ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 84, No. 6 (November, 2016), 2183–2213

CREDIBILITY OF CONFIDENCE SETS IN NONSTANDARD ECONOMETRIC PROBLEMS

ULRICH K. MÜLLER
Princeton University, Princeton, NJ 08544, U.S.A.

ANDRIY NORETS
Brown University, Providence, RI 02912, U.S.A.

Econometrica, Vol. 84, No. 6 (November, 2016), 2183–2213

CREDIBILITY OF CONFIDENCE SETS IN NONSTANDARD ECONOMETRIC PROBLEMS

BY ULRICH K. MÜLLER AND ANDRIY NORETS[1]

Confidence intervals are commonly used to describe parameter uncertainty. In nonstandard problems, however, their frequentist coverage property does not guarantee that they do so in a reasonable fashion. For instance, confidence intervals may be empty or extremely short with positive probability, even if they are based on inverting powerful tests. We apply a betting framework and a notion of bet-proofness to formalize the "reasonableness" of confidence intervals as descriptions of parameter uncertainty, and use it for two purposes. First, we quantify the violations of bet-proofness for previously suggested confidence intervals in nonstandard problems. Second, we derive alternative confidence sets that are bet-proof by construction. We apply our framework to several nonstandard problems involving weak instruments, near unit roots, and moment inequalities. We find that previously suggested confidence intervals are not bet-proof, and numerically determine alternative bet-proof confidence sets.

KEYWORDS: Confidence sets, betting, Bayes, conditional coverage, recognizable subsets, invariance, nonstandard econometric problems, unit roots, weak instruments, moment inequalities.

1. INTRODUCTION

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

Let us consider several examples. First, suppose we are faced with the single observation $X \sim \mathcal{N}(\theta, 1)$, where it is known that $\theta > 0$. (This is a stylized version of constructing an interval based on an asymptotically normal estimator with values close to the boundary of the parameter space.) Since $[X-1.96, X+1.96]$ is a 95% confidence interval without the restriction on θ, the set $[X-1.96, X+1.96] \cap (0, \infty)$ forms a 95% confidence interval. In fact, it

[1]We thank an anonymous referee, 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 and discussions. The second author gratefully acknowledges support from the NSF via Grant SES-1260861.

 DOI: 10.3982/ECTA14023

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 $X < -1.96$. An empty confidence set realization may be interpreted as evidence of misspecification. However, the set can also be arbitrarily short if X is just very slightly larger than -1.96.

As a second illustration, consider a homoscedastic instrumental variable (IV) regression in a large sample. Suppose that there is one endogenous variable and three instruments, and the concentration parameter is 12, so that the first stage F statistic is only rarely larger than 10 (see a survey by Stock, Wright, and Yogo (2002) for definitions). The 95% Anderson and Rubin (1949) interval is then empty approximately 1.2% of the time. Moreover, it is also very short with positive probability; for instance, it is shorter than the usual two-stage least squares interval (but not empty) approximately 2.7% of the time. Applied researchers faced with such short intervals would presumably conclude that the data were very informative, and report and interpret the interval in the usual manner. But intuitively, weak instruments decrease the informational content of data, rendering these conclusions quite suspect. The same holds for all confidence sets that are empty and, by continuity, very short with positive probability.[2]

As a third illustration, let us 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. Consider the distribution described in Table I. 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 an 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. Thus, the example demonstrates

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

TABLE I
DISTRIBUTION OF X CONDITIONAL ON θ

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

that confidence sets that solve classical decision problems, minimize an average expected length, or invert likelihood ratio tests do not necessarily provide reasonable descriptions of parameter uncertainty.

Our last example demonstrates that even when confidence sets are never empty, the set is not necessarily reasonable. It is due to Cox (1958) and involves a normal observation with random but observed variance. To be specific, suppose we observe (Y, S), where $Y|S = \mathcal{N}(\theta, S^2)$, $\theta \in \mathbb{R}$ and, say, $S = 1$ with probability $1/2$, and $S = 5$ with probability $1/2$. (This is a stylized version of conducting inference about a linear regression coefficient when the design matrix is random with known distribution, as in Phillips and Hansen's (1990) and Stock and Watson's (1993) cointegrating regressions, for example.) 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 smaller expected length. Yet, this second interval understates the degree of uncertainty relative to the nominal level whenever $S = 5$, since its coverage over the draws with $S = 5$ is only about 91%.

Following Buehler (1959) and Robinson (1977), we consider a formalization of "reasonableness" of a confidence set by a betting scheme: Suppose an inspector does not know the true value of θ either, but sees the data and the confidence set of level $1 - \alpha$. For any realization, the inspector can choose to object to the confidence set by claiming that she does not believe that the true value of θ is contained in the set. Suppose a correct objection yields her a payoff of unity, while she loses $\alpha/(1 - \alpha)$ for a mistaken objection, so that the odds correspond to the level of the confidence interval. Is it possible for the inspector to be right on average with her objections no matter what the true parameter is, that is, can she generate positive expected payoffs uniformly over the parameter space? Surely, if the confidence set is empty with positive probability, the inspector could choose to object only to those realizations, and the answer must be yes. Similarly, it is not hard to see that in the example involving (Y, S), the inspector should object whenever $S = 5$ to generate uniformly positive expected winnings. The possibility of uniformly positive expected winnings may thus usefully serve as a formal indicator for the "reasonableness" of confidence sets.

The analysis of set estimators via betting schemes, and the closely related notion of a relevant or recognizable subset, goes back to Fisher (1956), Buehler (1959), Wallace (1959), Cornfield (1969), Pierce (1973), and Robinson (1977). The main result of this literature is that a set is "reasonable" or bet-proof (uniformly positive expected winnings are impossible) if and only if it is a superset of a Bayesian credible set with respect to some prior. 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 can hence be shown to be bet-proof. In nonstandard problems, however, whether a given set is bet-proof is usually far from clear and the literature referenced above provides little guidance beyond several specific examples. Since much recent econometric research has been dedicated to the derivation of inference in nonstandard problems, it is important to develop a practical framework to analyze the bet-proofness of set estimators in these settings. We develop a set of theoretical results and numerical algorithms to address this problem.

First, we propose to *quantify* the degree of unreasonableness by the largest possible expected winnings of the inspector and obtain theoretical results that simplify the corresponding numerical calculations. We find that popular confidence intervals for inference with a single weak instrument, for autoregressive roots near unity and a version of Imbens and Manski's (2004) problem are quite unreasonable.

Second, we develop a generic approach to the construction of appealing bet-proof sets. Specifically, we propose a method for determining the confidence set that minimizes a weighted average length criterion, subject to the inclusion of a Bayesian credible set, which guarantees bet-proofness. 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. As an illustration, we apply this constructive recipe to determine "reasonable" confidence sets in the nonstandard inference problems mentioned above. From our perspective, these sets are a more compelling description of parameter uncertainty, and thus 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. Section 3 develops an approach to the construction of bet-proof confidence sets that minimize a weighted average of expected length. Our methodology is extended to invariant problems in Section 4. Applications of the methodology are presented in Section 5. Section 6 concludes. Proofs are collected in the Appendix. Additional information and figures are contained in the Supplemental Material (Müller and Norets (2016a)).

2. BET-PROOF SETS

2.1. *Definitions and Notation*

Suppose the distribution of the data $X \in \mathcal{X}$ given parameter $\theta \in \Theta$, $P(\cdot|\theta)$, has density $p(\cdot|\theta)$ with respect to a σ-finite measure ν. The parameter of interest is $\gamma = f(\theta) \in \Gamma$ for a given surjective function $f : \Theta \mapsto \Gamma$. We assume that $\mathcal{X}$, Θ, and Γ are subsets of Euclidean spaces with Borel σ-algebras.

We formally define a set by a rejection probability function $\varphi : \Gamma \times \mathcal{X} \mapsto [0, 1]$, where $\varphi(\gamma, x)$ is the probability that γ is not included in the set when $X = x$ is observed. The function φ defines a $1 - \alpha$ confidence set if

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

(equivalently, the function $\varphi(\gamma_0, \cdot)$ defines a level α test of $H_0 : f(\theta) = \gamma_0$, for all $\gamma_0 \in \Gamma$). Hereafter, we assume $0 < \alpha < 1$.

As described in the Introduction, we follow Buehler (1959) and others and study the “reasonableness” of the confidence set φ via a betting scheme: For any realization of $X = x$, an inspector can choose to object to the set described by φ. 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 γ, and she loses $\alpha/(1 - \alpha)$ otherwise. For a given betting strategy b and parameter θ, the expected loss of the inspector is thus

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

If there exists a strategy b such that $R_\alpha(\varphi, b, \theta) < 0$ for all $\theta \in \Theta$, then the inspector is right on average with her objections for any parameter value, and one might correspondingly call such a φ “unreasonable.” Buehler (1959) used $\{-1, 0, 1\}$ as a betting strategy space. 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 nonnegativity on bets in what follows. For technical reasons, it is useful to allow for values of b also in $(0, 1)$, so that the set of possible betting strategies is the set B of all measurable mappings $b : \mathcal{X} \mapsto [0, 1]$.

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

2.2. *Bet-Proofness and Similarity*

Proving analytically that a given confidence set is not bet-proof seems hard in general. One apparently new general result is as follows.

THEOREM 1: *Suppose a confidence set* $\varphi(\gamma, x)$ *is similar* ($\int \varphi(f(\theta), x) p(x|\theta)\, d\nu(x) = \alpha, \forall \theta \in \Theta$) *and there exists* $\mathcal{X}_0 \subset \mathcal{X}$, *such that for any* $x \in \mathcal{X}_0$, φ *includes the whole parameter space* ($\varphi(\gamma, x) = 0, \forall x \in \mathcal{X}_0, \forall \gamma \in \Gamma$). *If* $P(\mathcal{X}_0|\theta) > 0$ *for all* θ *in* Θ, *then* φ *is not bet-proof.*

Intuitively, the set $\varphi(f(\theta), X)$ might be considered unappealing because it overcovers when $X \in \mathcal{X}_0$ and undercovers when $X \notin \mathcal{X}_0$. Similar confidence sets that are equal to the whole parameter space with positive probability or, in other words, sets that satisfy the theorem's conditions on $\mathcal{X}_0$, are part of the standard toolbox in the weak instruments literature (Anderson and Rubin (1949), Staiger and Stock (1997), Kleibergen (2002), Moreira (2003), Andrews, Moreira, and Stock (2006), and Mikusheva (2010)). Thus, the sets proposed in this literature are too short whenever they are not equal to the whole parameter space.

