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.

1 Introduction

In empirical economics, parameter uncertainty is usually described by confidence sets. By definition, a confidence set of level $1-\alpha$ covers the true parameter θ with probability of at least $1-\alpha$ in repeated samples, for all true values of θ. This definition, however, does not guarantee that confidence sets are compelling descriptions of parameter uncertainty. For instance, confidence intervals may be empty or unreasonably short with positive probability, even if they are based on inverting powerful tests, or if they are chosen to minimize average expected length. At least for some realizations of the data such confidence sets thus understate the uncertainty about θ, so that applied researchers are led to draw erroneous conclusions.

Let us consider three examples. First, suppose we are faced with the single observation $X \sim \mathcal{N}(\theta, 1)$, where it is known that $\theta > 0$. (This is a stylized version of constructing an interval based on an asymptotically normal estimator with values close to the boundary of the parameter space.) Since $[X - 1.96, X + 1.96]$ is a 95% confidence interval without the restriction on θ, the set $[X - 1.96, X + 1.96] \cap (0, \infty)$ forms a 95% confidence interval. In fact, it is the confidence set that is obtained by "inverting" the uniformly most powerful unbiased test of the hypotheses $H_0 : \theta = \theta_0$, that is it collects all parameter values θ_0 that are not rejected by the test with critical region $|X - \theta_0| > 1.96$. Yet, the resulting set is empty whenever $X < -1.96$, and arbitrarily short if X is just very slightly larger than -1.96.

As a second illustration, consider a homoskedastic instrumental variable (IV) regression in a large sample. Suppose that there is one endogenous variable and three instruments, and the concentration parameter is 12, so that the first stage F statistic is only rarely larger than 10. The Anderson and Rubin (1949) interval is then empty approximately 1.2% of the time. Moreover, it is also very short with positive probability; for instance, it is shorter than the usual two-stage least squares interval (but not empty) approximately 2.7% of the time. Applied researchers faced with such short intervals would presumably conclude that the data was very informative, and report and interpret the interval in the usual manner. But intuitively, weak instruments decrease the informational content of data, rendering these conclusions quite suspect. The same holds for all confidence sets that are empty and, by continuity, very short with positive probability.[1]

A third illustration is due to Cox (1958) and involves a normal observation with random

[1]Further examples include intervals based on Guggenberger, Kleibergen, Mavroeidis and Chen's (2012) subset Anderson-Rubin statistic, intervals based on Stock and Wright's (2000) GMM S-statistic, Stoye's (2009) interval for a set-identified parameter, Wright's (2000) and Müller and Watson's (2013) confidence sets for cointegrating vectors and Elliott and Müller's (2007) interval for the date of a structural break in a linear regression.

but observed variance. To be specific, suppose we observe (Y, S), where $Y|S = \mathcal{N}(\theta, S^2)$, $\theta \in \mathbb{R}$ and, say, $S = 1$ with probability 1/2, and $S = 5$ with probability 1/2. (This is a stylized version of conducting inference about a linear regression coefficient when the design matrix is random with known distribution.) A natural 95% confidence set is then given by $[Y - 1.96S, Y + 1.96S]$. But the interval $[Y - 2.58S, Y + 2.58S]$ if $S = 1$ and $[Y - 1.70S, Y + 1.70S]$ if $S = 5$ is also a 95% confidence interval, and it has shorter expected length. Yet, this second interval understates the degree of uncertainty relative to the nominal level whenever $S = 5$, since its coverage over the draws with $S = 5$ is only about 91%.

A potentially attractive formalization of "reasonableness" of a confidence set as a description of parameter uncertainty is obtained by considering a betting scheme: Suppose an inspector does not know the true value of θ either, but sees the data and the confidence set of level $1 - \alpha$. For any realization, the inspector can choose to object to the confidence set by claiming that she does not believe that the true value of θ is contained in the set. Suppose a correct objection yields her a payoff of unity, while she loses $\alpha/(1-\alpha)$ for a mistaken objection, so that the odds correspond to the level of the confidence interval. Is it possible for the inspector to be right on average with her objections no matter what the true parameter is, that is, can she generate positive expected payoffs uniformly over the parameter space? Surely, if the confidence set is empty with positive probability, the inspector could choose to object only to those, and the answer must be yes. Similarly, it is not hard to see that in the example involving (X, S), the inspector should object whenever $S = 5$ to generate uniformly positive expected winnings. The main idea of the paper is to employ the possibility of uniformly positive expected winnings as a formal indicator for the "reasonableness" of confidence sets, and to study inference in non-standard problems from this perspective.[2]

The idea of analyzing set estimators via betting schemes and closely related notions of relevant or recognizable subsets is not new. See, for example, Fisher (1956), Buehler (1959), Wallace (1959), Cornfield (1969), Pierce (1973), and Robinson (1977). We consider it worthwhile to revisit some ideas of this literature, because much recent econometric research has been dedicated to the derivation of valid inference in non-standard problems. It is sometimes difficult to gauge the "reasonableness" of confidence sets in non-standard problems, since they are not empty with positive probability, and there is no obvious failure to condition on ancillary statistics (in contrast to the example involving (Y, S)). A first objective of this paper is thus to *quantify* the degree of unreasonableness by determining the largest possible

[2]We focus on non-standard problems, since in the standard problem of inference about an unrestricted mean of a normal variate with known variance, which arises as the limiting problem in well behaved parametric models, the usual interval is unobjectionable.

expected winnings of the inspector. We illustrate the first objective on popular confidence intervals for the largest autoregressive root near unity and a parameter near a boundary and show that they are quite unreasonable.

A second objective of this paper is more constructive. Is it possible to derive confidence sets that are unobjectionable by construction? As was already realized by the literature referenced above, any bet-proof set must be a superset of a Bayesian credible set relative to some prior. This suggests that attractive bet-proof confidence sets may be obtained by selecting a prior that induces a given type of credible set (such as the highest posterior density (HPD) set) to have frequentist coverage, that is for the credible set to become a confidence set. The main new theoretical result of this paper is to show that such a prior typically exists, at least when the parameter space is discretized to a finite set. Previous results for Bayesian sets with frequentist coverage are either for particular families of distributions, for invariant problems, or for (higher order) asymptotic equivalence of coverage and credibility in locally asymptotically normal (LAN) models. In contrast, our existence result is entirely generic in the sense that it applies to any (discretized) inference problem under very mild regularity conditions.

For problems without nuisance parameters, the HPD credible set with coverage inducing prior is always (i) similar (that is it has exact coverage for all parameter values), (ii) reasonable according to our betting criterion, and (iii) length-optimal (no uniformly shorter confidence interval can exist). In our view, these are attractive properties, so we recommend these sets for applied work. As an illustration, we determine an HPD set with frequentist coverage for the largest autoregressive root local to unity, and for a parameter near a boundary.

For problems that involve nuisance parameters, the HPD confidence set might overcover for some parameter values. Such problems might then require an alternative approach for the construction of attractive bet-proof sets, which we leave to further research.

The remainder of the paper is organized as follows. Section 2.1 formally introduces the betting problem and defines bet-proof sets. In Section 2.2, we show that similar confidence sets that are equal to the whole parameter space with positive probability are not bet-proof and discuss the implications for the popular confidence intervals from the weak IV literature. Section 2.3 describes our quantification of "unreasonableness" of non-bet-proof sets. In Section 2.4, we provide the characterization of bet-proof sets as supersets of Bayesian credible sets. In Section 3.1, we show the existence of a prior distribution for which a given type of credible set has frequentist coverage. Section 4.1 discusses the application of the theoretical results obtained for a finite parameter space to problems with a continuous parameter space. In Section 4.2, we extend the framework to prediction sets. Applications of the methodology are presented in Section 5. Section 6 concludes. Proofs are collected in Section 7.

