Credibility of Confidence Sets in Nonstandard Econometric Problems*

Ulrich K. Müller and Andriy Norets
Princeton University

September 2012

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 with positive probability, even if they are based on inverting powerful tests, or if they are chosen to minimize average expected length. 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 several nonstandard problems involving a parameter near a boundary, weak instruments, near unit roots, and moment inequalities. We find that most 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, nonstandard econometric problems, weak instruments, unit roots, moment inequalities.

*We 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 comments.

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, consider the problem of constructing a confidence set about θ based on the single observation $X \sim \mathcal{N}(\theta, 1)$, where it is known that $\theta > 0$. 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 $X < -1.96$, which surely does not constitute a reasonable description of parameter uncertainty.

Alternatively, one might approach the problem of set estimation as a decision problem, where the action space consists of all (measurable) sets. Assume a loss function that is the sum of two components: a unit loss if the reported set does not contain the true parameter, and a term that is linear in the length of the set. Decision rules that are optimal in the sense of minimizing a weighted average (over different parameter values) of risk, that is Bayes risk, might then still be empty with positive probability. As an illustration, consider the distribution described in Table 1. If the component that penalizes length (here: cardinality) has a coefficient strictly between 1/2 and 0.95/0.975, then the decision rule that minimizes the simple average of risk under θ_1 and θ_2 is given by the set that equals $\{\theta_i\}$ for $X = i$, $i = 1, 2$, and the empty set if $X = 3$. Intuitively, the draw $X = 3$ contains relatively little information about the parameter, so attempting to cover all plausible parameter values is too expensive in terms of the second component in the loss function. Indeed, this set also minimizes the simple average of the expected length among all 95% confidence sets, as may be checked by solving the corresponding linear program, and it also corresponds to the inversion of the most powerful 5% level tests.

Table 1: Distribution of X conditional on θ

$\theta \backslash x$	1	2	3
θ_1	0.950	0.025	0.025
θ_2	0.025	0.950	0.025

The possibility of empty confidence sets is not the only concern. In the first example, suppose the realization of X is just very slightly larger than -1.96. The confidence set then becomes a very small interval around zero, misleadingly conveying an extremely small degree of parameter uncertainty. Another well-known example was studied by Cox (1958) and involves the observation (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. 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 arguably understates the degree of uncertainty about θ whenever $S = 5$ relative to the nominal level, since averaged over draws where $S = 5$, coverage 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 always right on average with her objections, 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.

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.[1] Indeed, some of the suggested confidence sets are routinely relied upon in applied work. In contrast to

[1]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.

some of the more artificial examples provided above, it is typically quite difficult to gauge the "reasonableness" of popular confidence sets in such frameworks, because 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 expected winnings of the inspector. In particular, we find that popular confidence intervals for inference with a single weak instrument, for the largest autoregressive root near unity and a version of Imbens and Manski's (2004) problem are quite unreasonable, while Rosen's (2008) confidence sets for a different moment inequality problem turn out to be unobjectionable.

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 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. A major 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. Interestingly, the proof mirrors exactly the proof of existence of a general equilibrium in an exchange economy. To the extent that frequentist coverage property is deemed a desirable property also in a more explicitly Bayesian context, this approach may alternatively be viewed as a previously unexplored prior selection criterion. When applied to the HPD credible set, this endogenous prior puts positive mass on all parameter values as long as there are no nuisance parameters. The resulting confidence set is thus similar, that is it has exact coverage for all parameter values.

For problems with nuisance parameters the existence result still goes through, but the endogenous prior might put mass on a few parameter values only. This may result in confidence sets whose coverage far exceeds the nominal level for other parameter values. As a potential alternative, we suggest starting with an exogenous prior, and show how to determine the confidence set that minimizes a weighted average length criterion, subject to the inclusion of a credible set relative to the exogenous prior.

In addition, we show how problems that are naturally invariant along some dimension can be cast into a form to which our results apply. This is useful, as invariance often reduces the dimension of the effective parameter space, which in turn simplifies the numerical determination of attractive confidence sets. We also discuss how our approch may be applied for the construction of "reasonable" prediction sets, that is set estimators that contain the realization of a yet unobserved random variable with prespecified probability.

As an illustration, we determine "reasonable" confidence sets in the non-standard inference problems mentioned above. From our perspective, these sets are a more compelling description of parameter uncertainty, and thus potentially attractive for use in applied work.

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. 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. The use of particular types of credible sets, such as equal tailed sets or HPD sets, is justified in Sections 3.2 and 3.3 from a betting perspective. Section 3.4 develops the alternative approach to construction of bet-proof confidence sets which minimize a weighted average of expected length. Section 4.1 discusses the application of the theoretical results obtained for a finite parameter space to problems with a continuous but compact parameter space. In Section 4.2, we extend the framework to prediction sets. We discuss invariant problems in Section 5. Applications of the methodology are presented in Section 6. Section 7 concludes.

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 briefly mention implications for continuous but compact parameter spaces. Section 5 shows that the finiteness or compactness assumptions can be relaxed in invariant problems.

Suppose the distribution of the data $X \in \mathcal{X}$ given parameter θ, $P(\cdot|\theta)$, has density $p(x|\theta)$ with respect to some generic σ-finite measure ν. Since the parameter space is discrete it is necessary 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). Hereinafter, 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.[2] 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 always right on average with her objections, and one might correspondingly denote such 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. 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] 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.

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 $\frac{1}{1-\alpha}\int_{\mathcal{X}_0\setminus\mathcal{X}_0}(\varphi(x,\theta)-\alpha)p(x|\theta)d\nu(x)$. By similarity,

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

Since $\varphi(x,\theta) = 0$ 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 over-covers when $X \in \mathcal{X}_0$ and under-covers when $X \notin \mathcal{X}_0$.

An example of a similar confidence set that is equal to the whole parameter space with positive probability is the Anderson and Rubin (1949) confidence set when applied to a weak instrument regression; see Section 6.4 below.

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^* that solves the following problem

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

where $\pi_j \geq 0$, $\sum_{j=1}^m \pi_j = 1$, are fixed weights. For uniform weights, $\pi_i = 1/m$, b^* 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^* 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.[3] Thus, a finding of $W(\pi)$ close to α for a given $1-\alpha$ confidence set φ indicates a very high degree of "unreasonableness".

[3]The randomness that determines the interval realizes before the inspection—cf. footnote 2.

The following Lemma provides an explicit characterization of b^*. Its proof is analogous to the proof of Theorem 5 below and is omitted for brevity.

Lemma 1 *There exists a betting strategy b^* of the form*

$$b^*(x) = \mathbf{1}[\sum_{j=1}^{m}(\pi_j + \kappa_j)p(x|\theta_j)(\varphi(\theta_j, x) - \alpha) \geq 0]$$

where $\kappa_j \geq 0$ are such that (i) $\int[\varphi(\theta_j, x) - \alpha]b^*(x)p(x|\theta_j)d\nu(x) \geq 0$ for all j; (ii) κ_j is zero if $\int[\varphi(\theta_j, x) - \alpha]b^*(x)p(x|\theta_j)d\nu(x) > 0$. This betting strategy solves (4).

The optimal strategy in Lemma 1 is alternatively 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).

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.

A reader impatient to see how standard confidence sets fare with respect to these measures of "unreasonableness" in a number of non-standard econometric problems may proceed directly to Section 6 and its subsections titled "Destructive results".

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 $\pi = (\pi_1, \ldots, \pi_m)'$, $\pi_i \geq 0$, $\sum_i \pi_i = 1$, 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, equivalently,

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

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 $1-\alpha$.

Proof. Let $S = \{(y_1, \ldots, y_m) : y_j = R(\varphi, b, \theta_j), j = 1, \ldots, m \text{ and } b \in B\}$. 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

$$\sum_{k=1}^{m} (\alpha - \varphi(\theta_k, \tilde{x}))p(\tilde{x}|\theta_k)\pi^\star_k < 0. \tag{6}$$

By Definition 2, (6) implies $r(\pi^\star, b^n_{\tilde{x}}) < 0$ for sufficiently large n, which is a contradiction. Thus, (6) 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

$$\sum_{k=1}^{m} (\alpha - \varphi(\theta_k, x))p(x|\theta_k)\pi^\star_k \geq 0 \tag{7}$$

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. ■

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.[4] In the following subsection, we prove that for families of credible sets that satisfy

[4] 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.

a continuity restriction, there always exists a prior that turns a $1-\alpha$ credible set into a $1-\alpha$ confidence set. Next, we discuss betting based justifications for particular types of credible sets such as equal-tailed or HPD sets. Together, these results provide an attractive recipe for finding bet-proof confidence sets: (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. In the last subsection of this section, we consider an alternative approach to the construction of bet-proof confidence sets, which seems particularly attractive for problems involving nuisance parameters.

3.1 Existence of prior

For credible sets the function $\varphi(\theta_j, x)$ is determined by the prior π. We will make the dependence on π explicit in the notation whenever necessary. For example, a right tailed credible set for $\theta_1 < \theta_2 < \cdots < \theta_m$ can be defined as follows

$$\varphi(\theta_j, x; \pi) = \begin{cases} \min\{\max\{0, [\alpha - \sum_{k=1}^{j-1} p(\theta_k|x)]/p(\theta_j|x)\}, 1\} & \text{if } p(\theta_j|x) > 0, \\ \mathbf{1}[\sum_{k=1}^{j} p(\theta_k|x) \leq \alpha] & \text{if } p(\theta_j|x) = 0, \end{cases} \tag{8}$$

where for $j = 1$, $\sum_{k=1}^{j-1} p(\theta_k|x) = 0$, and the posterior probabilities $p(\theta_k|x)$ are functions of π. The theorem below requires the rejection probabilities $\int \varphi(\theta_j, x; \pi)p(x|\theta_j)d\nu(x)$, $j = 1, \ldots, m$ to be continuous functions of the prior π. This surely holds if $\varphi(\theta_j, x; \pi)$ itself is a continuous function of the prior π for almost all x. For instance, if $\sum_{k=1}^{j} p(\theta_k|x) = \alpha$ is a zero probability event for all j and π, then this holds in example (8). An HPD credible set is almost surely continuous as well as long as all likelihood ratios $p(\theta_j|x)/p(\theta_k|x)$, $j \neq k$ are continuously distributed.

Theorem 3 *Suppose $\varphi(\theta_j, x; \pi)$ defines a $1-\alpha$ credible set for any prior π. 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 π. Then there exists $\pi^\star$ such that $\varphi(\theta_j, x; \pi^\star)$ defines a $1-\alpha$ confidence set, that is*

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

Proof. 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. ■

The theorem trivially holds also for a general parameter of interest $\gamma = f(\theta)$ and $\varphi(f(\theta), x; \pi)$ as long as the suitably redefined $z(\pi)$ is continuous. Note that if f is one-to-one, the confidence set for $\gamma = f(\theta)$ can be taken to be the set $\varphi(\theta, x; \pi^\star)$ for θ transformed by f. Thus, it does not matter in which parameterization one determines frequentist coverage inducing priors— the priors do depend on the parameterization (if thought of as approximations to continuous distributions) but the resulting confidence sets do not.

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, there is a literature on higher order asymptotic equivalence of coverage and credibility, see a monograph on the subject by Datta and Mukerjee (2004) and references therein.

Potentially, many types of credible sets, including HPD and equal tailed sets, can satisfy the continuity requirement of the theorem. In the following subsections, we provide justifications for choosing a certain type of credible set.

3.2 Symmetry and equal tailed credible intervals

Suppose the econometrician cares about symmetry properties of the confidence interval for a one-dimensional parameter of interest. Specifically, suppose it would be "unreasonable" if the inspector could generative uniformly positive winnings with bets that claim the true parameter to be above the upper bound, or with bets that claim the true parameter value to be below the lower bound, with payoffs equal to $(1, -(\alpha/2)/(1-\alpha/2))$ in either case. In such a setting, the inspector's strategy consists of two components $(b_l, b_r) \in [0,1]^2$. A suitable extension of Lemma 2 is straightforward: Any confidence interval that is bet-proof relative to this enlarged betting space is a superset of an equal tailed $1-\alpha$ credible set. Thus, to the extent that such symmetry considerations are deemed important, one might apply Theorem 3 and determine a prior that leads to frequentist coverage of the equal tailed credible set.

3.3 Set length and HPD sets

This section formalizes the preference for shorter set estimators from a betting perspective. Suppose the inspector can challenge the set proposed by the econometrician, $\varphi(\theta, x)$, by an alternative set $\varphi'(\theta, x)$ of the same "length", $\sum_{j=1}^m \varphi'(\theta_j, x) = \sum_{j=1}^m \varphi(\theta_j, x)$ for all x. If only $\varphi'(\theta, x)$ happens to include the true parameter then the inspector wins 1; if only $\varphi(\theta, x)$ happens to include the true parameter then the inspector loses 1; if both intervals include or do not include the true parameter then the inspector wins 0. The set of possible bets is

$$\Psi'(\varphi) = \{\varphi'(\cdot,\cdot) : 0 \le \varphi' \le 1, \sum_{j=1}^m \varphi'(\theta_j, x) = \sum_{j=1}^m \varphi(\theta_j, x) \, \forall x \in \mathcal{X}, \, \varphi' \text{ is measurable}\}.$$

The expected inspector's loss at θ is

$$L(\varphi, \varphi', \theta) = \int [\varphi'(\theta, x) - \varphi(\theta, x)] p(x|\theta) d\nu(x).$$

It seems natural to rule out a set φ if there is an alternative set, φ', that has the same length and uniformly higher coverage ($L(\varphi, \varphi', \theta) < 0$ for all θ in Θ).

A set $\varphi(\theta, x)$ is an HPD set for a given prior π if for any $\varphi'(\theta, x)$ such that $\sum_{j=1}^m \varphi'(\theta_j, x) = \sum_{j=1}^m \varphi(\theta_j, x)$ and all x,

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

Lemma 3 *If for a set φ there exists no alternative set with the same length and uniformly higher coverage ($\forall \varphi' \in \Psi'(\varphi)$, $\exists \theta_j$ with $L(\varphi, \varphi', \theta_j) \ge 0$) then $\varphi(\theta, x)$ is an HPD set for some prior $\pi^\star$ at all x at which $[c - \varphi(\theta, x)] p(x|\theta)$ is d.u.s. continuous (in the sense of Definition 2) for all $\theta \in \Theta$ and $c \in [0,1]$.*

Conversely, if φ is an HPD set for some prior then there is no other set of the same length and uniformly higher coverage.

