Bootstrap Diagnostics for Irregular Estimators

Isaiah Andrews, *MIT and NBER*
Jesse M. Shapiro, *Harvard University and NBER**

January 2024

Abstract

Empirical researchers frequently rely on normal approximations in order to summarize and communicate uncertainty about their findings to their scientific audience. When such approximations are unreliable, they can lead the audience to make misguided decisions. We propose to measure the failure of the conventional normal approximation for a given estimator by the total variation distance between a bootstrap distribution and the normal distribution parameterized by the point estimate and standard error. For a wide class of decision problems and a class of uninformative priors, we show that a multiple of the total variation distance bounds the mistakes which result from relying on the conventional normal approximation. In a sample of recent empirical articles that use a bootstrap for inference, we find that the conventional normal approximation is often poor. We suggest and illustrate convenient alternative reports for such settings.

keywords: statistical communication, weak identification, Bayes bootstrap
JEL codes: C18, C44, D81

1 Introduction

Empirical researchers often summarize their findings using a point estimate and standard error. Such summaries can be justified by conventional asymptotic approximations. In re-

*E-mail: iandrews@mit.edu, jesse_shapiro@fas.harvard.edu. We acknowledge funding from the National Science Foundation (SES-1654234, SES-1949047) and the Semester Undergraduate Program for Economics Research at Harvard. We thank our dedicated research assistants for their contributions to this project, as well as the many authors who helped us to work with their code, data, and bootstrap replicates. We thank Anna Mikusheva, Jon Roth, and participants at the Princeton Day of Statistics for helpful comments.

cent decades economists have documented many reasons that such approximations can fail, including weak identification (Staiger and Stock 1997, Stock and Wright 2000), parameters near a boundary (Andrews 1999, Andrews 2001), highly nonlinear models (Dagenais and Dufour 1991, Andrews and Mikusheva 2016), non-differentiabilities (Hirano and Porter 2012, Fang and Santos 2019), and others. While diagnostic tools exist to measure the quality of conventional approximations in some cases (e.g., Stock and Yogo 2005), these diagnostic tools differ across settings, and in many important cases no widely accepted diagnostic tool exists.[1]

In this paper, we propose to diagnose the failure of conventional approximations using the total variation distance between a bootstrap distribution and the normal distribution parameterized by the point estimate and standard error. Our proposal is particularly easy to implement in the many settings where a bootstrap is used for inference, and we find that it implies substantial departures from normality in a large share of a sample of recent empirical articles.

We ground our proposal in a model of statistical communication following Andrews and Shapiro (2021). An analyst makes a report about a parameter of interest to an audience of agents, each of whom faces a decision problem with a bounded loss function. We consider agents whose priors belong to a density ratio neighborhood of a reference prior. For such agents, well-known results imply that reporting the reference posterior that results from updating the reference prior is as good—in terms of the loss it induces—as reporting the full data.

In many settings analysts do not report full posterior distributions but instead make do with simpler summaries. It is particularly common to report a point estimate and standard error, which can be interpreted as parameterizing a normal approximation to the posterior. We show that, across all agents in the audience, the increase in the expected loss from receiving any distributional summary—rather than the full data—is bounded by a multiple of the total variation distance between the reported distribution and the reference posterior. Hence, if the reference posterior is approximately normal, centered at the analyst's point estimate with standard deviation close to the analyst's standard error, it is essentially

[1]For the problem of detecting weak identification, for instance, the widely adopted rule of thumb which deems instruments weak when the first-stage $F-$statistic is smaller than 10 offers performance guarantees only for linear models with homoskedastic errors and a single endogenous regressor. Alternatives have been proposed by Montiel Olea and Pflueger (2013) for non-homoskedastic linear models, by Sanderson and Windmeijer (2016) and Lewis and Mertens (2022) for models with multiple endogenous variables, and by Andrews (2018) for nonlinear models.

without loss to summarize the data using the point estimate and standard error. If the reference posterior instead departs from the conventional approximation, summarizing using the point estimate and standard error creates scope for poor decision-making.

Evaluating the quality of the reported (e.g., normal) summary requires specifying a reference prior. We suggest a flexible class of Dirichlet process priors supported on the observed values of the data. In the limit as a reference prior in this class becomes uninformative, the corresponding reference posterior converges to the Bayes bootstrap distribution for the parameter of interest (Rubin 1981, Gasparini 1995). The increase in loss from reporting a point estimate and standard error, rather than the full data, is thus bounded by a multiple of the total variation distance between the implied normal distribution and the Bayes bootstrap distribution, plus a remainder that goes to zero as the prior becomes uninformative.

Under our assumptions, then, the analyst can evaluate the quality of the normal approximation (or any other proposed approximation) in a given application by comparison to the Bayes bootstrap distribution. The Bayes bootstrap distribution can be conveniently sampled in a wide range of applications. While the Bayes bootstrap distribution itself is often continuous, any finite sample of bootstrap replicates is necessarily discrete, and so will have total variation distance to any continuous (e.g., normal) distribution equal to one. We therefore propose to approximate the total variation distance using the signed Kolmogorov distance, which is a metric on distributions equal to the sum of the largest positive and negative distances between the two distributions' CDFs. The signed Kolmogorov distance is a lower bound on the total variation distance, and the two are equal for continuous distributions whose densities cross at most twice. In our applications, we show that this condition is typically not refuted, and we repeat our analysis using an alternative approach based on a kernel smoothing of the bootstrap replicates.

We outline a practical approach based on our findings. The analyst samples from a Bayes bootstrap distribution, and computes the signed Kolmogorov distance between the bootstrap sample and the normal distribution parameterized by the point estimate and standard error. If the distance is small then the analyst reports the point estimate and standard error. If the distance is large, the analyst selects an improved report, for instance a plot of the bootstrap distribution, or a table showing the mean and standard deviation of the normal distribution closest (in signed Kolmogorov distance) to the bootstrap distribution. An Online Appendix shows conditions for the asymptotic equivalence, in signed Kolmogorov distance, between the Bayes bootstrap and other bootstraps, and in our applications we find that conclusions based on the Bayes bootstrap distribution are often similar to those based on the nonparametric bootstrap distribution.

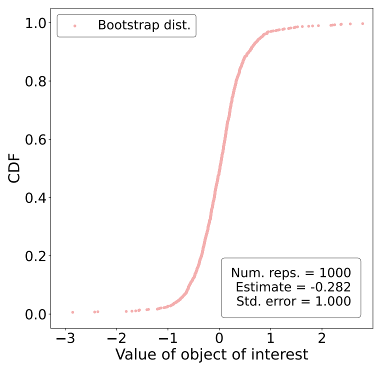
Figure 1 illustrates our proposed approach in a hypothetical example. The analyst samples from a bootstrap distribution (Panel A). The analyst computes the signed Kolmogorov distance between the bootstrap distribution and the normal distribution parameterized by the point estimate and standard error. The signed Kolmogorov distance can be read directly from a p-p plot (Panel B), which shows the relationship between the bootstrap CDF (y-axis) and the normal CDF (x-axis). In the hypothetical example, the median of the normal distribution is only at the 25th percentile of the bootstrap distribution, indicating that the bootstrap distribution is not centered on the point estimate. Moreover, 50 percent of the mass of the bootstrap distribution is contained between the median and 75th percentile of the normal distribution, indicating that the bootstrap distribution is less dispersed than the normal distribution. The analyst can simply report the distribution of the bootstrap replicates (Panel A), or, if a more compact report is desired, replace the default normal report (Panel C) with one that minimizes the signed Kolmogorov distance to the distribution of the bootstrap replicates (Panel D). The analyst can summarize the findings in a table (Panel E).

To facilitate adoption of our proposals, we have made available a Python package and web app called `BootstrapReport`. These tools take as inputs a point estimate, standard error, and set of bootstrap replicates, and return the diagnostic and alternative reports that we propose.

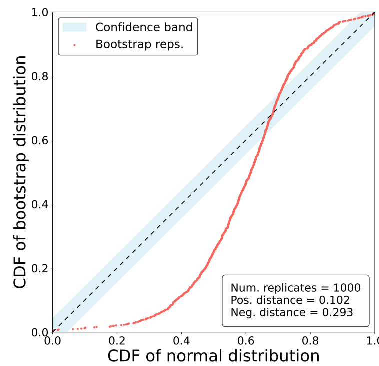
We apply our proposed approach to the universe of articles in the 2021 *American Economic Review* which bootstrap some object of interest, and for which we were able to recover bootstrap replicates. These articles cover a wide range of fields and methods. Our proposed approach applies readily to all of them. For many objects of interest, the conventional normal approximation is far from the bootstrap distribution. Even within a given article, the quality of the approximation can differ meaningfully across different objects of interest. The distance-minimizing normal approximation to the bootstrap distribution provides a superior visual fit and, we think, a better way to use two numbers to communicate the information in the data about the parameter of interest.

Figure 1: Illustration in a Hypothetical Example

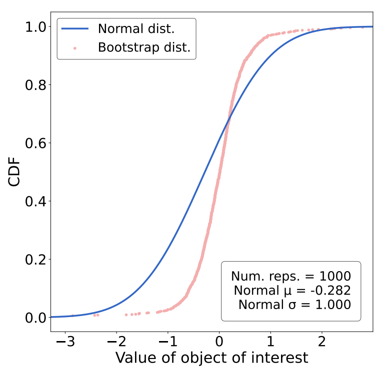
Panel A: CDF of bootstrap distribution



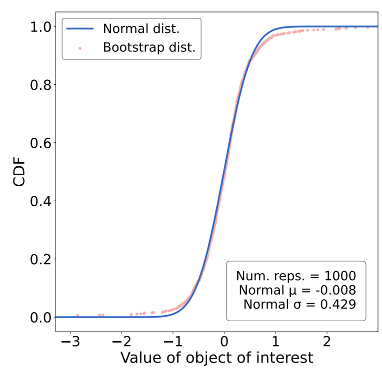
Panel B: Distance from default normal



Panel C: Default normal distribution



Panel D: Closer normal distribution



Panel E: Analyst's report

	Default	*Improved*
Mean	−0.282	*-0.008*
(Standard deviation)	(1.000)	*(0.429)*
[SK distance]	*[0.395]*	*[0.041]*

Notes: The figure illustrates our proposed approach in a hypothetical example. The analyst samples from a bootstrap distribution (Panel A). The analyst compares the distribution of the bootstrap replicates to the distribution of the default normal report, parameterized by the point estimate and bootstrap standard error (Panel B). If the distance between the two distributions is large, the analyst can replace the default normal report (Panel C) with the normal report closest, in signed Kolmogorov distance, to the bootstrap distribution (Panel D). The analyst can summarize the findings in a table (Panel E, with *italics* denoting values not reported in current practice).

Practically, our principal contribution is to propose an automated and practically appealing diagnostic for the quality of the conventional normal approximation in a wide range of settings. In addition to settings of correct specification and point identification, our approach directly covers settings of potential misspecification (by focusing on a pseudo-true parameter), and settings of partial identification (by focusing on the bounds of the identified set). The diagnostic we propose can be reported as a "third number" in addition to the point estimate and standard error commonly reported in empirical research (Athey and Imbens 2023).

Methodologically, our principal contribution is to cast the problem of estimation and uncertainty quantification for potentially irregular estimators as a communication problem. Rather than controlling frequentist criteria (e.g., size distortion) via an asymptotic approximation (e.g., Stock and Yogo 2005, Andrews 2018), our approach instead considers the potential for bad decision-making based on an approximation to a reference posterior from a finite sample of data.[2] Interpreting the usual point estimate and standard error as a description of a posterior aligns with how we (and, we think, others) often consume empirical research. Casting the problem as one of communication allows us to abstract from many details about the underlying model and estimator that might otherwise call for specialized treatments, and to obtain guarantees that hold in finite samples.

Empirically, we contribute what is to our knowledge the first census of failures of the conventional normal approximation in empirical economics that does not restrict attention to estimates of linear models, as well as the first large-scale census of bootstrap replicates. We expect our data and findings to be of interest to future researchers interested in issues of estimation and inference in economics.

While our approach neither targets nor offers frequentist guarantees, some of its elements can be given a frequentist interpretation. In particular, the signed Kolmogorov distance that we propose as a diagnostic converges to zero with the sample size when the bootstrap distribution is asymptotically normal. Asymptotic normality of the bootstrap distribution is, in turn, closely connected to asymptotic normality of the estimator (Mammen 1992). A large value of the signed Kolmogorov distance therefore suggests a failure of conventional

[2] Our focus on communicating a useful summary of data to an audience of decision-makers is related to the literatures on omniprediction (e.g., Gopalan et al. 2021) and sequential calibration (e.g., Noarov et al. 2023), though our setting, problem statement and, consequently, proposed solutions, are different.

asymptotic approximations in multiple respects, and so may be used analogously to other bootstrap diagnostics proposed in the literature (e.g., Beran 1997, Zhan 2018). In contrast, the frequentist interpretation of the improved reports we consider seems more delicate. It is well known that Bayesian and frequentist methods can disagree in irregular inference situations with, for instance, Bayesian credible sets differing substantially from frequentist confidence sets (Ham and Woutersen 2011, Moon and Schorfheide 2012, Kline and Tamer 2016, Kitagawa et al. 2020). We illustrate this difference in our applications, where we also discuss connections to existing proposals (e.g., Efron 1982b, Hall 1992) to improve the informativeness of bootstrap reports.

Section 2 introduces our communication framework and presents results bounding the increase in expected loss from reporting a distributional summary rather than the full data. Section 3 specializes these results to a particular class of reference priors connected to the Bayesian bootstrap. Section 4 lays out our proposed implementation, including the use of signed Kolmogorov distance to approximate total variation distance. Section 5 presents the findings from our census of articles in the 2021 *American Economic Review* that use the bootstrap. An Appendix contains proofs of all results stated in the main text. An Online Appendix contains additional theoretical results and additional findings from the bootstrap census.

2 Regret Bounds for Statistical Reports

In this section we lay out our abstract framework, which follows ideas in Andrews and Shapiro (2021). We use the framework to derive a bound on the regret from reporting an approximate posterior belief on the parameter of interest.

2.1 Model of Scientific Communication

Consider an analyst who observes data $X \in \mathcal{X}$ drawn according to some distribution $P \in \Delta(\mathcal{X})$, where $\Delta(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$. The analyst reports some function of the data $c(X)$ to an audience of agents, where $c: \mathcal{X} \to \mathcal{S}$ for a signal space $\mathcal{S}$. For a set of possible decisions $\mathcal{D}$ and a parameter space Θ, each agent is endowed with a loss function $L: \mathcal{D} \times \Theta \to [0,\lambda]$, so that an agent with loss function $L \in \mathcal{L}$ who makes decision

$d \in \mathcal{D}$ realizes loss $L(d,\theta) \in [0,\lambda]$ when the true parameter value is $\theta \in \Theta$. Each agent is further endowed with a prior $\pi \in \Pi \subseteq \Delta(\Theta \times \Delta(\mathcal{X}))$ which describes the relationship between the quantity of interest and the distribution of X. The family of priors Π is dominated by a σ-finite measure, and we interpret π as the density with respect to this measure. We identify each agent with their loss function and prior and write the audience as $\mathcal{L} \times \Pi$.[3]

For a given agent $(L,\pi) \in \mathcal{L} \times \Pi$, an optimal decision given data $X \in \mathcal{X}$ is one that minimizes the posterior expected loss $\mathrm{E}_\pi[L(d,\theta)|X]$ under their posterior distribution $\pi(\theta|X)$. We measure the expected cost, given data $X \in \mathcal{X}$, of taking some particular decision d, rather than an optimal decision, by the *posterior regret*

$$R(d;X,L,\pi) \equiv \mathrm{E}_\pi[L(d,\theta)|X] - \inf_{d' \in \mathcal{D}} \mathrm{E}_\pi[L(d',\theta)|X].$$

If the analyst sends report $c(X)$, the agent's optimal decision $d_{\pi(\theta|c(X))}$ minimizes the posterior expected loss under their posterior distribution $\pi(\theta|c(X))$:

$$d_{\pi(\theta|c(X))} \in \underset{d \in \mathcal{D}}{\operatorname{argmin}} \mathrm{E}_\pi[L(d,\theta)|c(X)].$$

If X cannot be recovered from $c(X)$, then $d_{\pi(\theta|c(X))}$ can differ from the agent's optimal decision given X, in which case the regret $R\left(d_{\pi(\theta|c(X))};X,L,\pi\right)$ is strictly positive.[4]

Averaging over possible realizations of the data X using the agent's prior, a given communication rule $c: \mathcal{X} \to \mathcal{S}$ induces the *communication regret*

$$R(c;L,\pi) = \mathrm{E}_\pi\left[R\left(d_{\pi(\theta|c(X))};X,L,\pi\right)\right].$$

The communication regret measures the increase in the agent's expected loss from observing only the report $c(X)$ rather than the full data X. It is a regret analogue of the communication risk defined in Andrews and Shapiro (2021), and so is also a rescaling of transparency as defined in Andrews, Gentzkow, and Shapiro (2020).

The timeline is summarized below:

[3] As the dimension of $\mathcal{D}$ is unrestricted, the assumption that all agents share a common decision space $\mathcal{D}$ is without loss of generality.

[4] When $\operatorname{argmin}_{d \in \mathcal{D}} \mathrm{E}_\pi[L(d,\theta)|c(X)]$ is not a singleton, we can take $d_{\pi(\theta|c(X))}$ to be any of its members. When $\operatorname{argmin}_{d \in \mathcal{D}} \mathrm{E}_\pi[L(d,\theta)|c(X)]$ is empty, the bounds we derive below continue to hold up to an arbitrarily small slack term. The proofs in the Appendix cover this more general case.

Timeline

- **Analyst** publicly commits to a reporting rule $c:\mathcal{X}\to\mathcal{S}$
- **Nature** determines $(\theta,P)\in(\Theta\times\Delta(\mathcal{X}))$
- **Analyst**
  - observes $X\sim P\in\Delta(\mathcal{X})$
  - publicly reports $c(X)\in\mathcal{S}$
- Each **agent** (L,π) in the **audience** $\mathcal{L}\times\Pi$
  - observes $c(X)$
  - chooses decision $d\in\mathcal{D}$
  - realizes loss $L(d,\theta)\in[0,\lambda]$

2.2 Class of Audience Priors

Without restrictions on the set of priors $\Pi\subseteq\Delta(\Theta\times\Delta(\mathcal{X}))$ it is difficult to obtain meaningful bounds on communication regret, since agents may have arbitrarily heterogeneous beliefs about how the parameter of interest relates to the distribution of the observed data. We instead focus on situations in which disagreement is of bounded magnitude and concerns only the parameter θ.