2 Bet-Proof Sets

2.1 Definitions and Notation

We start by assuming a finite parameter space, $\Theta = \{\theta_1, \ldots, \theta_m\}$, with θ_j itself the parameter of interest. Throughout the text, we discuss extensions of theoretical results to a general parameter of interest $\gamma = f(\theta) \in \Gamma$. In Section 4, we discuss implications for compact and unbounded 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. For simplicity, 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, we define a rejection probability function $\varphi : \Theta \times \mathcal{X} \mapsto [0,1]$, which is the probability that θ_j is not included in the set when $X = x$ is observed. The function φ defines a $1-\alpha$ confidence set if

$$\int [1 - \varphi(\theta_j, x)] p(x|\theta_j) d\nu(x) \geq 1 - \alpha, \, \forall \theta_j \in \Theta \tag{1}$$

(equivalently, the function $\varphi(\theta_j, \cdot)$ defines a level α test of $H_0 : \theta = \theta_j$, for all θ_j). Hereafter, we assume $0 < \alpha < 1$.

As described in the introduction, we follow Buehler (1959) and others and study the "reasonableness" of the confidence set φ via a betting scheme: For any realization of $X = x$, an inspector can choose to object to the set described by φ.[3] We denote the inspector's objection by $b(x) = 1$, and $b(x) = 0$ otherwise. If the inspector objects, then she receives 1 if φ does not contain θ, and she loses $\alpha/(1-\alpha)$ otherwise. For a given betting strategy b and parameter θ, the expected loss of the inspector is

$$R(\varphi, b, \theta) = \frac{1}{1-\alpha} \int [\alpha - \varphi(\theta, x)] b(x) p(x|\theta) d\nu(x). \tag{2}$$

If there exists a strategy b such that $R(\varphi, b, \theta) < 0$ for all $\theta \in \Theta$, then the inspector is right on average with her objections for any parameter value, and one might correspondingly call such

[3] If φ is randomized ($0 < \varphi < 1$), then the inspector examines the set *before* the randomization is realized. Inspections *after* realization of randomization may be modelled in our set-up by making the randomization device part of the observed data x.

a φ "unreasonable". Buehler (1959) considered a larger strategy space of $b(x) \in \{-1, 0, 1\}$. Intuitively, negative b allow the inspector to express the objection that the confidence set is "too large". However, since the definition of confidence sets involves an inequality that explicitly allows for conservativeness, we follow Robinson (1977) and impose non-negativity on bets in most of what follows. For technical reasons, it is useful to allow for values of b also in $(0, 1)$, so that the set of possible betting strategies is the set B of all measurable mappings $b : \mathcal{X} \mapsto [0, 1]$.

Definition 1 *If for any bet $b \in B$, $R(\varphi, b, \theta) \geq 0$ for some θ in Θ then φ is* bet-proof *at level $1 - \alpha$.*

The requirement of bet-proofness can also be deduced from purely frequentist considerations not involving a betting game. A betting strategy $b : \mathcal{X} \mapsto \{0, 1\}$ defines a subset $\mathcal{X}_b$ of the sample space where $b(x) = 1$. If b delivers uniformly positive winnings for φ then the coverage of φ conditional on $\mathcal{X}_b$ must be strictly less than the nominal level $1 - \alpha$ uniformly over the parameter space. If this is the case, then $\mathcal{X}_b$ is called a negatively biased recognizable subset. Even before the betting setup was introduced in Buehler (1959), the existence of recognizable subsets had been considered an unappealing property for confidence sets; see for example, Fisher (1956); Wallace (1959), Pierce (1973), and Robinson (1977) are also relevant.

2.2 Bet-proofness and Similarity

Proving analytically that a given confidence set is not bet-proof seems hard in general. One general result, however, is as follows: Suppose there exists a subset of the sample space, $\mathcal{X}_0 \subset \mathcal{X}$, such that for any $x \in \mathcal{X}_0$ the confidence set includes the whole parameter space ($\varphi(\theta, x) = 0$, $\forall x \in \mathcal{X}_0$). If φ is similar and $P(\mathcal{X}_0|\theta) > 0$ for all θ in Θ then φ is not bet-proof. To see this note that for $b(x) = \mathbf{1}[x \notin \mathcal{X}_0]$ the expected winnings can be written as $\int_{\mathcal{X}\setminus\mathcal{X}_0}(\varphi(\theta, x) - \alpha)p(x|\theta)d\nu(x)/(1 - \alpha)$. By similarity,

$$0 = \int_{\mathcal{X}_0}(\varphi(\theta, x) - \alpha)p(x|\theta)d\nu(x) + \int_{\mathcal{X}\setminus\mathcal{X}_0}(\varphi(\theta, x) - \alpha)p(x|\theta)d\nu(x). \tag{3}$$

Since $\varphi(\theta, x) = 0$ on $\mathcal{X}_0$ and $P(\mathcal{X}_0|\theta) > 0$, the first term on the right hand side of (3) is strictly negative for $\alpha > 0$ and the winnings are uniformly positive. Intuitively, the set $\varphi(\theta, X)$ might be considered unappealing because it overcovers when $X \in \mathcal{X}_0$ and undercovers when $X \notin \mathcal{X}_0$.

Similar confidence sets that are equal to the whole parameter space with positive probability or, in other words, sets that satisfy the above conditions on $\mathcal{X}_0$ are a part of the standard toolbox in the weak instruments literature (Anderson and Rubin (1949), Staiger and Stock (1997), Kleibergen (2002), Moreira (2003), Andrews, Moreira, and Stock (2006), and Mikusheva (2010)). Thus, it can be argued that most of the sets proposed in this literature are too short whenever they are not equal to the whole parameter space. A quantification of the unreasonableness of the popular confidence sets in the weak IV literature and the construction of bet-proof sets are left to further research.

2.3 Quantifying Unreasonableness of Non-Bet-Proof Sets

If a given confidence set φ is not bet-proof, we propose to measure the degree of its "unreasonableness" by the magnitude of inspector's winnings. Specifically, we consider an optimal betting strategy $b^\star$ that solves the following problem

$$W(\pi) = \sup_{b \in B:\, R(\varphi,b,\theta) \leq 0,\, \forall \theta} -\sum_{j=1}^{m} R(\varphi, b, \theta_j)\pi_j, \tag{4}$$

where $\pi_j \geq 0$, $\sum_{j=1}^m \pi_j = 1$, are fixed weights. For uniform weights, $\pi_i = 1/m$, $b^\star$ maximizes average expected winnings subject to the requirement that expected winnings are non-negative at all parameter values. If $\pi_i = 1$ and $\pi_j = 0$ for $j \neq i$, then $b^\star$ maximizes expected winnings at θ_i subject to uniform non-negativity of expected winnings. A calculation shows that for any confidence set φ of level $1-\alpha$, $0 \leq W(\pi) \leq \alpha$. The maximal expected winnings α can be obtained for the "completely unreasonable" confidence set that is equal to the parameter space with probability $1-\alpha$ and empty with probability α.[4] Thus, a finding of $W(\pi)$ close to α for a given $1-\alpha$ confidence set φ indicates a very high degree of "unreasonableness".

The following lemma provides an explicit characterization of $b^\star$.