2.3. *Quantifying the Unreasonableness of Non-Bet-Proof Sets*

For a non-bet-proof confidence set φ, we propose to measure the degree of its "unreasonableness" by the magnitude of inspector's winnings. Specifically, we consider an optimal betting strategy $b^\star$ that solves the following problem:

$$W(\Pi) = \sup_{b \in B: R_\alpha(\varphi, b, \theta) \leq 0, \forall \theta} -\int R_\alpha(\varphi, b, \theta)\, d\Pi(\theta), \tag{3}$$

where Π is a probability measure on Θ. Thus, $b^\star$ maximizes Π-average expected winnings subject to the requirement that expected winnings are nonnegative at all parameter values. Lemma 4 shows that for any $1 - \alpha$ confidence set, $W(\Pi) \leq \alpha$. 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. Thus, a finding of $W(\pi)$ close to α indicates a very high degree of "unreasonableness."

The following lemma provides a sufficient condition for the form of $b^\star$.

LEMMA 1: *Suppose* $b^\star(x) = \mathbf{1}[\int [\varphi(f(\theta), x) - \alpha] p(x|\theta)\, d(\Pi + K)(\theta) \geq 0]$, *where* K *is a finite measure on* Θ, $\int [\varphi(f(\theta), x) - \alpha] b^\star(x) p(x|\theta)\, d\nu(x) \geq 0$ *for all* $\theta \in \Theta$, *and* $K(\{\theta : \int [\varphi(f(\theta), x) - \alpha] b^\star(x) p(x|\theta)\, d\nu(x) > 0\}) = 0$. *Then,* $b^\star$ *solves* (3).

The optimal strategy in Lemma 1 is recognized as the inspector behaving like a Bayesian with a prior proportional to $\Pi + K$: She objects whenever the posterior probability that θ is excluded from the set φ exceeds α. Lemma 1 is useful for the numerical determination of $W(\Pi)$ in some applications.

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 nonnegative 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 $0 < \alpha' < 1$ simply by 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.

The appeal of bet-proofness can also be argued on the basis of purely frequentist considerations that do not involve a betting game. A betting strategy $b : \mathcal{X} \mapsto \{0, 1\}$ defines a subset $\mathcal{X}_b$ of the sample space where $b(x) = 1$. By the definition of R_α, the noncoverage of φ under θ conditional on $\mathcal{X}_b$ equals $\alpha - (1-\alpha)R_\alpha(\varphi, b, \theta)/q(b, \theta)$, where $q(b, \theta) = \int_{\mathcal{X}_b} p(x|\theta)\, d\nu(x)$ is the probability of betting. Thus, if b delivers uniformly positive winnings, then the coverage of φ conditional on $\mathcal{X}_b$ is strictly less than the nominal level $1-\alpha$ uniformly over the parameter space, and $\mathcal{X}_b$ 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 of confidence sets; see, for example, Fisher (1956); Wallace (1959), Pierce (1973), and Robinson (1977) are also relevant. If b delivers uniformly positive expected winnings under $\alpha' > \alpha$ payoffs, then the coverage of φ conditional on $\mathcal{X}_b$ is uniformly below $1-\alpha'$. The program (3) may thus also be seen as a particular strategy of determining recognizable subsets. Our preferred measure for the unreasonableness of non-bet-proof sets, the magnitude of the expected winnings $-R_{\alpha'}(\varphi, b, \theta)$ for various values of α', provides information both on the existence of recognizable subsets at a certain level of negative bias and on the probability of such objectionable realizations. In Müller and Norets (2016a), we presented plots of the probability of betting $q(b, \theta)$ and the conditional noncoverage probability for the applications considered in this paper.

3. CONSTRUCTION OF BET-PROOF CONFIDENCE SETS

The literature on betting and confidence sets showed that a set is bet-proof at level $1-\alpha$ if and only if it is a superset of a Bayesian $1-\alpha$ credible set; see, for example, Buehler (1959), Pierce (1973), and Robinson (1977). This characterization suggests that in a search of bet-proof confidence sets, one may restrict attention to supersets of Bayesian credible sets.[3] For completeness, we formally state the sufficiency of Bayesian credibility for bet-proofness.

LEMMA 2: *Suppose φ is a superset of a $1-\alpha$ credible set for some prior Π on Θ, that is*,

$$\int \varphi\bigl(f(\theta), x\bigr) p(x|\theta)\, d\Pi(\theta) \Big/ \int p(x|\theta)\, d\Pi(\theta) \leq \alpha \quad \forall x. \tag{4}$$

Then, *φ is bet-proof at level $1-\alpha$*.

To derive appealing bet-proof confidence sets, it is necessary to introduce additional criteria that rule out unnecessarily conservative sets. Specifically, we propose to first specify a prior Π^0 and a type of credible set (highest posterior density (HPD), one sided, or equal tailed) and to 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) among all such sets, minimizes a weighted average expected volume criterion.[4]

The volume of a set φ at realization x is $\int(1-\varphi(\gamma, x))\, d\gamma$, and its expected volume under θ equals $E_\theta[\int(1-\varphi(\gamma, x))\, d\gamma] = \int(\int(1-\varphi(\gamma, x))\, d\gamma) p(x|\theta)\, d\nu(x)$. The weighted average expected volume of a set φ equals $\int E_\theta[\int(1-\varphi(\gamma, x))\, d\gamma]\, dF(\theta)$, where F is a finite measure. In order to solve for a confidence set that minimizes this criterion, we exploit the relationship between volume minimizing 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 provides an explicit form for the best tests, which is useful for the derivation of numerical algorithms that approximate the minimum average expected volume sets. The existence result exploits the insights of Wald (1950) and Lehmann (1952) on the existence of least favorable distributions in testing problems. In practice, a least favorable distribution Λ can some-

[3]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.

[4]Müller and Norets (2016b) proposed an alternative construction of set estimators with frequentist and Bayesian properties based on coverage inducing priors. The approach proposed here is more generally applicable as it yields attractive sets also in the presence of nuisance parameters.

times be determined analytically, or one can resort to numerical approximations.

THEOREM 2: *Let $S^0(x)$ be a subset of the parameter of interest space.*

(a) *Suppose, for all $\gamma \in \Gamma$, there exists a probability distribution Λ_γ on Θ with $\Lambda_\gamma(\{\theta : f(\theta) = \gamma\}) = 1$ and constants $\mathrm{cv}_\gamma \geq 0$, $0 \leq \kappa_\gamma \leq 1$ such that*

$$\varphi_0(\gamma, x) = \begin{cases} 0 & \text{if } \gamma \in S^0(x), \\ \kappa_\gamma & \text{if } \int p(x|\theta)\, dF(\theta) = \mathrm{cv}_\gamma \int p(x|\theta)\, d\Lambda_\gamma(\theta) \\ & \text{and } \gamma \notin S^0(x), \\ \mathbf{1}\Big[\int p(x|\theta)\, dF(\theta) > \mathrm{cv}_\gamma \int p(x|\theta)\, d\Lambda_\gamma(\theta)\Big] & \\ & \text{otherwise}, \end{cases} \tag{5}$$

is a level α test of $H_{0,\gamma} : f(\theta) = \gamma$, and $\mathrm{cv}_\gamma(\int E_\theta[\varphi_0(f(\theta), X)]\, d\Lambda_\gamma(\theta) - \alpha) = 0$. Then for any level $1-\alpha$ confidence set φ for γ satisfying $\varphi(\gamma, x) = 0$ for $\gamma \in S^0(x)$,

$$\begin{aligned} &\int E_\theta\Big[\int \big(1 - \varphi(\gamma, x)\big)\, d\gamma\Big]\, dF(\theta) \\ &\quad \geq \int E_\theta\Big[\int \big(1 - \varphi_0(\gamma, x)\big)\, d\gamma\Big]\, dF(\theta). \end{aligned} \tag{6}$$

(b) *Suppose $f(\theta)$ is continuous and that either Θ is compact, or for any closed and bounded subset of the sample space $\mathcal{A} \subset \mathcal{X}$, $P(\mathcal{A}|\theta) \to 0$ whenever $\|\theta\| \to \infty$. Then for any $\gamma \in \Gamma$, there exist Λ_γ, cv_γ, and κ_γ as specified in part* (a).

4. 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. 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 moment inequality example from Imbens and Manski (2004) and further studied by Woutersen (2006), Stoye (2009), and Hahn and Ridder (2009). We also return to this example in Section 5.3.

EXAMPLE: A stylized asymptotic version of the problem consists of a bivariate normal observation

$$X^* = \begin{pmatrix} X_U^* \\ X_L^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu + \Delta \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \tag{7}$$

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. Formally, $\gamma = f(\theta^*) = \mu + \lambda\Delta$, where $\theta^* = (\mu, \Delta, \lambda)' \in \mathbb{R} \times \mathbb{R}_+ \times [0, 1]$. The objective is to construct a confidence set φ^* for γ. When both X_U^* and X_L^* 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 φ^* to shift by the same a.

More generally, 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, $g(a^{-1}, g(a, x^*)) = g(a^{-1} \circ a, x^*) = x^*$ 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 $x_1^* = g(a, x_2^*)$ for some $a \in A$, for all $x^*, x_1^*, x_2^* \in \mathcal{X}^*$. Suppose there exist measurable functions $U : \mathcal{X}^* \mapsto A$ and $\bar{U} : \Theta^* \mapsto A$ such that

$$\theta^* = \bar{g}\big(\bar{U}(\theta^*), \bar{T}(\theta^*)\big) \quad \text{for all } \theta^* \in \Theta^*, \tag{8}$$

$$x^* = g\big(U(x^*), T(x^*)\big) \quad \text{for all } x^* \in \mathcal{X}^*. \tag{9}$$

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, CONTINUED: $\theta = \bar{T}(\theta^*) = (0, \Delta, \lambda)'$, $A = \mathbb{R}$, $g(a, X^*) = (X_U^* + a, X_L^* + a)'$, $\bar{g}(a, \theta^*) = (\mu + a, \Delta, \lambda)'$, $X = T(X^*) = (X_U^* - X_L^*, 0)'$, $U(X^*) = X_L^*$, and $\bar{U}(\theta^*) = \mu$.