The proof of the lemma is analogous to the proof of Lemma 2 and is relegated to the appendix.

In the following theorem, we combine the betting games from Lemmas 2 and 3 and the prior existence theorem of Section 3.1. Thus, the inspector may bet against the true value being in the set φ, and at the same time she may bet that the true value is rather included in an alternative, equally long set of her choice. Any similar confidence set that is bet-proof in this extended betting game must be a $1-\alpha$ HPD credible set under a prior that induces frequentist coverage, at least under weak regularity conditions.

Theorem 4 *Assume $[c-\varphi(\theta,x)]p(x|\theta)$ is d.u.s. continuous (in the sense of Definition 2) for all θ and $c \in [0,1]$ for ν-almost all x.*

(i) Suppose φ is bet-proof in the extended betting game, that is for any $b \in B$ and $\varphi' \in \Psi'(\varphi)$

$$R(\varphi, b, \theta_j) + L(\varphi, \varphi', \theta_j) \geq 0 \tag{9}$$

for some $j \in \{1, \ldots, m\}$. Then φ is an HPD set of level of at least $1-\alpha$ relative to some prior a.s. ν.

(ii) Suppose that in addition to (i), φ is a similar $1-\alpha$ level confidence set. Then φ is an HPD set of level $1-\alpha$ relative to some prior a.s. ν.

(iii) Suppose ν is the Lebesgue measure. 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 there exists a prior for which the $1-\alpha$ HPD credible set satisfies the premises of parts (i) and (ii).

Proof. (i) Inequality (9) and an application of the minimax theorem as in Lemmas 2 and 3 implies the existence of a least favorable prior, $\pi^\star$, such that for any (b, φ'),

$$\sum_{j=1}^{m} [R(\varphi, b, \theta_j) + L(\varphi, \varphi', \theta_j)]\pi_j^\star \geq 0. \tag{10}$$

Set $\varphi' = \varphi$ and $b = b_{\tilde{x}}^n$ defined in the proof of Lemma 2. By the argument in Lemma 2, φ defines a superset of a $1-\alpha$ credible set. Next, set $b = 0$ and $\varphi = \varphi_{\tilde{x}}^n$ defined in the proof of Lemma 3. By the argument in Lemma 3, φ is an HPD set.

(ii) Suppose instead that φ has Bayesian credibility strictly higher than $1-\alpha$ for all $x \in \mathcal{X}_1$, $\nu(\mathcal{X}_1) > 0$: $\sum_{j=1}^m \varphi(\theta_j, x)p(x|\theta_j)\pi_j^\star < \alpha \sum_{j=1}^m p(x|\theta_j)\pi_j^\star$. Furthermore, $\sum_{j=1}^m \varphi(\theta_j, x)p(x|\theta_j)\pi_j^\star \leq \alpha \sum_{j=1}^m p(x|\theta_j)\pi_j^\star$ for $x \in \mathcal{X} \backslash \mathcal{X}_1$. Integrating these inequalities gives

$$\sum_{j=1}^{m} \int \varphi(\theta_j, x)p(x|\theta_j)\pi_j^\star d\nu(x) < \alpha,$$

which is a contradiction to φ being a similar confidence set of level $1-\alpha$.

(iii) If the likelihood ratio $p(x|\theta_i)/p(x|\theta_j)$ is a continuous random variable for any $i \neq j$ then ties in the posterior probabilities ($p(\theta_i|x) = p(\theta_j|x)$) happen with probability zero. An HPD credible set $\varphi(\cdot, \cdot; \pi)$ is a continuous function of 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 $1-\alpha$.

Next, let us show that $\pi^\star_j > 0$ for any j and $\varphi(\cdot, \cdot; \pi^\star)$ is a similar $1-\alpha$ confidence set. If $\pi^\star_j = 0$ 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, $\pi^\star_j > 0$ for all j. Since $\varphi(\cdot, \cdot; \pi)$ is a $1-\alpha$ credible set

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

Integration of the last display implies

$$\sum_{j=1}^{m} \left[\int \varphi(x, \theta_j; \pi^\star) p(x|\theta_j) d\nu(x) \right] \pi^\star_j = \alpha. \tag{11}$$

Since the coverage of $\varphi(\cdot, \cdot; \pi^\star)$ is at least $1-\alpha$, $\int \varphi(x, \theta_j; \pi^\star) p(x|\theta_j) d\nu(x) \leq \alpha$. Because $\pi^\star_j > 0$ for all j the equality in (11) can hold only if $\int \varphi(x, \theta_j; \pi^\star) p(x|\theta_j) d\nu(x) = \alpha$ for all j or, in other words, $\varphi(\cdot, \cdot; \pi^\star)$ is similar.

Finally, let us demonstrate property (9). Since $\varphi(\cdot, \cdot; \pi^\star)$ is a $1-\alpha$ HPD credible set for $\pi^\star$,

$$\sum_k (\varphi'(\theta_k, x) - \varphi(\theta_k, x; \pi^\star)) p(x|\theta_k) \pi^\star_k \geq 0 \text{ and } \sum_k (\alpha - \varphi(\theta_k, x; \pi^\star)) p(x|\theta_k) \pi^\star_k \geq 0$$

for any $\varphi' \in \Psi'(\varphi)$ and $x \in \mathcal{X}$. Adding the first inequality to the second one multiplied by $b(x) \geq 0$ and integrating with respect to ν yields inequality (10), which implies (9). ■

Similar to Lemma 2 and Theorem 3, Lemma 3 extends to a general parameter of interest $\gamma = f(\theta)$. However, Theorem 4 can be extended only partially if f is not one-to-one. Specifically, part (iii) no longer implies the premise of part (ii), that is the HPD confidence set may no longer be similar. The reason is that with f not one-to-one, 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. Indeed, an alternative approach presented in the following section delivers more appealing bet-proof confidence sets in our applications with nuisance parameters.

3.4 Exogenous prior and weighted average length

This section introduces an alternative approach to the construction of bet-proof confidence sets. The approach is first to specify a prior π^0 and a type of credible sets (HPD, one sided, or equal tailed) and then find a set that (i) has $1-\alpha$ frequentist coverage; (ii) includes the specified $1-\alpha$ credible set with respect to π^0 for all x; and (iii) minimizes the average expected volume. This procedure might be more appealing to researchers who take Bayesian inference seriously since they might be cautious about endogenizing the choice of a prior distribution. Also, the procedure might be useful for problems in which the endogenous prior and the resulting sets turn out to be obviously unappealing. This concern appears to be especially relevant for models with nuisance parameters. In these models, the endogenous prior can load all the weight on one particular value of the nuisance parameter and the resulting confidence sets can be quite unattractive. See Sections 6.4-6.6 for illustrations.

The choice of the credible set type can be based on the structure of the problem. For example, in the minimum of normal means problem (Section 6.5) it is natural to use a one-sided interval; in the transformed weak instruments problem (Section 6.4), the HPD set is more appropriate as there is no natural order on the angle parameter. Betting based considerations, as described in Sections 3.2 and 3.3, can also be relevant for choosing the interval type. (Trivially, if it was known that $\theta \sim \pi^0$, then φ is bet-proof in the extended games of Sections 3.2 and 3.3 if and only if φ includes the equal-tailed or HPD credible set relative to π^0.)

In order to solve for the minimum average expected volume sets we exploit the relationship between such sets and the inversion of best tests first noticed by Pratt (1961). The following theorem translates the insight of Pratt (1961) to our setting. It also provides an explicit form for the best tests. This is achieved by an extension of the generalized Neyman-Pearson lemma (Theorem 3.6.1 in Lehmann and Romano (2005)) to the construction of constrained tests, which might be of independent interest. The explicit form of the best test is particularly useful for the derivation of numerical algorithms that approximate the minimum average expected volume sets. Section 5.2 describes further generalizations and simplifications under invariance.

Theorem 5 *Let $S^0(x)$ be a subset of the parameter of interest space Γ (for example, a $1-\alpha$ credible set) and $g(x) = \sum_{i=1}^m \pi_i^0 p(x|\theta_i)$. There exists the best level α test of $H_0 : f(\theta) = \gamma_j$ against $H_1 : X \sim g$ that has the form*

$$\varphi(\gamma_j, x) = \begin{cases} 0 & \text{if } \gamma_j \in S^0(x), \\ \mathbf{1}[g(x) > \sum_{i:f(\theta_i)=\gamma_j} \kappa_{ij} p(x|\theta_i)] & \text{otherwise}, \end{cases} \tag{12}$$

where $\kappa_{ij} = 0$ if $E_{\theta_i}[\varphi(\gamma_j, X)] < \alpha$. The inversion of these tests yields a level $1-\alpha$ confidence set $\varphi(\gamma_j, x)$ that minimizes π^0-weighted average expected length among supersets of $S^0(x)$.

Proof. Claim 1: Let l be the number of distinct values of γ_j. If the tests $\varphi(\gamma_j, x)$, $j = 1, \cdots, l$ exist then the corresponding sets indeed minimize weighted average expected length, which is given by

$$\sum_{i=1}^{m} \pi_i^0 \int [\sum_{j=1}^{l} (1 - \varphi(\gamma_j, x))] p(x|\theta_i) d\nu(x) = \sum_{j=1}^{l} \int (1 - \varphi(\gamma_j, x)) g(x) d\nu(x).$$

Suppose $\tilde{\varphi}$ is an alternative set of level $1-\alpha$ satisfying $\tilde{\varphi}(\gamma_j, x) = 0$ if $\gamma_j \in S^0(x)$. As in the proof of the Neyman-Pearson lemma

$$0 \leq \int [\varphi(\gamma_j, x) - \tilde{\varphi}(\gamma_j, x)][g(x) - \sum_{i: f(\theta_i)=\gamma_j} \kappa_{ij} p(x|\theta_i)] d\nu(x).$$

Using shorthand notation for integrals, we have $\int \varphi p_i \leq \alpha$, $\int \tilde{\varphi} p_i \leq \alpha$, and $\kappa_{ij} = 0$ whenever $\int \varphi p_i < \alpha$, so that

$$\sum_{i: f(\theta_i)=\gamma_j} \kappa_{ij} (\int \varphi p_i - \int \tilde{\varphi} p_i) \geq 0.$$

Thus, $\int \varphi(\gamma_j, x) g(x) d\nu(x) \geq \int \tilde{\varphi}(\gamma_j, x) g(x) d\nu(x)$ for all j and the claim follows.

Claim 2: $\varphi(\gamma_j, x)$ exist. Fix j and without loss of generality let $\{i : f(\theta_i) = \gamma_j\} = \{1, 2, \ldots, m_j\}$.

Define $\Psi = \{\phi : 0 \leq \phi \leq 1, \phi(x) = 0 \text{ if } \gamma_j \in S^0(x)\}$ and $\Psi(c) = \{\phi \in \Psi, \int \phi g \geq c\}$ for $c \in [0, 1]$. Consider the following minimax problem

$$V(c) = \inf_{\phi \in \Psi(c)} \sup_{\pi} \sum_{i=1}^{m_j} \pi_i \int \phi p_i = \sup_{\pi} \sum_{i=1}^{m_j} \pi_i \int \phi^c p_i = \inf_{\phi \in \Psi(c)} \sum_{i=1}^{m_j} \pi_i^c \int \phi p_i. \tag{13}$$

Since $\{\phi : 0 \leq \phi \leq 1\}$ is compact in the topology of weak convergence (Theorem A.5.1 in Lehmann and Romano (2005)) so is $\Psi(c)$. The set $\{(y_1, \ldots, y_{m_j}) : y_i = \int \phi p_i, \phi \in \Psi(c)\}$ is convex, bounded, and closed. Thus, the minimax theorem of Ferguson (1967) (Theorem 2.9.1, p. 82) applies, π^c and ϕ^c exist, and (13) is justified.

Let $\bar{c} = \int_{x: \gamma_j \notin S^0(x)} g$. For $c \in [0, \bar{c}]$, $\Psi(c)$ is non-empty. If $\alpha > V(\bar{c}) = \max_i \int_{\gamma_j \notin S^0(x)} \mathbf{1}[g(x) > 0] d\nu(x)$ then claim 2 holds with $\kappa_{ij} = 0$ for all i (that is, $\{x : \gamma_j \in S^0(x)\} \cup \{x : g(x) = 0\}$ forms an acceptance region of probability of at least $1-\alpha$). Suppose $0 = V(0) < \alpha \leq V(\bar{c})$. $V(c)$ is continuous on $[0, \bar{c}]$. To see this, note that the maximum theorem implies $\inf_{\phi \in \Psi(c)} \sum_{i=1}^{m_j} \pi_i \int \phi p_i$ to be a continuous function of (c, π). As (c, π) take values in a compact set, this implies that also the supremum over π is a continuous function of c.

By the intermediate value theorem there exists c_α such that $V(c_\alpha) = \alpha$. Immediate implications of this fact are

$$\int \phi^{c_\alpha} p_i \leq \alpha, \forall i = 1, \cdots, m_j \text{ and } \pi_i^{c_\alpha} = 0 \text{ if } \int \phi^{c_\alpha} < \alpha. \tag{14}$$

Also, ϕ^{c_α} solves the following problem

$$\sup_{\phi \in \Psi, \int \phi \sum_{i=1}^{m_j} \pi_i^{c_\alpha} p_i \leq \alpha} \int \phi g.$$

If this were not the case then there would exist $\tilde{\phi} \in \Psi$ with $\int \tilde{\phi} \sum_{i=1}^{m_j} \pi_i^{c_\alpha} p_i \leq \alpha$ such that $\int \tilde{\phi} g > \int \phi^{c_\alpha} g$. But then $\hat{\phi}(x) = \max\{0, \tilde{\phi}(x) - \int (\tilde{\phi} - \phi^{c_\alpha}) g\} \in \Psi(c_\alpha)$ would deliver a value lower than α in the minimax problem, which is a contradiction. Thus, ϕ^{c_α} is a most powerful test of H_0 : the density of X is $\sum_{i=1}^{m_j} \pi_i^{c_\alpha} p_i$ against H_1 : the density of X is g among all tests in Ψ.