Lemma 1 *Suppose* $b^\star(x) = \mathbf{1}[\sum_{j=1}^m (\pi_j + \kappa_j) p(x|\theta_j)(\varphi(\theta_j, x) - \alpha) \geq 0]$, *where* $\kappa_j \geq 0$ *are such that (i)* $\int [\varphi(\theta_j, x) - \alpha] b^\star(x) p(x|\theta_j) d\nu(x) \geq 0$ *for all* j*; (ii)* κ_j *is zero if* $\int [\varphi(\theta_j, x) - \alpha] b^\star(x) p(x|\theta_j) d\nu(x) > 0$*. Then,* $b^\star$ *solves (4).*

The optimal strategy in Lemma 1 is recognized as the inspector behaving like a Bayesian with a prior proportional to $\pi_j + \kappa_j$: She objects whenever the posterior probability that θ is excluded from the set φ exceeds α. The characterization of Lemma 1 is useful for the numerical determination of the maximal average expected winnings (4).

[4] The randomness that determines the interval realizes before the inspection, see also footnote 3.

Most of this paper is concerned with the implication of bet-proofness relative to bets whose payoff corresponds to the level $1-\alpha$ of the confidence set. To shed further light on the severity and nature of the violation of bet-proofness, it is interesting to explore the possibility and extent of uniformly non-negative expected winnings also under less favorable payoffs for the inspector. Specifically, assume that a correct objection still yields her a payoff of unity, but she now has to pay $\alpha'/(1-\alpha')$ for a mistaken objection, where $\alpha' > \alpha$. If the inspector can still generate uniformly positive winnings under these payoffs, then the confidence set φ is not bet-proof even at the level $1-\alpha' < 1-\alpha$. Note that if a confidence set is empty with positive probability, then the inspector can generate positive expected winnings for any $\alpha' < 1$ by simply objecting only to realizations that lead to an empty set φ. In particular, the "completely unreasonable" level $1-\alpha$ confidence set that is empty with probability α still yields maximal expected winnings equal to α. In other problems, however, such as in Cox's example of a normal mean problem with random but observed variance mentioned in the introduction, there exists a cut-off $\bar{\alpha}' < 1$ such that no uniformly positive winnings are possible under any odds with $\alpha' > \bar{\alpha}'$.

The optimal betting strategy under such modified payoffs still follows from Lemma 1 with α replaced by α', as its proof does not depend on φ being a level $1-\alpha$ confidence set.

2.4 Credibility and Bet-Proof Sets

It turns out that bet-proof sets can be characterized in terms of Bayesian credible sets. Let us introduce the notation first and then provide the characterization. For a prior $\pi = (\pi_1, \ldots, \pi_m)'$, $\pi_j \geq 0$, $\sum_j \pi_j = 1$, the posterior distribution is defined as

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

A $1-\alpha$ credible set is defined by any $\varphi \in [0,1]$ such that

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

or

$$\sum_{j=1}^m (\alpha - \varphi(\theta_j, x))p(x|\theta_j)\pi_j = 0, \ \forall x. \tag{5}$$

The following lemma generalizes a result in Robinson (1977) from finite support of X to more general distributions for X. Pierce (1973) proves a related result when b can be negative.

Definition 2 *A function* $g : \mathcal{X} \mapsto \mathbb{R}$ *is* directionally upper semi- (d.u.s.) continuous *at* x_0 *if there exists a measurable* $A_{x_0} \subset \mathcal{X}$ *such that for any ball with center at* x_0 *and radius* ϵ, $B_\epsilon(x_0)$, $\nu(A_{x_0} \cap B_\epsilon(x_0)) > 0$ *and the restriction of* g *to* A_{x_0} *is upper semi-continuous at* x_0.

Lemma 2 *Suppose* φ *is bet-proof at level* $1 - \alpha$. *Then there exists a prior* $\pi^\star$ *for which* φ *describes a superset of a* $1-\alpha$ *credible set for all* $x_0 \in \mathcal{X}$ *at which* $(\alpha - \varphi(\theta, x))p(x|\theta)$ *is d.u.s. continuous at* x_0 *for all* $\theta \in \Theta$.

Conversely, if φ *is a superset of a* $1-\alpha$ *credible set for some prior* $\pi^\star$, *then* φ *is bet-proof at level* $1 - \alpha$.

Thus, up to the relatively minor technical qualification of Lemma 2, any "reasonable" level $1 - \alpha$ confidence set is a superset of a credible set of the same level relative to some prior.

Now suppose the parameter of interest is given by $f(\theta)$ for some known function $f : \Theta \mapsto \Gamma$. For instance, if the distribution of X depends on $\theta = (\gamma, \delta)$, where $\gamma \in \Gamma$ is the parameter of interest, and δ is a nuisance parameter, then $f(\theta) = \gamma$. A confidence set (or credible set) is now a function $\varphi : \Gamma \times \mathcal{X} \mapsto [0, 1]$ that satisfies (1) (or (5)) with $\varphi(\theta_j, x)$ replaced by $\varphi(f(\theta_j), x)$. Thus, Lemma 2 continues to hold with the continuity requirement in the sense of Definition 2 now imposed on the function $(\alpha - \varphi(f(\theta), x))p(x|\theta)$ for all $\theta \in \Theta$.

3 Construction of Appealing Bet-Proof sets

The characterization of bet-proof sets from the previous subsection suggests that in a search of reasonable confidence sets one may restrict attention to supersets of Bayesian credible sets.[5] It is necessary to introduce additional criteria that would rule out unnecessarily conservative sets. Allowing for negative bets ($b \in [-1, 1]$) is a natural way to proceed. An obvious extension of Lemma 2 to this case implies that subject to minor technical qualifications, a set is bet-proof at level $1 - \alpha$ if and only if it is a Bayesian credible set of level $1 - \alpha$ with respect to some prior.

In the following 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 bet-proof confidence sets:

[5] One might also question the appeal of the frequentist coverage requirement. We find Robinson's (1977) argument fairly compelling: In a many-person setting, frequentist coverage guarantees that the description of uncertainty cannot be highly objectionable *a priori* to any individual, as the prior weighted expected coverage is no smaller than $1 - \alpha$ under all priors.

(i) choose a type of credible set suitable for the problem at hand, for example, HPD if shorter sets are desirable, (ii) find 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.

3.1 Existence of Prior

Theorem 3 *Suppose $\varphi(\theta_j, x; \pi)$ 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*

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

Theorem 3 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 standard problems, see a monograph on the subject by Datta and Mukerjee (2004) and references therein. Note that our result in Theorem 3 does not appeal to invariance and 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 $\varphi(\theta_j, x; \pi)$ 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 corollary suggests that the use of HPD sets in Theorem 3 is particularly attractive.

Corollary 1 *If the likelihood ratio $p(X|\theta_i)/p(X|\theta_j)$ is a continuous random variable for all $\theta_i \neq \theta_j$, then an HPD set $\varphi(\theta_j, x; \pi)$ satisfies the following. (i) The continuity conditions of Theorem 3 are satisfied. (ii) The prior $\pi^\star$ puts positive mass on all values of the parameter and the resulting set $\varphi(\theta_j, x; \pi^\star)$ is a similar confidence set of level $1-\alpha$. (iii) There does not exist a level $1-\alpha$ confidence set that is uniformly shorter than $\varphi(\theta_j, x; \pi^\star)$, that is if $\sum_{j=1}^m \varphi'(\theta_j, x) \geq \sum_{j=1}^m \varphi(\theta_j, x; \pi^\star)$ for all $x \in \mathcal{X}$ and $\sum_{j=1}^m \varphi'(\theta_j, x) > \sum_{j=1}^m \varphi(\theta_j, x; \pi^\star)$ for all $x \in \mathcal{X}_l$ with $\nu(\mathcal{X}_l) > 0$, then $\int \varphi'(\theta_j, x) p(x|\theta_j) d\nu(x) > \alpha$ for some j.*

Thus, the HPD set with coverage inducing prior emerges as a particularly compelling description of uncertainty: it is bet-proof against bets that object to sets that understate or overstate degree of parameter uncertainty; it covers each θ_j with exact probability $1-\alpha$; and there does not exist a $1-\alpha$ confidence set that is uniformly shorter.