Assumption 1. *For some commonly known reference prior $\pi^*\in\Delta(\Theta\times\Delta(\mathcal{X}))$ and scalar $\rho\geq 1$, the class of audience priors Π satisfies $\Pi\subseteq\Pi(\rho,\pi^*)$ for*

$$\Pi(\rho,\pi^*)=\left\{\pi\in\Delta(\Theta\times\Delta(\mathcal{X})): \begin{array}{c}\textit{for all } P\in\Delta(\mathcal{X}),\ \theta,\ \theta'\in\Theta,\\ \pi(\theta,P)=\pi^*(P|\theta)\pi(\theta),\ \frac{\pi(\theta)}{\pi(\theta')}\leq\rho\cdot\frac{\pi^*(\theta)}{\pi^*(\theta')}\end{array}\right\}.$$

Assumption 1 states that agents in the audience agree on the conditional distribution of P given θ; that is, $\pi(P|\theta)=\pi^*(P|\theta)$ for all $\pi\in\Pi$. Agents may disagree on the marginal distribution $\pi(\theta)$, but such disagreement is bounded to lie within a density ratio neighborhood of some reference prior. To interpret this assumption we discuss two examples.

Example. (Parametric Model) Consider the case where agents agree on a parametric model for P with a finite-dimensional parameter θ, $P=P_\theta$. In this case $\pi(P|\theta)$ is a degenerate distribution at P_θ for all $\pi \in \Pi$, and the assumption that $\pi(P|\theta) = \pi^*(P|\theta)$ means that agents agree on the parametric model. The prior neighborhood $\Pi(\rho,\pi^*)$ also bounds the likelihood ratios of π relative to those of π^*, and the restrictiveness of this assumption depends on ρ. In the case where $\rho = 1$, $\Pi(1,\pi^*) = \{\pi^*\}$ and there is no prior disagreement. In the opposite case where $\rho \to \infty$, the requirement that $\pi \in \Pi(\rho,\pi^*)$ reduces to the requirement that π and π^* assign zero mass to the same sets (i.e., that π and π^* are mutually absolutely continuous).

Example. (Semiparametric Model) Consider the case where agents agree on a nonparametric model for P parameterized by ψ, so that $P = P_\psi$, where ψ may be infinite-dimensional. Agents' loss functions depend on ψ only via the finite-dimensional parameter $\theta = \theta(\psi)$. To ensure that $\pi(P|\theta) = \pi^*(P|\theta)$ in this context, it suffices that agents have homogeneous conditional priors on $\psi|\theta$, $\pi(\psi|\theta) = \pi^*(\psi|\theta)$. Kessler, Hoff, and Dunson (2015) advocate prior families of this form for non-parametric Bayesian estimation.

2.3 Bounds on Regret

Assumption 1 allows us to derive useful regret bounds. As a first step, we observe that reporting the reference posterior $\pi^*(\theta|X)$ allows all agents to achieve zero regret.

Proposition 1. *Under Assumption 1, if the analyst reports the reference posterior, $c^*(X) = \pi^*(\theta|X)$, then the posterior regret $R\left(d_{\pi(\theta|c^*(X))};X,L,\pi\right) = 0$ for all $X \in \mathcal{X}$ and all $(L,\pi) \in \mathcal{L} \times \Pi(\rho,\pi^*)$.*

All proofs are given in the Appendix.

Proposition 1 holds because, given $\pi^*(\theta|X)$, any agent with prior $\pi \in \Pi(\rho,\pi^*)$ can compute their posterior $\pi(\theta|X)$ by reweighting the reference posterior,

$$\pi(\theta|X) \propto \frac{\pi(\theta)}{\pi^*(\theta)} \pi^*(\theta|X).$$

Because $\pi(\theta|X)$ corresponds to the posterior belief that the agent would hold if they had observed the data directly, the agent can take the posterior-optimal decision following any report $\pi^*(\theta|X)$. Proposition 1 is a variant of an observation that has been made many times in the Bayesian statistics and econometrics literatures, including by Raiffa and

Schlaifer (1961), Hildreth (1963), and Geweke (1997). Motivated by this observation, we focus on communication rules that take $\mathcal{S} \subseteq \Delta(\Theta)$.

In many contexts the analyst does not report a reference posterior distribution, but instead reports another summary $c(X)$ which can be interpreted as a distribution on Θ. It is particularly common to report a frequentist point estimate and standard error for θ, a report which can be motivated by a conventional asymptotic approximation, suggesting a normal distribution on Θ. Such an approximation has the appeal of being easy to calculate in many situations, and easy to communicate as it is fully described by only two numbers.

Example. (Parametric Model, continued) Suppose that the analyst computes a maximum likelihood estimator $\hat{\theta}$ for θ. In sufficiently regular models, standard arguments imply that the maximum likelihood estimator will be approximately normally distributed in large samples, $\hat{\theta}(X) \approx^d N(\theta,\Sigma)$, and also imply consistent estimators $\widehat{\Sigma}(X)$ for Σ. If the analyst reports $\left(\hat{\theta}(X),\widehat{\Sigma}(X)\right)$, this may also be interpreted as reporting $c(X) = N\left(\hat{\theta}(X),\widehat{\Sigma}(X)\right) \in \Delta(\Theta)$.

One would hope that if the analyst's report is "close" to the reference posterior, agents will still be able to achieve low regret. Our main result shows that this is the case. To state this result, let $\mathcal{S}(\pi^*) \subseteq \Delta(\Theta)$ denote the set of distributions on Θ that are absolutely continuous with respect to the reference prior π^*, and so can be interpreted as approximate posteriors. Define the naive posterior $\hat{\pi}(\theta|c(X))$ under a prior $\pi \in \Pi$ as the distribution obtained by acting as if $c(X)$ were the reference posterior,

$$\hat{\pi}(\theta|c(X)) \propto \frac{\pi(\theta)}{\pi^*(\theta)} c(\theta|X),$$

where we write $c(\theta|X)$ for the density of $c(X)$ at θ. For $d_{\hat{\pi}(\theta|c(X))}$ the corresponding optimal decision, we have the following bound on the posterior regret.

Theorem 1. *Under Assumption 1, for any agent* $(L,\pi) \in \mathcal{L} \times \Pi(\rho,\pi^*)$*, any* $c : \mathcal{X} \to \mathcal{S}(\pi^*)$*, and any* $X \in \mathcal{X}$*,*

$$0 \leq R\left(d_{\hat{\pi}(\theta|c(X))};X,L,\pi\right) \leq (4\rho\lambda) TV(c(X),\pi^*(\theta|X)),$$

for $TV(\mu,\nu)$ *the total variation distance between measures* μ,ν*.*

For an agent that takes the analyst's report as a literal description of the reference posterior, Theorem 1 states that the agent's posterior regret is bounded above by a constant multiple of the total variation distance between the report and the reference posterior. The constant is proportional to the bound λ on the value of an agent's loss. The constant is also proportional to the size ρ of the density ratio neighborhood implied by Assumption 1.

Example. (Common Priors) Suppose that $\rho=1$ so that all agents share the reference prior. It is then appealing for the analyst to report the reference posterior. In modern applications of Bayesian statistics, exact computation of the posterior is typically infeasible. Theorem 1 establishes a sense in which agents in the audience can reasonably base decisions on a reported numerical approximation $c(X)$ to the reference posterior $\pi^*(\theta|X)$, provided that the two distributions are close in total variation.

For an agent that updates via Bayes rule based on the analyst's report, an analogous bound applies to the agent's communication regret.

Corollary 1. *Under Assumption 1, for any agent* $(L,\pi)\in\mathcal{L}\times\Pi(\rho,\pi^*)$ *and any* $c:\mathcal{X}\to\mathcal{S}(\pi^*)$,

$$0\leq R(c;L,\pi)\leq(4\rho\lambda)E_\pi[TV(c(X),\pi^*(\theta|X))].$$

Corollary 1 follows directly from Theorem 1. Intuitively, treating the analyst's report as a literal description of the reference posterior can only increase an agent's communication regret, relative to updating via Bayes rule. Controlling total variation distance therefore provides (via Theorem 1) a route to reducing a bound on the posterior regret for an agent that takes the analyst's report literally, and (via Corollary 1), a route to reducing a bound on the communication regret for a more sophisticated agent.

In models satisfying standard regularity conditions, Theorem 1 implies the asymptotic Bayes sufficiency of an efficient estimator and its estimated asymptotic variance for the audiences we study.[5]

Example. (Parametric Model, continued) Again suppose the analyst reports the maximum likelihood estimator and estimated asymptotic variance, $c(X)=N\left(\hat{\theta}(X),\widehat{\Sigma}(X)\right)$ for all $X\in\mathcal{X}$. In sufficiently regular models the Bernstein-von Mises Theorem (see, e.g., Theorem 12.1 of Ghosal and van der Vaart 2017) implies that for any continuous reference prior π^*

[5] We call a reporting rule "Bayes sufficient" with respect to an audience $\mathcal{L}\times\Pi$ if the reporting rule is as good, for any member of that audience, as reporting the full data X. This notion of sufficiency is a weakening of the concept of marginal sufficiency for θ with respect to Π as defined in Raiffa and Schlaifer (1961).

with a continuous, positive density, $TV(c(X),\pi^*(\theta|X))\to_p 0$ as the sample size grows. By Theorem 1 we therefore have that $R\left(d_{\hat{\pi}(\theta|c(X))};X,L,\pi\right)\to_p 0$ for any $\pi\in\Pi(\rho,\pi^*)$ and $L\in\mathcal{L}$ for finite $\rho\geq 1$, $\lambda\geq 0$. This establishes a sense in which a report consisting of the maximum likelihood estimator and its estimated variance is asymptotically Bayes sufficient for the audience $\mathcal{L}\times\Pi(\rho,\pi^*)$: for sufficiently large samples, reporting the maximum likelihood estimator together with the estimated variance allows each audience member to obtain nearly the same expected loss as reporting the full data.[6] A variety of arguments in the literature imply other senses in which the maximum likelihood estimator is asymptotically sufficient (see, e.g., Efron 1982a and Le Cam and Yang 2000).

Example. (Semiparametric Model, continued) Asymptotic sufficiency of efficient estimates is not limited to low-dimensional parametric models. Let $\hat{\theta}(X)$ be an asymptotically regular estimator of θ which achieves the semiparametric efficiency bound, and let $\widehat{\Sigma}(X)$ be an estimator for its asymptotic variance (see Chapter 25 of van der Vaart 1998). Then similar results (and caveats) obtain to those in the parametric case, provided that the model P_ψ and the reference prior π^* satisfy the conditions for a semiparametric Bernstein-von Mises Theorem (see, e.g., Theorem 12.8 of Ghosal and van der Vaart 2017).[7]

Theoretically, these examples provide a potential justification for some common statistical reports, and for this reason may be of independent interest. Practically, economists may not wish to rely on asymptotic approximations of the kind that we invoke in the examples, as these approximations have been found to break down in important applications. Our results suggest gauging the quality of an asymptotic approximation (or any other approximation) by measuring the total variation distance between the approximating distribution and a reference posterior. Operationalizing this suggestion requires specifying a choice of reference prior, which is the task we turn to in the next section.

[6] This is not simply a consequence of consistency of $\hat{\theta}(X)$ for θ, since our setup implicitly allows that the loss function may depend on the sample size n. For example, for a decision space $\mathcal{D}=\Theta=\mathbb{R}$ the set of loss functions $\mathcal{L}$ includes $L_n(d,\theta)=\lambda\cdot 1\left\{|d-\theta|>n^{-\frac{1}{2}}\right\}$, which is concerned with errors of order $\sqrt{n}$. Consequently simply reporting a consistent estimator of θ, or even a $\sqrt{n}$-consistent estimator, does not by itself suffice to ensure that the regret converges to zero.

[7] These conditions do not require that every prior in the audience satisfies the Bernstein-von Mises Theorem. In particular, the Bernstein-von Mises Theorem requires that the marginal density of the reference prior $\pi^*(\theta)$ is continuous, whereas the audience $\Pi(\rho,\pi^*)$ may include priors with discontinuous marginal densities.

3 A Class of Uninformative Reference Priors

Calculating the regret bounds derived in Section 2 requires a choice of reference prior π^*. In this section, we propose a class of reference priors closely linked to the Bayes bootstrap (Rubin 1981). The Bayes bootstrap distribution corresponds to a posterior distribution under a noninformative prior on the distribution of the data. While this "Bayes bootstrap prior" is improper, and so cannot serve as a reference prior π^*, we show that a large class of proper priors deliver nearly the same posterior distribution. We propose to use priors in this class as default choices for the reference prior.

3.1 Sampling Model

To develop the results in this section we impose two substantive restrictions on our general framework, both of which are common in applications of the bootstrap. The first restriction is that the data $X=(X_1,...,X_n)$ consist of $n\in\mathbb{N}$ observations, each drawn independently from a distribution $P_0\in\Delta(\mathcal{X}_0)$ supported on the sample space $\mathcal{X}_0$.[8] Thus, we have that $X\in\mathcal{X}_0^n=\mathcal{X}$ is distributed according to $P=\times_{i=1}^n P_0\in\Delta(\mathcal{X}_0)^n\subset\Delta(\mathcal{X})$. The second restriction is that the parameter θ is a function of the distribution P_0 of an individual observation, so $\theta=\theta(P_0)$ where $\theta:\Delta(\mathcal{X}_0)\to\Theta$ is a commonly known mapping. Under these restrictions any prior $\pi\in\Pi$ is fully described by the implied marginal prior $\pi(P_0)$ on $\Delta(\mathcal{X}_0)$. We next discuss two examples.

Example. (GMM Estimands) The setting of this section is an instance of the semiparametric example discussed in Section 2, where we take $\psi=P_0$ and assume that the data are an i.i.d. sample from P_0. Our results therefore cover inference on semiparametric target parameters, such as the estimands in moment condition models. Suppose in particular that we have an auxiliary model which implies that, for moment conditions $\phi(X_i,\vartheta)$ which depend on the data X and a parameter ϑ, the moments have mean zero at the true (P_0,ϑ) pair, $\mathrm{E}_{P_0}[\phi(X_i,\vartheta)]=0$. It is common to estimate ϑ using the generalized method of moments (GMM, Hansen 1982). For a weighting matrix $W(P_0)$ that may depend on

[8] The assumption of i.i.d. sampling allows for situations in which the data are clustered, since we may take X_i to be the data for cluster i and n the number of clusters.

P_0, the GMM estimand (i.e., the population analogue of the GMM estimator) is

$$\theta(P_0)\in\underset{\vartheta\in\Theta}{\operatorname{argmin}}\mathrm{E}_{P_0}[\phi(X_i,\vartheta)]'W(P_0)\mathrm{E}_{P_0}[\phi(X_i,\vartheta)]. \tag{1}$$

Note that $\theta(P_0)$ is well-defined irrespective of correct specification of the auxiliary model, that is, irrespective of whether there in fact exists a value of ϑ such that $\mathrm{E}_{P_0}[\phi(X_i,\vartheta)]=0$.[9] Therefore $\theta(P_0)$ can be interpreted as a pseudo-true parameter value.

Example. (Partially Identified Parameters) The restriction that $\theta=\theta(P_0)$ implies that θ is point-identified. However, the results developed in this section can also be applied to set-identified models under a restriction on the priors in the audience. Suppose that audience members have loss functions $\widetilde{L}:\mathcal{D}\times\Gamma\to[0,\lambda]$ that depend on a parameter $\gamma\in\Gamma$. If there exists some $\theta=\theta(P_0)$ which is sufficient for γ under each agent's prior, in the sense that $\pi(\gamma|\theta,P_0)=\pi(\gamma|\theta)$, then the analysis above applies directly, taking $L(d,\theta)=\mathrm{E}_{\pi(\gamma|\theta)}\left[\widetilde{L}(d,\gamma)\right]$. Importantly, this formulation allows the marginal prior $\pi(\gamma)$ to differ arbitrarily across agents. The formulation includes situations where, for example, $\gamma\in\mathbb{R}$ is a partially identified scalar parameter, $\theta\in\mathbb{R}$ is an upper bound on the identified set for γ (which implies that θ is point-identified by definition), and an agent's posterior belief $\pi(\gamma|\theta)$ on Γ, given θ, is a truncation of the agent's marginal prior $\pi(\gamma)$ to $(-\infty,\theta]$, where the prior $\pi(\gamma)$ may differ across agents.

3.2 Default Reference Priors and Bootstrap Regret Bounds

We make the following assumption on the reference prior π^*.

Assumption 2. *The reference prior π^* takes the form*

$$\pi^*(P_0)=DP(\alpha n,P_n^\omega)$$

where $DP(\cdot,\cdot)$ denotes a Dirichlet process, $\alpha>0$ is a scalar, $\omega\in\Delta(\{1,...,n\})$ are weights, and $P_n^\omega=\sum_{i=1}^n\omega_i\delta_{X_i}$ is the ω-weighted empirical distribution of the data, for δ_{X_i} the Dirac mass at X_i.

[9] When the argmin of (1) is non-unique, $\theta(P_0)$ applies some selection rule. Similarly, when the infimum is not achieved, we may define $\theta(P_0)$ arbitrarily.

Assumption 2 states that the reference prior is a Dirichlet process.[10] The parameter α controls the informativeness of the prior, with lower values corresponding to a less informative prior. The parameter P_n^ω controls the central tendency of the prior and assigns mass only to distributions supported on the observed data points $\{X_1,...,X_n\}$. The requirement to choose centering measures of the form P_n^ω becomes considerably less restrictive as the sample size n grows large.[11] We restrict attention to Dirichlet process priors for tractability, but they also have other attractive properties, such as posterior consistency for P_0.

Well-known conjugacy results for Dirichlet processes imply that for $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical distribution of the data, the reference posterior given X is

$$\pi^*(P_0|X) = DP\left((1+\alpha)n, \frac{\alpha}{\alpha+n}P_n^\omega + \frac{n}{\alpha+n}P_n\right).$$

The Bayes bootstrap posterior $\pi_0^B(P_0|X)$ corresponds to the limit of $\pi^*(P_0|X)$ as $\alpha \to 0$, i.e., $\pi_0^B(P_0|X) = DP(n,P_n)$. Results in Online Appendix B imply that for ω the uniform weights, the limit of the marginal prior $\pi^*(\theta(P_0))$ on Θ is uniform on a set of values supported by the observed data. For a given choice of ω and α, the reference priors specified in Assumption 2 are proper prior distributions and imply marginal priors $\pi^*(\theta(P_0))$ on Θ which can be sampled using a procedure analogous to the Bayes bootstrap. We illustrate this sampling in the Online Appendix in the context of our empirical applications.[12]

Importantly for our purposes, Assumption 2 implies that the reference and Bayes bootstrap posteriors are close in total variation for small α.

Proposition 2. *Under Assumption 2, for* $\zeta(\alpha,\omega,n) = TV\left(\pi^*(P_0|X), \pi_0^B(P_0|X)\right)$ *we have*

$$\lim_{\alpha\to 0} \sup_{\omega \in \Delta(\{1,...,n\})} \zeta(\alpha,\omega,n) = 0 \text{ for all } n.$$

Assumptions 1 and 2, together with Proposition 2, imply a bound on posterior regret

[10] See Chapter 4 of Ghosal and van der Vaart (2017) for a textbook treatment of Dirichlet processes.