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

$$\varphi^*\big(f(\theta^*), x^*\big) = \varphi\big(\hat{g}\big(a, f(\theta^*)\big), g\big(a, x^*\big)\big) \quad \text{for all } a \in A,\ \theta^* \in \Theta^* \text{ and } x^* \in \mathcal{X}^*. \tag{10}$$

Similarly, one might also be willing to restrict bets to satisfy $b(x^*) = b(g(a, x^*))$ for all $a \in A$ 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 $a \in A$.

We denote the density of $X = T(X^*)$ under $\theta^* = \theta \in \Theta = \bar{T}(\Theta^*)$ by $p(x|\theta)$ with respect to measure ν. The following lemma summarizes implications of imposing invariance in the analysis of bet-proofness.

LEMMA 3: *Consider an invariant inference problem.*

(i) *For any invariant set φ^*, 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) *For any given invariant set φ^*, define*

$$\varphi\big(f(\theta), x\big) = E_\theta\big[\varphi^*\big(f(\theta), X^*\big)|T\big(X^*\big) = x\big]. \tag{11}$$

The frequentist coverage of $\varphi^*(f(\theta^*), X^*)$ under θ^* satisfies

$$\begin{aligned}
&\int \big[1 - \varphi^*\big(f\big(\theta^*\big), x^*\big)\big] p^*\big(x^*|\theta^*\big)\, d\nu^*\big(x^*\big) \\
&\quad = \int \big[1 - \varphi\big(f(\theta), x\big)\big] p(x|\theta)\, d\nu(x),
\end{aligned} \tag{12}$$

and for $\theta = \bar{T}(\theta^*)$, the expected loss of the inspector from an invariant bet b satisfies

$$\begin{aligned}
&\frac{\int \big[\alpha - \varphi^*\big(f\big(\theta^*\big), x^*\big)\big] b\big(x^*\big) p^*\big(x^*|\theta^*\big)\, d\nu^*\big(x^*\big)}{1-\alpha} \\
&\quad = \frac{\int \big[\alpha - \varphi\big(f(\theta), x\big)\big] b(x) p(x|\theta)\, d\nu(x)}{1-\alpha}.
\end{aligned} \tag{13}$$

(iii) *If $\hat{g}(U(g(a, x^*))^{-1} \circ a, \gamma) = \hat{g}(U(x^*)^{-1}, \gamma)$ for all $x^* \in \mathcal{X}^*$, $\gamma \in \Gamma$, and $a \in A$ (which holds, for example, when the parameter of interest is not affected by $\hat{g}$ or when $g(a_1, x^*) = g(a_2, x^*)$ implies $a_1 = a_2$), then for any given set $\psi(\gamma, x)$, the set $\psi^*(\gamma, x^*) = \psi(\hat{g}(U(x^*)^{-1}, \gamma), T(x^*))$ is invariant, and $\psi^*(\gamma, x) = \psi(\gamma, x)$ for all $x = T(x^*) \in \mathcal{X} = T(\mathcal{X}^*)$.*

As shown in part (i) of the lemma, the distribution of $\varphi^*(f(\theta^*), X^*)$ only depends on $\theta = \bar{T}(\theta^*)$, 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 (12) and expected betting losses (13) are equivalent to (1) and (2) of Section 2.1. Thus, the results obtained in Section 2 carry over to invariant problems with this definition of φ, Θ, and X. In particular, the largest average expected winnings under invariant bets are obtained by the strategy of Lemma 1, and Lemma 2 shows that an invariant set φ^* is bet-proof relative to invariant bets if φ in (11) derived from φ^* is a superset of a $1-\alpha$ credible set in the $(\mathcal{X}, \Theta)$ problem.

Using the invariance of φ^* and (9), we can rewrite $\varphi(f(\theta), x) = E_\theta[\varphi^*(\hat{g}(U(X^*)^{-1}, f(\theta)), T(X^*))|T(X^*) = x]$. Then, the credibility level of φ may be given the following limited information interpretation in the original (X^*, Θ^*) problem: It is the probability that a Bayesian having a prior Π on $\theta \in \Theta$ and observing only $X = T(X^*)$ would assign to the event that the set φ^* includes the realization of the random variable $\hat{g}(U(X^*)^{-1}, f(\theta))$. This may be used constructively: For any $T(X^*) = x \in \mathcal{X}$, one could determine, say, the shortest set or equal-tailed interval $S^0(x) \subset \Gamma$ of credibility level $1-\alpha$ in this sense. Under the assumption of part (iii) of the lemma, the set $S^0(x^*) = \hat{g}(U(x^*), S^0(T(x^*)))$ for $x^* \in \mathcal{X}^*$ is an invariant set, and by Lemma 2 and (13), it is bet-proof against invariant bets. In the special case where the parameter of interest is unaffected by the transformations, $\hat{g}(\gamma, a) = \gamma$, $\varphi(f(\theta), x) = \varphi^*(f(\theta), x)$, and $S^0(x)$ reduces to the usual credible set in the $(\mathcal{X}, \Theta)$ problem. Either way, the construction of $S^0(x^*)$ only requires the specification of a prior on Θ, but not on the original parameter space Θ^*.

EXAMPLE, CONTINUED: $\hat{g}(U(X^*)^{-1}, f(\theta)) = \Delta\lambda - X_L^*$. The limited information Bayesian updates his prior on $\theta = (0, \Delta, \lambda)'$ based on $X = x$, and assigns credibility to any set $\varphi^*(\cdot, x)$ according to the probability that the posterior weighted (over θ) mixture of normals $\Delta\lambda - X_L^*|X = x \sim \mathcal{N}(\Delta\lambda + \frac{1}{2}(x - \Delta), \frac{1}{2})$ takes on values in φ^*. From this, one can easily determine the shortest set of credibility $1-\alpha$, or the interval $S^0(x) = [l^0(x), u^0(x)]$ such that exactly $\alpha/2$ of this mixture of normals probability is below and above the interval endpoints. The interval $S^0(x^*) = \hat{g}(U(x^*), S^0(T(x^*))) = [l^0(x_U^* - x_L^*) + x_L^*, u^0(x_U^* - x_L^*) + x_L^*]$ then is invariant and bet-proof against invariant bets.

As in Section 3, one can augment this credible set to induce coverage under all θ^* in a way that minimizes weighted average volume. The following theorem describes the form of this augmentation.

THEOREM 3: *Consider an invariant inference problem. Let $S^0(x^*)$ be an invariant subset of the parameter of interest space Γ, that is, $\gamma \in S^0(x^*)$ implies $\hat{g}(a, \gamma) \in S^0(g(a, x^*))$ for all $a \in A$ and $x^* \in \mathcal{X}^*$. Suppose that either*:

(a) *the parameter of interest is invariant, $\hat{g}(a, \gamma) = \gamma$ for all $a \in A$ and $\gamma \in \Gamma$, and there exists φ_0 as defined in Theorem* 2(a) *when applied to $(X, \theta, S^0(x))$; or*

(b) *the assumption in Lemma* 3(iii) *and the following three conditions hold*:

(b.i) *the random vector $(X, Y) = (T(X^*), \hat{g}(U(X^*)^{-1}, f(\theta)))$ under $\theta^* = \theta \in \Theta$ has density $\tilde{p}(x, y|\theta)$ with respect to $\nu(x) \times \mu(y)$, where μ is Lebesgue measure on Γ*;

(b.ii) *for any invariant set* φ^*, $\int(1-\varphi^*(\gamma, g(a,x)))\,d\gamma = g_l(a)\int(1-\varphi^*(\gamma, x))\,d\gamma$ *for all* $a \in A$ *and* $x^* \in \mathcal{X}^*$ *and some function* $g_l : A \mapsto \mathbb{R}_+$ *such that* $h_\theta(x) = E_\theta[g_l(U(X^*))|X = x]$ *exists for* ν*-almost all* x;

(b.iii) *for a finite measure* F *on* Θ, *there exists a probability distribution* Λ *on* Θ *and constants* $\mathrm{cv} \geq 0$ *and* $0 \leq \kappa \leq 1$ *such that*

$$\varphi_0(\gamma, x) = \begin{cases} 0 & \text{if } \gamma \in S^0(x), \\ \kappa & \text{if } \mathrm{cv}\displaystyle\int \tilde{p}(x, \gamma|\theta)\,d\Lambda(\theta) = \int h_\theta(x) p(x|\theta)\,dF(\theta) \\ & \text{and } \gamma \notin S^0(x), \\ \mathbf{1}\left[\mathrm{cv}\displaystyle\int \tilde{p}(x, \gamma|\theta)\,d\Lambda(\theta) < \int h_\theta(x) p(x|\theta)\,dF(\theta)\right] & \\ & \text{otherwise}, \end{cases} \tag{14}$$

satisfies $\iint \varphi_0(\gamma, x)\tilde{p}(x, \gamma|\theta)\,d\gamma\,d\nu(x) \leq \alpha$ *for all* $\theta \in \Theta$, *and* $\mathrm{cv}(\iiint \varphi_0(\gamma, x) \times \tilde{p}(x, \gamma|\theta)\,d\gamma\,d\nu(x)\,d\Lambda(\theta) - \alpha) = 0$.