It is straightforward to show that the Neyman-Pearson lemma, specifically, parts (i) and (ii) of Theorem 3.2.1 in Lehmann and Romano (2005), holds when the tests are restricted to be zero in a certain fixed set. Therefore, ϕ^{c_α} has to be of the form (12) up to a set of ν-measure zero. ∎

4 Extensions

4.1 Continuous parameter space

Suppose the parameter space Θ is compact, $p(x|\theta)$ is continuous in θ for every x, and there exists $\bar{p}(x)$ such that $p(x|\theta) \leq \bar{p}(x)$ for any θ in Θ and $\int \bar{p}(x) d\nu(x) < \infty$. Then $p(\cdot|\theta)$ is uniformly continuous in θ under the $L_1(\nu)$ distance. Consider a fine grid on Θ. For this grid we can construct $1 - \alpha$ bet-proof confidence sets as described in the previous sections. By uniform continuity of $p(\cdot|\theta)$, for any $\epsilon > 0$ and a sufficiently fine grid on Θ the coverage of the set is at least $1 - \alpha - \epsilon$ for any θ in Θ. At the same time, the set remains bet-proof at level $1 - \alpha$ by definition.

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 easily 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 Invariance

Many statistical problems have a structure that is invariant to certain transformations of data and parameters. Common examples include inference about location and/or scale. It seems reasonable to impose invariance properties on the solutions of such problems. Imposing invariance often simplifies problems and reduces their dimension. In our framework, it also lets us handle problems with unbounded parameter spaces. Berger's (1985) textbook provides an introduction to the use of invariance in statistical decision theory. The theoretical developments below are illustrated by the following example.

Example 1 *Inference about the minimum of two normal means.*

Section 6.5 motivates the following problem

$$X^* = \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu + \Delta \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \right),$$

where ρ and σ are known. The objective is to construct a one-sided confidence set $(-\infty, u(X^*)]$ for $\gamma = f((\mu, \Delta)') = \min(\mu + \Delta, \mu)$.

When both X_1^* and X_2^* are shifted by an arbitrary constant a it is clear that the structure of the problem does not change and in the absence of reliable a priori information about μ we would expect $u(X^*)$ to shift by the same a.

Formally, suppose the distribution of the data $X^* \in \mathcal{X}^*$ given parameter $\theta^* \in \Theta^*$ has a density $p^*(x^*|\theta^*)$ with respect to a generic measure ν^*. Consider a group of transformations in the sample space, indexed by $a \in A$, $g : A \times \mathcal{X}^* \mapsto \mathcal{X}^*$, and a corresponding group $\bar{g} : A \times \Theta^* \mapsto \Theta^*$ on the parameter space. The inverse element is denoted by a^{-1}, that is $g(a^{-1}, g(a, x^*)) = x^*$ and $\bar{g}(a^{-1}, \bar{g}(a, \theta^*)) = \theta^*$. Let $T : \mathcal{X}^* \mapsto \mathcal{X}^*$ and $\bar{T} : \Theta^* \mapsto \Theta^*$ be maximal invariants of these groups: (i) $T(X^*) = T(g(a, X^*))$ for any $a \in A$ and (ii) if $T(X_1^*) = T(X_2^*)$ then $X_1^* = g(a, T(X_2^*)$ for some $a \in A$. Suppose there exist measurable functions $U : \mathcal{X}^* \mapsto A$ and $\bar{U} : \Theta^* \mapsto A$ such that

$$\theta^* = \bar{g}(\bar{U}(\theta^*), \bar{T}(\theta^*)) \text{ for all } \theta^* \in \Theta^* \tag{15}$$

$$x^* = g(U(x^*), T(x^*)) \text{ for all } x^* \in \mathcal{X}^*. \tag{16}$$

The inference problem is said to be *invariant* if for all $a \in A$ and $\theta^* \in \Theta^*$ the density of $g(a, X^*)$ is $p^*(\cdot|\bar{g}(a, \theta^*))$ whenever the density of X^* is $p^*(\cdot|\theta^*)$. In other words, the distribution of $g(a, X^*)$ under θ^* is the same as the distribution of X^* under $\bar{g}(a, \theta^*)$. Note that the distribution of $T(X^*)$ then only depends on θ^* via $\bar{T}^*(\theta^*)$.

Example 2 *Inference about minimum of means continued.* $\theta^* = (\Delta, \mu)'$, $f(\theta^*) = \mu + \min(0, \Delta)$, $A = \mathbb{R}$, $g(a, X^*) = (X_1^* + a, X_2^* + a)'$, $\bar{g}(a, \theta^*) = (\Delta, \mu + a)'$, $T(X^*) = (X_1^* - X_2^*, 0)'$, $U(X^*) = X_2^*$, $\bar{T}(\theta^*) = (\Delta, 0)'$ *and* $\bar{U}(\theta^*) = \mu$.

5.1 Bet-proofness under invariance

Under invariance, it seems natural to restrict attention to set estimators $\varphi^* : \Theta^* \times \mathcal{X}^* \mapsto [0, 1]$ that satisfy the same invariance, that is with $f(\theta^*) \in \Gamma$ the parameter of interest, it should hold that

$$\varphi^*(f(\theta^*), x^*) = \varphi(f(\bar{g}(a, \theta^*)), g(a, x^*)) \text{ for all } a \in A,\ \theta^* \in \Theta^* \text{ and } x^* \in \mathcal{X}^*. \tag{17}$$

Similarly, one might also be willing to restrict bets to satisfy

$$b(x^*) = b(g(a, x^*)) \text{ for all } a \in A \text{ and } x^* \in \mathcal{X}^*.$$

Intuitively, if an inspector objects to the confidence set at $X^* = x^*$, then she should also object at $X^* = g(a, x^*)$, for any $a \in A$.

The following lemma summarizes implications of imposing invariance in the problem of finding bet-proof confidence sets.

Lemma 4 *Consider an invariant inference problem and an invariant set estimator φ^*.*

(i) The distribution of $\varphi^*(f(\theta^*), X^*)$ under θ^* is the same as the distribution of $\varphi^*(f(\bar{T}(\theta^*), g(U(X^*), T(X^*)))$ under $\bar{T}(\theta^*)$.

(ii) Define $\theta = \bar{T}(\theta^*)$, $X = T(X^*)$, and

$$\begin{aligned}\varphi(f(\theta), X) &= E_\theta[\varphi^*(f(\theta), g(U(X^*), X))|X] \\ &= E_\theta[\varphi^*(f(\bar{g}(U(X^*)^{-1}, \theta)), X)|X],\end{aligned} \tag{18}$$

where the second equality follows by invariance of φ^*. Denote the density of X given θ with respect to a measure ν by $p(x|\theta)$. The frequentist coverage of $\varphi^*(f(\theta^*), X^*)$ under θ^* satisfies

$$\int [1 - \varphi^*(f(\theta^*), x^*)] p(x^*|\theta^*) d\nu^*(x^*) = \int [1 - \varphi(f(\theta), x)] p(x|\theta) d\nu(x) \tag{19}$$

and the expected loss of the inspector from an invariant bet b can be computed as

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

Proof. (i) By invariance, the distribution of $g(\bar{U}(\theta^*), X^*)$ under $\bar{T}(\theta^*)$ is the same as the distribution of X^* under $\bar{g}(\bar{U}(\theta^*), \bar{T}(\theta^*)) = \theta^*$, where the last equality follows from (15). Therefore, the distribution of $\varphi^*(f(\theta^*), X^*)$ under θ^* is the same as the distribution of $\varphi^*(f(\theta^*), g(\bar{U}(\theta^*), X^*))$ under $\bar{T}(\theta^*)$. By invariance of φ^* and (15), $\varphi^*(f(\theta^*), g(\bar{U}(\theta^*), X^*)) = \varphi^*(f(\bar{T}(\theta^*)), X^*)$. Replacing X^* by $g(U(X^*), T(X^*))$ in the latter expression, which can be done by (16), completes the proof of the claim.

(ii) By part (i) of the lemma, the coverage, $E_{\theta^*}[1 - \varphi^*(f(\theta^*), X^*)]$ is equal to $E_\theta[1 - \varphi^*(f(\theta, g(U(X^*), X))]$. The formula for the frequentist coverage follows immediately from the law of iterated expectations.

Next, let us obtain the formula for the expected loss. The argument in the proof of (i) applied to

$$[\alpha - \varphi^*(f(\theta^*), X^*)] b(X^*) \tag{20}$$

shows that the distribution of (20) under θ^* is the same as the distribution of $[\alpha - \varphi^*(f(\bar{T}(\theta^*)), g(U(X^*), T(X^*)))] b(g(\bar{U}(\theta^*), X^*))$ under $\bar{T}(\theta^*)$. By invariance of b, $b(g(\bar{U}(\theta^*), X^*)) = b(X^*) = b(T(X^*))$, where the last equality follows by (16) and invariance. Thus, the expected loss can be computed as

$$E_\theta[(\alpha - \varphi^*(f(\theta), g(U(X^*), X))) b(X)]/(1-\alpha).$$

An application of the law of iterated expectations to the last display completes the proof of the claim. ■

A shown in part (i) of the lemma, the distribution of $\varphi^*(f(\theta^*), X^*)$ only depends on θ, which makes $\Theta = \bar{T}(\Theta^*)$ the effective parameter space. Similarly, the maximal invariant X can be thought of as the effective data.

Furthermore, with φ as defined in part (ii) of the lemma, the expressions for coverage and expected betting losses are formally equivalent to (1) and (2) of Section 2.1. Thus, when Θ is finite, the results obtained in Sections 2 and 3 carry over to invariant problems with this definition of φ, Θ, and X. In particular, under suitable continuity conditions on φ, an application of Lemma 2 shows that an invariant set φ^* is bet-proof relative to invariant bets if and only if φ derived from φ^* is a superset of a $1-\alpha$ credible set. The credibility of φ here is defined as in Section 2.4, that is by equation (5), and with a posterior computed from a prior π on θ and the likelihood $p(x|\theta)$. This credibility of φ may be given the following limited information interpretation in the original problem: The probability that a Bayesian with a prior π on θ and access to the observation X only would assign to the event that the set φ^* includes the realization of the random variable $f(\bar{g}(U(X^*)^{-1}, \theta))$ is equal to $1-\alpha$. Thus, for a particular family of sets φ^* (for example, shortest or one-sided sets) that are indexed by π and $1-\alpha$ credible in the limited information sense, Theorem 3 applies and ensures the existence of a prior $\pi^\star$ that induces frequentist coverage (19) at the level $1-\alpha$.

Example 3 *Inference about minimum of means continued.*

In this problem, $X = T(X^*) = (X_1^* - X_2^*, 0)'$, $\theta = (\Delta, 0)'$,

$$\varphi^*(f(\bar{g}(U(X^*)^{-1}, \theta)), X) = \mathbf{1}[\bar{g}(U(X^*)^{-1}, f(\theta)) \leq u(X)]$$

and, with $\Delta \in \{\Delta_1, \ldots, \Delta_m\}$, $\Theta = \{(\Delta_j, 0)'\}_{j=1}^m$. *For* φ *in (18) to be a* $1-\alpha$ *credible set relative to a prior* π *for* Δ, $u(x)$ *has to solve*

$$\begin{aligned}\alpha &= \sum_{j=1}^m p(\theta_j|x) E_{\theta_j}[\mathbf{1}[f(\bar{g}(U(X^*)^{-1}, \theta_j)) \leq u(X)]|X = x] \\ &= \sum_{j=1}^m p(\Delta_j|x_1^* - x_2^*) E_{\theta_j}[\mathbf{1}[\min(\Delta_j - X_2^*, -X_2^*) \leq u((X_1^* - X_2^*, 0)')]|X_1^* - X_2^* = x_1^* - x_2^*]\end{aligned}$$

where $p(\Delta_j|x_1^* - x_2^*) = p(x_1^* - x_2^*|\Delta_j)\pi_j / \sum_{k=1}^m p(x_1^* - x_2^*|\Delta_k)\pi_k$ *and* $p(x_1^* - x_2^*|\Delta_k)$ *is the density of* $X_1^* - X_2^* \sim \mathcal{N}(\Delta_k, 1 + \sigma^2 - 2\rho\sigma)$. *The value for* u *evaluated at a generic* (x_1^*, x_2^*) *equals* $u((x_1^*, x_2^*)') = u((x_1^* - x_2^*, 0)') + x_2^*$ *via translation invariance of* φ^* *and, thus,* u. *Further, under* $\theta = (\Delta, 0)'$,

$$\begin{pmatrix} X_1^* - X_2^* \\ X_2^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \Delta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 + \sigma^2 - 2\rho\sigma & \rho\sigma - \sigma^2 \\ \rho\sigma - \sigma^2 & \sigma^2 \end{pmatrix}\right)$$

so that

$$X_2^*|X_1^* - X_2^* = x_1^* - x_2^* \sim \mathcal{N}\left(\frac{\rho\sigma - \sigma^2}{1+\sigma^2-2\rho\sigma}(x_1^* - x_2^* - \Delta), \sigma^2 - \frac{(\rho\sigma-\sigma^2)^2}{1+\sigma^2-2\rho\sigma}\right)$$
$$\sim \mathcal{N}(\lambda(x_1^* - x_2^* - \Delta), \sigma^2_{X_2^*|X})$$

and u *satifies*

$$\alpha = 1 - \sum_{j=1}^{m} p(\Delta_j|x_1^* - x_2^*)\Phi\left(\frac{\min(\Delta_j, 0) - u((x_1^* - x_2^*, 0)') - \lambda(x_1^* - x_2^* - \Delta_j)}{\sigma^2_{X_2^*|X}}\right)$$

where Φ *is the c.d.f. of a standard normal. The right hand side of this equation depends smoothly on* u *and the prior for* Δ*. It follows that the solution* u *is a continuous function of the prior. Thus, the continuity assumption on* φ *in Theorem 3 holds and there exists a prior for* Δ *for which* u *defines a left-tailed* $1-\alpha$ *confidence set. We compute such a prior in Section 6.5.*

5.2 Exogenous prior and weighted average length under invariance

In Section 3.4, we introduced an approach to the construction of bet-proof confidence sets which minimize the average expected length under the constraint of credibility with respect to an exogenously specified prior.