[11] In particular, Appendix A proves that if $\mathcal{X}_0$ is a Polish space and P_0 has full support, then for every $Q_0 \in \Delta(\mathcal{X}_0)$ and almost every sequence of draws $\{X_1,X_2,...\}$ from P_0 there exists a sequence of weights ω_n such that $P_n^{\omega_n}$ converges weakly to Q_0 as $n \to \infty$.

[12] Chamberlain and Imbens (2003) propose to measure the informativeness of the marginal prior on $\theta(P_0)$ in the Bayes bootstrap by applying the Bayes bootstrap to subsamples of the data. Although we are not aware of a proper prior π^* on $\Delta(\mathcal{X}_0)$ for which this procedure samples the marginal prior, the procedure seems similar in spirit to the one we illustrate in the Online Appendix, including in its dependence on the observed data.

which approaches the total variation distance between $c(X)$ and the Bayes bootstrap posterior as α becomes small.

Theorem 2. *Under Assumptions 1 and 2, for any π^* with $\pi^*(P_0)=DP(\alpha n, P_n^\omega)$, any agent $(L,\pi)\in\mathcal{L}\times\Pi(\rho,\pi^*)$, any $c:\mathcal{X}\to\mathcal{S}(\pi^*)$, and any $X\in\mathcal{X}$,*

$$0\leq R\left(d_{\hat{\pi}(\theta|c(X))};X,L,\pi\right)\leq(4\rho\lambda)\left(TV\left(c(X),\pi^B(\theta|X)\right)+\zeta(\alpha,\omega,n)\right).$$

Theorem 2 states that, for priors in a density ratio neighborhood of the reference priors defined in Assumption 2, we can bound the posterior regret for the naive-optimal decision $d_{\hat{\pi}(\theta|c(X))}$ by a constant multiple of the total variation distance between $c(X)$ and the Bayes bootstrap posterior, plus a remainder that goes to zero with α.[13] In order to compute $\hat{\pi}(\theta|c(X))$, agents in the audience need to interpret the marginal reference prior $\pi^*(\theta)$. Theoretical results in Online Appendix B and numerical results in Online Appendix D establish that the marginal reference prior is uninformative in a variety of respects. In order to compute the bound in Theorem 2 the analyst needs the Bayes bootstrap posterior, which can be conveniently sampled via the Bayes bootstrap procedure. In the next section, we discuss the bootstrap procedure along with other important aspects of our proposed implementation.

4 Implementation with a Scalar Parameter

In this section we lay out a practical approach to applying our ideas in the case of a scalar parameter of interest $\theta\in\mathbb{R}$. We assume that the analyst begins with a default report $c_0:\mathcal{X}\to\mathcal{S}\subseteq\mathcal{S}(\pi^*)$, where the signal space $\mathcal{S}$ describes the richness with which the analyst will report to the reader. For example, if $\mathcal{S}$ is the set of univariate normal distributions, then any report $c(X)\in\mathcal{S}$ is fully described by two numbers (the mean and standard deviation). If $\mathcal{S}=\mathcal{S}(\pi^*)$ is the set of all distributions dominated by $\pi^*(\theta)$, then describing it may require a richer language, such as a plot. The leading example of a default report is a point estimate $\hat{\theta}$ and standard error $\hat{\sigma}_{\hat{\theta}}$, which we interpret as taking $c_0(X)=N\left(\hat{\theta},\hat{\sigma}^2_{\hat{\theta}}\right)$.

[13] There is no direct analogue of the bound in Corollary 1 in this setting because the support of the reference prior π^*, and consequently of all priors $\pi\in\Pi$, depends on the realization of the data X.

4.1 Sample from a Bootstrap Distribution

Given a default report, the first step is to create a sample of J bootstrap draws $\hat{\eta} = \frac{1}{J}\sum_j \delta_{\hat{\theta}_j} \in \Delta(\Theta)$ for the parameter of interest θ, for $\delta_{\hat{\theta}_j}$ the Dirac mass at $\hat{\theta}_j$. In the case of a weighted bootstrap, this is done by sampling J vectors of weights $\{W_1,...,W_J\}$ i.i.d. according to some distribution and, for each draw $W_j \in \Delta(\{1,...,n\})$, constructing the weighted empirical distribution $P_n^{W_j} = \sum_{i=1}^n W_{j,i}\delta_{X_i}$ and calculating the implied value $\hat{\theta}_j = \theta\left(P_n^{W_j}\right)$. The Bayes bootstrap takes $W_j \sim Dirichlet\left(\frac{1}{n},...,\frac{1}{n}\right)$, and in this case the distribution function of $\hat{\eta}$ converges uniformly to that of $\pi^B(\theta|X)$ as $J \to \infty$ by the Glivenko-Cantelli Theorem (van der Vaart 1998, Theorem 19.1).

We recommend using the Bayes bootstrap, but researchers may alternatively substitute other bootstrap schemes. Appendix C provides conditions for the asymptotic equivalence of two bootstrap distributions under the metric we consider. We read the conditions for the large-sample equivalence of the Bayes bootstrap and the nonparametric bootstrap (which takes $W_j \sim \frac{1}{n} Multinomial\left(\left(\frac{1}{n},...,\frac{1}{n}\right);n\right)$) as fairly mild, but read the conditions for the large-sample equivalence of the Bayes bootstrap with the parametric bootstrap, or other bootstrap schemes that rely on correct specification or exact normality of particular statistics, as more restrictive.

4.2 Approximate the Total Variation Distance

The next step is to approximate the total variation distance $TV\left(c_0(X),\pi^B(\theta|X)\right)$ between the default report and the reference posterior. We do this by way of a lower bound because convergence of $\hat{\eta}$ to $\pi^B(\theta|X)$ in terms of distribution functions does not imply convergence in total variation.[14]

Definition 1. The **signed Kolmogorov distance** between measures $\mu,\nu \in \Delta(\mathbb{R})$ is

$$SK(\mu,\nu) = \sup_{t\in\mathbb{R}}(F_\mu(t) - F_\nu(t))_+ + \sup_{t\in\mathbb{R}}(F_\nu(t) - F_\mu(t))_+$$

for $F_\mu(\cdot),F_\nu(\cdot)$ the CDFs associated with measures μ,ν and $(A)_+ = \max\{A,0\}$.

The signed Kolmogorov (SK) distance between two distributions is found by adding

[14] Indeed, when $\hat{\eta}$ is discrete while $c_0(X)$ is atomless, $TV(c_0(X),\hat{\eta}) = 1$.

together the largest positive and negative vertical distances between the distributions' CDFs.[15] This distance has several convenient properties.

Proposition 3. *The signed Kolmogorov distance is a metric on* $\Delta(\mathbb{R})$. *In addition, it is a lower bound on total variation distance,*

$$SK(\mu,\nu) \leq TV(\mu,\nu) \text{ for all } \mu,\nu \in \Delta(\mathbb{R}).$$

This inequality binds, with $SK(\mu,\nu) = TV(\mu,\nu)$, *when* μ *and* ν *are continuous with densities whose difference changes sign at most twice.*

We find that the data are consistent with the condition that the densities for μ and ν cross at most twice in a large majority of our applications.

We propose to calculate the signed Kolmogorov distance $SK(c_0(X),\hat{\eta})$ between the default report $c_0(X)$ and the bootstrap distribution $\hat{\eta}$. The signed Kolmogorov distance $SK(c_0(X),\hat{\eta})$ can be read off of a p-p plot of the distributions of $\hat{\eta}$ and $c_0(X)$, and is in this sense trivial to compute. When $\hat{\eta}$ is sampled from the Bayes bootstrap distribution $\pi^B(\theta|X)$, convergence of $F_{\hat{\eta}}$ to $F_{\pi^B(\theta|X)}$ as $J \to \infty$, together with the fact that signed Kolmogorov distance is a metric, implies that

$$SK(c_0(X),\hat{\eta}) \to_p SK\big(c_0(X),\pi^B(\theta|X)\big) \text{ as } J \to \infty.$$

Hence, as the number of bootstrap draws grows large, the analyst will consistently recover the SK distance (and, under a condition on the density, the total variation distance) between the default report and the Bayes bootstrap distribution. A more involved alternative to the SK distance, which we illustrate in the Online Appendix for our applications, is to compute $TV(c_0(X),\tilde{\eta})$ for $\tilde{\eta}$ a smoothed version of $\hat{\eta}$.

4.3 Improve the Report

As a final step, the analyst may improve their report. If $SK(c_0(X),\hat{\eta})$ is small, the analyst may simply report $c_0(X)$ and $SK(c_0(X),\hat{\eta})$. If $SK(c_0(X),\hat{\eta})$ is large, the analyst may

[15] The signed Kolmogorov distance is distinct from what Filion (2015) terms the "signed Kolmogorov-Smirnov test."

wish to improve upon $c_0(X)$. The bound in Theorem 2, together with the approximation result in Proposition 3, suggests selecting the report

$$c^*(X) \in \underset{s \in \mathcal{S}}{\operatorname{argmin}} SK(s,\hat{\eta}) \tag{2}$$

that minimizes the SK distance between the report and the bootstrap distribution. If $\mathcal{S}$ is the set of all normal distributions, $c^*(X)$ corresponds to the normal distribution closest, in SK distance, to the bootstrap distribution $\hat{\eta}$. If the minimized distance $SK(c^*(X),\hat{\eta})$ is small, the analyst may consider $c^*(X)$ a good summary, and report $c^*(X)$ alongside $SK(c^*(X),\hat{\eta})$. Alternatively, if $SK(c^*(X),\hat{\eta})$ remains large, the analyst may relax the constraints on the set of possible reports. In particular, if $\mathcal{S}$ is unrestricted then reporting $c^*(X)=\hat{\eta}$, for example by plotting the bootstrap distribution, is always a solution to (2), and seems a natural choice of report.

5 Bootstrap Distributions in the *American Economic Review*

To demonstrate applicability to a wide range of economic settings, we applied the approach in Section 4 to a census of papers in the 2021 *American Economic Review* that use a bootstrap. In this section we describe our procedures and findings.

5.1 Sample from a Bootstrap Distribution

We used a Google Scholar query to identify papers published in the *American Economic Review* in 2021 that use a bootstrap. For each paper, we identified the main objects of interest, which we define to be objects for which a quantitative or a qualitative description appears in the abstract or introduction. We focus on objects of interest for which the bootstrap is used for inference. We excluded from our census papers that are primarily methodological, papers that use the bootstrap only to calculate a p-value, or papers that use a bootstrap exclusively for objects reported in appendices. Appendix Table 1 lists the papers we include in our census along with the number of objects of interest in each paper that we include in our analysis.

The papers in the census cover a range of topics including public economics, labor economics, macroeconomics, behavioral economics, industrial organization, and development economics. The objects of interest include parameters describing technology, welfare calculations from a structural model, impulse responses, and transformations of regression coefficients.

For each paper, we attempted to reproduce the bootstrap replicates using the published replication code and data. When this was not feasible (e.g., due to confidential data), we contacted the authors to request the bootstrap replicates. We succeeded in obtaining bootstrap replicates covering 81 objects of interest across 14 articles, with only 1 article for which we were unable to obtain the replicates. For each object of interest, the bootstrap replicates define a distribution $\hat{\eta}$.

Across the papers in the census, the most common form of bootstrap is the nonparametric bootstrap (10 articles), followed by the parametric bootstrap (3 articles).[16] Only 1 article used the Bayes bootstrap. Our main theoretical results cover the Bayes bootstrap. Appendix C establishes conditions for the asymptotic equivalence of Bayes and other bootstraps in SK distance. Appendix Figure 1 shows a scatterplot of the SK distance between the default normal report and nonparametric bootstrap replicates (y-axis) versus the SK distance between the default normal report and Bayes bootstrap replicates (x-axis), across the objects of interest for which we are able to implement a Bayes bootstrap using the authors' original code and data. In no case do we statistically reject the equality of the two SK distances.

5.2 Approximate the Total Variation Distance

We take the default report to be $c_0(X) = N\left(\hat{\theta}, \hat{\sigma}^2_{\hat{\theta}}\right)$, where $\hat{\theta}$ is the reported estimate of the object of interest θ, and $\hat{\sigma}_{\hat{\theta}} = \sqrt{\mathrm{Var}_{\hat{\eta}}(\theta)}$ is the bootstrap standard error. All of the original articles in the census report a point estimate $\hat{\theta}$, and 10 out of 14 report a bootstrap standard error. The remaining 4 out of 14 articles report a bootstrap confidence interval. To better capture such situations, and because of other known limitations of the bootstrap standard error (e.g., Hahn and Liao 2021), Appendix Figure 2 shows results where we take the variance of the default normal to match the difference between the 97.5th and 2.5th percentiles of $\hat{\eta}$ rather than the standard deviation of $\hat{\eta}$.

[16] Our classification treats block bootstraps as nonparametric, and asymptotic bootstraps (which assume that some statistics are exactly Gaussian) as parametric. In all but 1 of the papers in the census, the authors' bootstrap procedure implicitly treats the data as i.i.d. across some observed units.

For each object of interest, we calculate the SK distance $SK(c_0(X),\hat{\eta})$ between the default report and the bootstrap distribution. Figure 2 shows the CDF of $SK(c_0(X),\hat{\eta})$ across the objects of interest in our census. The figure shows that the distance between $c_0(X)$ and $\hat{\eta}$ varies substantially across the objects of interest in our census.

To illustrate, Figure 3 shows a series of p-p plots comparing the bootstrap distribution $\hat{\eta}$ to the default report $c_0(X)$. Each plot reports the maximum positive and negative vertical distance between the CDF of the bootstrap replicates and of the default report. The sum of these distances is $SK(c_0(X),\hat{\eta})$. For each paper in our census, we depict two p-p plots, corresponding to the objects with the smallest and largest $SK(c_0(X),\hat{\eta})$ among the objects of interest in the paper. The figure shows that $SK(c_0(X),\hat{\eta})$ can differ meaningfully even across objects of interest reported in the same paper.

The SK distance is an approximation to the TV distance. The approximation is exact when the densities for the two measures cross at most twice, which implies that the difference in CDFs changes direction at most twice. Appendix Table 2 shows that this hypothesis is statistically consistent with the distribution of the replicates in all but 1 of the 81 of the objects of interest in our census. As an additional sensitivity analysis, Appendix Figure 3 shows the CDF of $TV(c_0(X),\tilde{\eta})$ across the objects of interest in our census, where $\tilde{\eta}$ is a smoothed distribution of bootstrap replicates.

5.3 Improve the Report

When $SK(c_0(X),\hat{\eta})$ is small, we recommend that researchers report $c_0(X)$ and $SK(c_0(X),\hat{\eta})$. When $SK(c_0(X),\hat{\eta})$ is large, researchers can alternatively report the mean and variance of the SK-distance-minimizing normal report, i.e., the distribution $c^*(X)$ that solves (2) for $\mathcal{S}$ the set of univariate normal distributions. Figure 2 compares the distributions of $SK(c^*(X),\hat{\eta})$ and $SK(c_0(X),\hat{\eta})$ across the objects of interest in our census. The share of objects with SK distance greater than 0.1 falls from 72 to 10 percent when we replace the default normal report $c_0(X)$ with the distance-minimizing normal report $c^*(X)$. In the remaining situations in which $SK(c^*(X),\hat{\eta})$ is nontrivial, we recommend that researchers report a richer summary than $c^*(X)$, such as the CDF of the bootstrap replicates.

To illustrate, Figure 4 shows a series of plots comparing the CDFs of $c_0(X)$, $c^*(X)$, and $\hat{\eta}$ across objects of interest in our census. Each row corresponds to a paper and focuses on the object of interest with the greatest value of $SK(c_0(X),\hat{\eta})$ among objects in that paper. The first column compares the distribution $\hat{\eta}$ of the bootstrap replicates to the distribution $c_0(X)$ of the default normal report. The second column compares the distribution $\hat{\eta}$ of the bootstrap replicates to the distribution $c^*(X)$ of the distance-minimizing normal report. Subjectively, we view $c^*(X)$ as a much better fit to the distribution of the replicates than is $c_0(X)$, especially in cases where $SK(c_0(X),\hat{\eta})$ is large.

The distributions implied by $c^*(X)$ and $c_0(X)$ often differ to a degree that seems economically meaningful. To illustrate, Figure 5 reports, across all objects of interest, the CDFs of the ratio of means $\mathrm{E}_{c^*(X)}[\theta]/\left|\mathrm{E}_{c_0(X)}[\theta]\right|$ and the ratio of standard deviations $\sqrt{\mathrm{Var}_{c^*(X)}(\theta)/\mathrm{Var}_{c_0(X)}(\theta)}$ between the distance-minimizing and default normal reports. In 33 percent of cases, the mean of the distance-minimizing normal report is more than 1.1 times or less than 0.9 times as large as the point estimate $\hat{\theta}$. In 42 percent of cases, the standard deviation of the distance-minimizing normal report is more than 1.1 times or less than 0.9 times as large as the bootstrap standard error $\hat{\sigma}_{\hat{\theta}}$.

Following Section 2.3, the interpretation of the report depends on the reference prior. Appendix Figure 4 shows the result of sampling from the reference prior and Bayes bootstrap posterior for a subset of the papers in our census.

5.4 Comparison to Bias-Corrected Confidence Intervals

Prior work in the frequentist tradition recommends reporting richer bootstrap information than the standard error. For example, Hall (1992, Appendix III) suggests reporting "confidence pictures" that consist of smoothed densities of the bootstrap replicates. A long tradition (e.g., Efron 1982b, Hahn and Liao 2021) advocates reporting confidence intervals based on percentile and other methods.

Our approach relates to these in recommending alternative reports to the bootstrap standard error, but differs in neither targeting nor achieving frequentist coverage. To illustrate the distinction, Appendix Figure 5 reproduces the plots in Figure 4, overlaying (in the left column of plots) the 95% bias-corrected bootstrap confidence interval (Efron 1982b), which is the default form of bootstrap confidence interval in Stata (2022), as well as (in the right column of plots) the 95% credible interval based on the improved normal report $c^*(X)$. In some cases, particularly those in which the bootstrap distribution is skewed, or is centered away from the reported point estimate, the bias-corrected confidence interval contains a fairly small share of the bootstrap replicates.[17] Such cases highlight a contrast between the goal of correct frequentist coverage and the goal of communicating the distribution of the replicates.

Appendix: Proofs

Proof of Proposition 1 By Bayes rule, $\pi(\theta|X) \propto \pi(X|\theta)\pi(\theta)$. Note, however, that $\pi(X|\theta) = \int P(X)\, d\pi(P|\theta)$, so Assumption 1 implies that $\pi(X|\theta) = \pi^*(X|\theta)$ for all $\pi \in \Pi(\rho,\pi^*)$ and all $X \in \mathcal{X}$. Hence, $\pi(\theta|X) \propto \pi^*(X|\theta)\pi(\theta) \propto \pi^*(\theta|X)\pi(\theta)/\pi^*(\theta)$, from which the conclusion follows. $\square$

We next state and prove three lemmas which we will use to prove Theorem 1.