Then the set $\varphi_0^*(\gamma, x^*) = \varphi_0(\hat{g}(U(x^*)^{-1}, \gamma), T(x^*))$ *is* (i) *invariant*, (ii) *satisfies* $\varphi_0^*(\gamma, x^*) = 0$ *for* $\gamma \in S^0(x^*)$, *and* (iii) *is of level* $1-\alpha$. *Furthermore, for any other set* φ^* *with these three properties*,

$$\begin{aligned} &\int E_\theta\left[\int \left(1 - \varphi_0^*(\gamma, X^*)\right) d\gamma\right] dF(\theta) \\ &\quad \leq \int E_\theta\left[\int \left(1 - \varphi^*(\gamma, X^*)\right) d\gamma\right] dF(\theta). \end{aligned} \tag{15}$$

The assumptions in part (a) cover cases where invariance reduces the parameter space, but leaves the parameter of interest unaffected; the near unit root application below is such an example. The determination of the weighted average volume minimizing set then simply amounts to applying Theorem 2 to the problem of observing the maximal invariant $X = T(X^*)$ with density indexed by $\theta \in \Theta = \bar{T}(\Theta^*)$, and the extension of the resulting test φ_0 to values of $x^* \notin \mathcal{X}$ via $\varphi_0^*(\gamma, x^*) = \varphi_0(\hat{g}(U(x^*)^{-1}, \gamma), T(x^*)) = \varphi_0(\gamma, T(x^*))$. Part (b) deals with cases where the invariance affects the parameter of interest. The assumption (b.ii) is relevant for volume changing transformations, such as those arising under a scale invariance. The only unknown in the form of the weighted expected volume minimizing set is a least favorable distribution Λ on the reduced parameter space Θ and critical value cv, which, in contrast to Theorem 2, are no longer indexed by the parameter of interest γ. In either case, note that the measure F only needs to be specified on the reduced parameter space Θ.

EXAMPLE, CONTINUED: $\hat{g}(a,\gamma) = \gamma + a$, $Y = \Delta\lambda - X_L^*$, and $g_l(a) = h_\theta(x) = 1$ so that for $\gamma \notin S^0(x)$, $\varphi_0(\gamma, x)$ equals 1 if

$$\text{cv} \cdot \frac{(2\pi)^{-1} \int \exp\left[-\frac{1}{2}\begin{pmatrix} x-\Delta \\ \gamma-\Delta\lambda \end{pmatrix}' \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x-\Delta \\ \gamma-\Delta\lambda \end{pmatrix}\right] d\Lambda(\theta)}{(2\pi)^{-1/2} 2^{-1/2} \int \exp\left[-\frac{1}{2}(x-\Delta)^2/2\right] dF(\theta)} \leq 1$$

for endogenously determined cv and distribution Λ on $\theta = (0, \Delta, \lambda) \in \{0\} \times \mathbb{R}_+ \times [0, 1]$.

5. APPLICATIONS

In this section, we consider the following nonstandard inference problems: (i) inference about the largest autoregressive root near unity, (ii) instrumental variable regression with a single weak instrument, (iii) inference for a set-identified parameter where the bounds of the identified set are determined by two moment equalities. First, for each of these problems, we explore whether previously suggested 95% confidence sets are bet-proof. For all problems, this turns out not to be the case. As discussed in Section 2.3, we compute maximal weighted average expected winnings to gauge the degree of unreasonableness. Next, we determine the "augmented credible set" along the lines of Sections 3 and 4 (implementation details are presented in Müller and Norets (2016a)).

In all examples, the parameter space is not naturally compact, even after imposing invariance, which potentially complicates numerical implementation. At the same time, most nonstandard inference problems are close to an unrestricted Gaussian shift experiment for most of the parameter space. In particular, inference about the largest autoregressive root becomes "almost" a Gaussian shift experiment for large degrees of mean reversion, inference with a weak instrument becomes "almost" a standard problem unless the instrument is quantitatively weak, and inference close to the boundary of the identified set becomes close to an unrestricted one-sided Gaussian inference problem as the identified set becomes large (see Elliott, Müller, and Watson (2015) for a formal discussion).

In the computations presented below, we therefore focus on the substantively nonstandard part of the parameter space. We find that in the unit root and weak instrument example, the credible set S^0, computed from a fairly vague prior, quickly converges to the standard Gaussian shift confidence set as the mean reverting parameter and concentration parameters increase. Correspondingly, the coverage of S^0 quickly converges to its nominal level, so that augmentation of S^0 is only necessary over a small compact part of the parameter space. In the set-identified problem, this Bernstein–von Mises approximation does not hold, and the credible set substantially undercovers—see

Moon and Schorfheide (2012) for further discussion of this effect. But given the convergence to the one-sided Gaussian problem, it makes sense to switch to the standard confidence interval for all realizations that, with very high probability, stem from the standard part of the parameter space, just as in Elliott, Müller, and Watson (2015). This approach formally amounts to setting S^0 in Theorems 2 and 3 equal to the union of the credible set, and this additional exogenous switching set.

5.1. *Autoregressive Coefficient Near Unity*

As in Stock (1991), Andrews (1993), Hansen (1999), and Elliott and Stock (2001), among others, suppose we are interested in the largest autoregressive root ρ of a univariate time series y_t,

$$y_t = m + u_t, \quad (1-\rho L)\phi(L)u_t = \varepsilon_t, \quad t = 1, \dots, T,$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_{p-1} z^{p-1}$, $\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$, and u_0 is drawn from the unconditional distribution. 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 $\rho = \rho_T = 1 - \gamma/T$, so that equivalently, γ is the parameter of interest. Under $m = T^{1/2}\mu$, the appropriate limiting experiment under $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$ in the sense of LeCam involves observing $X^*(\cdot) = \mu + J(\cdot)$, where J is a stationary Ornstein–Uhlenbeck on the unit interval with mean reversion parameter γ.

This limit experiment is translation invariant. Thus, the discussion in Section 4 applies, with $\theta^* = (\gamma, \mu)$, $g(x^*, a) = x^* + a$, and $\hat{g}(\gamma, a) = \gamma$. One choice of maximal invariants are $\theta = T(\theta^*) = (\gamma, 0)$ and $X = T(X^*) = X^*(\cdot) - X^*(0)$, so that $X(s) = Z(e^{-\gamma s} - 1)/\sqrt{2\gamma} + \int_0^s \exp[-\gamma(s-r)]\,dW(r)$ with $W(\cdot)$ a standard Wiener process independent of $Z \sim \mathcal{N}(0,1)$. The density of X relative to the measure of a standard Wiener process is (cf. Elliott (1999))

$$p(x|\theta) = \sqrt{\frac{2}{2+\gamma}} \exp\left[-\frac{\gamma}{2}\big(x(1)^2 - 1\big) - \frac{\gamma^2}{2}\int_0^1 x(s)^2\,ds + \frac{\gamma\left(\gamma \int_0^1 x(s)\,ds + x(1)\right)^2}{2(2+\gamma)} \right] \tag{16}$$

for $\gamma \geq 0$. The same limiting problem may also be motivated using the framework in Müller (2011), without relying on an assumption of Gaussian u_t.

As pointed out by Mikusheva (2007), care must be taken to ensure that confidence sets for γ imply uniformly valid confidence sets for ρ outside the local-to-unity region. We focus on two such intervals that are routinely computed in applied work: (i) Andrews's (1993) level α confidence sets that are based on the $\alpha/2$ and $1-\alpha/2$ quantiles of the OLS estimator $\hat{\gamma} = -\frac{1}{2}(\bar{X}^*(1)^2 - \bar{X}^*(0)^2 - 1)/\int_0^1 \bar{X}^*(s)^2\,ds$ with $\bar{X}^*(s) = X^*(s) - \int_0^1 X^*(r)\,dr$; (ii) Hansen's (1999) equal-tailed inversions of tests of $H_0: \gamma = \gamma_0$ based on the t-statistic $\hat{t} = (\hat{\gamma} - \gamma_0)/(\int_0^1 \bar{X}^*(s)^2\,ds)^{-1/2}$. Note that both these sets are translation invariant.

5.1.1. *Quantifying Violations of Bet-Proofness*

Applying (13) in Lemma 3(ii) yields that the expected loss under invariant bets are those in the problem of observing X with density (16), indexed only by γ. We numerically approximate the optimal betting strategy of Lemma 1 and impose nonnegativeness on the grid $\gamma \in \{0, 0.25, \dots, 200\}$. To avoid artificial endpoint effects at the upper bound, we restrict the inspector to never object to an interval with upper endpoint larger than 200. Under that restriction, any betting strategy yields uniformly nonnegative expected winnings under $\gamma > 200$ (because any objection against a set that excludes the values $\gamma > 200$ is necessarily correct for true values $\gamma > 200$).

Figure 1 plots the expected winnings as a function of γ when the inspector seeks to maximize the weighted average of the expected winnings with a nearly flat Π with density proportional to $(100+\gamma)^{-1.1}\mathbf{1}[\gamma \geq 0]$. For small γ, both Andrews's (1993) and Hansen's (1999) intervals allow for substantial expected winnings, even under fairly unfavorable payoffs. The normalization by the realized information in Hansen's t-statistic approach seems to somewhat reduce the extent of expected winnings. Still, these results indicate that both intervals are not compelling descriptions of uncertainty about the value of γ.

5.1.2. *Bet-Proof Confidence Set*

We apply the approach discussed in Section 4. Specifically, we construct $S^0(x)$ as 95% HPD set of γ under prior Π given observation X, and extend

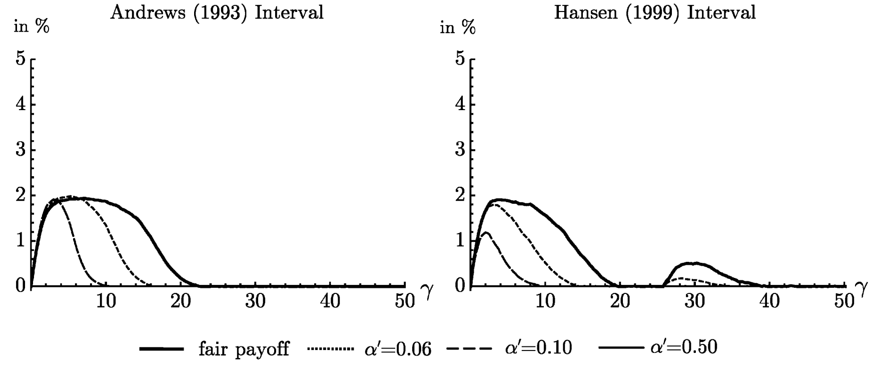

FIGURE 1.—Autoregressive coefficient near unity: expected winnings.