Up to this corollary, all theoretical results including Lemma 2 and Theorem 3 can be extended to the case of a general parameter of interest $\gamma = f(\theta)$. When f is not one-to-one, however, the similarity of φ in Corollary 1 does not necessarily hold. The prior that induces frequentist coverage of the HPD set might have zero mass for some θ_j. The HPD confidence set might thus be overly conservative for these θ_j's. Therefore, problems with nuisance parameters require an alternative approach for construction of attractive bet-proof sets, which is an important subject for future work.

4 Extensions

4.1 Continuous Parameter Space

4.1.1 Approximations for Compact Parameter Space

Suppose the parameter space $\tilde{\Theta}$ is continous but compact, the likelihood $\tilde{p}(x|\tilde{\theta})$ is continuous in $\tilde{\theta}$ for every x, and there exists $\bar{p}(x)$ such that $\tilde{p}(x|\tilde{\theta}) \leq \bar{p}(x)$ for any $\tilde{\theta}$ in $\tilde{\Theta}$ and $\int \bar{p}(x) d\nu(x) < \infty$. Then $\tilde{p}(\cdot|\tilde{\theta})$ is uniformly continuous in $\tilde{\theta}$ under the $L_1(\nu)$ distance. Consider a fine partition of $\tilde{\Theta}$ into m convex sets $\tilde{\Theta}_j$ with center point θ_j. For the center points, we can construct $1-\alpha$ bet-proof confidence set $\varphi(\theta_j, x)$ as described in the previous sections. Now define $\tilde{\varphi}(\tilde{\theta}, x) = \varphi(\theta_j, x)$ for $\tilde{\theta} \in \tilde{\Theta}_j$, that is, $\tilde{\Theta}_j$ is covered by $\tilde{\varphi}(\tilde{\theta}, x)$ if and only if θ_j is covered by $\varphi(\theta_j, x)$. By uniform continuity of $\tilde{p}(\cdot|\tilde{\theta})$, for any $\epsilon > 0$ and a sufficiently fine partition on $\tilde{\Theta}$ the coverage of $\tilde{\varphi}(\tilde{\theta}, x)$ is at least $1-\alpha-\epsilon$ for any $\tilde{\theta} \in \tilde{\Theta}$. At the same time, $\tilde{\varphi}(\cdot, x)$ is bet-proof at level $1-\alpha$, since it is a superset of the credible set $\varphi(\cdot, x)$ of level $1-\alpha$.

4.1.2 Unbounded Parameter Space and Weighted Coverage

In this subsection, we obtain a version of Theorem 3 that is applicable to an unbounded and continuous parameter space when hierarchical Bayesian priors and weighted average coverage are used.

Consider a general 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 the Lebesgue measure on $\tilde{\Theta}$ and $\pi = (\pi_1, \ldots, \pi_m)$, $\pi_j \geq 0$, $\sum_j \pi_j = 1$. The mixture density

$$\sum_{j=1}^{m} \pi_j \cdot \psi_j(\tilde{\theta}) \tag{6}$$

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 3, 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 (6). The set is defined as an explicit function of π (ψ_j's are considered 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

$$\varphi(\theta_j, x; \pi) = \frac{\int \tilde{\varphi}(\tilde{\theta}, x; \pi) \tilde{p}(x|\tilde{\theta}) \psi_j(\tilde{\theta}) d\tilde{\theta}}{p(x|\theta_j)} \tag{7}$$

be the posterior non-coverage 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

$$\int \left[\int [1 - \tilde{\varphi}(\tilde{\theta}, x; \pi) \tilde{p}(x|\tilde{\theta})] d\nu(x) \right] \psi_j(\tilde{\theta}) d\tilde{\theta} = \int [1 - \varphi(\theta_j, x; \pi)] p(x|\theta_j) d\nu(x). \tag{8}$$

Note that the right-hand side (r.h.s.) of (8) is the same as the coverage of θ_j in the discrete model of Section 2. Also, Bayesian credibility of the set $\tilde{\varphi}(\tilde{\theta}, x; \pi)$ at level $1 - \alpha$ implies credibility (5) in the discrete model of Section 2 of the set $\varphi(\theta_j, x; \pi)$ defined in (7). Thus, if the weighted coverage rate in (8) is a continuous function of π, then Theorem 3 implies the existence of $\pi^\star$ such that (8) evaluated at $\pi^\star$ is at least $1 - \alpha$ for all models θ_j, $j = 1, \ldots, m$.

Thus, general unbounded spaces may be discretized in a finite number of base probability densities ψ_j of arbitrary support, and Theorem 3 implies the existence of a mixture of these base densities such that the HPD set has ψ_j-weighted average coverage equal to $1-\alpha$, for all $j = 1, \ldots, m$. This may be specialized to base distributions with supports that form a partition of $\tilde{\Theta}$. If in an application, the $\pi^{\star}$-mixture prior leads to sets with a coverage rate that varies little on each support, then the resulting set $\tilde{\varphi}$ also has nearly point-wise coverage on $\tilde{\Theta}$.

4.2 Predictive Sets

The issue of how to describe uncertainty appropriately also arises in a forecasting setting. Our framework and theoretical results can be straightforwardly extended to the problem of constructing "reasonable" prediction sets.

Suppose that the econometrician is interested in describing 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, $y \to 1 - \varphi_p(y, x)$ 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 (5), respectively. Also, the expected loss to the inspector from a bet b against φ_p is correspondingly given by $R(\varphi, b, \theta)$ defined in (2). Therefore, the characterization of bet-proof sets in terms of Bayesian credible sets (Lemma 2) and the existence of a prior that guarantees frequentist nominal coverage for credible sets (Theorem 3) also hold for prediction sets.

5 Applications

In this section, we illustrate our methodology on two nonstandard inference problems: (i) inference for a parameter near a boundary of the parameter space and (ii) inference about the largest autoregressive root near unity. First, we are "destructive" and explore whether natural or previously suggested 95% confidence intervals are bet-proof. For all problems this turns out not to be the case. As discussed in Section 2.3, we compute maximal weighted average expected winnings to gauge the degree of unreasonableness. Next, we turn to being "constructive" and determine for each example the HPD set with frequentist coverage.

In our examples, the parameter space is not naturally compact. In the destructive calculations we discretize the most interesting part of the parameter space by a grid, and avoid artificial end-point effects by restricting bets to be zero whenever the MLE of the parameter of interest is outside the grid.[6] In the constructive calculations we implement the mixture approach discussed in Section 4.1.2, where the baseline densities are uniform on a fine partition of the most interesting part of the parameter space, plus one very slowly decaying density with unbounded support for the remaining part of the parameter space. Numerical calculations suggest that the resulting HPD set comes very close to controlling size uniformly over the original continuous parameter space (within 0.2% and 0.3% in the boundary and near unit root examples, respectively). Implementation details are discussed in Appendix C.