Lemma 1. *Under Assumption 1, for any $\pi \in \Pi(\rho,\pi^*)$, any $c: \mathcal{X} \to \mathcal{S}(\pi^*)$, and any $X \in \mathcal{X}$,*

$$TV(\hat{\pi}(\theta|c(X)),\pi(\theta|X)) \leq (2\rho) TV(c(X),\pi^*(\theta|X)).$$

Proof of Lemma 1 Let μ denote a dominating measure for $\{\pi(\theta): \pi \in \Pi(\rho,\pi^*)\}$. For a given $\pi \in \Pi(\rho,\pi^*)$, define $r_\pi(\theta) = \pi(\theta)/\pi^*(\theta)$ as the Radon-Nikodym derivative of $\pi(\theta)$ with respect to $\pi^*(\theta)$, and note that by Assumption 1, $\sup_{\theta,\theta' \in \Theta} r_\pi(\theta)/r_\pi(\theta') \leq \rho$. By Assumption 1, the definition of $\hat{\pi}$, and Bayes rule, respectively, we have that

$$\hat{\pi}(\theta|X) = \frac{r_\pi(\theta)c(\theta|X)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}, \pi(\theta|X) = \frac{r_\pi(\theta)\pi^*(\theta|X)}{\int r_\pi(\theta')\pi^*(\theta'|X)d\mu(\theta')}.$$

Hence,

$$TV(\hat{\pi}(\theta|X),\pi(\theta|X)) = \frac{1}{2}\int \left| \frac{r_\pi(\theta)c(\theta|X)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')} - \frac{r_\pi(\theta)\pi^*(\theta|X)}{\int r_\pi(\theta')\pi^*(\theta'|X)d\mu(\theta')} \right| d\mu(\theta) \leq$$

$$\frac{1}{2}\int \frac{r_\pi(\theta)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')} |c(\theta|X) - \pi^*(\theta|X)| d\mu(\theta) +$$

$$\frac{1}{2}\int \left| \frac{1}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')} - \frac{1}{\int r_\pi(\theta')\pi^*(\theta'|X)d\mu(\theta')} \right| r_\pi(\theta)\pi^*(\theta|X)d\mu(\theta).$$

[17] The bias-corrected bootstrap confidence interval contains fewer than 80 percent of the bootstrap replicates for 10 percent of the objects of interest in the census. By contrast, a 95 percent centered credible interval based on the distance-minimizing normal report $c^*(X)$ never contains fewer than 80 percent of the bootstrap replicates.

Note that $\int c(\theta|X)d\mu(\theta)=1$ by definition, so $\frac{r_\pi(\theta)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}\leq\rho$ for all θ. This implies that

$$\frac{1}{2}\int\frac{r_\pi(\theta)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}|c(\theta|X)-\pi^*(\theta|X)|d\mu(\theta)\leq\rho\cdot TV(c(X),\pi^*(\theta|X)).$$

Note, further, that

$$\frac{1}{2}\int\left|\frac{1}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}-\frac{1}{\int r_\pi(\theta')\pi^*(\theta'|X)d\mu(\theta')}\right|r_\pi(\theta)\pi^*(\theta|X)d\mu(\theta)=$$

$$\frac{1}{2}\left|\frac{1}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}-\frac{1}{\int r_\pi(\theta')\pi^*(\theta'|X)d\mu(\theta')}\right|\int r_\pi(\theta)\pi^*(\theta|X)d\mu(\theta)=$$

$$\frac{1}{2}\left|\frac{\int r_\pi(\theta)\pi^*(\theta|X)d\mu(\theta)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}-1\right|\leq\frac{1}{2}\int\frac{r_\pi(\theta)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}|\pi^*(\theta|X)-c(\theta'|X)|d\mu(\theta)\leq$$
$$\rho\cdot TV(c(X),\pi^*(\theta|X)),$$

where we have again used that $\frac{r_\pi(\theta)}{\int r_\pi(\theta')c(\theta'|X)d\mu(\theta')}\leq\rho$ for all θ. $\square$

Lemma 2. *For any agent $(L,\pi)\in\mathcal{L}\times\Pi(\rho,\pi^*)$, any $c:\mathcal{X}\to\mathcal{S}(\pi^*)$, any $X\in\mathcal{X}$, any $\kappa>0$, and any $\hat{d}\in\mathcal{D}$ such that*

$$\mathrm{E}_{\hat{\pi}(\theta|c(X))}\left[L\left(\hat{d},\theta\right)\right]\leq\inf_{d\in\mathcal{D}}\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)]+\kappa,$$

we have that

$$R\left(\hat{d};X,L,\pi\right)\leq(2\lambda)TV(\hat{\pi}(\theta|c(X)),\pi(\theta|X))+\kappa.$$

Proof of Lemma 2 The definition of total variation distance implies that

$$\sup_{f:\Theta\to[0,\lambda]}\left\{\mathrm{E}_{\pi(\theta|X)}[f(\theta)]-\mathrm{E}_{\hat{\pi}(\theta|c(X))}[f(\theta)]\right\}=\lambda\cdot TV(\hat{\pi}(\theta|c(X)),\pi(\theta|X)).$$

Therefore for any $d\in\mathcal{D}$ we have that

$$\mathrm{E}_{\pi(\theta|X)}[L(d,\theta)]\leq\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)]+\lambda\cdot TV(\hat{\pi}(\theta|c(X)),\pi(\theta|X)).$$

It follows that, for any $\kappa^*>0$ and $d^*\in\mathcal{D}$ such that

$$\mathrm{E}_{\pi(\theta|c(X))}[L(d^*,\theta)]\leq\inf_{d\in\mathcal{D}}\mathrm{E}_{\pi(\theta|c(X))}[L(d,\theta)]+\kappa^*,$$

we must have

$$\mathrm{E}_{\pi(\theta|X)}\left[L\left(\hat{d},\theta\right)\right]-\mathrm{E}_{\pi(\theta|X)}[L(d^*,\theta)]\leq$$

$$\mathrm{E}_{\hat{\pi}(\theta|c(X))}\left[L\left(\hat{d},\theta\right)\right]-\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d^*,\theta)]+(2\lambda)TV(\hat{\pi}(\theta|c(X)),\pi(\theta|X)).$$

Now, by the definition of d^*,

$$\mathrm{E}_{\pi(\theta|X)}\left[L\left(\hat{d},\theta\right)\right]-\mathrm{E}_{\pi(\theta|X)}[L(d^*,\theta)]\geq R\left(\hat{d};X,L,\pi\right)-\kappa^*,$$

while by the definition of $\hat{d}$,

$$\mathrm{E}_{\hat{\pi}(\theta|c(X))}\left[L\left(\hat{d},\theta\right)\right]-\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d^*,\theta)]\leq\kappa.$$

Since we can set κ^* arbitrarily close we zero, we thus have that

$$R\left(\hat{d};X,L,\pi\right)\leq(2\lambda)TV(\hat{\pi}(\theta|c(X)),\pi(\theta|X))+\kappa,$$

as we aimed to show. $\square$

Lemma 3. *Under Assumption 1, for any agent $(L,\pi)\in\mathcal{L}\times\Pi(\rho,\pi^*)$, any $c:\mathcal{X}\to\mathcal{S}(\pi^*)$, any $X\in\mathcal{X}$, any $\kappa>0$, and any $\hat{d}\in\mathcal{D}$ such that*

$$\mathrm{E}_{\hat{\pi}(\theta|c(X))}\left[L\left(\hat{d},\theta\right)\right]\leq\inf_{d\in\mathcal{D}}\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)]+\kappa,$$

we have

$$R\left(\hat{d};X,L,\pi\right)\leq(4\rho\lambda)TV(c(X),\pi^*(\theta|X))+\kappa.$$

Proof of Lemma 3 The result is immediate from Lemmas 1 and 2. $\square$

Proof of Theorem 1 The result is immediate from Lemma 3, noting that if

$$\hat{d}=d_{\hat{\pi}(\theta|c(X))}\in\underset{d\in\mathcal{D}}{\operatorname{argmin}}\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)|c(X)]$$

then we can take $\kappa=0$. $\square$

Proof of Corollary 1 In the case where $\operatorname{argmin}_{d\in\mathcal{D}}\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)]$ is non-empty, the result follows from Theorem 1 and the law of iterated expectations. Here we instead prove the result using Lemma 3, which does not require the existence of an optimal decision. Specifically, note that by the law of iterated expectations,

$$R(c;L,\pi)=\mathrm{E}_\pi\left[\inf_{d\in\mathcal{D}}\mathrm{E}_{\pi(\theta|c(X))}[R(d;X,L,\pi)]\right]\leq \mathrm{E}_\pi[R(\delta(c(X));X,L,\pi)]$$

for all functions $\delta:\mathcal{S}(\pi^*)\to\mathcal{D}$. Note, however, that for any $\kappa>0$ we can choose δ such that

$$\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(\delta(c(X)),\theta)]\leq\inf_{d\in\mathcal{D}}\mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)]+\kappa$$

for all X. For such a choice of δ, however, Lemma 3 implies

$$\mathrm{E}_\pi[R(\delta(c(X));X,L,\pi)]\leq(4\rho\lambda)\mathrm{E}_\pi[TV(c(X),\pi^*(\theta|X))]+\kappa.$$

Since we can take κ arbitrarily close to zero, the result is immediate. $\square$

Proof of Proposition 2 By Pinsker's inequality,

$$TV\left(\pi^*(P_0|X),\pi_0^B(P_0|X)\right)\leq\sqrt{\frac{1}{2}KL(\pi_0^B(P_0|X),\pi^*(P_0|X))}.$$

As discussed in Section 3.2, $\pi^*(P_0|X)=DP\left((1+\alpha)n,\frac{\alpha}{n+\alpha}P_n^\omega+\frac{n}{n+\alpha}P_n\right)$ while $\pi_0^B(P_0|X)=DP(n,P_n)$, so the problem reduces to that of bounding the KL divergence between a $DP(n,P_n)$ and a $DP\left((1+\alpha)n,\frac{\alpha}{n+\alpha}P_n^\omega+\frac{n}{n+\alpha}P_n\right)$. Since both base measures are discrete and supported on $\{X_1,...,X_n\}$, this is simply the KL divergence between the finite-dimensional Dirichlet distributions $\mathrm{Dir}(\beta)$ and $\mathrm{Dir}\left(\tilde{\beta}\right)$, for $\beta_i=1$ for all i and $\tilde{\beta}_i=\alpha\omega_i+1$. Using the form of the Dirichlet density, however,

$$KL\left(\mathrm{Dir}(\beta),\mathrm{Dir}\left(\tilde{\beta}\right)\right)=$$

$$\log\left(\frac{\Gamma(n)}{\Gamma((1+\alpha)n)}\right)+\alpha n\cdot\psi(n)+\sum_{i=1}^n\left(\log\left(\frac{\Gamma(\alpha\omega_i+1)}{\Gamma(1)}\right)-\alpha\omega_i\cdot\psi(1)\right),$$

where Γ and ψ denote the gamma and digamma functions, respectively. Since $\psi(x)\equiv\frac{\partial}{\partial x}\log(\Gamma(x))$, however, the definition of the derivative implies that

$$\lim_{\alpha\to0}\frac{\log\left(\frac{\Gamma(n)}{\Gamma((1+\alpha)n)}\right)+\alpha n\cdot\psi(n)}{\alpha}=0 \text{ and } \lim_{\alpha\to0}\frac{\log\left(\frac{\Gamma(\alpha\omega_i+1)}{\Gamma(1)}\right)-\alpha\omega_i\cdot\psi(1)}{\alpha}=0.$$

Since the second equality holds for all $\omega_i \in \mathbb{R}$, it is immediate that

$$\lim_{\alpha \to 0} \sup_{\omega \in \Delta\{1,\dots,n\}} \frac{KL\left(\pi_0^B(P_0|X), \pi^*(P_0|X)\right)}{\alpha} = 0,$$

from which the result follows. □

Proof of Theorem 2 We again prove the stronger result that under Assumptions 1 and 2, for any π^* with $\pi^*(P_0) = DP(\alpha n, P_n^\omega)$, any agent $(L,\pi) \in \mathcal{L} \times \Pi(\rho, \pi^*)$, any $c: \mathcal{X} \to \mathcal{S}(\pi^*)$, any $X \in \mathcal{X}$, and any $\hat{d} \in \mathcal{D}$ such that

$$\mathrm{E}_{\hat{\pi}(\theta|c(X))}\left[L\left(\hat{d},\theta\right)\right] \leq \inf_{d \in \mathcal{D}} \mathrm{E}_{\hat{\pi}(\theta|c(X))}[L(d,\theta)] + \kappa,$$

we have that

$$0 \leq R\left(\hat{d}; X, L, \pi\right) \leq (4\rho\lambda)\left(TV\left(c(X), \pi^B(\theta|X)\right) + \zeta(\alpha,\omega,n)\right) + \kappa.$$

To see this, note that for any such $\hat{d}$, Lemma 3 implies that

$$R\left(\hat{d}; X, L, \pi\right) \leq (4\lambda\rho) TV(c(X), \pi^*(\theta|X)) + \kappa.$$

Since total variation distance is a metric, the triangle inequality implies that

$$TV(c(X), \pi^*(\theta|X)) \leq TV\left(c(X), \pi^B(\theta|X)\right) + TV\left(\pi^B(\theta|X), \pi^*(\theta|X)\right),$$

from which the result is immediate using the definition of $\zeta(\alpha,\omega,n)$. □

Proof of Proposition 3 To show that SK is a metric, note first that SK is symmetric and non-negative by construction, and that for any $\mu, \nu, \tau \in \Delta(\mathbb{R})$,

$$\sup_{t \in \mathbb{R}} (F_\mu(t) - F_\nu(t))_+ \leq \sup_{t \in \mathbb{R}} (F_\mu(t) - F_\tau(t))_+ + \sup_{t \in \mathbb{R}} (F_\tau(t) - F_\nu(t))_+,$$

so $SK(\mu,\nu) \leq SK(\mu,\tau) + SK(\tau,\nu)$, and SK satisfies the triangle inequality. Finally, note that $SK(\mu,\nu) = 0$ if and only if $F_\mu(t) = F_\nu(t)$ for all t.

To show that SK is a lower bound on TV, observe that for any $\mu, \nu \in \Delta(\mathbb{R})$,

$$SK(\mu,\nu) = \sup_{t_L, t_R \in \mathbb{R}} \left| \Pr_\mu(t \in [t_L, t_R]) - \Pr_\nu(t \in [t_L, t_R]) \right| \leq TV(\mu,\nu)$$

where the inequality follows from the definition of TV.

To show that SK and TV agree for continuous distributions μ and ν whose densities cross at most twice, note first that if the distributions are identical then SK and TV coincide trivially, so suppose the distributions are not identical. It follows that their densities must cross at least once, so suppose without loss of generality that the density of μ is initially above that of ν. First consider the case where the densities cross exactly once. Then there exists $t^* \in \mathbb{R}$ such that $\mu(A) - \nu(A) \geq 0$ for all $A \subseteq (-\infty, t^*]$ while $\mu(A) - \nu(A) \leq 0$ for all $A \subseteq [t^*, \infty)$. Note that in this case, we have by construction that $F_\mu(t) - F_\nu(t) \geq 0$ for all t. Consequently,

$$TV(\mu,\nu) = \mu((-\infty,t^*]) - \nu((-\infty,t^*]) = SK(\mu,\nu).$$

Next consider the case where the densities cross exactly twice. In this case there exist $t_L^*, t_U^* \in \mathbb{R}$ such that $\mu(A) - \nu(A) \geq 0$ for all $A \subseteq (-\infty, t_L^*] \cup [t_U^*, \infty)$ while $\mu(A) - \nu(A) \leq 0$ for all $A \subseteq [t_L^*, t_U^*]$. Note that in this case, $F_\mu(t) - F_\nu(t)$ is weakly increasing for $t \leq t_L^*$, weakly decreasing for $t \in [t_L^*, t_U^*]$, and again weakly increasing for $t \geq t_U^*$. Consequently,

$$SK(\mu,\nu) = (F_\mu(t_L^*) - F_\nu(t_L^*)) + (F_\nu(t_U^*) - F_\mu(t_U^*)) =$$

$$\nu([t_L^*, t_U^*]) - \mu([t_L^*, t_U^*]) = TV(\mu,\nu),$$

so $SK(\mu,\nu) = TV(\mu,\nu)$, as desired. $\square$

References

Andrews, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

——— (2001): "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, 100, 337–348.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2020): "Transparency in Structural Research," *Journal of Business and Economic Statistics*, 38, 711–722.

ANDREWS, I., AND A. MIKUSHEVA (2016): "A Geometric Approach to Nonlinear Econometric Models," *Econometrica*, 84, 1249–1264.

ANDREWS, I., AND J. M. SHAPIRO (2021): "A Model of Scientific Communication," *Econometrica*, 89, 2117–2142.

ATHEY, S., AND G. W. IMBENS (2023): "In Search of the Third Number: What to Report Beyond Point Estimates and Standard Errors," `https://youtu.be/i-K9a6UhtsE?si=-rBsjVgY2p3c_DLc`, March, Berkeley Initiative for Transparency in the Social Sciences Keynote Lecture; video accessed December 2023.

BERAN, R. (1997): "Diagnosing Bootstrap Success," *Annals of the Institute of Statistical Mathematics*, 49, 1–24.

BILLINGSLEY, P. (1995): *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, New York: J. Wiley and Sons, 3rd edition.

CHAMBERLAIN, G., AND G. W. IMBENS (2003): "Nonparametric Applications of Bayesian Inference," *Journal of Business and Economic Statistics*, 21, 12–18.

DAGENAIS, M. G., AND J.-M. DUFOUR (1991): "Invariance, Nonlinear Models, and Asymptotic Tests," *Econometrica*, 59, 1601–1615.

EFRON, B. (1982a): "Maximum Likelihood and Decision Theory," *Annals of Statistics*, 10, 340–356.

——— (1982b): *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics: Society for Industrial and Applied Mathematics.

FANG, Z., AND A. SANTOS (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86, 377–412.

FILION, G. J. (2015): "The Signed Kolmogorov-Smirnov Test: Why it Should Not be Used," *GigaScience*, 4.

GASPARINI, M. (1995): "Exact Multivariate Bayesian Bootstrap Distributions of Moments," *Annals of Statistics*, 23, 762–768.

GEWEKE, J. (1997): "Posterior Simulators in Econometrics," in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress* ed. by D. M. Kreps and K. F. Wallis, Volume 3: Cambridge: Cambridge University Press, 128–165.