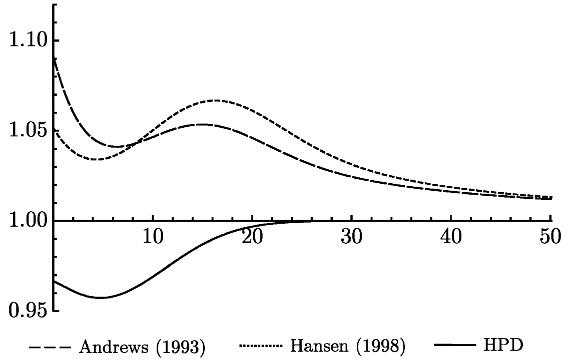

FIGURE 2.—Autoregressive coefficient near unity: expected length of sets relative to expected length of augmented credible set.

it to an invariant HPD set $S^0(x^*)$ via $S^0(x^*) = S^0(x^* - x^*(0))$. The assumptions of Theorem 3(a) hold, so we apply and numerically implement the construction of Theorem 2 with $F = \Pi$ to obtain the weighted average expected length minimizing augmentation of $S^0(x^*)$. Note that with γ the only parameter in the problem, Λ_γ in Theorem 2 is degenerate, so determining the set φ_0 defined there only requires computation of the critical values cv_γ that induce coverage. By Lemma 3(ii) and Lemma 2, the resulting confidence set φ_0^* is bet-proof against translation invariant bets. Without augmentation, the set S^0 undercovers at some γ. However, S^0 has (at least) nominal coverage for all $\gamma \geq 26$, so the augmented credible set differs from S^0 only in its inclusion of values of $\gamma \leq 26$ ($\text{cv}_\gamma = 0$ for $\gamma > 26$ in Theorem 2), which makes its numerical determination entirely straightforward.

Figure 2 plots the expected length of the Andrews (1993) and Hansen (1999) intervals, and of the HPD set S^0, relative to the expected length of this augmented credible set. For small γ, the HPD set S^0 is up to 3% shorter on average than the augmented credible set. At the same time, the augmented credible set is uniformly shorter in expectation than the Andrews (1993) and Hansen (1999) intervals, with a largest difference of 11% and 7%, respectively. As such, the augmented credible set seems a clearly preferable description of uncertainty about the degree of mean reversion of y_t.

5.2. *Weak Instruments*

A large body of work is dedicated to deriving inference methods that remain valid in the presence of weak instruments—see, for instance, Staiger and Stock (1997), Moreira (2003), and Andrews, Moreira, and Stock (2006, 2008). Following Chamberlain (2007), we showed in Müller and Norets (2016a) that in the case of a single endogenous variable and single instrument, the relevant asymptotic problem may be usefully reparameterized as

$$X^* = \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \rho \sin\phi \\ \rho\cos\phi \end{pmatrix}, I_2\right) \tag{17}$$

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 discussed 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 (17) (also see Chamberlain (2007) for related arguments). In particular, the groups of transformations on the parameter space, the sample space, and the parameter of interest space Γ are 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 $f(\theta) = 0$. Note that the AR confidence set is invariant with respect to g. Thus, after imposing invariance, Lemma 3 shows that the problem is effectively indexed only by the nuisance parameter $\rho \geq 0$.

5.2.1. *Quantifying Violations of Bet-Proofness*

As noted in Section 2.2, the AR interval cannot be bet-proof. Figure 3 quantifies its unreasonableness for ρ restricted to the grid $R = \{0, 0.2, 0.4, 0.6, \ldots, 8\}$. As a baseline, we specify the weight function Π of the inspector to be uniform on R (left panel). To study the sensitivity of the results to this choice, we also derive the envelope of the expected winnings, that is, for each value of $\rho \in R$, we set Π to a point mass at ρ, and report the expected winnings at that point (right panel).

Under considerably unfavorable payoffs $\alpha' \in \{0.1, 0.5\}$, the expected winnings in Figure 3 are indistinguishable from zero. Still, under fair payoffs, the 95% AR interval appears to be a very unreasonable description of uncertainty, since for ρ close to zero, the inspector can generate expected winnings very close to the maximum of 5%.

5.2.2. *Bet-Proof Confidence Set*

We construct the shortest invariant 95% credible set $S^0(x^*)$ as discussed below Lemma 3 under the prior Π with density proportional to $(100 + \rho)^{-1.1}\mathbf{1}[\rho \geq 0]$. We find that $S^0(x^*)$ undercovers under small ρ. We thus apply Theorem 3(b)

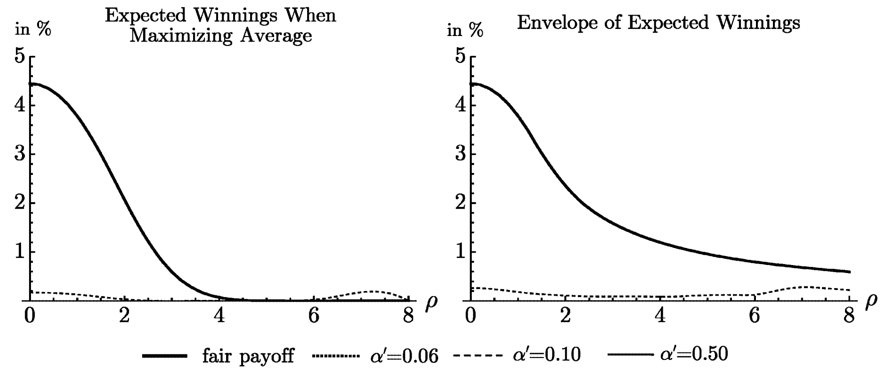

FIGURE 3.—Weak instruments: expected winnings.

with $F = \Pi$, and a numerical calculation reveals Λ in (14) to be a point mass at $\rho = 0$. The left panel of Figure 4 shows the boundary of the critical region of the null hypothesis $H_0 : \gamma = 0$ that is implied by $S^0(x^*)$, and by the augmented credible set. Points with $x_1^* = 0$ are always in the acceptance region of the augmented credible set. The confidence interval is constructed from these rejection regions via rotational invariance; see the right panel of Figure 4. Both the AR and the augmented credible intervals are equal to the parameter space for small realizations of $\|X^*\|$, and they also essentially coincide for large values of $\|X^*\|$. However, for most of the intermediate values of $\|X^*\|$, the augmented credible interval is considerably longer than the AR interval, for example, 20% longer at $\|X^*\| = 2.5$. In practice, AR intervals are usually reported only when there is no overwhelming evidence of a strong instrument, that is, when $\|X^*\|$ is

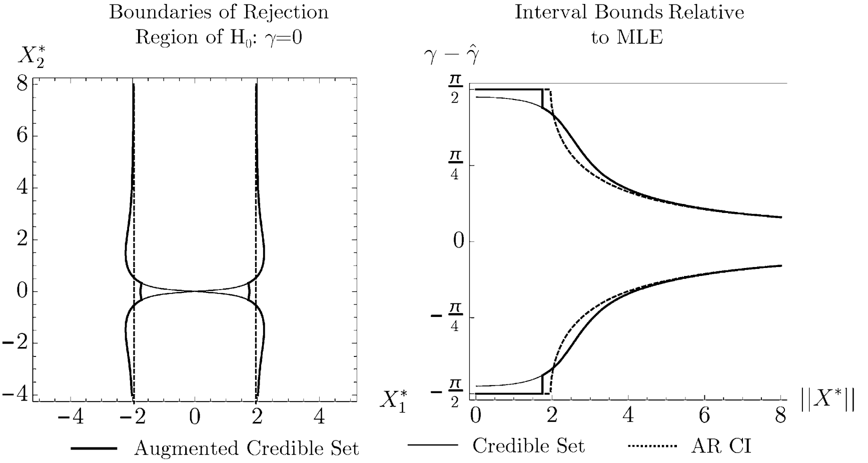

FIGURE 4.—Weak instruments: intervals.

not very large. But for such values, the augmented credible interval provides a more conservative and more reasonable description of parameter uncertainty.

5.3. *Imbens–Manski Problem*

We now return to the moment inequality problem from Imbens and Manski (2004), the running Example of Section 4 above. Imbens and Manski (2004) observed 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 suggested 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 (2015) derived 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 5 shows the Stoye and EMW intervals (we discuss the credible set and augmented credible set in Section 5.3.2 below).

5.3.1. *Quantifying Violations of Bet-Proofness*

The results are based on the grid $\Delta \in D = \{0, 0.2, \dots, 10\}$. It turns out that in this problem, the specification of Π plays a very minor role in the sense that the expected winnings under a Π uniform on Δ and with point mass at $\lambda = 1/2$ are numerically nearly indistinguishable from the envelope of expected winnings for the Stoye and EMW sets. Figure 6 plots these expected winnings as a function of Δ under $\lambda \in \{0, 1\}$ (left panel) and at $\lambda = 1/2$ (right panel). Recall that $\lambda = 1/2$ corresponds to the true parameter being in the middle of the

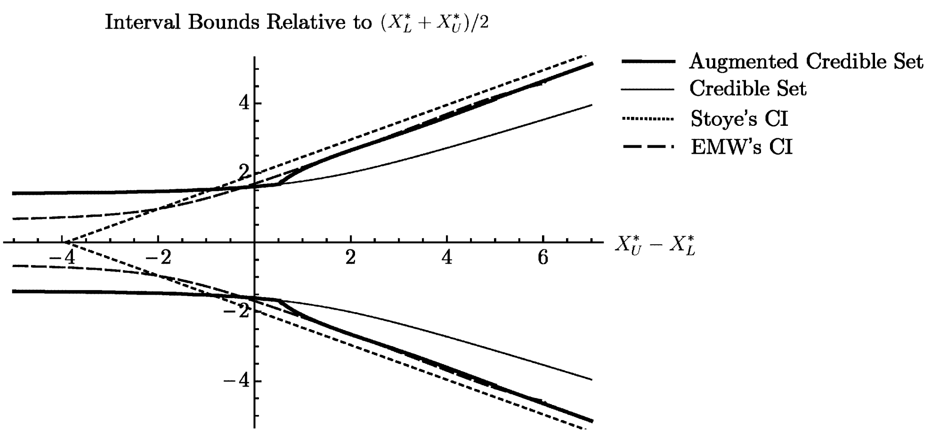

FIGURE 5.—Imbens–Manski problem: intervals.