5.1 Parameter Near a Boundary

Consider a model in the LAN family, but suppose plausible values for the parameter of interest are close to the boundary of the parameter space. The bound on the parameter space may arise naturally, such as the non-negativity of variances, or it might be the result of a priori knowledge about parameter values, such as time discount factors being smaller than unity. After suitable normalizations, the relevant limiting experiment in the sense of Le Cam (1972) then becomes[7]

$$X \sim \mathcal{N}(\theta, 1), \quad \theta > 0. \tag{9}$$

Alternatively, interest in inference about θ in (9) may arise without reference to Limits of Experiments theory because one seeks to derive a confidence set based on a given asymptotically normal estimator. As discussed in the introduction, a "standard" 95% confidence interval is given by $[x - 1.96, x + 1.96] \cap (0, \infty)$. Alternatively, one might consider inverting the generalized likelihood ratio statistic (gLR) for testing $H_0 : \theta = \theta_0$,

$$\mathrm{gLR}(\theta_0, x) = \frac{\sup_{\theta>0} \exp[-\frac{1}{2}(x-\theta)^2]}{\exp[-\frac{1}{2}(x-\theta_0)^2]},$$

so that the 95% interval becomes $\{\theta_0 : \mathrm{gLR}(\theta_0, x) < \mathrm{cv}(\theta_0)\}$, with $\mathrm{cv}(\theta_0)$ equal to the 95% quantile of $\mathrm{gLR}(\theta_0, X)$ under $X \sim \mathcal{N}(\theta_0, 1)$.

[6]The reported results are fairly robust to finer grids and/or different endpoints.

[7]The original model may include additional nuisance parameters, but as long as these are not close to their respective bounds, asymptotically uniformly most powerful (unbiased) tests of $H_0 : \theta = \theta_0 > 0$ are still based on the scalar component of the MLE that corresponds to the parameter of interest; see Chapter 15.2 of van der Vaart (1998).

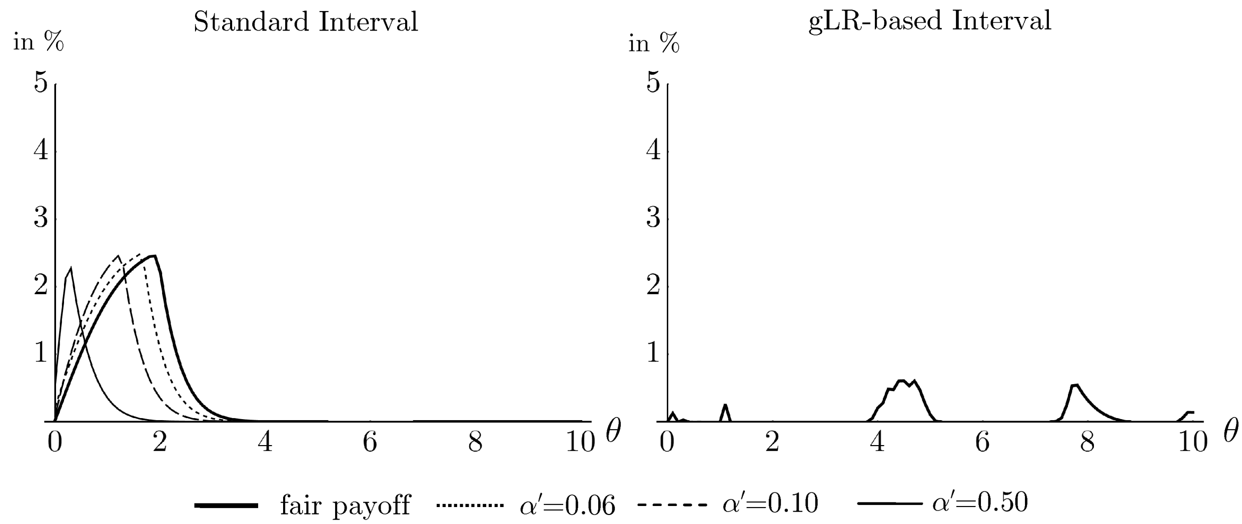

Figure 1: Parameter Near a Boundary: Optimal Expected Winnings

5.1.1 Destructive Results

Figure 1 shows expected winnings of the inspector against these two intervals as a function of θ when the inspector seeks to maximize a simple average of expected winnings over an equal-spaced grid subject to the constraint that the inspector never loses in expectation for any true θ. Different lines correspond to different payoff functions for the inspector, with the thick black line corresponding to the "fair" payoff $(1, -0.05/0.95)$ for correct and incorrect objections, respectively, and the thinner lines to payoffs $(1, -\alpha'/(1-\alpha'))$, with $\alpha' \in \{0.06, 0.1, 0.5\}$. Recall from Section 2.3 that expected winnings for the inspector cannot be larger than the nominal level of 5% for any 95% confidence set. Relative to this number, the standard confidence interval is seen to be quite unreasonable, with maximal expected winnings close to the boundary of roughly 2.5%. Winnings of this magnitude are obtainable even if the payoff function for the inspector is quite unfavorable, pointing to the severity of the issue—the reason being, of course, that with $X \sim \mathcal{N}(0,1)$, 2.5% of all draws lead to an empty confidence set, for which an objection is never mistaken. In contrast, the gLR confidence intervals are much less objectionable in this metric, with no positive winnings for $\alpha' \in \{0.06, 0.1, 0.5\}$.

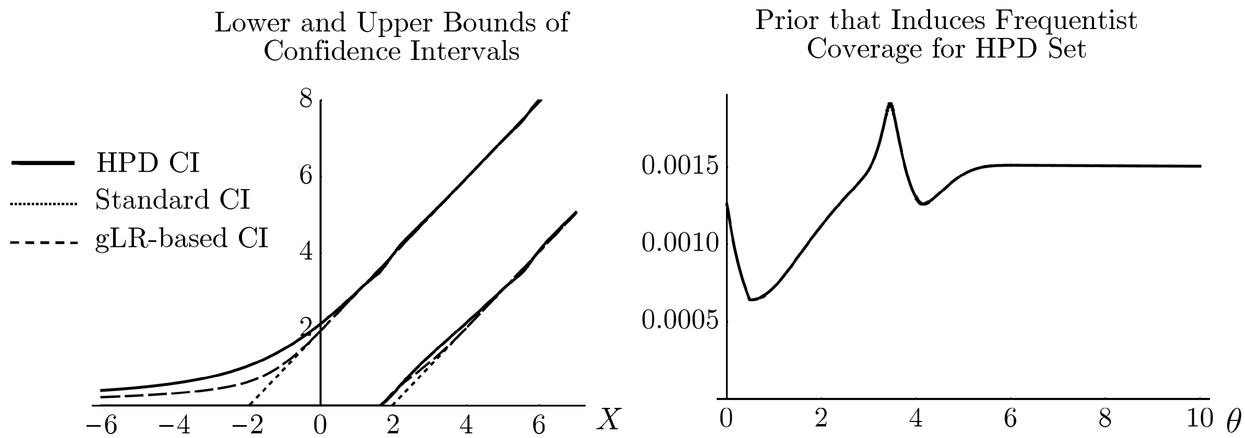

Figure 2: Parameter Near a Boundary: Constructive Results

5.1.2 Constructive Results

The right panel in Figure 2 shows the prior that (approximately) induces frequentist coverage of the HPD credible set, and the left panel compares the resulting confidence interval with the standard and gLR intervals. By construction, the HPD confidence set is never empty, even for very small realizations of X, and it is seen to become very close to the standard confidence interval for realizations away from the bound.