GHOSAL, S., AND A. VAN DER VAART (2017): *Fundamentals of Nonparametric Bayesian Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

GOPALAN, P., A. T. KALAI, O. REINGOLD, V. SHARAN, AND U. WIEDER (2021): "Omnipredictors," *arXiv preprint arXiv:2109.05389*.

HAHN, J., AND Z. LIAO (2021): "Bootstrap Standard Error Estimates and Inference," *Econometrica*, 89, 1963–1977.

HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, New York, NY: Springer.

HAM, J., AND T. WOUTERSEN (2011): "Calculating Confidence Intervals for Continuous and Discontinuous Functions of Estimated Parameters," *IZA Institute of Labor Economics Discussion Paper*, 5816.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HILDRETH, C. (1963): "Bayesian Statisticians and Remote Clients," *Econometrica*, 31, 422–438.

HIRANO, K., AND J. R. PORTER (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769–1790.

KESSLER, D. C., P. D. HOFF, AND D. B. DUNSON (2015): "Marginally Specified Priors for Non-parametric Bayesian Estimation," *Journal of the Royal Statistical Society, Series B*, 77, 35–58.

KITAGAWA, T., J. L. MONTIEL OLEA, J. PAYNE, AND A. VELEZ (2020): "Posterior Distribution of Nondifferentiable Functions," *Journal of Econometrics*, 217, 161–175.

KLINE, B., AND E. TAMER (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7, 329–366.

Le Cam, L., and G. L. Yang (2000): *Asymptotics in Statistics*, Springer Series in Statistics, New York, NY: Springer.

Lewis, D., and K. Mertens (2022): "A Robust Test for Weak Instruments with Multiple Endogenous Regressors," *Federal Reserve Bank of New York Staff Reports*, No. 1020.

Mammen, E. (1992): "Bootstrap and Asymptotic Normality," in *When Does Bootstrap Work? Asymptotic Results and Simulations* ed. by E. Mammen, New York, NY: Springer, 9–30.

Montiel Olea, J. L., and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics*, 31, 358–369.

Moon, H. R., and F. Schorfheide (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782.

Noarov, G., R. Ramalingam, A. Roth, and S. Xie (2023): "High-Dimensional Prediction for Sequential Decision Making," *arXiv preprint arXiv:2310.17651.*

Raiffa, H., and R. Schlaifer (1961): *Applied Statistical Decision Theory*, Boston, MA: Harvard University Graduate School of Business Administration.

Rubin, D. B. (1981): "The Bayesian Bootstrap," *Annals of Statistics*, 9, 130–134.

Sanderson, E., and F. Windmeijer (2016): "A Weak Instrument F-test in Linear IV Models with Multiple Endogenous Variables," *Journal of Econometrics*, 190, 212–221.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stata Corporation (2022): "bootstrap postestimation," July, `https://web.archive.org/web/20220727215836/https://www.stata.com/manuals/rbootstrappostestimation.pdf`, Accessed on 07 August 2023.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

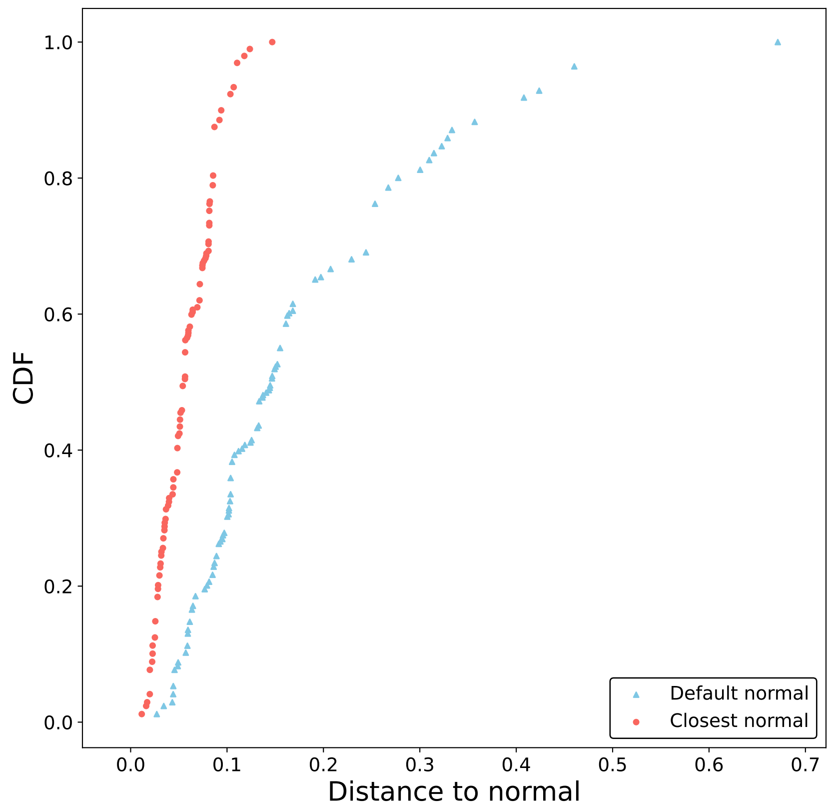
Stock, J. H., and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor* *of Thomas Rothenberg* ed. by D. W. Andrews and J. H. Stock, Cambridge: Cambridge University Press, 80–108.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes, with Applications to Statistics*, Springer Series in Statistics, New York, New York: Springer, 1st edition.

Zhan, Z. (2018): "Detecting Weak Identification by Bootstrap," https://web.archive.org/web/20230806184309/https://sites.google.com/site/zhaoguozhan/, Accessed on 07 August 2023.

Figure 2: Distribution of Signed Kolmogorov Distance to Normal Report



Notes: The plot is a weighted empirical CDF. The unit of analysis is an object of interest and, for each paper, we weight each object of interest by the inverse of the number of objects of interest associated with the paper. For each object of interest we calculate the signed Kolmogorov distance between the distribution of bootstrap replicates and the default normal report, whose mean is given by the point estimate and whose standard deviation is given by the bootstrap standard error. We also calculate the signed Kolmogorov distance between the distribution of the bootstrap replicates and the closest normal report, whose mean and standard deviation are chosen to minimize the signed Kolmogorov distance. The plot shows the weighted empirical CDF of each of these two distances across the objects of interest in our census.

Figure 3: Comparison of Bootstrap Distribution to Default Normal Report

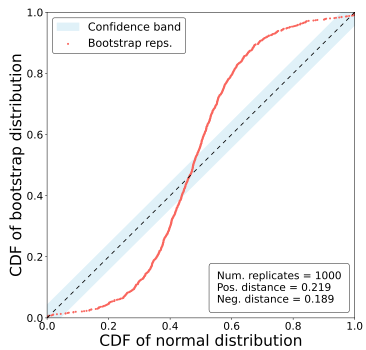
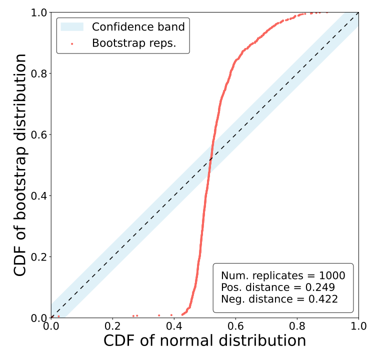
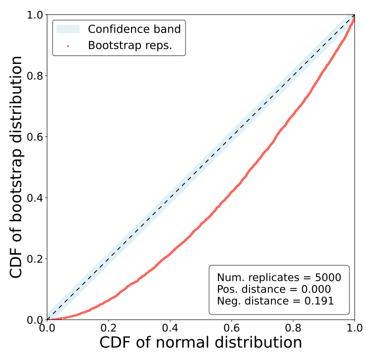
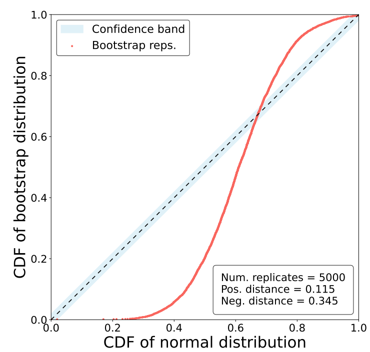
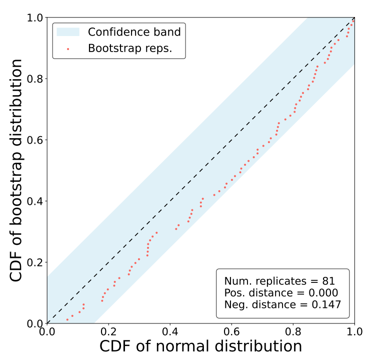
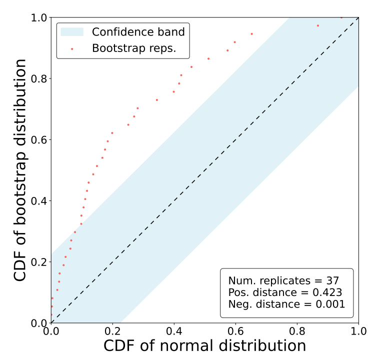
Paper	**Object with...** least SK distance	greatest SK distance
1		
2		
3		

Figure 3 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

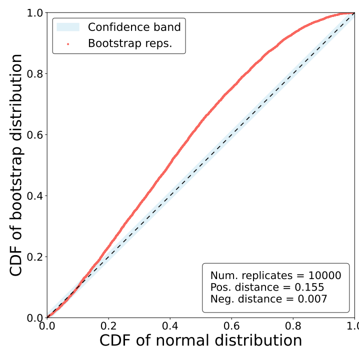
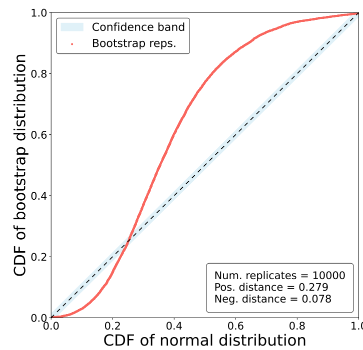
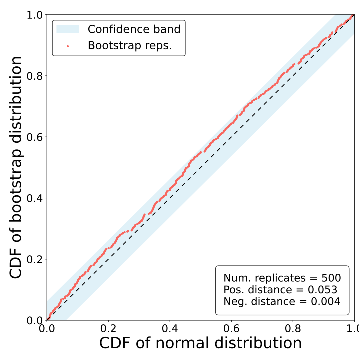
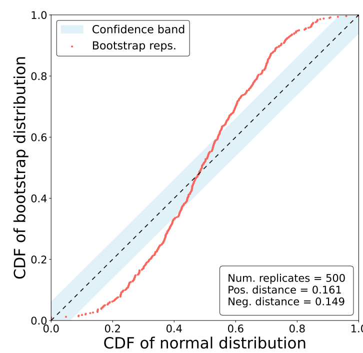
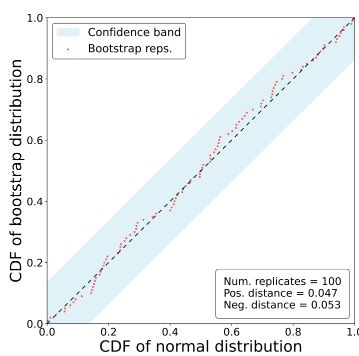
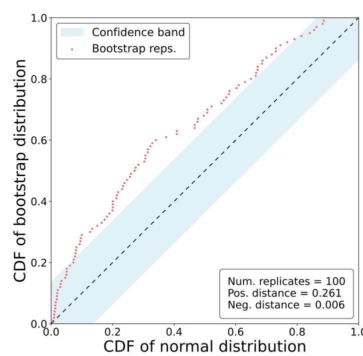
Paper	**Object with...** **least SK distance**	**greatest SK distance**
4		
5		
6		

Figure 3 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

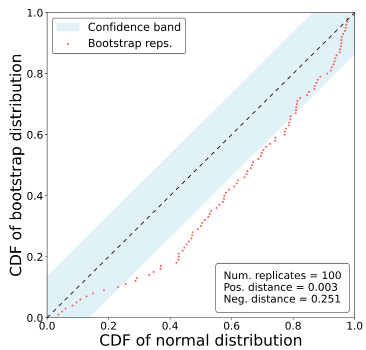
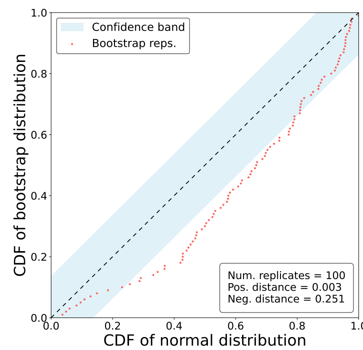
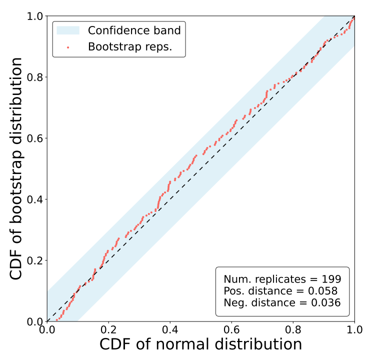
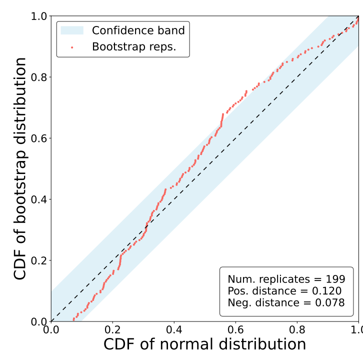
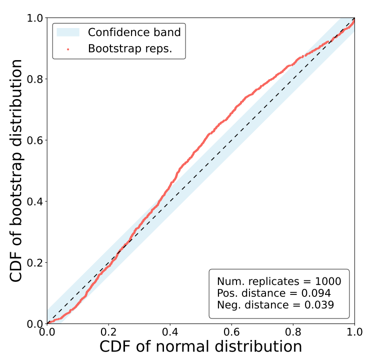
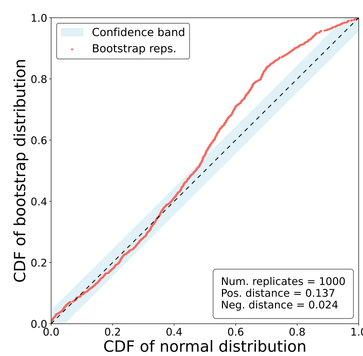
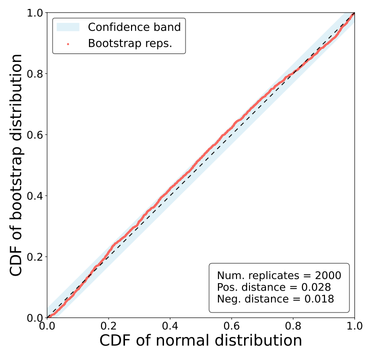
Paper	**Object with...** least SK distance	greatest SK distance
7		
8		
9		

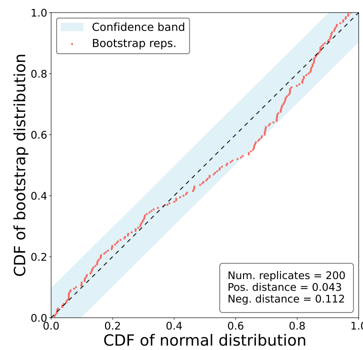
Figure 3 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

Paper **Object with...**

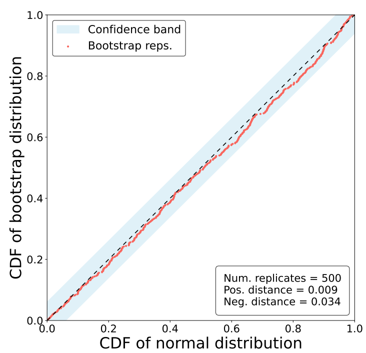
least SK distance **greatest SK distance**

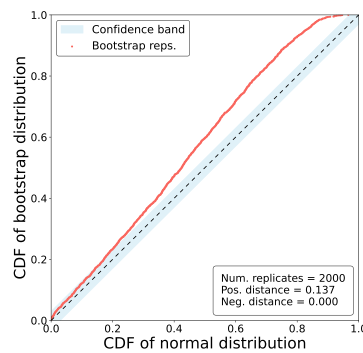
10



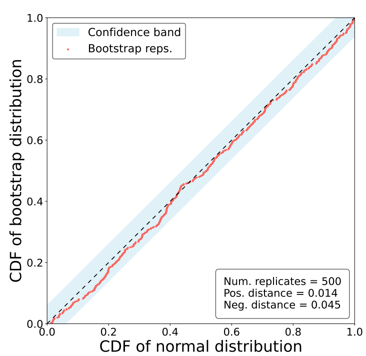


11





12



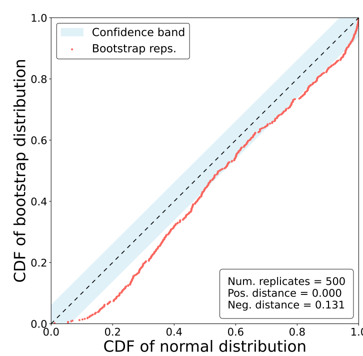
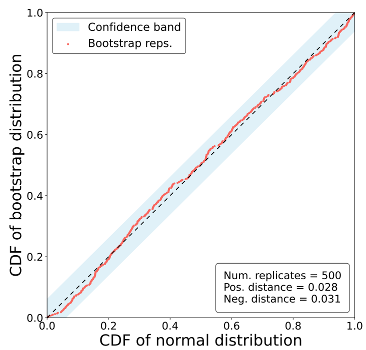
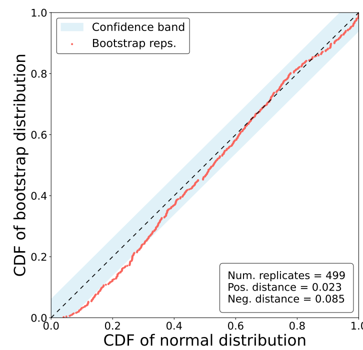


Figure 3 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

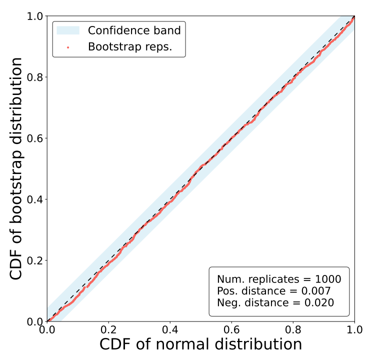
Paper	**Object with...**	
	least SK distance	**greatest SK distance**

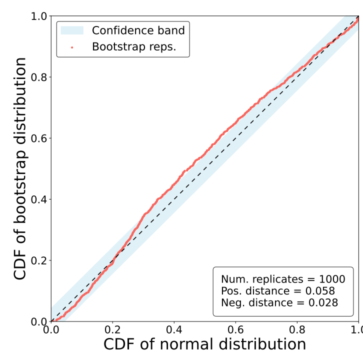
13





14





Notes: Each row corresponds to a paper in our bootstrap census. Each plot is a p-p plot comparing the distribution of the bootstrap replicates to the distribution of the default normal report whose mean is given by the point estimate and whose standard deviation is given by the bootstrap standard error. The shaded region is a uniform confidence band designed to contain the empirical CDF of the bootstrap replicates with probability at least 0.95 whenever the true bootstrap distribution is given by the default normal report. Each plot legend reports the maximum positive and negative vertical distances between the two distributions. Each row includes two plots, one for the object of interest with the smallest sum of maximum positive and negative distances ("least SK distance") and one for the object of interest with the largest sum of distances ("greatest SK distance"). Rows (papers) are in descending order according to their greatest signed Kolmogorov distance.