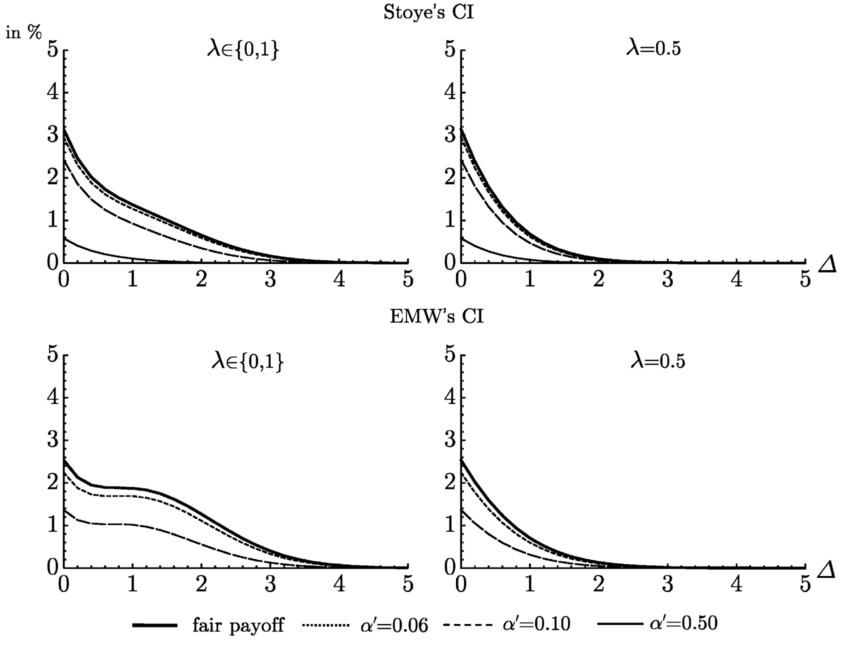

FIGURE 6.—Imbens–Manski problem: expected winnings as function of Δ and λ.

identified set; this configuration induces the largest coverage of the confidence intervals, which in turn reduces the expected winnings of the inspector.

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 unfavorable 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, also EMW's interval is far from bet-proof, with expected winnings that are, if anything, even larger than for Stoye's set, at least for $1-\alpha'$ close to the nominal level. The reason why EMW's intervals are unreasonable becomes readily apparent by inspection of Figure 5. For $X_U^* - X_L^* < -2$, EMW's interval has endpoints $\frac{1}{2}(X_U^* + X_L^*) \pm c$, where $c < 1$. But even under $\Delta = 0$, so that $\frac{1}{2}(X_U^* + X_L^*) \sim \mathcal{N}(\gamma, 1/2)$, the probability that this interval covers γ 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.

5.3.2. *Bet-Proof Confidence Set*

We construct the shortest invariant 95% credible set $S^0(x^*)$ as discussed below Lemma 3 under the prior Π proportional to $(100 + \Delta)^{-1.1}\mathbf{1}[\Delta \geq 0]$ on Δ, and conditional on Δ, λ is uniform on $[0, 1]$. The interval for γ under this prior has coverage above the nominal level for $\Delta = 0$, but it very substantially undercovers for larger Δ. As discussed above, the inference problem converges to a one-sided Gaussian shift experiment as $\Delta \to \infty$. We correspondingly impose in the application of Theorem 3 that for $X_U^* - X_L^* > 5$, $S^0(x^*)$ also contains the interval $[X_L^* - 1.645, X_U^* + 1.645]$, which guarantees that coverage of $S^0(x^*)$ converges to the nominal level as $\Delta \to \infty$. Setting $F = \Pi$, we numerically approximate Λ in (14) to determine the weighted expected length minimizing coverage inducing augmentation φ_0^* of this set. As can be seen from Figure 5, the resulting "augmented credible set" connects smoothly with the standard Gaussian shift interval at $X_U^* - X_L^* = 5$.

6. CONCLUSION

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

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

We provide remedies for this problem. On the one hand, we develop a numerical approach that quantifies the degree of "unreasonableness" of a given confidence set. This can serve as a criterion to choose among previously derived sets. On the other hand, we derive confidence sets that are reasonable by construction. Specifically, we suggest enlarging a credible set relative to a prespecified prior by some minimal amount to induce frequentist coverage. In combination, these results allow the determination of sets that credibly describe parameter uncertainty in nonstandard econometric problems.

APPENDIX: PROOFS AND AUXILIARY RESULTS

LEMMA 4: *For any confidence set φ of level $1 - \alpha$, $0 \leq W(\Pi) \leq \alpha$.*

PROOF:

$$
\begin{aligned}
&-(1-\alpha)R_\alpha(\varphi, b, \theta) \\
&\quad \leq \int_{\varphi(f(\theta),x)\geq\alpha} \bigl[\varphi\bigl(f(\theta), x\bigr) - \alpha\bigr] p(x|\theta)\, d\nu(x) \\
&\quad \leq \min\bigl\{(1-\alpha)P\bigl(\varphi\bigl(f(\theta), X\bigr) \geq \alpha|\theta\bigr),
\end{aligned}
$$

$$\alpha\big(1-P\big(\varphi\big(f(\theta),X\big)\geq\alpha|\theta\big)\big)\big\}$$
$$\leq\alpha(1-\alpha). \qquad \textit{Q.E.D.}$$

PROOF OF THEOREM 1: Note that for $b(x)=\mathbf{1}[x\notin X_0]$, the expected winnings can be written as $\int_{\mathcal{X}\setminus\mathcal{X}_0}(\varphi(f(\theta),x)-\alpha)p(x|\theta)\,d\nu(x)/(1-\alpha)$. By similarity,

$$0=\int_{\mathcal{X}_0}\big(\varphi\big(f(\theta),x\big)-\alpha\big)p(x|\theta)\,d\nu(x)+\int_{\mathcal{X}\setminus\mathcal{X}_0}\big(\varphi\big(f(\theta),x\big)-\alpha\big)p(x|\theta)\,d\nu(x). \tag{18}$$

Since $\varphi(f(\theta),x)=0$ on X_0 and $P(X_0|\theta)>0$, the first term on the right-hand side of (18) is strictly negative for $\alpha>0$ and the winnings are uniformly positive. *Q.E.D.*

PROOF OF LEMMA 1: Consider an alternative strategy $b\in B$ that delivers uniformly nonnegative winnings. By definition of $b^\star$,

$$\int\big[b^\star(x)-b(x)\big]\int\big[\varphi\big(f(\theta),x\big)-\alpha\big]p(x|\theta)\,d(\Pi+K)(\theta)\,d\nu(x)\geq 0.$$

It follows that

$$\begin{aligned}&\int\big[b^\star(x)-b(x)\big]\int\big[\varphi\big(f(\theta),x\big)-\alpha\big]p(x|\theta)\,d\Pi(\theta)\,d\nu(x)\\&\quad\geq-\iint b^\star(x)\big[\varphi\big(f(\theta),x\big)-\alpha\big]p(x|\theta)\,d\nu(x)\,dK(\theta)\\&\qquad+\iint b(x)\big[\varphi\big(f(\theta),x\big)-\alpha\big]p(x|\theta)\,d\nu(x)\,dK(\theta).\end{aligned}$$

The first expression on the right-hand side of this inequality is equal to zero by the definition of $b^\star(x)$. The second expression is nonnegative as b delivers uniformly nonnegative winnings. Thus, the winnings from $b^\star(x)$ are at least as large as the winnings from b. *Q.E.D.*

PROOF OF LEMMA 2: Note that $\varphi(\cdot,\cdot)$ being a superset of a $1-\alpha$ credible set for Π implies

$$\int\big(\alpha-\varphi\big(f(\theta),x\big)\big)p(x|\theta)\,d\Pi(\theta)\geq 0$$

for any x. Multiplication of this inequality by any $b(x) \geq 0$ and integration with respect to ν gives

$$\int \big(\alpha - \varphi\big(f(\theta), x\big)\big) b(x) p(x|\theta)\, d\nu(x)\, d\Pi(\theta) \geq 0.$$

Therefore, $R_\alpha(\varphi, b, \theta) \geq 0$ for some $\theta \in \Theta$. *Q.E.D.*

PROOF OF THEOREM 2: Without loss of generality, we can assume F to be a probability measure.

(a) Let $p_1(x) = \int p(x|\theta)\, dF(\theta)$ and $p_{0,\gamma}(x) = \int p(x|\theta)\, d\Lambda_\gamma(\theta)$. By definition of φ_0 and the fact that $\varphi_0(\gamma, x) = \phi(\gamma, x) = 0$ for $\gamma \in S^0(x)$,

$$\int \big(\varphi_0(\gamma, x) - \phi(\gamma, x)\big)\big(p_1(x) - \mathrm{cv}_\gamma\, p_{0,\gamma}(x)\big)\, d\nu(x) \geq 0. \tag{19}$$

Since $\phi(\gamma, \cdot)$ is of level α under $H_{0,\gamma}$, $\int \phi(\gamma, x) p(x|\theta)\, d\nu(x) \leq \alpha$ for all $\theta \in \Theta$ with $f(\theta) = \gamma$, it also has a rejection probability no larger than α under $p_{0,\gamma}$, $\int \phi(\gamma, x) p_{0,\gamma}\, d\nu(x) = \int [\int \phi(\gamma, x) p(x|\theta)\, d\nu(x)]\, d\Lambda_\gamma(\theta) \leq \alpha$. Thus $\mathrm{cv}_\gamma \int (\varphi_0(\gamma, x) - \phi(\gamma, x)) p_{0,\gamma}(x)\, d\nu(x) \geq 0$. Therefore, (19) implies that, for all γ,

$$\int \big(\varphi_0(\gamma, x) - \phi(\gamma, x)\big) p_1(x)\, d\nu(x) \geq 0$$

or, equivalently,

$$\begin{aligned} &\int \big(1 - \varphi_0(\gamma, x)\big) p_1(x)\, d\nu(x) \\ &\quad \leq \int \big(1 - \phi(\gamma, x)\big) p_1(x)\, d\nu(x) \end{aligned} \tag{20}$$

for all γ.