Let us consider extensions of this approach to invariant problems. In addition to the invariance notation introduced above, let $\hat{g} : A \times \Gamma \mapsto \Gamma$ denote a group of transformations on the parameter of interest space. For any fixed $\gamma \in \Gamma$ let $\pi^*(\theta^*, \gamma)$ denote a probability density on Θ^* with respect to some measure η, which will serve as a weighting function in the definition of the alternative of the tests to be inverted. The following lemma shows that under suitable invariance assumptions one can solve for the best test $\varphi^*(\gamma_0, x^*)$ only once for an arbitrary $\gamma_0 \in \Gamma$ and then obtain $\varphi^*(\gamma, x^*)$ for any $\gamma \in \Gamma$ by invariance. The proof is relegated to Appendix A.

Lemma 5 *Assume that for all* $a \in A$
(i) if $\theta^* \sim \pi^*(\cdot, \gamma)$ *then* $\bar{g}(a, \theta^*) \sim \pi^*(\cdot, \hat{g}(a, \gamma))$;
(ii) $f(\theta^*) = \gamma$ *is equivalent to* $f(\bar{g}(a, \theta^*)) = \hat{g}(a, \gamma)$;
(iii) $\hat{g}(a, S^0(x^*)) = S^0(g(a, x^*))$ *or, equivalently,* $\gamma \in S^0(x^*)$ *implies* $\hat{g}(a, \gamma) \in S^0(g(a, x^*))$.

If $\varphi^*(\gamma, x^*)$ *is the best test of* $H_0 : f(\theta^*) = \gamma$ *against* $H_1 : x^* \sim \int p^*(x^*|\theta^*)\pi^*(\theta^*, \gamma)d\eta(\theta^*)$ *among tests equal to zero on* $\{x^* : \gamma \in S^0(x^*)\}$*, then for all* $a \in A$

$$\varphi^*(\hat{g}(a, \gamma), x^*) = \varphi^*(\gamma, g(a^{-1}, x^*))$$

is the best test of $H_0 : f(\theta^*) = \hat{g}(a,\gamma)$ *against* $H_1 : x^* \sim \int p^*(x^*|\theta^*)\pi^*(\theta^*, \hat{g}(a,\gamma))d\eta(\theta^*)$ *among tests equal to zero on* $\{x^* : \hat{g}(a,\gamma) \in S^0(x^*)\}$.

In the formulation and proof of Theorem 5, we assumed that Γ is a finite set. While this assumption is in line with our other theoretical results, it is easy to relax under invariance and, thus, extend the applicability of the methodology discussed in that subsection. First, let us define a weighted expected volume of a set φ^* suitable for invariant problems,

$$\int [1 - \varphi^*(\gamma, x^*)]\pi^*(\theta^*, \gamma)d\omega(\gamma)p^*(x^*|\theta^*)d\nu^*(x^*)d\eta(\theta^*), \tag{21}$$

where ω is a measure on Γ and $\pi^*(\theta^*, \gamma)$ is a weighting function. This definition reflects the idea that it might be useful to allow the volume of a set to depend on θ^* if we entertain that $X^* \sim p^*(\cdot|\theta^*)$. The function $\pi^*(\theta^*, \gamma)$ accomplishes exactly that. As an example, consider defining an operational notion of "length" for an infinite on the left interval. For this problem, Pratt (1961) suggests making $\pi^*(\theta^*, \gamma)$ proportional to $\mathbf{1}[\gamma > f(\theta^*)]$. This weighting assigns zero length to all γ's below the "true" value. In addition to introducing desirable dependence of the measure of length on the "true" parameter value, $\pi^*(\theta^*, \gamma)$ also has to be an invariant density of θ^* given γ w.r.t. η so that Theorem 5 can be generalized in the following corollary under invariance and volume (21). We illustrate below how a suitable $\pi^*(\theta^*, \gamma)$ may be found in applications by continuing the discussion of the minimum of means example.

Corollary 1 *Assume that*
(i) for any $\gamma \in \Gamma$ *the set* $\{\theta^* : f(\theta^*) = \gamma\}$ *is finite;*
(ii) for any $\gamma \in \Gamma$, $\pi^*(\theta^*, \gamma)$ *is a density on* Θ^* *with respect to* η;
(iii) assumptions (i)-(iii) of Lemma 5 hold.
Then, Theorem 5 and Lemma 5 hold with (x^*, θ^*, p^*) *replacing* (x, θ, p),

$$g(x^*) = \int p^*(x^*|\theta^*)\pi^*(\theta^*, \gamma)d\eta(\theta^*)$$

and the weighted expected volume defined in (21) *with an arbitrary finite measure* ω.

Example 4 *Inference about minimum of means continued.*

With $\hat{g}(a,\gamma) = \gamma + a$, *assumption (ii) of Lemma 5 holds. Assumption (i) of Corollary 1 holds because* $\Delta \in \{\Delta_1, \ldots, \Delta_m\}$.

Since we are interested in a left-tailed interval, we can specify $\pi^*((\mu, \Delta), \gamma)$ *as follows: a uniform distribution on* $\{\Delta_1, \ldots, \Delta_m\}$ *for* Δ *and conditional on* Δ, μ *is a normal truncated to* $(-\infty, \gamma - \min(0, \Delta))$ *with mean* $\gamma - \min(0, \Delta)$ *and a large fixed variance. Then assumptions (i) of Lemma 5 and (ii) of Corollary 1 hold.*

In order to obtain a suitable S^0 that satisfies assumption (iii) of Lemma 5 one could specify a uniform prior on $\{\Delta_1, \ldots, \Delta_m\}$ for Δ and find the corresponding left-tailed invariant credible set as described in Example 3 of Section 5.1. As discussed there, any superset of S^0 is bet-proof with respect to invariant bets.

6 Applications

In this section, we consider the following six nonstandard inference problems: (i) inference for a parameter near the boundary of the parameter space, (ii) inference about the largest autoregressive root near unity, (iii) joint inference about the date and magnitude of a moderate structural break in a time series model, (iv) instrumental variable regression with a single weak instrument, (v) inference for a set-identified parameter where the upper bound of the set is determined by two competing moment inequalities; (vi) inference for a set-identified parameter where the bounds of the identified set are determined by two moment equalities. First, we are 'destructive' and explore whether previously suggested 95% confidence sets are bet-proof. For all but the first moment inequality problem (v) 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, and, for the problems involving a nuisance parameter, minimum average expected length confidence sets along the lines of Sections 3.4 and 5.2 above.[5]

In all examples, the parameter space is not naturally compact, even after imposing invariance considerations. At the same time, in most examples almost all of the natural parameter space leads to an inference problem that is very close to an unrestricted Gaussian shift experiment. For instance, inference with weak instruments becomes "almost" a standard problem unless the instruments are quantitatively weak (say, a concentration parameter below 8), and inference about the largest autoregressive root becomes "almost" a Gaussian shift experiment for large degrees of mean reversion.

In the computations presented below, we therefore work with a grid that focusses on the substantively non-standard part of the parameter space. In the destructive calculations we avoid artificial end-point effects by restricting bets to be zero whenever the MLE of the parameter of interest is outside the grid. In the constructive calculations one could presumably adapt our solutions on the grid to a continuous and unbounded parameter space by appropriately smoothing and by switching to the standard confidence interval in the almost

[5]Implementation details are discussed in Appendix C.

standard part of the parameter space, similar to the approach of Elliott, Müller, and Watson (2012). Preliminary calculations suggest that with the endogenous HPD prior on the grid smoothly extended to a density with respect to Lebesgue measure, one does indeed obtain confidence sets that come close to controlling size uniformly. Details are omitted for brevity.

6.1 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[6]

$$X \sim \mathcal{N}(\theta, 1), \quad \theta > 0. \tag{22}$$

Whether or not $\theta = 0$ is part of the parameter space is immaterial, since the distribution $\mathcal{N}(\theta, 1)$ converges in total variation to the distribution $\mathcal{N}(0, 1)$ as $\theta \to 0$. Alternatively, interest in inference about θ in (22) 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 $[x - 1.96, x + 1.96] \cap (0, \infty)$.

The numerical results below are based on the grid $\theta \in \{0, 0.1, 0.2, \cdots, 10\}$.

6.1.1 Destructive results

Figure 1 shows how much the inspector can win in repeated draws from (22) with the uncertainty about θ described by the standard interval, under the constraint that the inspector never loses in expectation for any true θ. The left panel shows maximal weighted average expected winnings with a weighting function that puts equal mass on all points of the grid. The right panel provides the envelope of the winnings, that is it plots the maximal expected winnings at $\theta = \theta_i$ with a weighting function that puts all mass at $\theta = \theta_i$, as a function of θ_i. Different lines in the upper panels 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

[6]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).

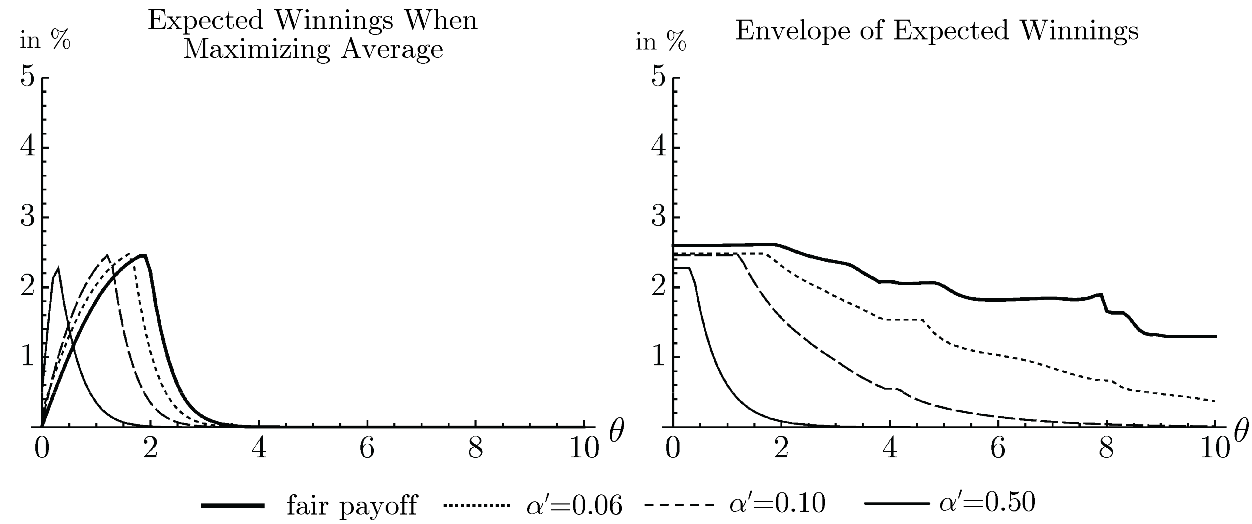

Figure 1: Parameter Near a Boundary: Destructive Results

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 valid 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.

Maximizing the simple average of winnings on the grid under fair payoffs, the inspector optimally objects whenever $X < 0$. Less favorable payoffs push the betting threshold to more negative values of X. With the objective of maximizing expected winnings at $\theta = \theta_i > 0$, fairly unintuitive betting strategies maximize the expected winnings, as they transport the winnings to larger values of θ.

6.1.2 Constructive results

The left panel in Figure 2 shows the prior that induces frequentist coverage of the HPD credible set (up to numerical approximation error, which is less than 0.1%), and the right panel compares the resulting confidence interval with the standard confidence interval. By construction, the HPD confidence set is never empty, even for very small realizations of X,

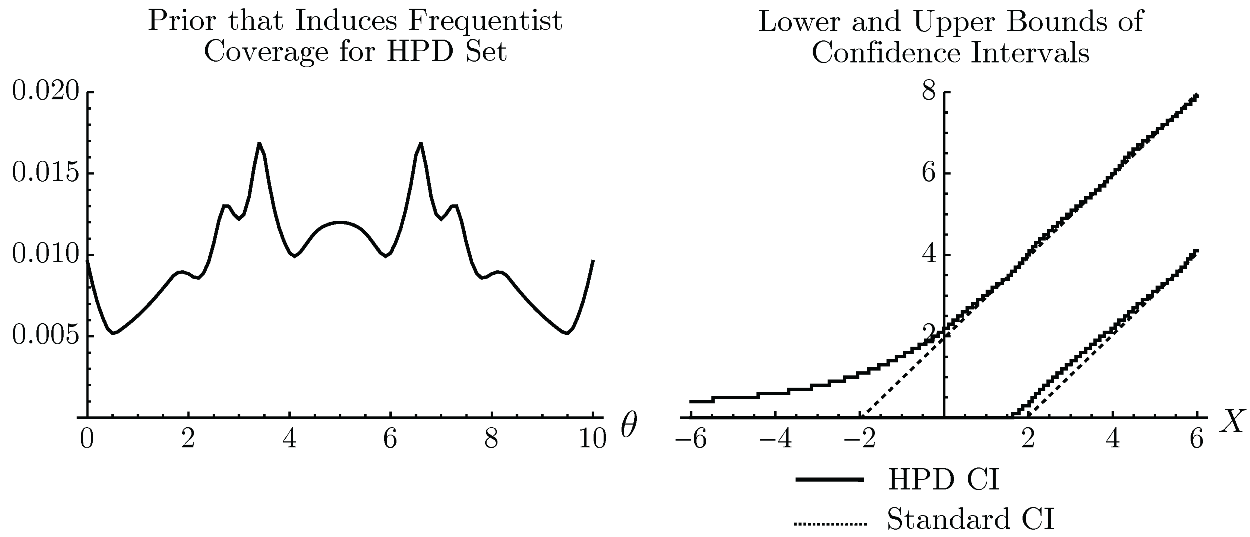

Figure 2: Parameter Near a Boundary: Constructive Results

and it is seen to become very close to the standard confidence set for realizations away from the bound.