Figure 4: Alternative Normal Reports

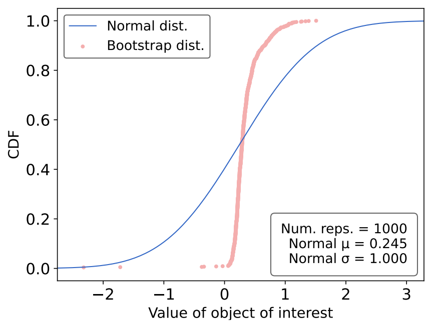
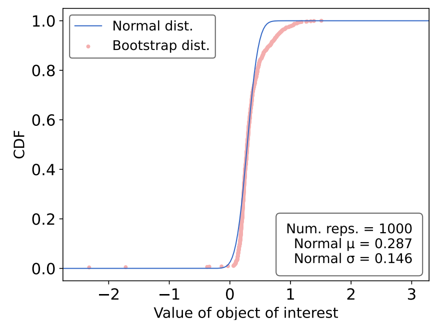
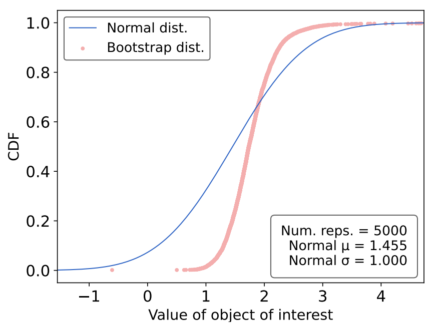
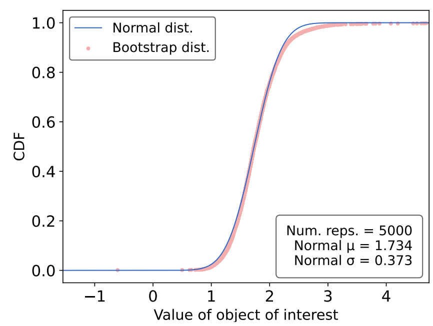
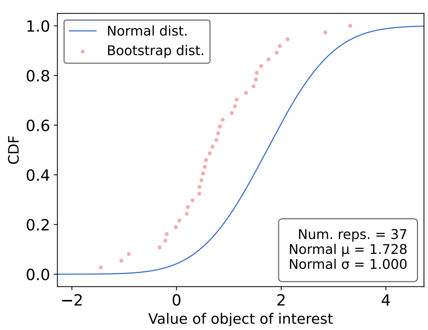
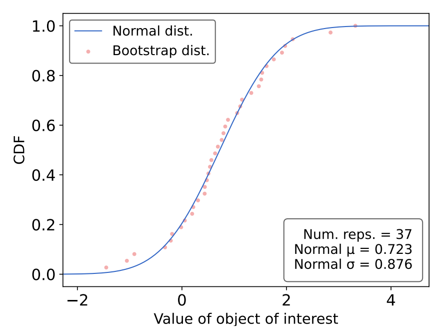
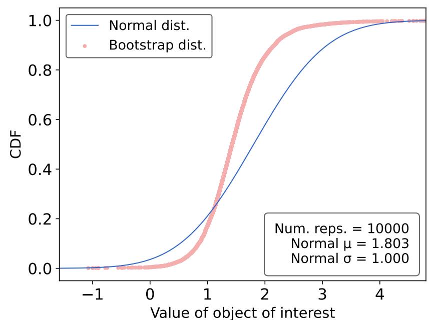
Paper	**Comparing bootstrap distribution to...**	
	default normal report	**distance-minimizing normal report**
1		
2		
3		

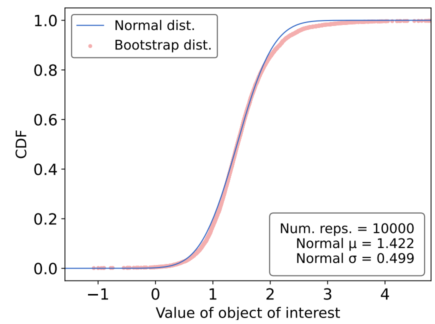
Figure 4 (continued): Alternative Normal Reports

Paper — **Comparing bootstrap distribution to...**

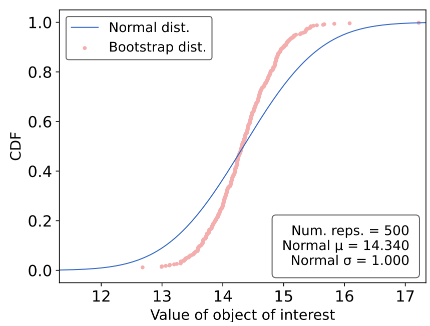
default normal report | **distance-minimizing normal report**

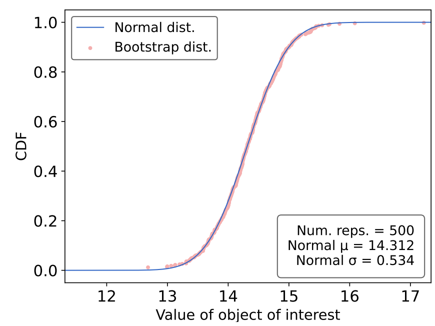
4



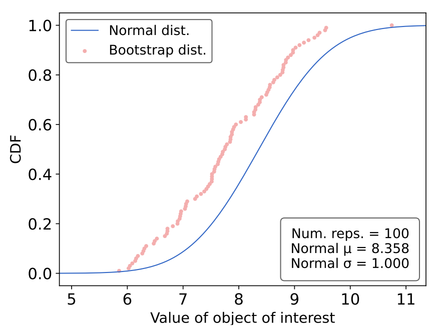


5





6



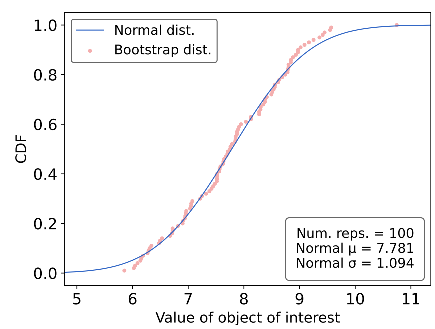
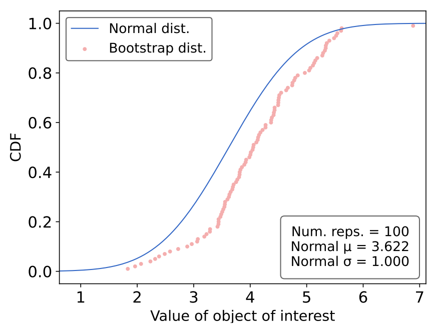


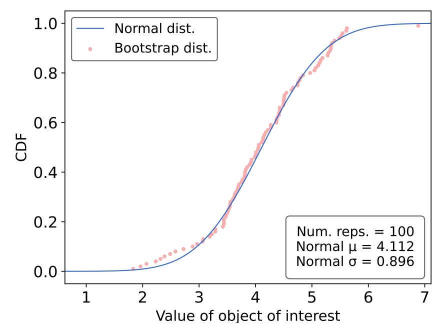
Figure 4 (continued): Alternative Normal Reports

Paper — **Comparing bootstrap distribution to...**

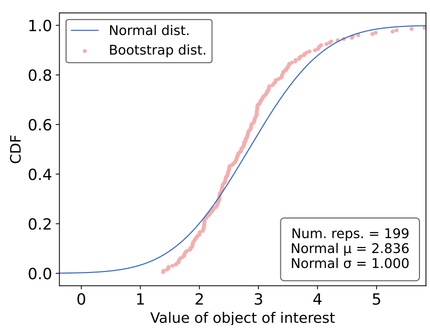
default normal report | **distance-minimizing normal report**

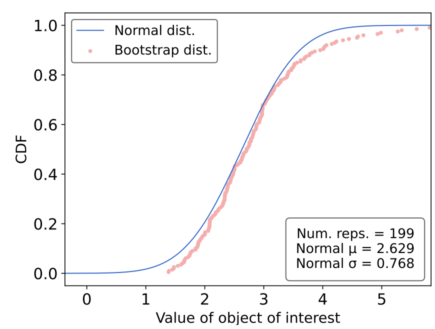
7



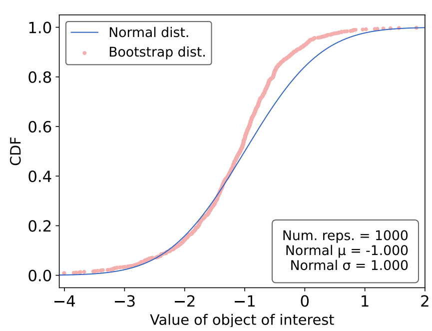


8





9



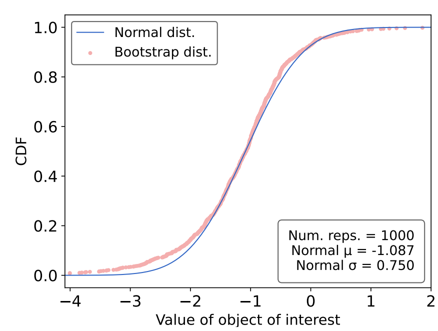
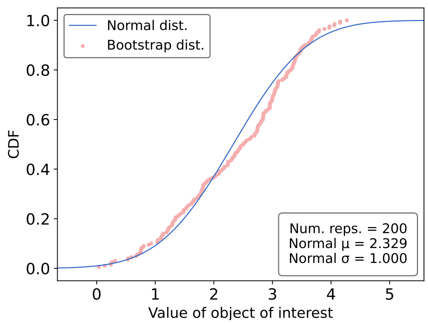
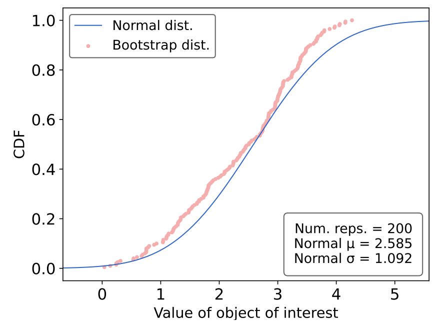


Figure 4 (continued): Alternative Normal Reports

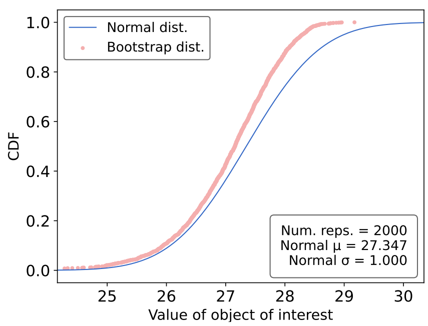
Paper	**Comparing bootstrap distribution to...**	
	default normal report	**distance-minimizing normal report**

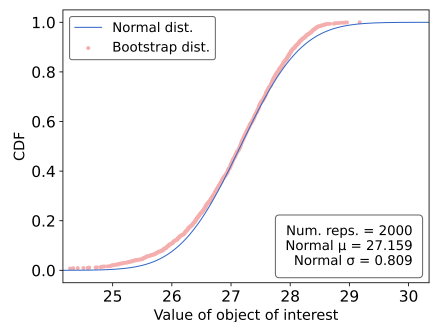
10



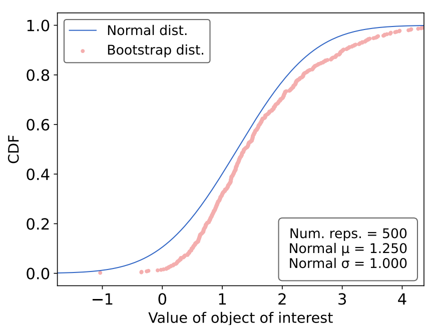


11





12



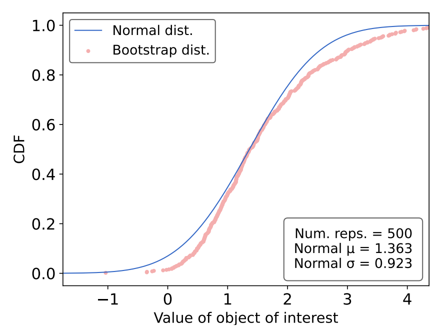
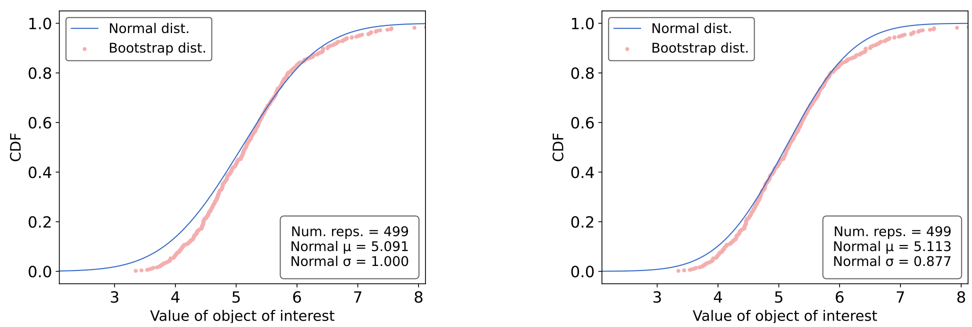


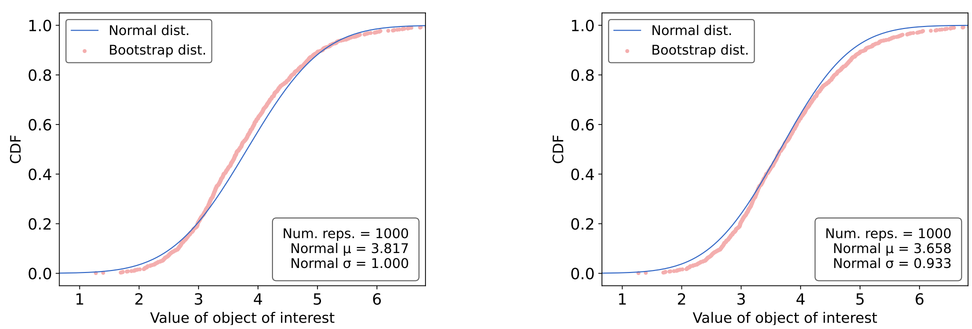
Figure 4 (continued): Alternative Normal Reports

Paper	**Comparing bootstrap distribution to...**	
	default normal report	**distance-minimizing normal report**

13



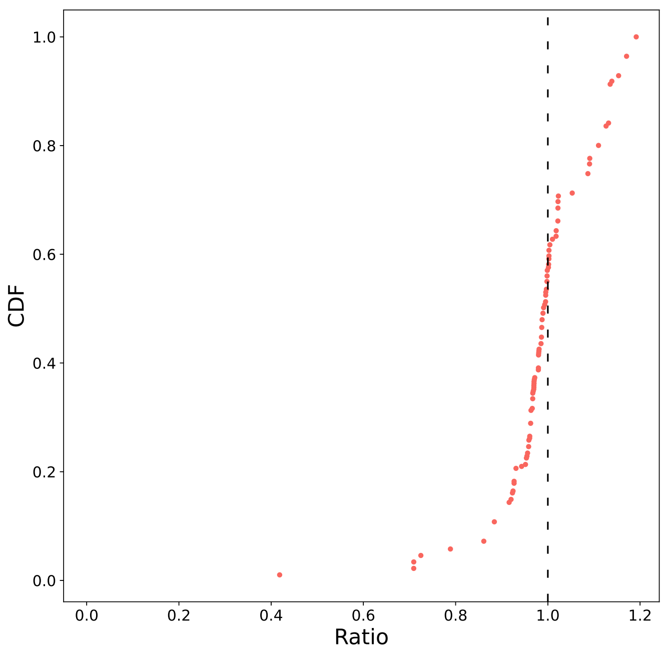
14



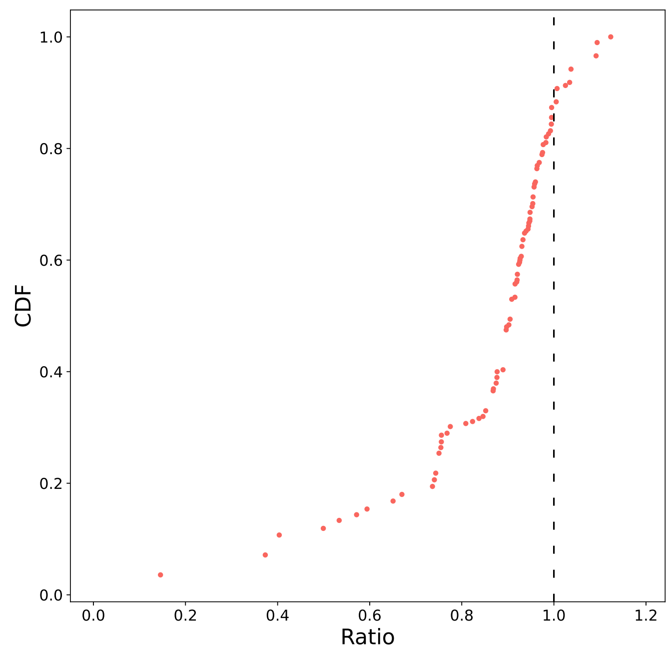
Notes: Each row corresponds to a paper in our bootstrap census and focuses on the object of interest with the greatest signed Kolmogorov distance as defined in Figure 3. The first column compares the empirical CDF of the bootstrap replicates to the CDF of the default normal report, whose mean is given by the point estimate and whose standard deviation is given by the bootstrap standard error. The second column compares the empirical CDF of the bootstrap replicates to the CDF of the distance-minimizing normal report, whose mean and standard deviation are chosen to minimize the signed Kolmogorov distance. Each plot legend reports the number of replicates and the mean and standard deviation of the normal. All values are normalized by dividing by the standard deviation of the bootstrap replicates. Rows (papers) are in descending order according to their greatest signed Kolmogorov distance.

Figure 5: Distance-Minimizing Normal Reports

Panel A: Ratio of mean of distance-minimizing normal report to point estimate



Panel B: Ratio of SD of distance-minimizing normal report to bootstrap standard error



Notes: Each plot is a weighted empirical CDF. The unit of analysis is an object of interest and, for each paper, we weight each object of interest by the inverse of the number of objects of interest associated with the paper. For each object of interest we calculate the distance-minimizing normal report, whose mean and standard deviation are chosen to minimize the signed Kolmogorov distance to the distribution of the bootstrap replicates. Panel A shows the weighted empirical CDF of the ratio of the mean of the distance-minimizing normal report to the point estimate. Panel B shows the weighted empirical CDF of the ratio of the standard deviation of the distance-minimizing normal report to the bootstrap standard error.