5.2 Largest Autoregressive Root

As in Stock (1991), Hansen (1999), Elliott and Stock (2001), and Mikusheva (2007), suppose we are interested in the largest autoregressive root ρ of the univariate time series y_t,

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

where $\phi(z) = 1 - \phi_1 z - \ldots - \phi_{p-1} z^{p-1}$, $\varepsilon_t \sim iid(0, \sigma^2)$, and $u_0 = O_p(1)$. Suppose it is known that the largest root ρ is close to unity, while the roots of ϕ are all bounded away from the complex unit circle. Formally, let $\rho = \rho_T = 1 - \theta/T$, so that equivalently, θ is the parameter of interest. The appropriate limiting 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 $X(s) = \int_0^s \exp[-\theta(s-r)]dW(r)$ with $W(r)$ a standard Wiener process. The density of X relative to the measure of a standard Wiener process is given by

$$p(x|\theta) = \exp[-\tfrac{1}{2}\theta(x(1)^2 - 1) - \tfrac{1}{2}\theta^2 \int_0^1 x(s)^2 ds]. \tag{10}$$

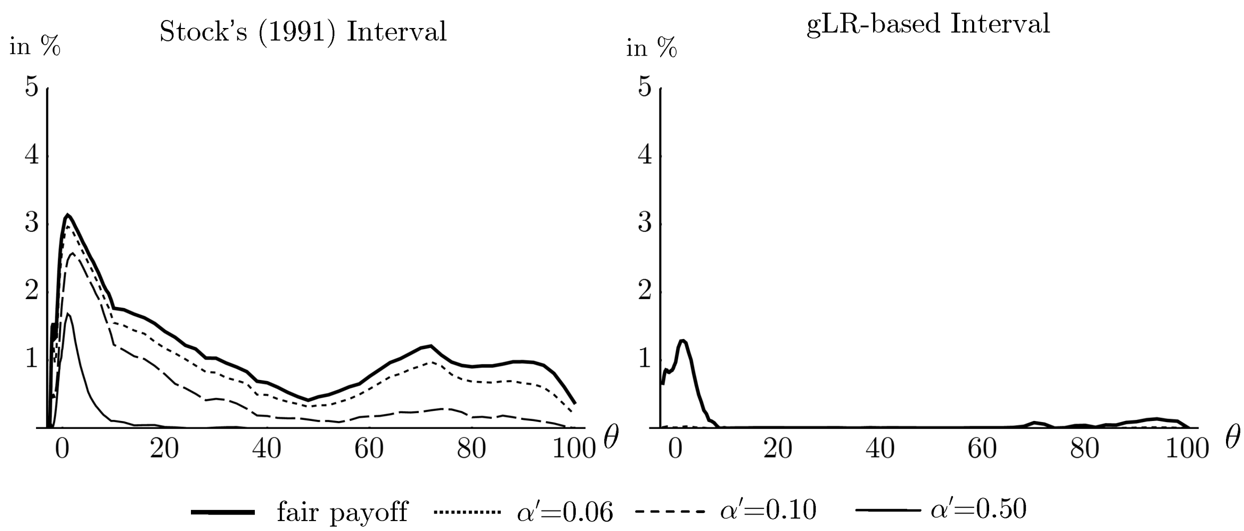

Figure 3: Largest Autoregressive Root: Optimal Expected Winnings

The same limiting problem may also be motivated using the framework in Müller (2011), without relying on an assumption of Gaussian u_t.

Stock (1991) suggested inverting the augmented Dickey-Fuller t-test (where the Dickey-Fuller regression contains a constant) for obtaining a confidence interval for θ. This corresponds asymptotically to inverting the statistic

$$\frac{\bar{X}(1)^2 - \bar{X}(0)^2 - 1}{2\sqrt{\int_0^1 \bar{X}(s)^2 ds}}, \tag{11}$$

where $\bar{X}(s) = X(s) - \int_0^1 X(r)dr$. Alternatively, one might again consider the generalized likelihood ratio statistic based on the density (10).

5.2.1 Destructive Results

Figure 3 shows the expected winnings as a function of θ when the inspector seeks to maximize a weighted average of the expected winnings on a grid. The grid and the weights are chosen to approximate a uniform distribution on $[-2.5, 120]$ as described in Section 8. Stock's (1991) confidence sets are seen to allow for very substantial winnings, even under fairly unfavorable payoffs and large θ. This might not be too surprising—there is no reason to expect that intervals based on the *Unit Root* t-statistic would yield standard intervals for $\theta \gg 0$, even if the problem there is fairly close to an unrestricted Gaussian shift experiment. The gLR-based intervals are again much less unreasonable.

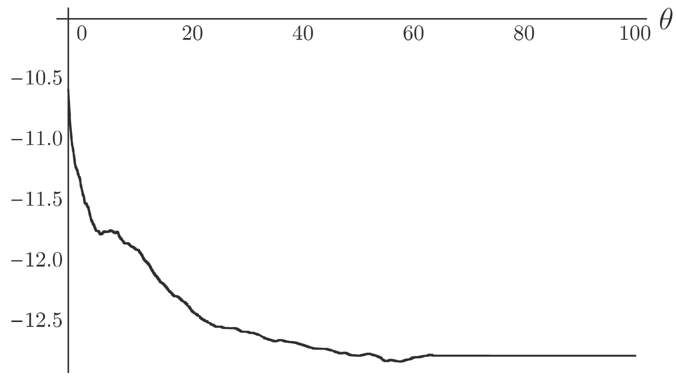

Figure 4: Largest Autoregressive Root: Log of Coverage Inducing Prior

5.2.2 Constructive Results

Figure 4 reports the shape of the prior that (nearly) induces coverage of the HPD credible set. It puts much more mass on small values of θ, counteracting the well known positive bias of the peak of the likelihood (that is the downward bias of the MLE of the largest root ρ).

6 Conclusion

By definition, the level of a confidence set is a pre-sample statement: at least $100(1-\alpha)\%$ of data draws yield a confidence set that covers the true value. But once the sample is realized, "unreasonable" confidence sets (as defined in the paper) understate the level of parameter uncertainty, at least for some draws. For a set to be reasonable, it has to contain a credible set of level $1-\alpha$ relative to some prior. A compelling description of parameter uncertainty in both the pre- and post-sample sense must therefore possess frequentist and conditional (Bayesian) properties.

Many popular confidence sets in non-standard problems do not have this property. At least occasionally, applied research based on such sets hence understates the extent of parameter uncertainty, and thus comes to misleading conclusions.

As a remedy, econometric theory should derive confidence sets that do not suffer from this problem. This can be done informally by comparing the "unreasonableness" of alternative confidence set constructions, and by selecting the most reasonable among them. In the few examples that we considered, intervals based on the generalized likelihood ratio statistic had relatively good properties in this regard.[8] Alternatively, one can derive confidence sets that

[8] In Cox's example of the introduction, the interval based on the generalized likelihood ratio statistic is

are reasonable by construction. For problems that do not involve nuisance parameters, an elegant approach is to identify the prior that induces frequentist coverage of the highest posterior density credible set. The resulting confidence set is then similar, bet-proof, length-optimal, and thus an attractive choice to credibly describe parameter uncertainty in non-standard problems.

References

Anderson, T. W., and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *The Annals of Mathematical Statistics*, 21, 570–582.

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

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

Buehler, R. J. (1959): "Some Validity Criteria for Statistical Inferences," *The Annals of Mathematical Statistics*, 30(4), pp. 845–863.

Cornfield, J. (1969): "The Bayesian Outlook and Its Application," *Biometrics*, 25(4), pp. 617–657.

Cox, D. R. (1958): "Some Problems Connected with Statistical Inference," *The Annals of Mathematical Statistics*, 29(2), pp. 357–372.

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

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

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

also bet-proof. However, it is easy to construct examples where confidence intervals based on the generalized likelihood ratio statistic are empty, and hence not close to being bet-proof.

Ferguson, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, 1st edn.

Fisher, S. R. A. (1956): *Statistical Methods and Scientific Inference*. Oliver & Boyd, first edition edn.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012): "On the Asymptotic Sizes of Subset AndersonRubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression," *Econometrica*, 80(6), 2649–2666.