6.2 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 $\phi(z) = 1 - \phi_1 z - \cdots - \phi_{p-1}z^{p-1}$, $\varepsilon_t \sim iid(0, \sigma^2)$ 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. 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

$$\frac{\bar{J}(1)^2 - \bar{J}(0)^2 - 1}{2\sqrt{\int_0^1 \bar{J}(s)^2 ds}} \tag{23}$$

where $\bar{J}(s) = J(s) - \int_0^1 J(r)dr$ and $J(s) = \int_0^s \exp[-\theta(s-r)]dW(r)$ with $W(r)$ a standard Wiener process. Alternatively, one might invert the nearly optimal DF-GLS t-test derived in Elliott, Rothenberg, and Stock (1996), which simply amounts to replacing $\bar{J}$ by J in (23). Finally, it might be considered more natural (although computationally more cumbersome) to invert tests of $H_0 : \theta = \theta_0$ using the quasi-MLE t-statistic under the assumption that $\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2)$, which in the Elliott, Rothenberg, and Stock (1996) framework corresponds asymptotically to inverting the statistics $(\hat{\theta} - \theta_0)/\sqrt{\hat{I}}$, where $\hat{\theta} = -\frac{1}{2}(J(1)^2 - 1)\hat{I}$ and $\hat{I} = 1/\int_0^1 J(s)^2 ds$.

Under the assumption $u_0 = O_p(1)$, the appropriate limiting experiment under $\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2)$ in the sense of LeCam (or, equivalently, in the sense of Müller (2011)) involves observing the Ornstein-Uhlenbeck process J on the unit interval $X = J(\cdot)$ with a density relative to the measure of a standard Wiener process equal to

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

The following numerical results are based on the grid $\theta \in \{-3, -2.95, \cdots, -0.05, 0, 0.5, 1, \cdots, 9.5, 10, 12, 14, \cdots, 100\}$.

6.2.1 Destructive results

The left three panels of Figure 3 show the expected winnings as a function of θ when the inspector seeks to maximize the simple average of the expected winnings on the grid, and the right panels show the corresponding probabilities of objections. Stock's (1991) confidence sets are seen to allow for very substantial winnings, even under fairly unfavorable payoffs. The DF-GLS t-test based intervals are less unreasonable, but still allow for substantial gains, also under strongly mean reverting series. 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. Finally, the inversion of the sequence of the quasi-MLE t-tests $H_0 : \theta = \theta_j$ yields much less objectionable confidence sets, with substantive gains only under the fair payoff scheme.

6.2.2 Constructive results

Figure 4 reports the shape of the prior that makes the HPD credible set into a confidence set. It puts much more mass on small values of θ, counteracting the well known positive bias of the peak of the likelihood at $\hat{\theta}$.

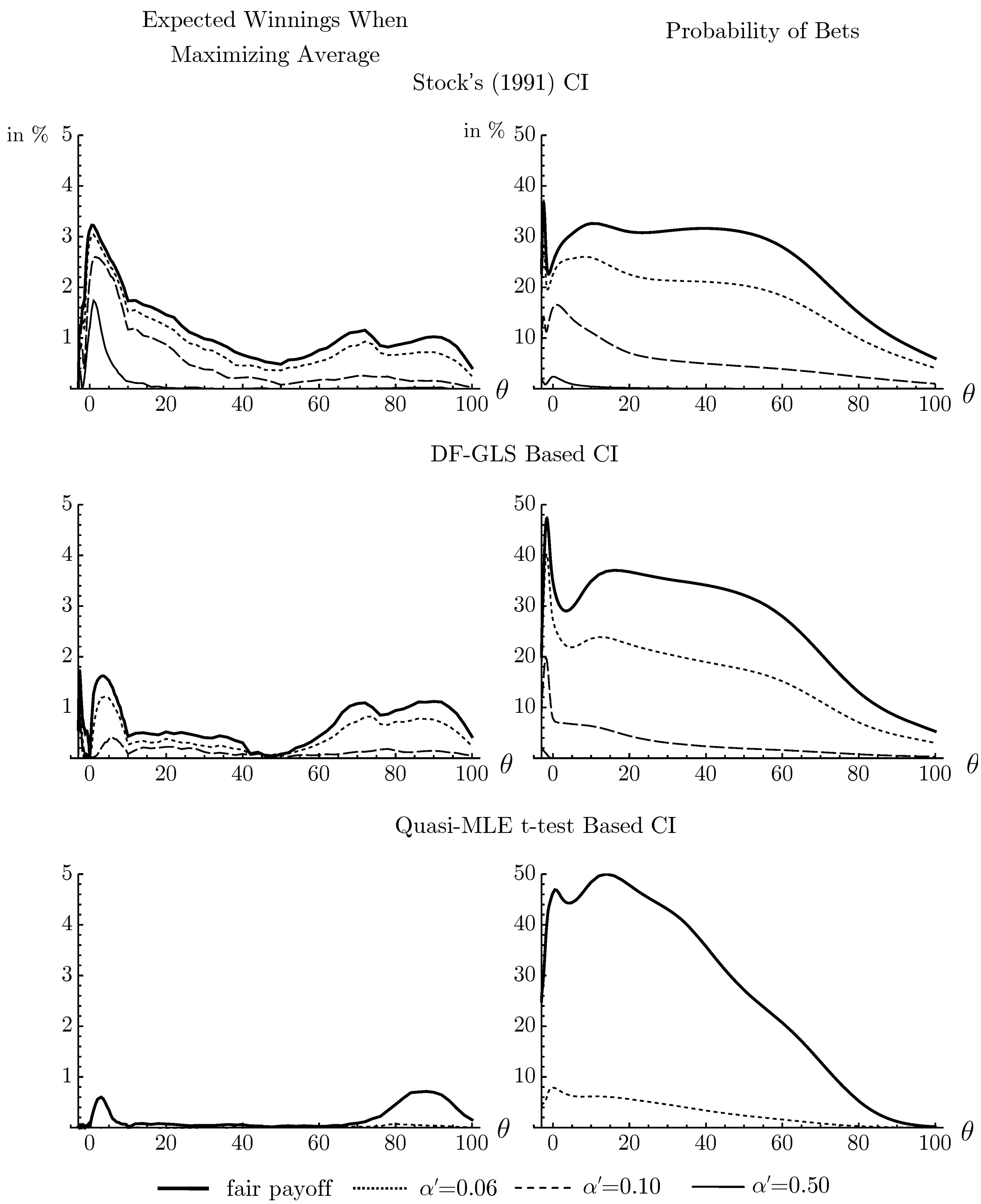

Figure 3: Largest Autoregressive Root: Destructive Results

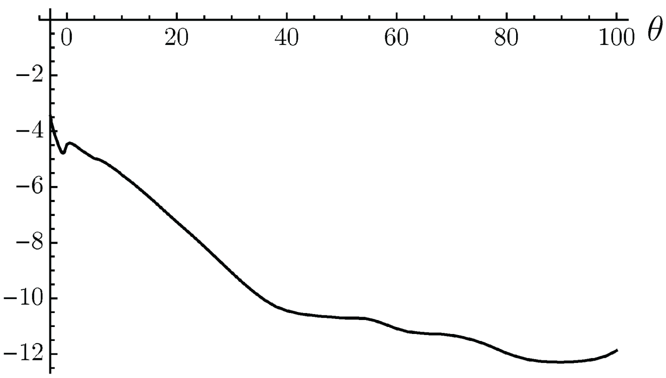

Figure 4: Largest Autoregressive Root: Log-Prior that Induces Frequentist Coverage of HPD Set

6.3 Break date and magnitude

In this section we consider the construction of a joint confidence set for the date and magnitude of a parameter shift in a time series model with T observations. As an illustration, consider first a 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

$$y_t = \mu + d\mathbf{1}[t \leq \tau] + \varepsilon_t$$

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 $d = d_T = \delta/\sqrt{T}$. Imposing translation invariance to deal with the nuisance parameter μ, we find that the partial sum process of the demeaned observations $\bar{y}_t = y_t - T^{-1}\sum_{s=1}^{T} y_s$ satisfies

$$T^{-1/2}\sum_{t=1}^{[sT]} \bar{y}_t \sim G(s) = W(s) - sW(1) + \delta(\min(\beta, s) - \beta s)$$

for any $s = t/T$, where W is a standard Wiener process, and $\beta = \tau/T$ is the break date measured in the fraction of the sample size. This suggests that the relevant observation in the limiting problem is $X = G(\cdot)$. Elliott and Müller (2009) formally show that this is indeed the relevant asymptotic experiment for a moderate structural break in well behaved parametric time series models. The density of X under $\theta = (\beta, \delta)$ relative to the measure of a standard Brownian Bridge $W(s) - sW(1)$ is given by

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

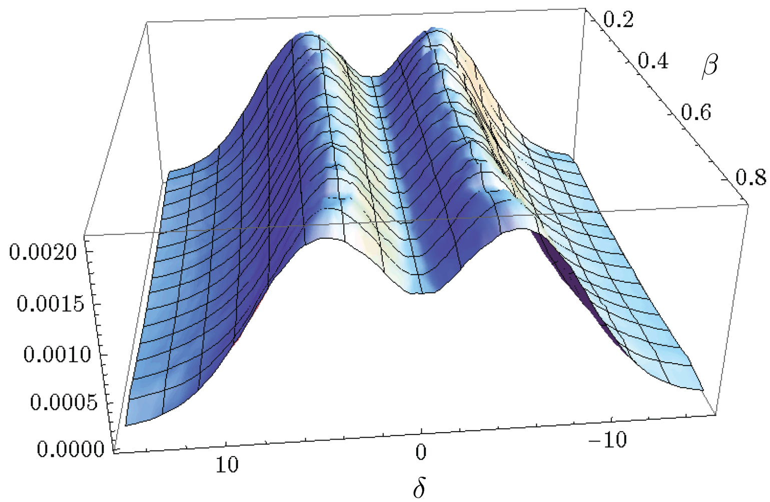

Figure 5: Break Date and Magnitude: Prior that Induces Frequentist Coverage of HPD Set

6.3.1 Constructive results

To the best of our knowledge, no asymptotically valid confidence set for the pair (β, δ) has been suggested in the literature. In fact, the construction of any valid confidence set that is not overly conservative seems a very difficult problem. But Theorem 3 provides a way forward: For any finitely discretized parameter space, there exist a prior that induces *exact* coverage of the HPD credible set (and, by Theorem 4, the resulting set is attractive from a betting point of view).

We implement this approach by imposing the grid $B \times D$ on (β, δ), where $B = \{0.15, 0.17, \cdots, 0.85\}$, and $D = \{-14.85, -14.55, -14.15, \cdots, 14.85\}$. The grid B restricts potential breaks to occur in the middle 70% of the sample, which is a standard assumption in the structural breaks literature. The largest absolute magnitude of the break is constrained to be less than 15; while this should cover the empirically most relevant part of the parameter space, it would nevertheless be desirable to allow for even larger breaks. As mentioned at the beginning of Section 6, we leave this extension to future research.

Figure 5 shows the prior π on $B \times D$ that induces the 95% HPD set to have 95% coverage. The symmetry around $\beta = 0.5$ and $\delta = 0$ is imposed in its computation.

The resulting form of the HPD confidence set is illustrated in Figure 6 for a random draw with true break date at $\beta = 0.35$ and $\delta = 8$. This particular draw of X has a second tent-like peak around $\beta = 0.6$, which leads to the inclusion of corresponding values in the HPD confidence set. If the parameter of interest was either β or δ in isolation, one could project

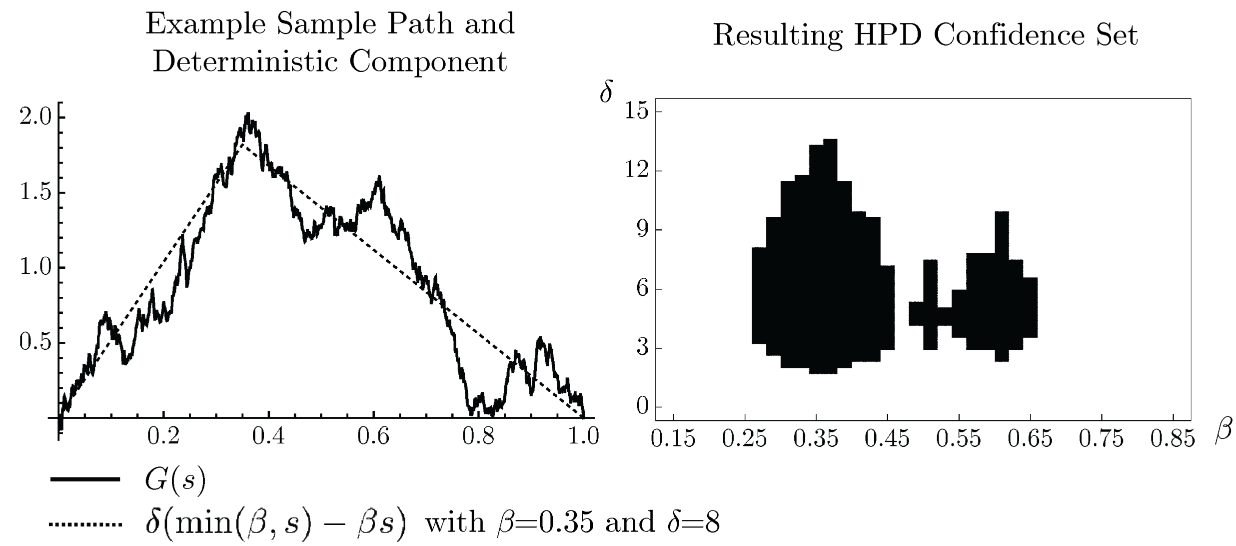

Figure 6: Break Date and Magnitude: Example of HPD Confidence Set

the joint confidence set on either axis to obtain valid confidence sets (cf. Dufour (1997)), although the resulting sets would in general be conservative.

6.4 Weak instruments

A large body of work is dedicated to deriving inference methods that remain valid even in the presence of weak instruments—see, for instance, Staiger and Stock (1997), Moreira (2003) and Andrews, Moreira and Stock (2006, 2008).

Applying the reparameterization from Chamberlain (2007) to the case of a single endogenous variable and a single instrument, we rewrite the relevant asymptotic problem as[7]

$$X^* = \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \rho \sin \phi \\ \rho \cos \phi \end{pmatrix}, I_2 \right) \tag{24}$$

with parameter $\theta^* = (\phi, \rho) \in [0, 2\pi) \times [0, \infty)$. The original coefficient of interest is a one-to-one transformation of $\gamma = f(\theta^*) = \mathrm{mod}(\phi, \pi) \in \Gamma = [0, \pi)$, with ρ a nuisance parameter that measures the strength of the instrument. In this parameterization, the popular Anderson and Rubin (1949) 5% level test of $H_0 : \phi = \phi_0$ rejects if $|X_1^* \cos \phi_0 - X_2^* \sin \phi_0| > 1.96$. Since this test is similar, its inversion yields a similar confidence set. Furthermore, note that this AR confidence set is equal to the parameter space $[0, \pi)$ whenever $||X^*|| < 1.96$. As we discussed

[7]We provide details in Appendix B.