Online Appendix for
"Bootstrap Diagnostics for Irregular Estimators"

Isaiah Andrews, *MIT and NBER*[1]
Jesse M. Shapiro, *Harvard University and NBER*

A Approximation Properties of Reference Priors

Proposition 4. *If $\mathcal{X}_0$ is a Polish space and P_0 has support $\mathcal{X}_0$, then for any $Q_0 \in \Delta(\mathcal{X}_0)$ and almost every sequence of draws $\{X_1, X_2, ...\}$ from P_0 there exists a sequence of weights ω_n such that $P_n^{\omega_n} \to_d Q_0$ as $n \to \infty$.*

Proof of Proposition 4 Observe, first, that the sequence $\{X_1, X_2, ...\}$ is dense in $\mathcal{X}_0$ with probability one. To see that this is the case, note that since $\mathcal{X}_0$ is Polish it is also a metric space for some metric m_0, and has countable dense subset $\{x_1, x_2, ...\}$. For any $\varepsilon > 0$ and $k \in \mathbb{N}$, note that for $\mathcal{B}_\varepsilon(x_k)$ the ε-ball around x_k, $\mathcal{B}_\varepsilon(x_k) = \{x \in \mathcal{X}_0 : m_0(x_k, x) < \varepsilon\}$, since P_0 has support $\mathcal{X}_0$ we must have $\Pr_{P_0}\{X_i \in \mathcal{B}_\varepsilon(x_k)\} > 0$. For $K \in \mathbb{N}$ let A_K be the event that the sequence $\{X_1, X_2, ...\}$ intersects $\mathcal{B}_\varepsilon(x_k)$ for all $k \in \{1, ..., K\}$, i.e.,

$$A_K = \{|\{k \in \{1, ..., K\} : \{X_1, X_2, ...\} \cap \mathcal{B}_\varepsilon(x_k) \neq \emptyset\}| = K\}.$$

For $P_0^\infty = \times_{i=1}^\infty P_0$, note that $\Pr_{P_0^\infty}\{A_K\} = 1$ for all K. Moreover, for any finite collection $\{K_1, ..., K_M\} \subset \mathbb{N}$ with $K_1 < ... < K_M$,

$$\Pr_{P_0^\infty}\{A_{K_1} \cap A_{K_2} ... \cap A_{K_M}\} = \Pr_{P_0^\infty}\{A_{K_M}\} = 1,$$

so the events A_K are mutually independent. Since $\sum_{K=1}^\infty \Pr_{P_0^\infty}\{A_K\} = \infty$ the second Borel-Cantelli Lemma (see, e.g., Theorem 4.4 in Billingsley 1995) implies that

$$\Pr_{P_0^\infty}\{\{X_1, X_2, ...\} \cap \mathcal{B}_\varepsilon(x_k) \neq \emptyset \text{ for all } k \in \mathbb{N}\} = 1.$$

Since m_0 is a metric, it follows that

$$\Pr_{P_0^\infty}\{\{X_1, X_2, ...\} \cap \mathcal{B}_{2\varepsilon}(x) \neq \emptyset \text{ for all } x \in \mathcal{X}_0\} = 1.$$

[1] E-mail: iandrews@mit.edu, jesse_shapiro@fas.harvard.edu.

Since we can repeat this argument for all $\varepsilon > 0$, it follows that $\{X_1, X_2, ...\}$ is dense in $\mathcal{X}_0$ almost surely.

Let us next fix a draw of $\{X_1, X_2, ...\}$ which is dense in $\mathcal{X}_0$. Theorem A.3 of Ghosal and van der Vaart (2017) implies that the set of weighted empirical measures

$$\mathcal{P} = \{P_n^{\omega} : n \in \mathbb{N}, \omega \in \Delta(\{1, ..., n\})\}$$

is dense in $\Delta(\mathcal{X}_0)$ in the weak topology. Since the sets

$$\mathcal{P}_n = \{P_{\tilde{n}}^{\omega} : \tilde{n} \in \{1, ..., n\}, \omega \in \Delta(\{1, ..., \tilde{n}\})\}$$

are increasing in n and $\cup_{n=1}^{\infty} \mathcal{P}_n = \mathcal{P}$, it follows that for any $Q_0 \in \Delta(\mathcal{X}_0)$ there exists a sequence of weights $\{\omega_n\}_{n=1}^{\infty}$ with $P_n^{\omega_n} \in \mathcal{P}_n$ for all n such that $P_n^{\omega_n} \to_d Q_0$, as we aimed to show. $\square$

B Interpretation of the Limiting Marginal Prior

Here we discuss the interpretation of the limiting value of the marginal prior on the parameter θ as the informativeness α of the reference prior tends to zero. We assume that the target parameter of interest is well behaved when evaluated on weighted empirical distributions where the weights approach a degenerate limit.

Assumption 3. *For any permutation $\varphi : \{1, ..., n\} \to \{1, ..., n\}$, any sequence $\omega_s \in \Delta(\{1, ..., n\})$ with $\omega_{s,i} > 0$ and $\lim_{s \to \infty} \omega_{s,i+1}/\omega_{s,i} = 0$ for all i, and $P_{n,s}^{\varphi,\omega} = \sum_n \delta_{X_i} \omega_{s,\varphi(i)}$, we have that*

$$\lim_{s \to \infty} \theta\left(P_{n,s}^{\varphi,\omega}\right) = \bar{\theta}\left(X_{\varphi(1)}, X_{\varphi(2)}, ..., X_{\varphi(n)}\right)$$

for some limit $\bar{\theta}(\cdot)$.

To simplify discussion, we further assume, without loss of generality, that $X_i \neq X_j$ for all $i \neq j$.[2] Assumption 3 holds for a wide range of target parameters.

Example. (Functions of a Mean) Suppose that $\theta(P_0) = \chi(\mathrm{E}_{P_0}[\varrho(X_i)])$ for known functions χ and ϱ, where $\chi(\cdot)$ is continuous at $\varrho(X_i)$ for $i \in \{1, ..., n\}$. In this case, it is immediate that Assumption 3 holds for $\bar{\theta}(x_1, x_2, ..., x_n) = \chi(\varrho(x_1))$. This example includes cases such

[2] If this is not the case, we can simply drop observation j and add its weight ω_j to the weight for observation i.

as minimum distance estimators where the target moments can be expressed as averages of separable functions of the observations.

Under Assumption 3, the limiting marginal prior as $\alpha \to 0$ takes a simple form.

Theorem 3. *Under Assumptions 2 and 3, as $\alpha \to 0$ the marginal reference prior $\pi^*(\theta)$ converges weakly to $\pi^\infty \in \Delta(\Theta)$ where*

$$\pi^\infty(\theta) = \sum_{\varphi \in \Phi} p(\varphi, \omega) 1\left\{\bar{\theta}\left(X_{\varphi(1)}, X_{\varphi(2)}, ..., X_{\varphi(n)}\right) = \theta\right\}$$

for Φ the set of permutations $\varphi : \{1, ..., n\} \to \{1, ..., n\}$ and

$$p(\varphi, \omega) = \prod_{i=1}^{n} \frac{\omega_{\varphi(i)}}{1 - \sum_{j<1} \omega_{\varphi(j)}}$$

the probability of drawing the sequence $\left(X_{\varphi(1)}, X_{\varphi(2)}, ..., X_{\varphi(n)}\right)$ when sampling $(X_1, ..., X_n)$ without replacement using weights $(\omega_1, ..., \omega_n)$.

Proof of Theorem 3 The stick-breaking representation of Dirichlet processes (see e.g. Theorem 4.12 of Ghosal and van der Vaart 2017) implies that we can write draws from the prior $\pi^*(P_0)$ as

$$P_0 = \sum_{k=1}^{\infty} V_k(\alpha) \delta_{\tilde{X}_k}$$

where $\delta_{\tilde{X}_k}$ is a Dirac mass at $\tilde{X}_k$, the random variables $\tilde{X}_k$ are drawn i.i.d. from P_n^ω, and $V_k(\alpha) = \left(1 - U_k^{\frac{1}{\alpha}}\right) \prod_{i=1}^{k-1} U_i^{\frac{1}{\alpha}}$ where the random variables U_i are i.i.d. standard uniform. Note that $\Pr\{U_j \in (0,1) \text{ for all } j\} = 1$, and that conditional on this event $V_k(\alpha) \in (0,1)$ for all k and all $\alpha > 0$, while $V_{k+1}(\alpha)/V_k(\alpha) \to 0$ as $\alpha \to 0$.

Let $\varphi(1) \in \{1, ..., n\}$ be the index for the observation in the original data with $X_{\varphi(1)} = \tilde{X}_1$. For $j \in \{2, ..., n\}$, let $k(j)$ be the smallest k such that $\tilde{X}_k$ is distinct from $\left\{X_{\varphi(1)}, ..., X_{\varphi(j-1)}\right\}$, and let $X_{\varphi(j)} = \tilde{X}_{k(j)}$. We can equivalently write

$$P_0 = P_n^{\varphi, W(\alpha)} = \sum_{i=1}^{n} \delta_{X_{\varphi(i)}} W_i(\alpha),\ W_i(\alpha) = \sum_{k=1}^{\infty} V_k(\alpha) 1\left\{\tilde{X}_k = X_{\varphi(i)}\right\}.$$

By construction $P_0 \in \Delta(\{X_1, ..., X_n\})$, and $W_i(\alpha) \in (0,1)$ with probability one for all $\alpha > 0$.

Moreover, as $\alpha \to 0$ we see that $W_{i+1}(\alpha)/W_i(\alpha) \to 0$ for all i, so

$$\lim_{\alpha \to 0} \theta\left(P_n^{\varphi,W(\alpha)}\right) = \bar{\theta}\left(X_{\varphi(1)}, X_{\varphi(2)}, \ldots, X_{\varphi(n)}\right)$$

by Assumption 3. The form of $p(\varphi,\omega)$ is then immediate from the definition of φ. $\square$

Example. (Functions of a Mean) In the case where $\theta(P_0) = \chi(\mathrm{E}_{P_0}[\varrho(X_i)])$ and $\omega = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$, the limiting marginal prior $\pi^\infty(\theta)$ corresponds to a draw from the empirical distribution of $\chi(\varrho(X_i))$. This prior is uninformative in several respects. For example, when $\chi(\cdot)$ is the identity function, Proposition 4.3 of Ghosal and van der Vaart (2017) implies that the limit of the prior variance is simply the sample variance,

$$\lim_{\alpha \to 0} \mathrm{Var}_{\pi^*}(\theta(P_0)) = \mathrm{Var}_{P_n}(\varrho(X_i)),$$

while the limit of the posterior variance is far smaller in large samples,

$$\lim_{\alpha \to 0} \mathrm{Var}_{\pi^*(\theta|X)}(\theta(P_0)) = \frac{1}{n+1} \mathrm{Var}_{P_n}(\varrho(X_i)).$$

C Equivalence of Bootstraps Under the SK Metric

This section provides sufficient conditions for two bootstraps, A and B, to coincide in SK distance as $n \to \infty$. To emphasize the flexibility of the asymptotic setting, we here make explicit that the data-generating process $P_{0,n} \in \Delta(\mathcal{X}_0)$ may depend on the sample size.

Our analysis assumes that the estimator $\hat{\theta}$ can be approximated by continuous functions of objects which converge in distribution.

Assumption 4. *The parameter space Θ is a subset of $\mathbb{R}$ and the estimator $\hat{\theta}$ can be written as $\hat{\theta} = a_n + b_n \theta_n\left(\hat{\beta}_n\right)$ for a_n and b_n non-random sequences of scalars, θ_n a sequence of functions, and $\hat{\beta}_n = \beta_n(P_n)$ a sequence of statistics with population analogues $\beta_{0,n} = \beta_n(P_{0,n})$. Moreover, for BL_K the set of Lipschitz functions with supremum norm and Lipschitz constant both bounded by K,*

(a) As $n \to \infty$, we have that $\hat{\beta}_n - \beta_{0,n} \to Z_\beta \sim Q_\beta \in \Delta(\mathcal{Z})$ for $\mathcal{Z}$ a normed vector space and Z_β a random variable such that $\|Z_\beta\|$ is continuously distributed.

(b) For all $\epsilon > 0$, there exists a constant $K(\epsilon) \in [0,\infty)$, a sequence of functions $\theta_{n,\epsilon} \in BL_{K(\epsilon)}$, and an open set $\mathcal{Z}_\epsilon$ such that for c_ϵ the $1-\epsilon$ quantile of $\|Z_\beta\|$,

$$\limsup_{n\to\infty} \sup_{z\in\mathcal{Z}\setminus\mathcal{Z}_\epsilon} 1\{|\theta_n(z+\beta_{0,n})-\theta_{n,\epsilon}(z)|>0\}=0 \text{ and } \sup_{z\in\mathcal{Z}:\|z\|\leq c_\epsilon} \Pr_{Q_\beta}\{Z_\beta+z\in\mathcal{Z}_\epsilon\}\leq\epsilon.$$

Assumption 4(a) states that the estimator can be represented as a function (or functional) of a statistic that converges in distribution, where we do not restrict the dimension of β beyond requiring convergence in distribution.[3] Assumption 4(b) requires that the estimator $\hat{\theta}$ be sufficiently continuous in $\hat{\beta}_n$, but is weaker than assuming continuity or differentiability of θ_n, and does not in general imply normality of $\hat{\theta}$.[4]

For instance, if θ is the maximum of two means, $\theta(P_{0,n})=\max\{E_{P_{0,n}}[X_{i,1}],E_{P_{0,n}}[X_{i,2}]\}$ where $(X_{i,1},X_{i,2})$ are bounded, Assumption 4 holds if we take $\hat{\beta}_n=\beta_n(P_n)=\sqrt{n}F_{P_n}$ to be the scaled empirical distribution function and $\beta_{0,n}=\beta_n(P_{0,n})=\sqrt{n}F_{P_{0,n}}$ to be the scaled distribution function of $P_{0,n}$, even though $\theta_n(\cdot)$ is non-differentiable and $\hat{\theta}$ need not be asymptotically normal in this case. Similarly, if θ corresponds to a ratio of regression coefficients, $\theta(P_{0,n})=\beta_1(P_{0,n})/\beta_2(P_{0,n})$, Assumption 4 holds under minimal conditions when we take $\hat{\beta}_n=\sqrt{n}(\beta(P_n))$ and $\beta_{0,n}=\sqrt{n}\beta(P_{0,n})$ for $\beta(P_n)=(\beta_1(P_n),\beta_2(P_n))$ and $\beta(P_{0.n})=(\beta_1(P_{0,n}),\beta_2(P_{0,n}))$ the sample and population regression coefficients, respectively. In this case $\theta_n(\cdot)$ is discontinuous, and $\hat{\theta}$ again need not be asymptotically normal.

We further assume that both bootstraps consistently recover the asymptotic distribution of $\hat{\beta}_n$ and deliver an asymptotically continuous distribution for $\hat{\theta}$.

Assumption 5. *For a bootstrap with distribution $\eta(X)$ and other objects as defined in Assumption 4, we have the following.*

(a)

$$\sup_{h\in BL_1}\left\{\mathrm{E}_{\eta_\beta(X)}\left[h\left(\beta-\hat{\beta}_n\right)\right]-\mathrm{E}_{Q_\beta}\left[h\left(\hat{\beta}_n-\beta_{0,n}\right)\right]\right\}\to_p 0. \tag{3}$$

[3]In cases where $\mathcal{Z}$ is infinite-dimensional and $\hat{\beta}_n$ need not be measurable for finite n, the statements below can be adapted by replacing the ordinary expectation E and convergence in probability $\to_p$ by the upper expectation E^* and convergence in outer probability $\to_{p^*}$ respectively (see Chapter 1 of van der Vaart and Wellner 1996).

[4]Assumption 4(b) is implied by local Lipschitz conditions such as those considered by Kitagawa et al. (2020).

(b) For $\mathcal{B}_{\varepsilon,n}(\theta)=\left\{\tilde{\theta}:\left|\tilde{\theta}-\theta\right|<\varepsilon/b_n\right\}$,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\mathrm{E}\left[\sup_{\theta\in\mathbb{R}}\eta_\theta(\mathcal{B}_{\varepsilon,n}(\theta)|X)\right]=0.$$

Assumption 5(a) states a sense in which the bootstrap consistently recovers the asymptotic distribution of the statistics $\hat{\beta}_n$. Importantly, Assumption 5(a) does not require that the bootstrap consistently recovers the asymptotic distribution of the estimator $\hat{\theta}$, or even that such a distribution exists. Assumption 5(b) states a sense in which the bootstrap implies an asymptotically continuous distribution for $\hat{\theta}$. Again, this continuous distribution need not coincide with a true sampling distribution for $\hat{\theta}$, so Assumption 5(b) allows for cases where the bootstrap need not be consistent, such as the examples discussed above.

If Assumption 4 holds for the estimator $\hat{\theta}$ and Assumption 5 holds for two bootstraps A and B, then the distributions over $\hat{\theta}$ implied by the two bootstraps A and B are asymptotically equivalent in SK.