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

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 (Methodological)*, 36(2), pp. 237–242.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), pp. 1781–1803.

Le Cam, L. (1972): "Limits of experiments," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, pp. 245–261, Berkeley, Calif. Univ. California Press.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*. Oxford University Press, USA.

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

Mikusheva, A. (2010): "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 157(2), 236–247.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Müller, U. K. (2011): "Efficient Tests under a Weak Convergence Assumption," *Econometrica*, 79, 395–435.

Müller, U. K., and M. W. Watson (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81.

PIERCE, D. A. (1973): “On Some Difficulties in a Frequency Theory of Inference,” *The Annals of Statistics*, 1(2), pp. 241–250.

ROBINSON, G. K. (1977): “Conservative Statistical Inference,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(3), pp. 381–386.

STAIGER, D., AND J. H. STOCK (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–586.

STOCK, J. H. (1991): “Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series,” *Journal of Monetary Economics*, 28, 435–459.

——— (2000): “A Class of Tests for Integration and Cointegration,” in *Cointegration, Causality, and Forecasting — A Festschrift in Honour of Clive W.J. Granger*, ed. by R. F. Engle, and H. White, pp. 135–167. Oxford University Press.

STOYE, J. (2009): “More on Confidence Intervals for Partially Identified Parameters,” *Econometrica*, 77, 1299–1315.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK.

WALLACE, D. L. (1959): “Conditional Confidence Level Properties,” *The Annals of Mathematical Statistics*, 30(4), pp. 864–876.

WRIGHT, J. H. (2000): “Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests,” *Journal of Business and Economic Statistics*, 18, 211–222.

7 Proofs

Proof of Lemma 1

Consider an alternative strategy $b \in B$ that delivers uniformly non-negative winnings. By definition of $b^\star$,

$$\int [b^\star(x) - b(x)] \sum_j [\varphi(\theta_j, x) - \alpha](\pi_j + \kappa_j) p(x|\theta_j) d\nu(x) \geq 0.$$

It follows that

$$\int [b^\star(x) - b(x)] \sum_j [\varphi(\theta_j, x) - \alpha]\pi_j p(x|\theta_j) d\nu(x) \geq -\sum_j \kappa_j \int b^\star(x)[\varphi(\theta_j, x) - \alpha] p(x|\theta_j) d\nu(x)$$

$$+\sum_j \kappa_j \int b(x)[\varphi(\theta_j, x) - \alpha]p(x|\theta_j)d\nu(x).$$

The first expression on the r.h.s. of this inequality is equal to zero by the definition of $b^\star(x)$. The second expression is non-negative as b delivers uniformly non-negative winnings. Thus, the winnings from $b^\star(x)$ are at least as large as the winnings from b.

Proof of Lemma 2

Let $S = \{(y_1, \dots, y_m) : y_j = R(\varphi, b, \theta_j), j = 1, \dots, m \text{ and } b \in B\}$. Note that S is bounded below. Since $R(\varphi, b, \theta_j)$ are linear in b and B is convex, S is convex. By a version of the minimax theorem from Ferguson (1967) (Theorem 2.9.1, p. 82), for

$$r(\pi, b) = \sum_{k=1}^{m} R(\varphi, b, \theta_k)\pi_k$$

there exists the value of the game V and a least favorable prior $\pi^\star$ such that

$$V = \inf_{b\in B} \sup_{\pi\in\Delta} r(\pi, b) = \sup_{\pi\in\Delta} \inf_{b\in B} r(\pi, b) = \inf_{b\in B} r(\pi^\star, b).$$

If $\varphi(\cdot,\cdot)$ is bet-proof at level $1-\alpha$ then for any $b \in B$ there exists j such that $R(\varphi, b, \theta_j) \geq 0$. Thus, $V \geq 0$.

Fix $\tilde{x}$ at which $(\alpha - \varphi(\theta, \tilde{x}))p(\tilde{x}|\theta)$ is d.u.s. continuous and let $b^n_{\tilde{x}}(x) = \mathbf{1}[x \in A_{\tilde{x}} \cap B_{1/n}(\tilde{x})]$, where $B_{1/n}(\tilde{x})$ is a ball with radius $1/n$ and center $\tilde{x}$ and $A_{\tilde{x}}$ is from Definition 2. For any n, $0 \leq V \leq r(\pi^\star, b^n_{\tilde{x}})$ by the definition of the least favorable prior. Let us assume contrary to the first claim of the lemma that

$$\sum_{k=1}^{m} (\alpha - \varphi(\theta_k, \tilde{x}))p(\tilde{x}|\theta_k)\pi^\star_k < 0. \tag{12}$$

By Definition 2, (12) implies $r(\pi^\star, b^n_{\tilde{x}}) < 0$ for sufficiently large n, which is a contradiction. Thus, (12) cannot hold and the first claim of the lemma follows immediately.

To prove the converse note that $\varphi(\cdot,\cdot)$ being a superset of a $1-\alpha$ credible set for $\pi^\star$ implies

$$\sum_{k=1}^{m} (\alpha - \varphi(\theta_k, x))p(x|\theta_k)\pi^\star_k \geq 0 \tag{13}$$

for any x. Multiplication of this inequality by any $b(x) \geq 0$ and integration with respect to ν gives $r(\pi^\star, b) \geq 0$. Therefore, $R(\varphi, b, \theta_k) \geq 0$ for some k.

Proof of Theorem 3

The claim of the theorem follows if for some $\pi^\star$, $z_j(\pi^\star) \leq 0$ for all $j \in \{1, \ldots, m\}$, which we establish below.

The problem is identical to the problem of proving existence of a general equilibrium in an exchange economy with m goods (π corresponds to the vector of prices and $z(\pi)$ corresponds to excess demand). Proposition 17.C.2 in Mas-Colell, Whinston, and Green (1995) implies the result if $z(\pi)$ is continuous, homogeneous of degree zero, and satisfies Walras' law, that is $\sum_{j=1}^{m} \pi_j z_j(\pi) = 0$ for any π.

It is clear that with a suitable extension of the domain, the posterior distribution is a function of π that is homogenous of degree zero for all x. Thus, also $\varphi(\theta_j, x; \pi)$ and $z(\pi)$ are homogenous of degree zero. Further, since $\varphi(\theta_j, x; \pi)$ defines a credible set, $\sum_{j=1}^{m} \pi_j z_j(\pi) = 0$ for all π. Continuity of $z(\pi)$ is assumed.

Proof of Corollary 1

(i) If the likelihood ratio $p(X|\theta_i)/p(X|\theta_j)$ is a continuous random variable for any $i \neq j$, then $p(x|\theta_j)$, $j = 1, \ldots, m$, have the same support equal to $\mathcal{X}$ and the posterior distribution is well defined for any prior π and ν-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 π whenever there are no ties in the posterior probabilities. The function $z(\pi)$ defined in Theorem 3 is therefore continuous in π and Theorem 3 implies that there exists a prior $\pi^\star$ for which $\varphi(\cdot, \cdot; \pi^\star)$ has coverage of at least $1 - \alpha$.