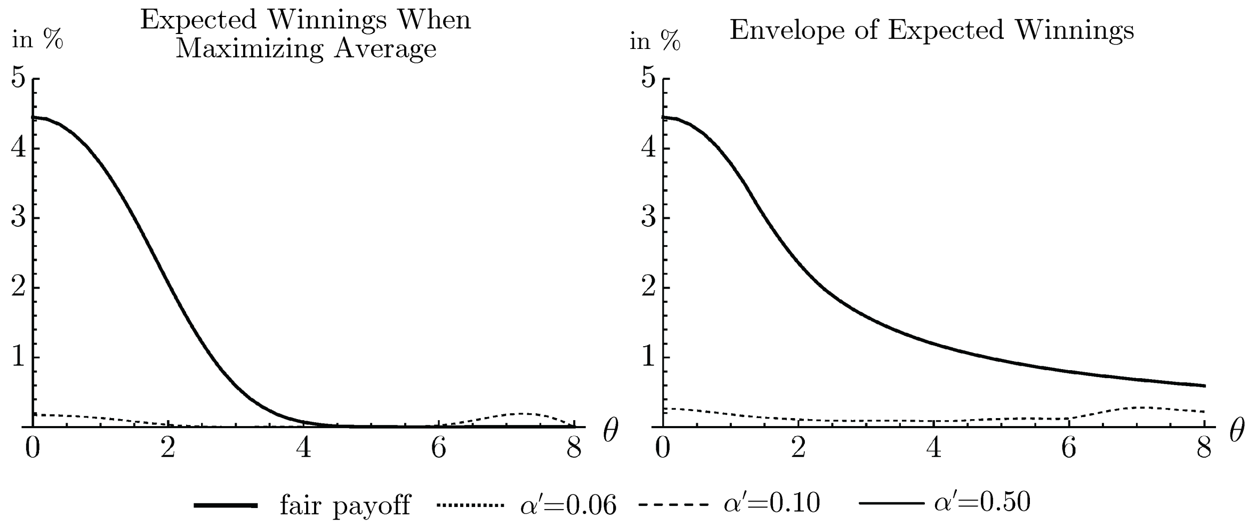

Figure 7: Weak Instruments: Destructive Results

in Section 2.2, these two observations already suffice to conclude that the AR confidence set cannot be bet-proof.

In the following results, we exploit the rotational symmetry of the problem in (24) (also see Chamberlain (2007) for related arguments). In particular, the groups of transformations on the parameters space, the sample space and the parameter of interest space Γ are, in the notation of Section 5, given by $g(a,X)=O(a)X$, $\bar{g}(a,\theta^*)=(\mathrm{mod}(\phi+a,2\pi),\rho)$ and $\hat{g}(a,\gamma)=\mathrm{mod}(\gamma+a,\pi)$, where $a\in A=[0,2\pi)$ and multiplication by the 2×2 matrix $O(a)$ rotates a 2×1 vector by the angle a. Thus, $X=T(X^*)=(0,||X^*||)'$, $||X^*||(\sin(U(X^*)),\cos(U(X^*)))=(X_1^*,X_2^*)$ (i.e., $U(X^*)\in[0,2\pi)$ is the angle of (X_1^*,X_2^*) expressed in polar coordinates), $\theta=\bar{T}(\theta^*)=(0,\rho)'$, $\bar{U}(\theta^*)=\phi$ and $f(\theta)=0$. Thus, after imposing invariance, the problem is effectively indexed only by the nuisance parameter $\rho\geq 0$. Note that the AR confidence set is also appropriately invariant. The following numerical results are based on the grid $\rho\in\{0,0.2,0.4,0.6,\cdots,8\}$.

6.4.1 Destructive results

As can be seen from Figure 7, there is a sense in which the AR interval is very unreasonable, since for ρ close to zero, the inspector can generate winnings very close to the maximum of 5%. At the same time, these winnings are quite fragile: under slightly less favorable payoffs, it becomes essentially impossible to generate uniformly positive expected winnings for the inspector. In other words, while the AR interval is not bet-proof at the nominal level of

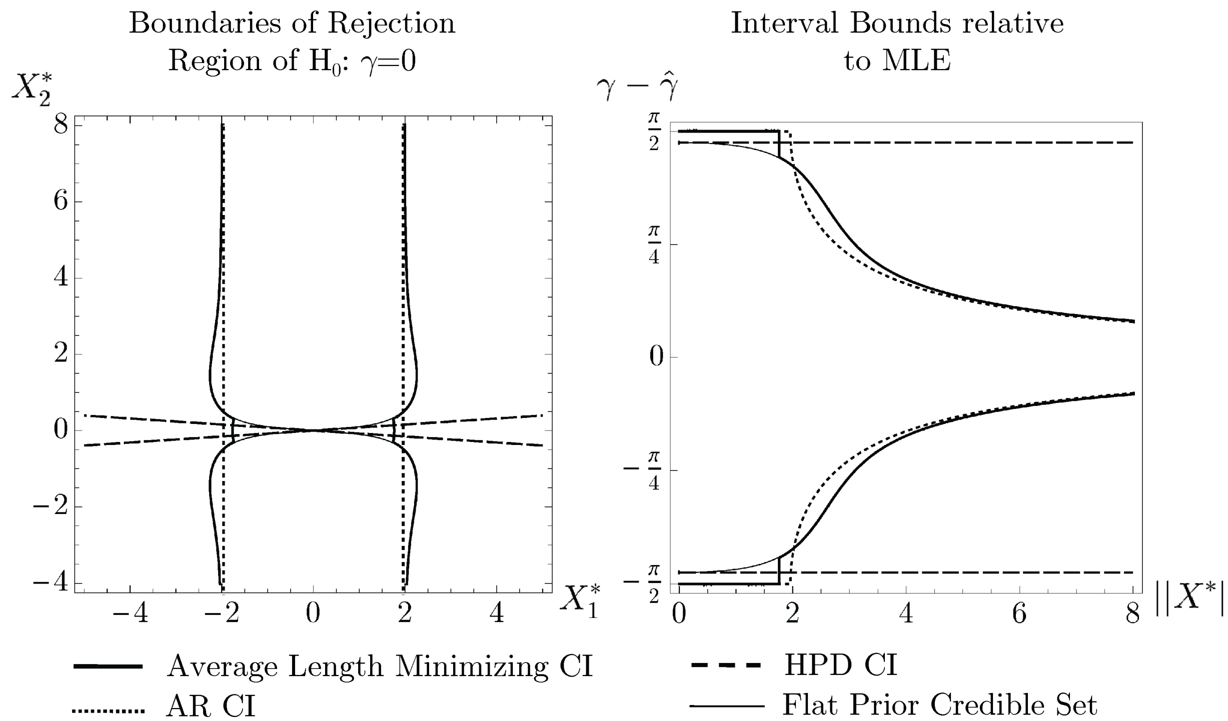

Figure 8: Weak Instruments: Constructive Results

95%, it seems very close to being reasonable at a slightly smaller level $1-\alpha'$, $\alpha' > 5\%$ (the expected winnings for $\alpha' \in \{0.1, 0.5\}$ in Figure 7 are indistinguishable from zero). Intuitively, for large values of ρ, the probability of the AR set to be equal to Γ becomes very small, so that the expected gains from betting whenever the AR set is not equal to Γ are tiny under fair odds, so that they quickly turn negative under less favorable odds. The numerical results in Figure 7 suggest that there does not exist another betting strategy that would still yield substantive gains when $\alpha' \geq 0.06$.

6.4.2 Constructive results

The prior that induces the HPD set to become a confidence set puts all mass at $\rho \to 0$. (Technically speaking, the HPD set is not defined for $\rho = 0$. But for any grid on ρ with the smallest value equal to $\epsilon > 0$, an HPD confidence set is obtained for the prior that puts all mass at ϵ.) The resulting HPD confidence set is quite unappealing: it always excludes the 5% of values of ϕ that are furthest away from the MLE, thus leading to very long confidence intervals even when ρ is large.

Following Sections 3.4 and 5.2, we therefore posit a flat prior for ρ on the grid and for ϕ on $[0, 2\pi)$, and construct the confidence interval that has minimum average expected length, subject to always containing the HPD set w.r.t. the flat prior. The left panel of Figure 8 shows the boundary of the implied critical regions for the null hypothesis $H_0 : \gamma = 0$ that is implied by these sets; points with $x_1^* = 0$ are always in the acceptance region. The confidence interval can be constructed from these rejection regions via rotational invariance; see the right panel of Figure 8.

Both the AR and the minimum average expected length interval are equal to the parameter space for small realizations of $||X^*||$,[8] and they also essentially coincide for large values of $||X^*||$. The fact that the AR interval requires only slight adjustments to become bet-proof is in line with the fragility of the destructive results of the previous subsection.

6.5 Minimum of Means Problem

As a simple motivation for the problem, suppose the parameter of interest β is known to satisfy the two moment inequalities

$$E(Y_1 - \beta) \geq 0,\ E(Y_2 - \beta) \geq 0.$$

With access to large independent samples from the populations of (Y_1, Y_2), suitably scaled sample averages become approximately normal by a central limit theorem. The relevant asymptotic problem is thus inference about $\gamma = \min(\mu + \Delta, \mu)$ based on the observation

$$X^* = \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu + \Delta \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \right). \tag{25}$$

This problem was introduced in Section 5, and it is a special case of the set-up in Rosen (2008), Andrews, Moreira, and Stock (2008), Andrews and Soares (2010) and Hirano and Porter (2011). In our following numerical work, we focus on the case where Y_1 and Y_2 are independent and of equal variance, so that $\rho = 0$ and $\sigma^2 = 1$ in (25). A popular way to form an interval about γ inverts tests of $H_0 : \gamma = \gamma_0$ using the test statistic

$$\inf_{t \in \mathbb{R}_+^2} \sum_{i=1}^{2} (X_i^* - \gamma_0 - t_i)^2 . \tag{26}$$

See Rosen (2008) for the computation of suitable critical values. We denote the resulting one-sided interval of the form $(-\infty, u(X^*)]$ as Rosen's interval.

[8] The minimum length interval is not similar, though, so (3) does not apply.

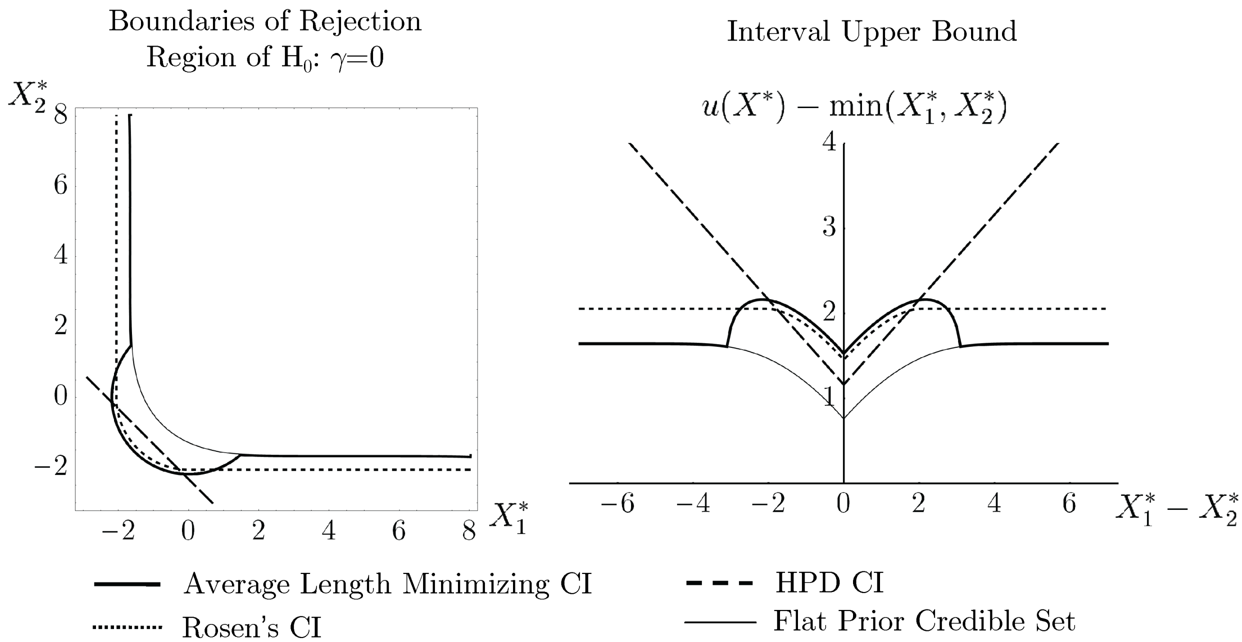

Figure 9: Minimum of Means Problem: Constructive Results

As discussed in Section 5, this problem is naturally thought of as being invariant to translations. After imposing invariance, the effective parameter space reduces to the difference of the two means Δ, for which we impose the grid $\Delta = \{-10, -9.8, \cdots, 10\}$.

6.5.1 Destructive results

It turns out that Rosen's intervals are bet-proof: they always contain the one-sided credible set relative to a flat prior on Δ (see Figure 9).

6.5.2 Constructive results

The prior that induces the one-sided credible set to have frequentist coverage puts all mass on $\Delta = 0$. This prior leads to an upper bound for the interval equal to $u(X^*) = \frac{1}{2}(X_1^* + X_2^*) + 1.645/\sqrt{2}$. While this upper bound is the smallest possible if $X_1^* = X_2^*$, the interval is arguably quite unappealing overall, as it covers values very much larger than $\min(X_1^*, X_2^*)$ when $|X_1^* - X_2^*|$ is large.