By Tonelli's theorem, we have

$$\begin{aligned} &\int\int \left[\int \big(1 - \phi(\gamma, x)\big)\, d\gamma\right] p(x|\theta)\, d\nu(x)\, dF(\theta) \\ &\quad = \int \left[\int\int \big(1 - \phi(\gamma, x)\big) p(x|\theta)\, dF(\theta)\, d\nu(x)\right] d\gamma \\ &\quad = \int \left[\int \big(1 - \phi(\gamma, x)\big) p_1(x)\, d\nu(x)\right] d\gamma \end{aligned} \tag{21}$$

and also

$$\int\!\!\int\left[\int\big(1-\varphi_0(\gamma,x)\big)\,d\gamma\right]p(x|\theta)\,d\nu(x)\,dF(\theta)$$
$$=\int\left[\int\big(1-\varphi_0(\gamma,x)\big)p_1(x)\,d\nu(x)\right]d\gamma, \tag{22}$$

where either integral might diverge. By (20), the integrand in (22) is weakly smaller than the one on the right-hand side of (21) for all γ, so that if the integral in (21) does not diverge, the desired inequality follows. If the integral does diverge, then there is nothing to prove.

(b) For a given γ, define $S=\{x:\gamma\in S^0(x)\}$, and let $p_1(x)=\int p(x|\theta)\,dF(\theta)$, as in the proof of part (a). Let Φ_S be the set of tests satisfying $\varphi(x)=0$ for $x\in S$. Suppose first that $\int_S p_1(x)\,d\nu(x)=1$ (so that any test $\varphi\in\Phi_S$ has power zero against p_1). Then $p_1(x)=0$ ν-almost surely, so one may choose Λ_γ arbitrarily, and set $\mathrm{cv}_\gamma=\kappa_\gamma=0$. So from now on, suppose $\int_S p_1(x)\,d\nu(x)<1$. Consider the testing problem

$$H_{0,\gamma}:f(\theta)=\gamma\quad\text{against}\quad H_1:\text{the density of } x \text{ is } p_1(x), \tag{23}$$

where tests φ are constrained to be in Φ_S. Define $\tilde{p}_1(x)=p_1(x)\mathbf{1}[x\notin S]/\omega$, where $\omega=1-\int_S p_1(x)\,d\nu(x)$, and consider the unconstrained testing problem

$$H_{0,\gamma}:f(\theta)=\gamma\quad\text{against}\quad H_1:\text{the density of } x \text{ is } \tilde{p}_1(x). \tag{24}$$

Suppose $\varphi_u(x)$ is a most powerful test for (24). Define $\varphi_c(x)=\varphi_u(x)\mathbf{1}[x\notin S]$, which is level α in (23). For any test $\varphi\in\Phi_S$ that has level α for $H_{0,\gamma}$ in (23) and (24),

$$\int\varphi p_1\,d\nu=\omega\int\varphi\tilde{p}_1\,d\nu\le\omega\int\varphi_u\tilde{p}_1\,d\nu=\int\varphi_c p_1\,d\nu.$$

Thus, $\varphi_c(x)$ is a most powerful test for (23), and it suffices to invoke previous results on the existence of a least favorable distribution in the unconstrained problem (24). Specifically, Wald's (1950) Theorem 3.14 implies the existence under compact $\Theta_{0,\gamma}=\{\theta:f(\theta)=\gamma\}$ (see the discussion in Lehmann (1952)). For non-compact Θ, the existence follows from Theorem 4 in Lehmann (1952) under the assumptions of the theorem. *Q.E.D.*

PROOF OF LEMMA 3: (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 (8). 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 (8), $\varphi^*(f(\theta^*),g(\bar{U}(\theta^*),X^*))=$

$\varphi^*(f(\bar{T}(\theta^*)), X^*)$. Replacing X^* by $g(U(X^*), T(X^*))$ in the latter expression, which can be done by (9), 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

$$[\alpha - \varphi^*(f(\theta^*), X^*)]b(X^*) \tag{25}$$

shows that the distribution of (25) 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 (9) and invariance. Thus, the expected loss can be computed as

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

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

(iii) First, let us show that the assumption $\hat{g}(U(g(a, x^*))^{-1} \circ a, \gamma) = \hat{g}(U(x^*)^{-1}, \gamma)$ follows from the uniqueness of the index a for the group action on x^* ($g(a_1, x^*) = g(a_2, x^*)$ for some x^* implies $a_1 = a_2$). By substituting $g(a, x^*)$ for x^* in (9), we obtain, for all $x^* \in X^*$,

$$g\big(a, x^*\big) = g\big(U\big(g\big(a, x^*\big)\big), T\big(g\big(a, x^*\big)\big)\big) = g\big(U\big(g\big(a, x^*\big)\big), T\big(x^*\big)\big),$$

where the last equality uses maximality of T. Furthermore, by applying the transformation $g(a, \cdot)$ to (9), we obtain

$$g\big(a, x^*\big) = g\big(a \circ U\big(x^*\big), T\big(x^*\big)\big).$$

Thus, we conclude $U(g(a, x^*)) = a \circ U(x^*)$, and $U(g(a, x^*))^{-1} \circ a = U(x^*)^{-1}$, which implies the desired result.

Now for the invariance of ψ^* when $\hat{g}(U(g(a, x^*))^{-1} \circ a, \gamma) = \hat{g}(U(x^*)^{-1}, \gamma)$,

$$\begin{aligned}
\psi^*\big(\hat{g}(a, \gamma), g\big(a, x^*\big)\big) &= \psi\big(\hat{g}\big(U\big(g\big(a, x^*\big)\big)^{-1}, \hat{g}(a, \gamma)\big), T\big(g\big(a, x^*\big)\big)\big) \\
&= \psi\big(\hat{g}\big(U\big(g\big(a, x^*\big)\big)^{-1} \circ a, \gamma\big), T\big(x^*\big)\big) \\
&= \psi\big(\hat{g}\big(U\big(x^*\big)^{-1}, \gamma\big), T\big(x^*\big)\big) = \psi^*\big(\gamma, x^*\big),
\end{aligned}$$

as was to be shown, where the second equality uses the maximality of T.

For the second claim, using (9) with $x^* = T(x^*) = x$ and the maximality of T, we obtain $x = T(x) = g(U(x), T(x)) = g(U(x), x)$. Thus, for all

$x = T(x^*) \in \mathcal{X}$,

$$\begin{aligned}\psi^*(\gamma, x) &= \psi^*\big(\hat{g}\big(U(x), \gamma\big), g\big(U(x), T(x)\big)\big) \\ &= \psi^*\big(\hat{g}\big(U(x), \gamma\big), x\big) \\ &= \psi\big(\hat{g}\big(U(x)^{-1} \circ U(x), \gamma\big), T(x)\big) \\ &= \psi(\gamma, x),\end{aligned}$$

where the first equality stems from the invariance of ψ^*, the third applies the definition of ψ^*, and the last uses the maximality of T. *Q.E.D.*

PROOF OF THEOREM 3: Without loss of generality, we can assume F to be a probability measure.

Claim (i) follows from Lemma 3(iii). For claim (ii), note that $\varphi_0(\gamma, x) = 0$ for $\gamma \in S^0(x)$ implies via $\varphi_0^*(\gamma, x^*) = \varphi_0(\hat{g}(U(x^*)^{-1}, \gamma), T(x^*))$ that $\varphi_0^*(\gamma, x^*) = 0$ if $\hat{g}(U(x^*)^{-1}, \gamma) \in S^0(T(x^*))$. Now by invariance of S^0, this latter condition equivalently becomes $\hat{g}(a \circ U(x^*)^{-1}, \gamma) \in S^0(g(a, T(x^*)))$ for any a, so setting $a = U(x^*)$ yields $\gamma \in S^0(g(U(x^*), T(x^*))) = S^0(x^*)$, where the last step applies (9). For claim (iii), for now only note that by Lemma 3(i), for any invariant set ψ^*, the coverage probability of $\psi^*(f(\theta^*), X^*)$ under θ^* is the same as the coverage probability $\psi^*(f(\theta), X^*)$ under $\theta = \bar{T}(\theta^*)$,

$$\int \psi^*\big(f\big(\theta^*\big), x^*\big) p^*\big(x^*|\theta^*\big)\, d\nu^*\big(x^*\big) = \int \psi^*\big(f(\theta), x^*\big) p^*\big(x^*|\theta\big)\, d\nu^*\big(x^*\big). \tag{26}$$

Let us first complete the proof for part (iii) and prove (15) under assumption (a). With $\hat{g}(a, \gamma) = \gamma$ for all $a \in A$, $\gamma \in \Gamma$, for any invariant set ψ^*,

$$\psi^*\big(\gamma, x^*\big) = \psi^*\big(\hat{g}\big(U\big(x^*\big)^{-1}, \gamma\big), T\big(x^*\big)\big) = \psi^*\big(\gamma, T\big(x^*\big)\big). \tag{27}$$

In particular, $\varphi_0^*(f(\theta), x^*) = \varphi_0(f(\theta), T(x^*))$ for all $\theta \in \Theta$, $x^* \in X^*$, so claim (iii) follows from (26) and the coverage property of φ_0.