(ii) Next, let us show that $\pi_j^\star > 0$ for any j and $\varphi(\cdot, \cdot; \pi^\star)$ is a similar $1 - \alpha$ confidence set. If $\pi_j^\star = 0$ for some j then θ_j is not in a $1 - \alpha$ HPD credible set for any x (as long as $\alpha > 0$) and $\varphi(\cdot, \cdot; \pi^\star)$ has zero coverage at θ_j. Thus, $\pi_j^\star > 0$ for all j. Since $\varphi(\cdot, \cdot; \pi)$ is a $1 - \alpha$ credible set

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

Integration of the last display implies

$$\sum_{j=1}^{m} \left[\int \varphi(\theta_j, x; \pi^\star) p(x|\theta_j) d\nu(x) \right] \pi_j^\star = \alpha. \tag{14}$$

Since the coverage of $\varphi(\cdot, \cdot; \pi^\star)$ is at least $1 - \alpha$, $\int \varphi(\theta_j, x; \pi^\star) p(x|\theta_j) d\nu(x) \leq \alpha$. Because $\pi_j^\star > 0$ for all j the equality in (14) can hold only if $\int \varphi(\theta_j, x; \pi^\star) p(x|\theta_j) d\nu(x) = \alpha$ for all j or, in other words, $\varphi(\cdot, \cdot; \pi^\star)$ is similar.

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

$$\sum_{j=1}^m \varphi(\theta_j, x; \pi^\star) p(x|\theta_j) \pi_j^\star < \sum_{j=1}^m \varphi'(\theta_j, x) p(x|\theta_j) \pi_j^\star.$$

Integrating this inequality with respect to ν yields $\sum_{j=1}^m \pi_j^\star \int (\varphi(\theta_j, x; \pi^\star) - \varphi'(\theta_j, x)) p(x|\theta_j) d\nu(x) < 0$. Since $\sum_{j=1}^m \pi_j^\star \int \varphi(\theta_j, x; \pi^\star) p(x|\theta_j) d\nu(x) = \alpha$ by part (ii), this implies that there exists j such that $\int \varphi'(\theta_j, x) p(x|\theta_j) d\nu(x) > \alpha$.

8 Implementation details

8.1 Destructive Results

For the parameter near a boundary problem, the destructive results are computed via linear programming. Specifically, the parameter space is discretized as $\theta_j \in \{0, 0.1, 0.2, \ldots, 10\}$, and the betting strategy space is discretized via disjoint sets $\mathcal{X}_i \subset \mathcal{X}$, so that the only possible $b(x)$ are of the form $b(x) = \sum_{i=1}^n b_i \mathbf{1}[x \in \mathcal{X}_i]$ with $b_i \in [0, 1]$. Specifically, $\mathcal{X}_i \in \{[-6, -5.9), [-5.9, -5.8), \ldots, [9.9, 10)\}$ (so that the inspector cannot bet if X is larger than the largest parameter value in the grid). From the definition in (2), the expected winnings of this betting strategy for a given θ and α' are

$$\frac{1}{1-\alpha'} \int [\varphi(\theta, x) - \alpha'] b(x) p(x|\theta) d\nu(x) = \frac{1}{1-\alpha'} \sum_{i=1}^n b_i \int_{\mathcal{X}_i} [\varphi(\theta, x) - \alpha'] p(x|\theta) d\nu(x).$$

The integrals $\int_{\mathcal{X}_i} [\varphi(\theta, x) - \alpha'] p(x|\theta) d\nu(x)$ can be computed analytically or numerically, depending on the problem under consideration. We experimented with coarser and finer grids and found very similar results.

In the autoregressive root near unity problem, a discretization of the sample space is not feasible, since Stock's (1991) interval is a function of three different statistics. We thus apply Lemma 1 directly to the grid $\theta_j \in \{-2.50, -2.45, -2.40 \ldots, 0, 0.5, 1, \ldots, 10, 12, 14, \ldots, 120\}$, with the π_j chosen so that the inspector approximately maximizes a weighted average of expected winnings relative to a weighting function that is uniform on $[-2.5, 120]$. The grid is chosen to be finer near 0 since the distribution of data is more sensitive to changes in θ in that region. The inspector cannot bet if the MLE $\hat{\theta}$ is smaller than -2.5, or larger than 120. The expected winnings for given values of κ_j and implied b from Lemma 1 are computed via

Monte Carlo integration with $N = 1,000$ independently identically distributed draws $X_{j,i}$, $i = 1, \ldots, N$ from $p(x|\theta_j)$, $j = 1, \ldots, m$ combined via importance sampling

$$R(\varphi, b, \theta_j) = (mN)^{-1} \sum_{l=1}^{m} \sum_{i=1}^{N} \frac{p(X_{l,i}|\theta_j)}{\bar{p}(X_{l,i})} b(X_{l,i}) \frac{\varphi(\theta_j, X_{i,l}) - \alpha'}{1 - \alpha'}, \tag{15}$$

where $\bar{p}(x) = m^{-1} \sum_{l=1}^{m} p(x|\theta_l)$. The values for κ_j that satisfy the conditions of Lemma 1 are determined by a simple iterative scheme: Posit $\kappa_j = \exp[\eta_j]$, initialize η_j at -5, and then iteratively adjust η_j as a function of whether or not expected winnings at θ_j are positive or negative.

8.2 Constructive Results

As discussed in the main text, we apply the approach of Section 4.1.2. For the boundary problem, the partition in uniform densities ψ_j is based on the grid $\{0, 0.0002, 0.0004, \ldots, 0.3, 0.3005, 0.3010, \ldots, 0.5, 0.502, 0.504, \ldots, 6\}$, and beyond the grid, we use an exponential density with parameter 0.001 with support on $[6, \infty)$. For the near unit root problem, the grid for the uniform densities is $\{-2.5, -2.496, -2.492, \ldots, -2, -1.98, -1.96, \ldots, 0, 0.02^2, 0.04^2, \ldots, 8^2\}$, and the exponential density with support on $[64, \infty)$ has parameter 0.0001. For a given draw of the data, the HPD set under the hierarchical prior (6) is of the form $\{\tilde{\theta} : \tilde{p}(\tilde{\theta}|x) > c(x)\}$, where $p(\tilde{\theta}|x)$ is the posterior and $c(x)$ is defined via $\int p(\tilde{\theta}|x)\mathbf{1}[\tilde{p}(\tilde{\theta}|x) > c(x)]d\tilde{\theta} = 0.95$. For a given draw of the data $X = x$, $c(x)$ is computed via the bisection method. This can be performed quite quickly, since the posterior is a mixture of normals. Thus, if $\tilde{p}(\tilde{\theta}|x)$ crosses $c(x)$ on a given interval $[\tilde{\theta}_l, \tilde{\theta}_u]$ the intersection point $\tilde{\theta}^\star$ and corresponding posterior probability $\int_{\tilde{\theta}_l}^{\tilde{\theta}^\star} \tilde{p}(\tilde{\theta}|x)dx$ is available in closed form.

The determination of the (weighted average) confidence set inducing prior is then computed as follows: Initialize $\pi_j = \exp(\eta_j)$ at some constant, determine the coverage rates of the HPD set given $\{\pi_j\}$ at each model θ_j, and iteratively adjust η_j proportional to the discrepancy between the nominal level and the estimated coverage. In the parameter near boundary problem, the coverage rate is determined by a Riemann approximation with a grid on x equal to $\{-6, -5.996, -5.992, \ldots, 14\}$. In the near unit root problem, the coverage rates are approximated via Monte Carlo integration and importance sampling as in (15), based on a total of 100,000 draws. Point-wise coverage is checked by computing coverage on 5 equal spaced points within each interval, and on the points $\{6.01, 6.02, 6.03, \ldots, 10\}$ and $\{8.02^2, 8.04^2, \ldots, 11^2\}$ in the boundary and near unit root problems, respectively. Due to the discontinuities in the prior, the HPD set is equal to a union of disjoint sets for some draws of X. This effect seems to essentially disappear as the grid becomes finer, though: With our (relatively fine) grids, the expected length of the convex hull (depicted in Figure 2 for the parameter on the boundary problem) is at most 0.03% and 0.4% longer than the actual HPD sets in the boundary and near unit root problems, respectively, under all considered parameter values.