We therefore also compute the confidence interval that minimizes the weighted expected length criterion subject to including the credible set relative to a flat prior, as discussed in detail in Section 5.2. Figure 9 describes the four intervals. The left panel shows the

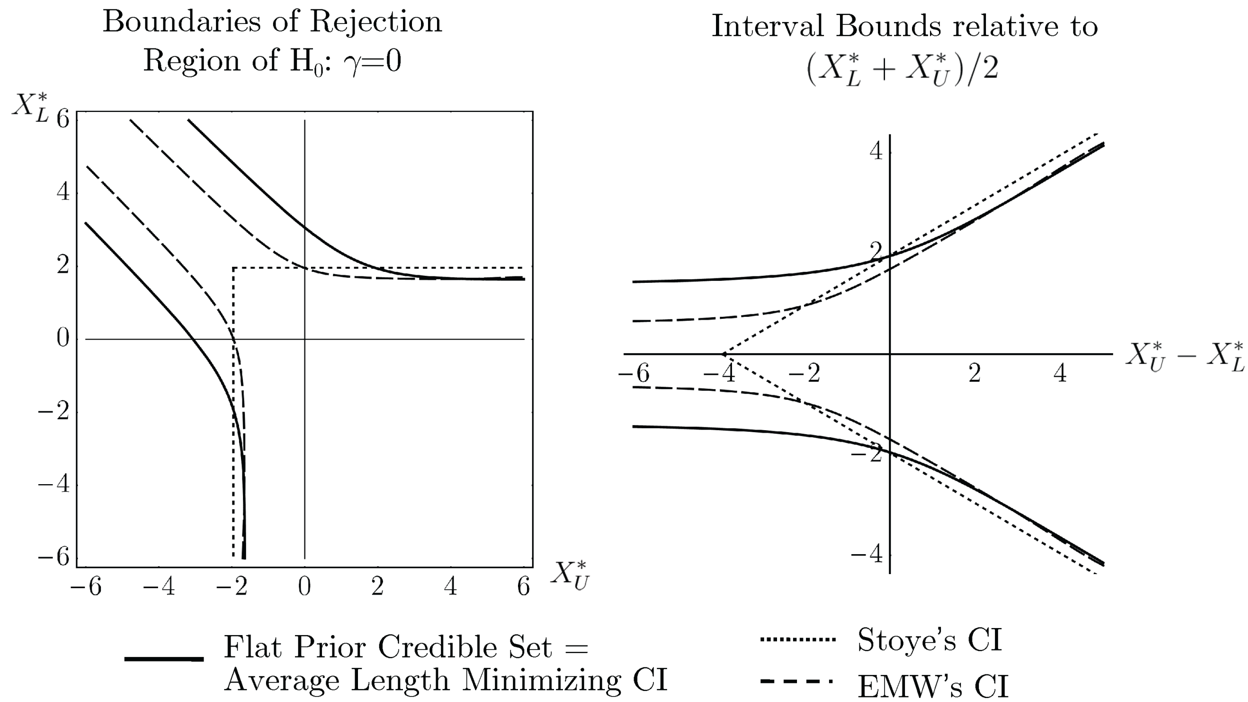

Figure 10: Imbens-Manski Problem: Constructive Results

boundaries of the rejection region of $H_0 : \gamma = 0$ that are implied by the intervals, and the right panel plots the difference of the interval upper bound from $\min(X_1^*, X_2^*)$. Relative to Rosen's interval, the average length minimizing interval has a slightly larger upper bound when $|X_1^* - X_2^*|$ is small, but equals the flat prior interval for larger values of $|X_1^* - X_2^*|$. Since the flat prior set has an upper bound very close to $\min(X_1^*, X_2^*) + 1.645$ for larger values of $|X_1^* - X_2^*|$, the average length minimizing interval thus achieves automatic "moment selection" without reliance on any diverging sequences.

6.6 Imbens-Manski Problem

Another moment inequality problem was introduced by Imbens and Manski (2004) and further studied by Woutersen (2006), Stoye (2009) and Hahn and Ridder (2011). The asymptotic version of the problem consists of a bivariate normal observation

$$X^* = \begin{pmatrix} X_U^* \\ X_L^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu + \Delta \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}\right) \tag{27}$$

where $\mu \in \mathbb{R}$ and $\Delta \geq 0$, and the parameter of interest γ is known to satisfy $\mu \leq \gamma \leq \mu + \Delta$. With $\Delta > 0$, γ is not point identified. We focus on the case where $\rho = 0$ and $\sigma^2 = 1$. Imbens and Manski (2004) observe that for large values of Δ, the natural confidence interval for γ is given by $[X_L^* - 1.645, X_U^* + 1.645]$.

Note, however, that this interval only has coverage of 90% if $\Delta = 0$. In absence of a consistent estimator for Δ, Stoye (2009) thus suggests using $[X_L^* - 1.96, X_U^* + 1.96]$ instead. This "Stoye" confidence set is empty whenever $X_U^* - X_L^* < -2 \cdot 1.96$, has exact coverage at $\Delta = 0$, and as $\Delta \to \infty$, has coverage converging to 97.5%. Elliott, Müller, and Watson (2012) derive a weighted average power maximizing test of $H_0 : \gamma = \gamma_0$ in this model. In contrast to Stoye's set, the inversion of this test always leads to an "EMW" confidence interval of positive length. Figure 10 shows the rejection regions of $H_0 : \gamma = 0$ and intervals of the Stoye and EMW sets (we discuss the minimum length region and interval in the constructive section below).

Note that this problem is translation invariant, just like the moment inequality problem above. In the following, we impose the grid $\Delta \in \{0, 0.2, \cdots, 10\}$.

6.6.1 Destructive results

Figure 11 quantifies the unreasonableness of the Stoye and EMW intervals, analogous to Figure 1 in the parameter near a boundary problem. Given that Stoye's interval is empty with positive probability, it is not surprising to see that the inspector can obtain uniformly positive expected winnings, even under very unfavorably payoffs. Note, however, that even with $\Delta = 0$, Stoye's interval is empty only with 0.28% probability. Most of the gains are rather generated by objections to intervals that are of positive length, but "too short".

Interestingly, EMW's interval is also far from bet-proof, with expected winnings that are, if anything, even larger than for Stoye's set. Part of the reason is that EMW's interval is numerically fairly close to being similar, which facilitates the "transport" of expected winnings to a different part of the parameter space. The reason why EMW's intervals are unreasonable becomes readily apparent by inspection of the right panel of Figure 10. For $X_U^* - X_L^* < -2$, EMW's interval has end-points $\frac{1}{2}(X_U^* + X_L^*) \pm c$, where $c < 1$. But even for $\Delta = 0$, the probability that this interval covers μ is less than 85%. This is just like Cox's example of the introduction: conditional on $X_U^* - X_L^* < -2$, the interval is obviously too short. Unconditionally, EMW's interval does achieve coverage, since also realizations with $X_U^* - X_L^*$ large occur often, and conditional on those realizations, the coverage is much above the nominal 95%.

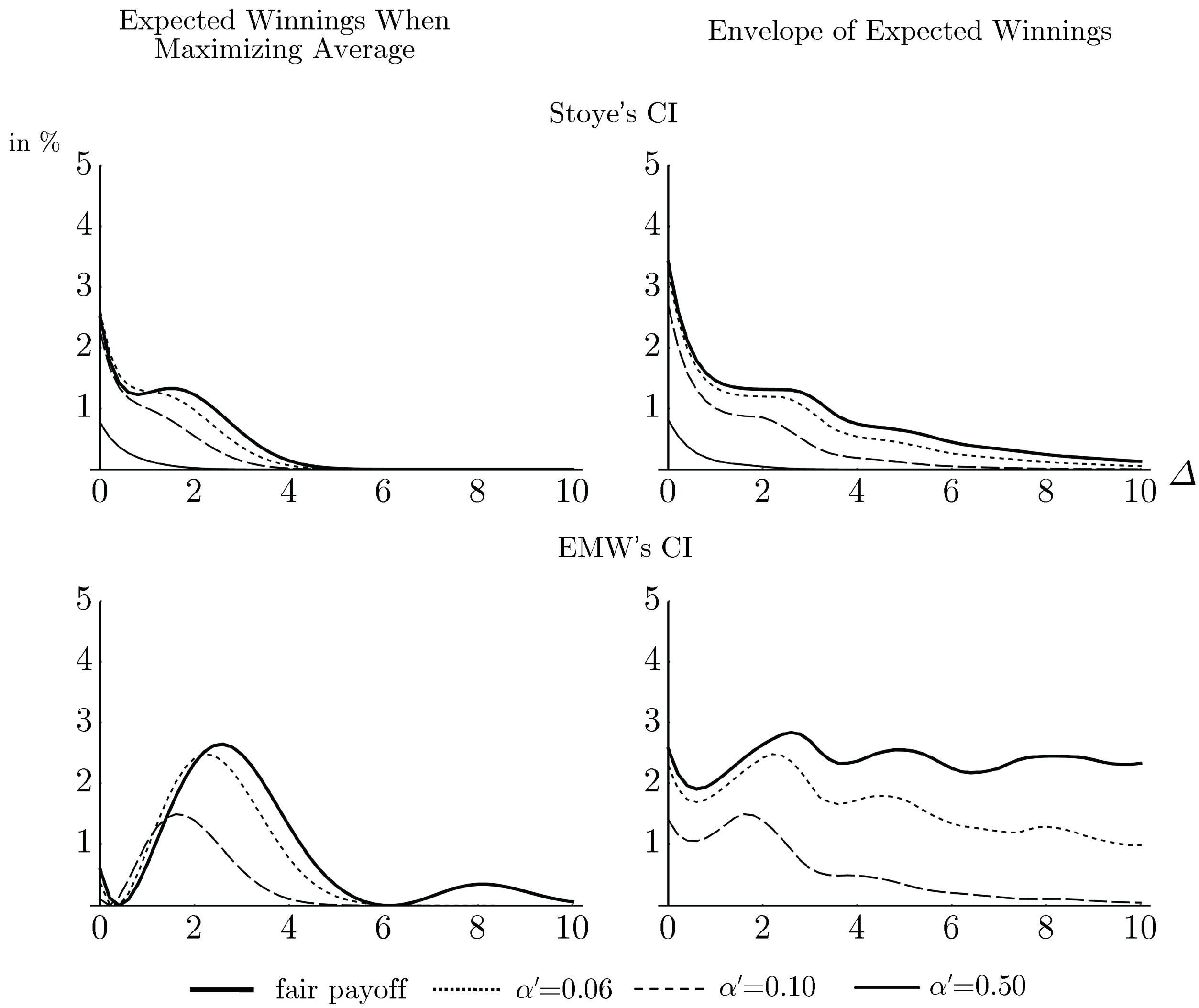

Figure 11: Imbens-Manski Problem: Destructive Results

6.6.2 Constructive results

The prior that makes the HPD set into a confidence set puts all mass at the largest value of Δ in the grid. Since the resulting interval depends crucially on our choice of grid, we do not pursue this solution further. Instead, we posit a prior that is flat on Δ, and conditional on Δ, γ is equal to either μ or $\mu + \Delta$ with probability 1/2. The equal tailed credible interval for γ under this prior turns out to have frequentist coverage for all values of Δ that are not too close to the upper bound. Thus, except for end-point effects that result from a finite grid on Δ, this credible set is trivially also the solution to the minimum length problem subject to inclusion of this credible set. As can be seen from Figure 10, the credible set, just like EMW's set, is numerically very close to $[X_L^* - 1.645, X_U^* + 1.645]$ for $X_U^* - X_L^* > 2$.

7 Conclusion

By definition, the level of a confidence set is a pre-sample statement: at least a fraction of $1 - \alpha$ data draws yield a confidence set that covers the true value. But once the sample is realized, "unreasonable" confidence sets 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.

This basic point was understood long ago by the literature surveyed in the introduction. The main theoretical contribution of this paper is to provide a practical implementation of this program. One approach to unifying Bayesian and frequentist properties is by appropriately endogenizing the prior: as we show, under weak conditions there exists a prior that induces frequentist coverage of the credible sets. In addition, we also provide results on how to enlarge the credible set relative to a fixed prior by some minimal amount to obtain the frequentist property. In combination, these recipes should enable the determination of sets that credibly describe parameter uncertainty in many nonstandard econometric problems.

A Some proofs

Proof of Lemma 3:

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

$$l(\pi, \varphi') = \sum_{k=1}^{m} L(\varphi, \varphi', \theta_k)\pi_k$$

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

$$V = \inf_{\varphi' \in \Psi'(\varphi)} \sup_{\pi \in \Delta} l(\pi, \varphi') = \sup_{\pi \in \Delta} \inf_{\varphi' \in \Psi'(\varphi)} l(\pi, \varphi') = \inf_{\varphi' \in \Psi'(\varphi)} l(\pi^\star, \varphi').$$

If for any $\varphi' \in \Psi'(\varphi)$ there exists j such that $L(\varphi, \varphi', \theta_j) \geq 0$ then $V \geq 0$.

Let us assume contrary to the first claim of the lemma that for some $\tilde{x}$, at which $[c - \varphi(\theta, x)]p(x|\theta)$ is d.u.s. continuous, there exists $\varphi_j^\star \in [0, 1]$, $j \in \{1, \ldots, m\}$, such that $\sum_{j=1}^{m} \varphi(\theta_j, x) = \sum_{j=1}^{m} \varphi_j^\star$ and

$$\sum_{j=1}^{m} [\varphi_j^\star - \varphi(\theta_j, \tilde{x}))]p(\tilde{x}|\theta_j)\pi_j^\star < 0. \tag{28}$$

Let $\varphi_{\tilde{x}}^n(\theta_j, x) = \varphi(\theta_j, x)$ for $x \notin A_{\tilde{x}} \cap B_{1/n}(\tilde{x})$ and $\varphi_{\tilde{x}}^n(\theta_j, x) = \varphi_j^\star$ for $x \in A_{\tilde{x}} \cap B_{1/n}(\tilde{x})$, 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, $0 \leq V \leq l(\pi^\star, \varphi_{\tilde{x}}^n)$ by the definition of the least favorable prior. By d.u.s. continuity, (28) implies $l(\pi^\star, \varphi_{\tilde{x}}^n) < 0$ for sufficiently large n, which is a contradiction. Thus, (28) cannot hold and the first claim of the lemma follows.