By (26) and (27), the assumed coverage of φ^* implies that $\alpha \geq \int \varphi^*(f(\theta), x^*) p^*(x^*|\theta)\, d\nu^*(x^*) = \int \varphi^*(f(\theta), x) p(x|\theta)\, d\nu(x)$ for all $\theta = T(\theta^*) \in \Theta$. The test φ^* thus satisfies the assumption about φ in Theorem 2(a). Also, again applying (27), for any invariant set ψ^*

$$\begin{aligned}&\int \left[\int \big(1 - \psi^*\big(\gamma, x^*\big)\big)\, d\gamma\right] p^*\big(x^*|\theta\big)\, d\nu^*\big(x^*\big) \\ &\quad = \int \left[\int \big(1 - \psi^*\big(\gamma, T\big(x^*\big)\big)\big)\, d\gamma\right] p^*\big(x^*|\theta\big)\, d\nu^*\big(x^*\big) \\ &\quad = \int \left[\int \big(1 - \psi^*(\gamma, x)\big)\, d\gamma\right] p(x|\theta)\, d\nu(x).\end{aligned}$$

Thus, inequality (15) reduces to claim (6) in Theorem 2(a).

Under assumption (b), the coverage probability of any invariant set ψ^* under $\theta = T(\theta^*)$ can be written as

$$E_\theta\big[\psi^*\big(f(\theta), X^*\big)\big] = E_\theta\big[\psi^*(Y, X)\big] = \iint \psi^*(\gamma, x)\tilde{p}(x, \gamma|\theta)\, d\gamma\, d\nu(x), \tag{28}$$

where the first equality uses (9). In particular, it thus follows from the assumption $\iint \varphi_0(\gamma, x)\tilde{p}(x, \gamma|\theta)\, d\gamma\, d\nu(x) \le \alpha$ for all $\theta \in \Theta$ and $\varphi_0^*(\gamma, x) = \varphi_0(\gamma, x)$ from Lemma 3(iii) that φ_0^* is of level $1 - \alpha$ on Θ^*.

Also, the expected length of an invariant set ψ^* under θ can be written as follows:

$$\begin{aligned}
&E_\theta\bigg[\int \big(1 - \psi^*\big(\gamma, X^*\big)\big)\, d\gamma\bigg] \\
&\quad = E_\theta\bigg[\int \big(1 - \psi^*\big(\gamma, g\big(U\big(X^*\big), T\big(X^*\big)\big)\big)\big)\, d\gamma\bigg] \\
&\quad = E_\theta\bigg[g_l\big(U\big(X^*\big)\big) \int \big(1 - \psi^*\big(\gamma, T\big(X^*\big)\big)\big)\, d\gamma\bigg] \\
&\quad = E_\theta\bigg[E_\theta\bigg[g_l\big(U\big(X^*\big)\big) \int \big(1 - \psi^*\big(\gamma, T\big(X^*\big)\big)\big) \,\bigg|\, T\big(X^*\big)\bigg]\, d\gamma\bigg] \\
&\quad = E_\theta\bigg[h_\theta(X) \int \big(1 - \psi^*(\gamma, X)\big)\, d\gamma\bigg] \\
&\quad = \int h_\theta(x)\bigg(\int \big(1 - \psi^*(\gamma, x)\big)\, d\gamma\bigg) p(x|\theta)\, d\nu(x),
\end{aligned}$$

where the first equality applies (9), the second assumption (b.ii), and the third the law of iterated expectations. Using this expression and Tonelli's theorem, the F-weighted expected length of any invariant set is equal to

$$\begin{aligned}
&\int E_\theta\bigg[\int \big(1 - \psi^*\big(\gamma, X^*\big)\big)\, d\gamma\bigg] dF(\theta) \\
&\quad = \iint \bigg[\int \big(1 - \psi^*\big(\gamma, x^*\big)\big)\, d\gamma\bigg] p\big(x^*|\theta\big)\, d\nu^*\big(x^*\big)\, dF(\theta) \\
&\quad = \int \bigg(\int \big(1 - \psi^*(\gamma, x)\big)\, d\gamma\bigg) p_1(x)\, d\nu(x),
\end{aligned} \tag{29}$$

where $p_1(x) = \int h_\theta(x) p(x|\theta)\, dF(\theta)$.

Now by construction of φ_0, using $\varphi_0^*(\gamma, x) = \varphi_0(\gamma, x)$ from Lemma 3(iii),

$$\iint \big(\varphi_0^*(\gamma, x) - \varphi^*(\gamma, x)\big)\big(p_1(x) - \mathrm{cv}\, \tilde{p}_0(x, \gamma)\big)\, d\gamma\, d\nu(x) \geq 0, \tag{30}$$

where $\tilde{p}_0(x, y) = \int \tilde{p}(x, y|\theta)\, d\Lambda(\theta)$. Since φ^* is of level $1 - \alpha$, (28) implies that $\iint \varphi^*(\gamma, x)\tilde{p}_0(x, \gamma)\, d\gamma\, d\nu(x) \leq \alpha$. Thus, $\mathrm{cv}(\iint \varphi_0(\gamma, x)\tilde{p}_0(x, \gamma|\theta)\, d\gamma\, d\nu(x) - \alpha) = 0$ implies $\mathrm{cv} \iint (\varphi_0^*(\gamma, x) - \varphi^*(\gamma, x))\tilde{p}_0(x, \gamma)\, d\nu(x)\, d\gamma \geq 0$. Therefore, (30) yields

$$\iint \big(\varphi_0^*(\gamma, x) - \varphi^*(\gamma, x)\big) p_1(x)\, d\gamma\, d\nu(x) \geq 0$$

or, equivalently,

$$\begin{aligned} &\iint \big(1 - \varphi_0^*(\gamma, x)\big) p_1(x)\, d\gamma\, d\nu(x) \\ &\quad \leq \iint \big(1 - \varphi^*(\gamma, x)\big) p_1(x)\, d\gamma\, d\nu(x), \end{aligned}$$

which in light of (29) was to be shown. *Q.E.D.*

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. [2184,2188,2200]

ANDREWS, D. W. K. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165. [2197-2199]

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. [2188,2199]

——— (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression With Weak Instruments," *Journal of Econometrics*, 146, 241–254. [2199]

BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. New York: Springer. [2191]

BUEHLER, R. J. (1959): "Some Validity Criteria for Statistical Inferences," *The Annals of Mathematical Statistics*, 30 (4), 845–863. [2185-2187,2189,2190]

CHAMBERLAIN, G. (2007): "Decision Theory Applied to an Instrumental Variables Model," *Econometrica*, 75 (3), 609–652. [2199,2200]

CORNFIELD, J. (1969): "The Bayesian Outlook and Its Application," *Biometrics*, 25 (4), 617–657. [2186]

COX, D. R. (1958): "Some Problems Connected With Statistical Inference," *The Annals of Mathematical Statistics*, 29 (2), 357–372. [2185]

ELLIOTT, G. (1999): "Efficient Tests for a Unit Root When the Initial Observation Is Drawn From Its Unconditional Distribution," *International Economic Review*, 40, 767–783. [2197]

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. [2184]

ELLIOTT, G., AND J. H. STOCK (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics*, 103, 155–181. [2197]

ELLIOTT, G., U. K. MÜLLER, AND M. W. WATSON (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811. [2196,2197,2202]

FISHER, S. R. A. (1956): *Statistical Methods and Scientific Inference* (First Ed.). London and Edinburgh: Oliver & Boyd. [2185,2189]

GUGGENBERGER, P., F. KLEIBERGEN, S. MAVROEIDIS, AND L. CHEN (2012): "On the Asymptotic Sizes of Subset Anderson–Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression," *Econometrica*, 80 (6), 2649–2666. [2184]

HAHN, J., AND G. RIDDER (2009): "Partial Identification and Confidence Intervals," Working Paper, UCLA. [2191]

HANSEN, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607. [2197-2199]

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [2186,2191,2202]

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70 (5), 1781–1803. [2188]

LEHMANN, E. L. (1952): "On the Existence of Least Favorable Distributions," *The Annals of Mathematical Statistics*, 23, 408–416. [2190,2207]

MIKUSHEVA, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452. [2197]

——— (2010): "Robust Confidence Sets in the Presence of Weak Instruments," *Journal of Econometrics*, 157 (2), 236–247. [2188]

MOON, H. R., AND F. SCHORFHEIDE (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782. [2197]

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048. [2188,2199]

MÜLLER, U. K. (2011): "Efficient Tests Under a Weak Convergence Assumption," *Econometrica*, 79, 395–435. [2197]

MÜLLER, U. K., AND A. NORETS (2016a): "Supplement to 'Credibility of Confidence Sets in Nonstandard Econometric Problems'," *Econometrica Supplemental Material*, 84, http://dx.doi.org/10.3982/ECTA14023. [2186,2189,2196,2199]

——— (2016b): "Coverage Inducing Priors in Nonstandard Inference Problems," *Journal of the American Statistical Association* (forthcoming). [2190]

MÜLLER, U. K., AND M. W. WATSON (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81. [2184]

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression With $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125. [2185]

PIERCE, D. A. (1973): "On Some Difficulties in a Frequency Theory of Inference," *The Annals of Statistics*, 1 (2), 241–250. [2186,2189,2190]

PRATT, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56 (295), 549–567. [2190]

ROBINSON, G. K. (1977): "Conservative Statistical Inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39 (3), 381–386. [2185-2187,2189,2190]

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [2188,2199]

STOCK, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459. [2197]

——— (2000): "A Class of Tests for Integration and Cointegration," in *Cointegration, Causality, and Forecasting—A Festschrift in Honour of Clive W. J. Granger*, ed. by R. F. Engle and H. White. Oxford: Oxford University Press, 135–167. [2184]

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820. [2185]

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20 (4), 518–529. [2184]

STOYE, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315. [2184,2191,2202]

WALD, A. (1950): *Statistical Decision Functions*. Oxford, England: John Wiley & Sons. [2190,2207]

WALLACE, D. L. (1959): "Conditional Confidence Level Properties," *The Annals of Mathematical Statistics*, 30 (4), 864–876. [2186,2189]

WOUTERSEN, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," Working Paper, Johns Hopkins University. [2191]

WRIGHT, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business & Economic Statistics*, 18, 211–222. [2184]

Dept. of Economics, Princeton University, Princeton, NJ 08544, U.S.A.; umueller@princeton.edu

and

Dept. of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, U.S.A.; andriy_norets@brown.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received December, 2015; final revision received May, 2016.