To prove the converse note that $\varphi(\cdot, \cdot)$ being an HPD set for some $\pi^\star$ implies

$$\sum_{k} (\varphi'(\theta_k, x) - \varphi(\theta_k, x))p(x|\theta_k)\pi_k^\star \geq 0 \tag{29}$$

for any $\varphi' \in \Psi'(\varphi)$ and $x \in \mathcal{X}$. Integrating this inequality with respect to ν gives $l(\pi^\star, \varphi') \geq 0$ for any $\varphi' \in \Psi'(\varphi)$. Therefore, $L(\varphi, \varphi', \theta_k) \geq 0$ for some k. ■

Proof of Lemma 5:

In order to obtain a contradiction suppose $\varphi^*(\gamma, x^*)$ is the best test of $H_0 : f(\theta^*) = \gamma$, and $\varphi^*(\hat{g}(a, \gamma), x^*)$ is not the best test of $H_0 : f(\theta^*) = \hat{g}(a, \gamma)$. The latter implies the existence of a level α test $\psi(x^*)$ which equals to zero whenever $\hat{g}(a, \gamma) \in S^0(x^*)$, and which has power strictly larger than $\varphi^*(\hat{g}(a, \gamma), x^*)$.

Given this $\psi(x^*)$, we can define another test $\psi(g(a,x^*))$. First, note that $\psi(g(a,x^*)) = 0$ whenever $\gamma \in S^0(x^*)$. This follows since $\hat{g}(a,\gamma) \in S^0(g(a,x^*))$ implies $\gamma \in S^0(x^*)$ by assumption (iii). Second, by invariance of $p^*(x^*|\theta^*)$,

$$\int \psi(g(a,x^*))p^*(x^*|\theta^*)d\nu(x^*) = \int \psi(x^*)p^*(x^*|\bar{g}(a,\theta^*))d\nu(x^*).$$

The right hand side of the above display is bounded above by α when $f(\bar{g}(a,\theta^*)) = \hat{g}(a,\gamma)$ as $\psi(x^*)$ is a level α test under this condition. Thus, $\psi(g(a,x^*))$ is a level α test under $f(\bar{g}(a,\theta^*)) = \hat{g}(a,\gamma)$, which is equivalent to $f(\theta^*) = \gamma$ by assumption (ii).

Next, note that

$$\begin{aligned}
&\int\int \varphi^*(\gamma,x^*)p^*(x^*|\theta^*)d\nu(x^*)\pi^*(\theta^*,\gamma)d\eta(\theta^*) \\
&= \int\int \varphi^*(\gamma,x^*)p^*(x^*|\bar{g}(a^{-1},\theta^*))d\nu(x^*)\pi^*(\theta^*,\hat{g}(a,\gamma))d\eta(\theta^*) \quad \text{(by invariance of } \pi^*\text{)} \\
&= \int\int \varphi^*(\gamma,g(a^{-1},x^*))p^*(x^*|\theta^*)d\nu(x^*)\pi^*(\theta^*,\hat{g}(a,\gamma))d\eta(\theta^*) \quad \text{(by invariance of } p\text{)} \\
&= \int\int \varphi^*(\hat{g}(a,\gamma),x^*)p^*(x^*|\theta^*)d\nu(x^*)\pi^*(\theta^*,\hat{g}(a,\gamma))d\eta(\theta^*) \quad \text{(by def-n of } \varphi^*(\hat{g}(a,\gamma),x^*)\text{)} \\
&< \int\int \psi(x^*)p^*(x^*|\theta^*)d\nu(x^*)\pi^*(\theta^*,\hat{g}(a,\gamma))d\eta(\theta^*) \quad \text{(by assumption on } \psi(x^*)\text{)} \\
&= \int\int \psi(x^*)p^*(x^*|\bar{g}(a,\theta^*))d\nu(x^*)\pi^*(\theta^*,\gamma)d\eta(\theta^*) \quad \text{(by invariance of } \pi^*\text{)} \\
&= \int\int \psi(g(a,x^*))p^*(x^*|\theta^*)d\nu(x^*)\pi^*(\theta^*,\gamma)d\eta(\theta^*) \quad \text{(by invariance of } p\text{)}
\end{aligned}$$

Thus, $\psi(g(a,x^*))$ has power strictly larger than $\varphi^*(\gamma,x^*)$ and it also satisfies the level and equality to zero on $\{x^* : \gamma \in S^0(x^*)\}$ constraint, which is a contradiction to $\varphi^*(\gamma,x^*)$ being the best test. ■

Proof of Corollary 1:

Claim 2 in the proof of Theorem 5 holds by assumption (i) of the present corollary. Claim 1 in the proof of Theorem 5 can be shown as follows. By Claim 2, $\varphi^*(\gamma,x^*)$ is UMP. Thus, for any $\gamma \in \Gamma$ and $\varphi^*(\gamma,x^*)$ satisfying the coverage and equality to zero on $\{x^* : \gamma \in S^0(x^*)\}$ constraint,

$$\begin{aligned}
&\int [1-\varphi^*(\gamma,x^*)] \int p^*(x^*|\theta^*)\pi^*(\theta^*,\gamma)d\eta d\nu(x^*) \\
&\leq \int [1-\varphi^*(\gamma,x^*)] \int p^*(x^*|\theta^*)\pi^*(\theta^*,\gamma)d\eta(\theta^*)d\nu(x^*).
\end{aligned}$$

Integrating this inequality with respect to ω and exchanging the order of integration, which is justified by the Fubini theorem and the non-negativity of the integrands, delivers the result. Note that without the assumed finiteness of ω the set volumes would be infinite and it would be impossible to make the volume comparison.

Finally, Lemma 5 holds since all its conditions are assumed to hold. ■

B Chamberlain's (2007) reparameterization of the weak instrument problem

The structural and reduced form equations are

$$\begin{aligned} y_{1,t} &= y_{2,t}\beta + u_{t,1} \\ y_{2,t} &= z_t\gamma + v_{t,2} \end{aligned}$$

with β the parameter of interest, and the reduced form for $y_{1,t}$ is given by

$$y_{1,t} = z_t\gamma\beta + v_{t,1}.$$

For nonstochastic z_t and $v_t = (v_{1,t}, v_{2,t})' \sim i.i.d.\mathcal{N}(0,\Omega)$ with Ω known, by sufficiency, the relevant data are effectively 2-dimensional

$$Y = \sum_{t=1}^{T} \begin{pmatrix} z_t y_{1,t} \\ z_t y_{2,t} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} S_z\gamma\beta \\ S_z\gamma \end{pmatrix}, \Omega S_z \right), \quad S_z = \sum_{t=1}^{T} z_t^2.$$

The reparameterization is $X^* = S_z^{-1/2}\Omega^{-1/2}Y$ and $S_z^{1/2}\Omega^{-1/2}(\gamma\beta,\gamma)' = \rho(\sin\phi,\cos\phi)'$. Inference about β based on Y, with Ω and S_z known and γ a nuisance parameter, is then transformed into inference about $\mathrm{mod}(\phi,\pi)$ in (24). For $\gamma \neq 0$ (or, equivalently, $\rho \neq 0$),

$$\beta = \frac{\left[\Omega^{1/2}(\sin\phi,\cos\phi)'\right]_1}{\left[\Omega^{1/2}(\sin\phi,\cos\phi)'\right]_2},$$

where $[a]_i$ stands for i-th coordinate of the vector a.

C Implementation details

C.1 Destructive Results

For all except the autoregressive root near unity problem, the destructive results are computed via linear programming. Specifically, the betting strategy space is discretized via disjoint sets

$\mathcal{X}_j \subset \mathcal{X}$, so that the only possible $b(x)$ are of the form $b(x) = \sum_{j=1}^n b_j \mathbf{1}[x \in \mathcal{X}_j]$ with $b_j \in [0,1]$. The expected winnings of this betting strategy for a given θ and α' are (cf. (2))

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

The integrals $A_j = \int_{\mathcal{X}_j} [\varphi(\theta,x) - \alpha'] p(x|\theta) d\nu(x)$ are computed analytically or numerically, depending on the problem.

In the parameter near boundary problem, the $\mathcal{X}_j$'s are the intervals $\{[-6,-5.9), [-5.9,-5.8), \ldots, [9.9,10)\}$, and with $\varphi(\theta,x) = \mathbf{1}[|x-\theta| > 1.96]$, A_j may be expressed in closed form as linear combination of the c.d.f. Φ of a standard normal evaluated at appropriate arguments. The upper bound of 10 in the $\mathcal{X}_j$ imposes that no bets may be made whenever the MLE is larger than the upper bound on the parameter space for θ.

For the weak instrument problem, define (ρ_X, ϕ_X) by $(X_1^*, X_2^*) = (\rho_X \sin\phi_X, \rho_X \cos\phi_X)$. Lemma 4 implies $\varphi(\theta, X) = E_\theta[\varphi^*(f(\theta), g(U(X^*), X)|X] = E_\rho[\varphi^*(0, (\rho_X, \phi_X))|\rho_X]$. The Jacobian determinant of the transformation $(\rho_X, \phi_X) \to (\rho_X \sin\phi_X, \rho_X \cos\phi_X) = X^{*\prime}$ is equal to $-\rho_X$. Thus, $p((\rho_X, \phi_X)|\theta) \propto |\rho_X| \exp[\rho\rho_X \cos\phi_X - \frac{1}{2}\rho_X^2]$ so that

$$p(\phi_X|\rho_X, \theta) \propto \exp[\rho\rho_X \cos\phi_X].$$

Also note that the AR interval can be written as follows

$$\varphi^*(0, (\rho_X, \phi_X)) = \mathbf{1}[\phi_X \in [\psi, \pi - \psi] \cup [\pi + \psi, 2\pi - \psi]],$$

where $\psi = \arcsin\min(1, z_\alpha/\rho_X)$. Thus

$$\varphi(\rho, \rho_X) = \frac{2\int_\psi^{\pi-\psi} \exp\{\rho\rho_X \cos\phi_X\} d\phi_X}{\int_0^{2\pi} \exp\{\rho\rho_X \cos\phi_X\} d\phi_X},$$

where the denominator is equal to 2π times the modified Bessel function of the first kind, $I_0(\rho\rho_X)$, which can be evaluated by standard software, and the numerator can be computed numerically. The integrals A_j are computed numerically on the sets $\mathcal{X}_j \in \{[0,0.2), [0.2,0.4), \ldots, [12.8,13), [13,\infty)\}$.

In the Imbens-Manski problem, the original parameter space is the triple $\theta^* = (\mu, \Delta, \kappa)' \in \mathbb{R} \times \mathbb{R}_+ \times [0,1]$, so that $\gamma = f(\theta^*) = \mu + \kappa\Delta$. Denote the lower and upper interval bounds by $l(X^*)$ and $u(X^*)$, respectively (where $l(X^*) = u(X^*)$ whenever the interval is empty). Analogous to the minimum means problem discussed in Section 5, translation invariance yields $\theta = \bar{T}(\theta^*) = (0, \Delta, \kappa)'$, $\bar{U}(\theta^*) = \mu$, $X = T(X^*) = (X_H^* - X_L^*, 0)'$, $U(X^*) = X_L^*$ and $f(\theta) = \kappa\Delta$, and we obtain

$$\begin{aligned}
\varphi(f(\theta), x) &= E_\theta[\varphi^*(f(\theta), g(U(X^*), X)|X = x] \\
&= \Phi\left(\frac{l(x) - \kappa\Delta + \lambda(x_H^* - x_L^* - \Delta)}{\sigma^2_{X_2^*|X}}\right) + 1 - \Phi\left(\frac{u(x) - \kappa\Delta + \lambda(x_H^* - x_L^* - \Delta)}{\sigma^2_{X_2^*|X}}\right).
\end{aligned}$$

As one might intuitively expect, the largest gains for the inspector are obtained for $\kappa \in \{0, 1\}$, and they are identical in either case. We thus set $\kappa = 0$ in the computation of $\varphi(f(\theta), x)$, and compute A_j numerically using $X_H^* - X_L^* \sim \mathcal{N}(\Delta, 2)$ over the sets $\mathcal{X}_j \in \{[-4, -3.8), [-3.8, -3.6), \ldots, [13.8, 14)\}$.

In the autoregressive root near unity problem, 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 and numerically determine appropriate κ_j, where expected winnings are computed using Monte Carlo integration. In detail, we 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.

C.2 Constructive Results

The confidence sets inducing HPD priors are computed as follows: Initialize $\pi_j = \exp(\eta_j)$ at some constant, estimate the coverage rate of the HPD set given $\{\pi_j\}$ by numerical integration, and iteratively adjust η_j by the discrepancy between the nominal level and the estimated coverage at θ_j. In the parameter near boundary problem, the numerical integration uses a Riemann approximation with a grid on x of step-size of approximately 10^{-2}. In the autoregressive root and structural break problems, the coverage rates are computed by Monte Carlo with 10,000 draws.

It is straightforward to check analytically that the priors mentioned in the main text do indeed induce coverage of the HPD sets in the weak instrument and moment inequality problems.

The minimum length confidence sets are obtained by inverting the corresponding best tests, which are determined with the algorithm developed in Elliott, Müller, and Watson (2012).

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., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

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.

——— (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 146, 241–254.

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.

CHAMBERLAIN, G. (2007): "Decision Theory Applied to an Instrumental Variables Model," *Econometrica*, 75(3), 609–652.

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.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

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.

——— (2009): "Pre and Post Break Parameter Inference," *Working Paper, Princeton University*.

ELLIOTT, G., U. K. MÜLLER, AND M. W. WATSON (2012): "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Working Paper, Princeton University.*

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

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

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.

HAHN, J., AND G. RIDDER (2011): "A Dual Approach to Confidence Intervals for Partially Identified Parameters," *Working Paper, UCLA.*

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

HIRANO, K., AND J. PORTER (2011): "Impossibility Results for Nondifferentiable Functionals," *Working Paper, Department of Economics, Univeristy of Wisconsin.*

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

JOSHI, V. M. (1974): "A Note on the Incompatibility of One-Sided Bayes and Frequency Confidence Intervals," *Journal of the Royal Statistical Society. Series B*, 36(2), 237–242.

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.

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses (Springer Texts in Statistics).* Springer, New York.

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.

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.

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

PRATT, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56(295), pp. 549–567.

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

ROSEN, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

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.

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.

WOUTERSEN, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," *Unpublished Mansucript, Johns Hopkins University.*