Proposition 5. *If Assumption 4 holds for* $\hat{\theta}$ *and Assumption 5 holds for bootstraps* A *and* B *with distributions* $\eta^A(X)$ *and* $\eta^B(X)$*, then*

$$SK\left(\eta_\theta^A(X),\eta_\theta^B(X)\right)\to_p 0.$$

Proof of Proposition 5 Let $\hat{\beta}_n^{*,A}$ and $\hat{\beta}_n^{*,B}$ denote independent draws from the bootstrap distributions $\eta_\beta^A(X)$ and $\eta_\beta^B(X)$, respectively. By the triangle inequality, Assumption 5(a) implies that

$$\sup_{h\in BL_1}\left\{\mathrm{E}_{\eta_\beta^A(X)}\left[h\left(\beta-\hat{\beta}_n\right)\right]-\mathrm{E}_{\eta_\beta^B(X)}\left[h\left(\beta-\hat{\beta}_n\right)\right]\right\}=$$

$$\sup_{h\in BL_1}\left\{\mathrm{E}\left[h\left(\hat{\beta}_n^{*,A}-\hat{\beta}_n\right)|X\right]-\mathrm{E}\left[h\left(\hat{\beta}_n^{*,B}-\hat{\beta}_n\right)|X\right]\right\}\to_p 0.$$

To prove Proposition 5, we first show that the bootstrap distributions for $\hat{\theta}_n^{*,A}=\theta_n\left(\hat{\beta}_n^{*,A}\right)$ and $\hat{\theta}_n^{*,B}=\theta_n\left(\hat{\beta}_n^{*,B}\right)$ are equivalent in bounded Lipschitz metric,

$$\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\hat{\theta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\theta}_n^{*,A}\right)|X\right]\right|\to_p 0. \tag{4}$$

To establish the equivalence (4), note that

$$\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\hat{\theta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\theta}_n^{*,B}\right)|X\right]\right|=$$

$$\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\tilde{\theta}_n\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)|X\right]-\mathrm{E}\left[h\left(\tilde{\theta}_n\left(\hat{\beta}_n^{*,B}-\beta_{0,n}\right)\right)|X\right]\right|$$

for $\tilde{\theta}_n(\beta)=\theta_n(\beta+\beta_{0,n})$. Note, next, that for n sufficiently large,

$$\left\{(z_1,z_2)\in\mathcal{Z}^2:\tilde{\theta}_n(z_1+z_2)\neq\theta_{n,\epsilon}(z_1+z_2)\right\}\subseteq\{(z_1,z_2):\|z_2\|\geq c_\epsilon\}\cup\mathcal{C}_\epsilon$$

$$\mathcal{C}_\epsilon=(\{(z_1,z_2):\|z_2\|<c_\epsilon\}\cap\{(z_1,z_2):z_1+z_2\in\mathcal{Z}_\epsilon\})$$

where $\mathcal{C}_\epsilon$ is open. We can write $\hat{\beta}_n^{*,A}-\beta_{0,n}=\hat{\beta}_n^{*,A}-\hat{\beta}_n+\hat{\beta}_n-\beta_{0,n}$, where Assumptions 5(a) and 4(a) imply that

$$\begin{pmatrix}\hat{\beta}_n^{*,A}-\hat{\beta}_n\\ \hat{\beta}_n-\beta_{0,n}\end{pmatrix}\to_d\begin{pmatrix}Z_\beta^*\\ Z_\beta\end{pmatrix},\ Z_\beta^*,Z_\beta\overset{i.i.d.}{\sim}F_\beta. \tag{5}$$

Hence,

$$\limsup_{n\to\infty}\mathrm{Pr}_{P_{0,n}}\left\{\tilde{\theta}_n\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\neq\theta_{n,\varepsilon}\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right\}\leq$$

$$\limsup_{n\to\infty}\mathrm{Pr}_{P_{0,n}}\left\{\left\|\hat{\beta}_n-\beta_{0,n}\right\|\geq c_\epsilon\right\}+\limsup_{n\to\infty}\mathrm{Pr}_{P_{0,n}}\left\{\left(\hat{\beta}_n^{*,A}-\hat{\beta}_n,\hat{\beta}_n-\beta_{0,n}\right)\in\mathcal{C}_\epsilon\right\}.$$

Note, however, that

$$\limsup_{n\to\infty}\mathrm{Pr}_{P_{0,n}}\left\{\left\|\hat{\beta}_n-\beta_{0,n}\right\|\geq c_\epsilon\right\}\leq\epsilon,\ \limsup_{n\to\infty}\mathrm{Pr}_{P_{0,n}}\left\{\left(\hat{\beta}_n^{*,A}-\hat{\beta}_n,\hat{\beta}_n-\beta_{0,n}\right)\in\mathcal{C}_\epsilon\right\}\leq\epsilon,$$

where the first inequality follows from (5) and the fact that $\|Z_\beta\|$ is continuously distributed, while the second follows from Assumption 4(a), the joint convergence (5), the Portmanteau Lemma (Lemma 2.2 of van der Vaart 1998), and the fact that $\mathcal{C}_\varepsilon$ is open. We thus have that

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\tilde{\theta}_n\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)-h\left(\theta_{n,\epsilon}\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)|X\right]\right|\right]\leq 2\epsilon,$$

so since the same also holds for $\hat{\beta}_n^{*,B}$,

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\theta_n\left(\hat{\beta}_n^{*,A}\right)\right)|X\right]-\mathrm{E}\left[h\left(\theta_n\left(\hat{\beta}_n^{*,B}\right)\right)|X\right]\right|\right]\leq$$

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\theta_{n,\epsilon}\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)|X\right]-\mathrm{E}\left[h\left(\theta_{n,\epsilon}\left(\hat{\beta}_n^{*,B}-\beta_{0,n}\right)\right)|X\right]\right|\right]+4\epsilon\leq$$

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{h\in BL_{K(\epsilon)}}\left|\mathrm{E}\left[h\left(\hat{\beta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\beta}_n^{*,B}\right)|X\right]\right|\right]+4\epsilon,$$

where we have used the fact that a composition of a function $BL_{K(\epsilon)}$ with one in BL_1 is necessarily in $BL_{K(\epsilon)}$ (where we assume without loss of generality that $K(\epsilon)\geq 1$). However,

$$\sup_{h\in BL_{K(\epsilon)}}\left|\mathrm{E}\left[h\left(\hat{\beta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\beta}_n^{*,B}\right)|X\right]\right|=K(\epsilon)\cdot \sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\hat{\beta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\beta}_n^{*,B}\right)|X\right]\right|,$$

so Assumption 5(a) implies that

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{h\in BL_1}\left|\mathrm{E}\left[h\left(\hat{\theta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\theta}_n^{*,B}\right)|X\right]\right|\right]\leq 4\epsilon.$$

Since we can repeat this argument for all $\epsilon>0$, we have verified (4).

It remains to translate convergence of $\left(\hat{\theta}_n^{*,A},\hat{\theta}_n^{*,B}\right)$ in bounded Lipschitz metric to convergence of $\left(\eta_\theta^A(X),\eta_\theta^B(X)\right)$ in SK metric. Let $\eta_{\theta_n}^A(X)$ denote the distribution of $\hat{\theta}_n^{*,A}$, and note that since SK is unchanged by linear reparameterization, $SK\left(\eta_\theta^A(X),\eta_\theta^B(X)\right)=SK\left(\eta_{\theta_n}^A(X),\eta_{\theta_n}^B(X)\right)$. Recall next that $F_{\eta_{\theta_n}^A(X)}\left(\tilde{\theta}\right)=\mathrm{E}_{\eta_{\theta_n}^A(X)}\left[1\left\{\theta\leq\tilde{\theta}\right\}\right]$, and note that for any $\upsilon\in(0,1)$ and each $\tilde{\theta}\in\mathbb{R}$, there exists a function $h_\upsilon\in BL_1$ such that $h_\upsilon(\theta)=\upsilon\cdot 1\left\{\theta\leq\tilde{\theta}\right\}$ for all $\theta\notin\left(\tilde{\theta}-\upsilon,\tilde{\theta}\right)$. Hence, for all $\upsilon\in(0,1)$

$$\sup_{\tilde{\theta}}\left|F_{\eta_{\theta_n}^A(X)}\left(\tilde{\theta}\right)-F_{\eta_{\theta_n}^B(X)}\left(\tilde{\theta}\right)\right|\leq$$

$$\upsilon^{-1}\cdot\sup_{h\in BL_1}\left|\mathrm{E}_{\eta_{\theta_n}^A(X)}[h(\theta)]-\mathrm{E}_{\eta_{\theta_n}^B(X)}[h(\theta)]\right|+$$

$$\sup_{\tilde{\theta}}\left(\mathrm{E}_{\eta_{\theta_n}^A(X)}\left[1\left\{\theta\in\left(\tilde{\theta}-\upsilon,\tilde{\theta}\right)\right\}\right]+\mathrm{E}_{\eta_{\theta_n}^B(X)}\left[1\left\{\theta\in\left(\tilde{\theta}-\upsilon,\tilde{\theta}\right)\right\}\right]\right).$$

Thus, we have that for all $\upsilon\in(0,1)$,

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{\tilde{\theta}}\left|F_{\eta_{\theta_n}^A(X)}\left(\tilde{\theta}\right)-F_{\eta_{\theta_n}^B(X)}\left(\tilde{\theta}\right)\right|\right]\leq$$

$$\upsilon^{-1} \cdot \limsup_{n\to\infty} \mathrm{E}\left[\sup_{h\in BL_1}\left|\mathrm{E}_{\eta^A_{\theta_n}(X)}[h(\theta)] - \mathrm{E}_{\eta^B_{\theta_n}(X)}[h(\theta)]\right|\right] +$$

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{\theta\in\mathbb{R}} \eta^A_\theta(\mathcal{B}_{\upsilon,n}(\theta)|X)\right] + \limsup_{n\to\infty} \mathrm{E}\left[\sup_{\theta\in\mathbb{R}} \eta^B_\theta(\mathcal{B}_{\upsilon,n}(\theta)|X)\right].$$

However, we have already established that the first term goes to zero for all $\upsilon \in (0,1)$, while the second and third terms go to zero as $\upsilon \to 0$ by Assumption 5(b). Hence, $\eta^A_{\theta_n}(X)$ and $\eta^B_{\theta_n}(X)$ converge in Kolmogorov metric. Since SK is bounded by twice the Kolmogorov distance

$$SK\left(\eta^A_\theta(X), \eta^B_\theta(X)\right) \leq 2 \sup_{\tilde{\theta}} \left| F_{\eta^A_\theta(X)}\left(\tilde{\theta}\right) - F_{\eta^B_\theta(X)}\left(\tilde{\theta}\right)\right|,$$

the conclusion of the proposition follows immediately. □

D Additional Findings from Bootstrap Census

Appendix Table 1: List of Papers in Bootstrap Census

Citation	Objects	Transmitted?
Abebe, Girum; Caria, A. Stefano; and Ortiz-Ospina, Esteban. Code for: “The Selection of Talent: Experimental and Structural Evidence from Ethiopia.” https://doi.org/10.1257/aer.20190586.	7	N
Adermon, Adrian; Lindahl, Mikael; and Palme, Mårten. Code for: “Dynastic Human Capital, Inequality, and Intergenerational Mobility.” https://doi.org/10.1257/aer.20190553.	2	Y
Bailey, Martha J.; Sun, Shuqiao; and Timpe, Brenden. Code for: “Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency.” https://doi.org/10.1257/aer.20181801.	6	Y
Bourreau, Marc; Sun, Yutec; and Verboven, Frank. Code for: “Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market.” https://doi.org/10.1257/aer.20190540.	20	Y
Braguinsky, Serguey; Ohyama, Atsushi; Okazaki, Tetsuji; and Syverson, Chad. Code for: “Product Innovation, Product Diversification, and Firm Growth: Evidence from Japan’s Early Industrialization.” https://doi.org/10.1257/aer.20201656.	2	N
Dinerstein, Michael; and Smith, Troy D. Code for: “Quantifying the Supply Response of Private Schools to Public Policies.” https://doi.org/10.1257/aer.20151723.	1	N
Finkelstein, Amy; Gentzkow, Matthew; and Williams, Heidi. Code for: “Place-Based Drivers of Mortality: Evidence from Migration.” https://doi.org/10.1257/aer.20190825.	3	Y
Goodman-Bacon, Andrew. Code for: “The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes.” https://doi.org/10.1257/aer.20171671.	2	Y

Appendix Table 1 (continued): List of Papers in Bootstrap Census

Citation	Objects	Transmitted?
Känzig, Diego R.. Code for: “The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements.” https://doi.org/10.1257/aer.20190964.	6	N
Køstol, Andreas R.; and Myhre, Andreas S. Code for: “Labor Supply Responses to Learning the Tax and Benefit Schedule.” https://doi.org/10.1257/aer.20201877.	5	Y
Mueller, Andreas I.; Spinnewijn, Johannes; and Topa, Giorgio. Code for: “Job Seekers’ Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias.” https://doi.org/10.1257/aer.20190808.	3	N
Reimers, Imke; and Waldfogel, Joel. Code for: “Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings.” https://doi.org/10.1257/aer.20200153.	7	N
Seibold, Arthur. Code for: “Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities.” https://doi.org/10.1257/aer.20191136.	4	Y
Weaver, Jeffrey. Code for: “Jobs for Sale: Corruption and Misallocation in Hiring.” https://doi.org/10.1257/aer.20201062.	13	Y

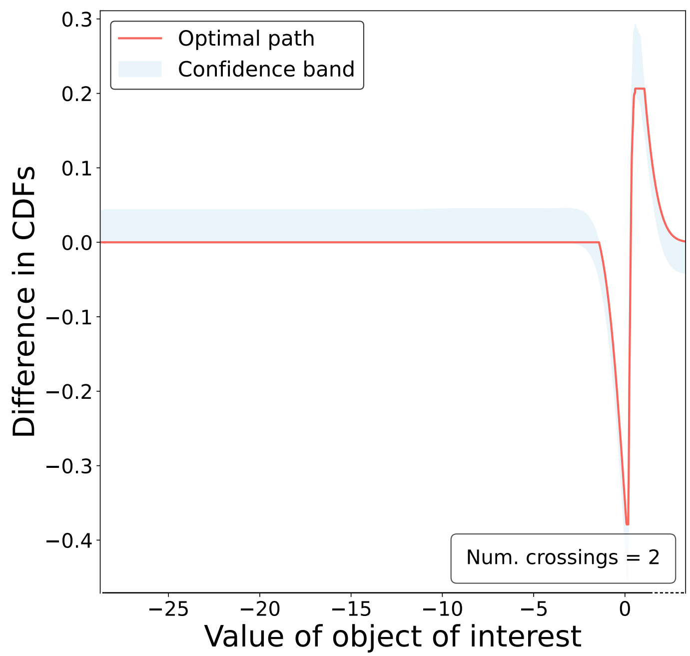
Notes: For each paper in our bootstrap census, the table reports an abbreviated citation for the code and data, the number of objects of interest for which we obtain replicates, and an indicator for whether or not we received a transmission of bootstrap replicates directly from the authors. Papers are in ascending alphabetical order by the first author’s last name. In each case the publisher is “Nashville, TN: American Economic Association,” the distributor is “Ann Arbor, MI: Inter-university Consortium for Political and Social Research,” and the publication year is 2021.

Appendix Table 2: Testing the Tightness of the SK Bound

Panel A: Minimum number of crossings

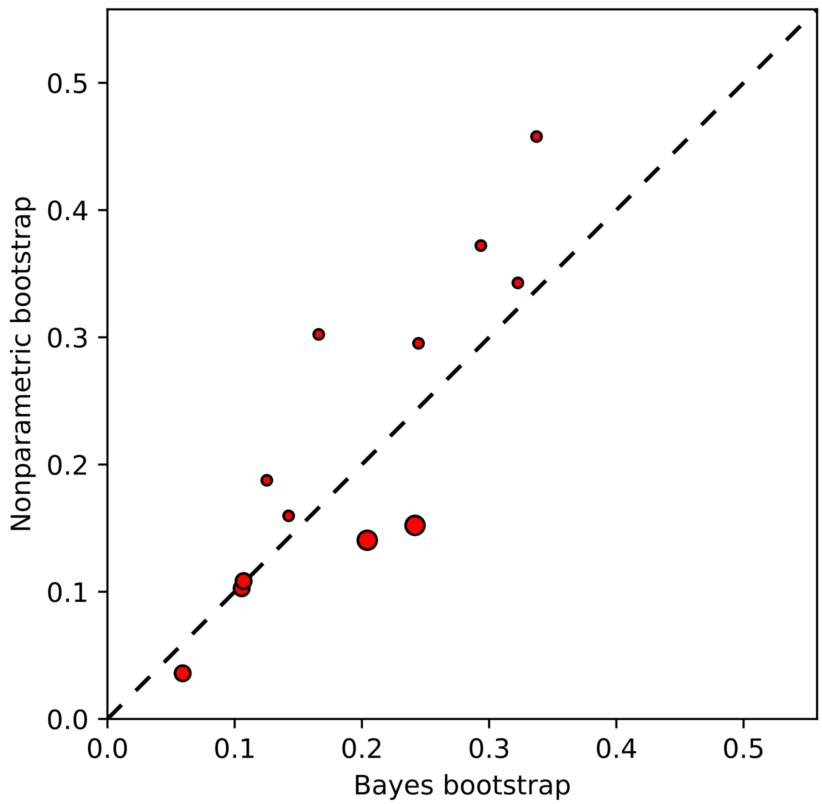
Crossings	Num. objects
0	37
1	34
2	9
3	1

Panel B: Illustration of algorithm



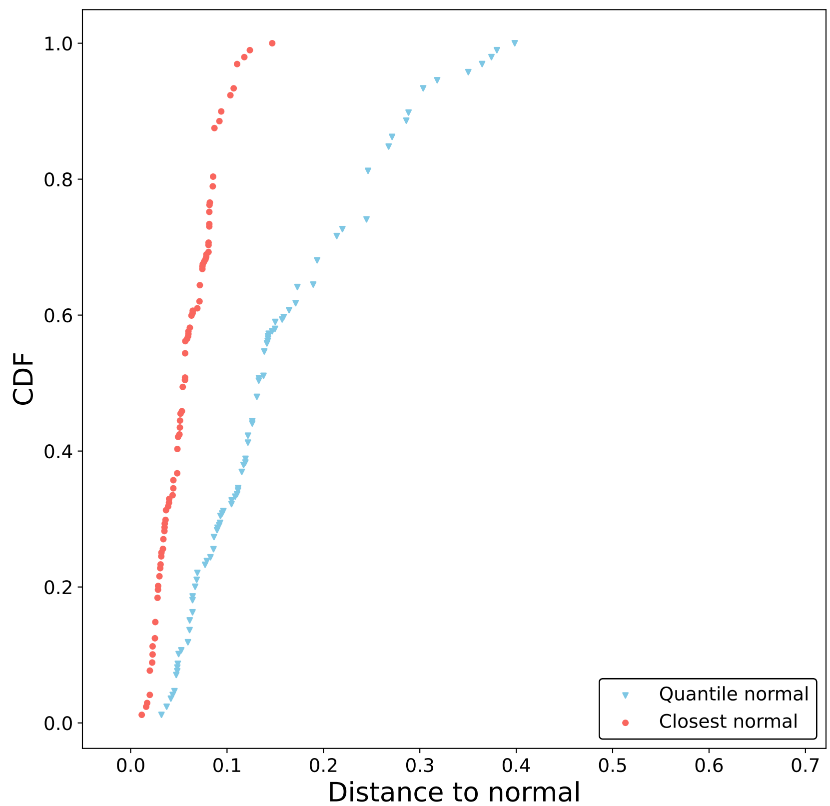
Notes: Panel A shows the minimum number of crossings between the density of the bootstrap distribution and the density of the default normal report that is statistically consistent with the authors' replicates. We compute the minimum number of crossings consistent with the replicates by finding the path through a 95% uniform (DKW) confidence band on the difference in CDFs between the bootstrap distribution and the default normal distribution that changes direction the least number of times. Panel B illustrates the algorithm used to determine this path. The algorithm traverses either the least increasing or greatest decreasing path through the uniform confidence band, changing direction only when necessary to stay within the band. The illustration is drawn for the object of interest with the largest signed Kolmogorov distance between the bootstrap replicates and the default normal report, with the values of replicates normalized by dividing by their standard deviation.

Appendix Figure 1: Signed Kolmogorov Distance, Nonparametric vs. Bayes Bootstrap



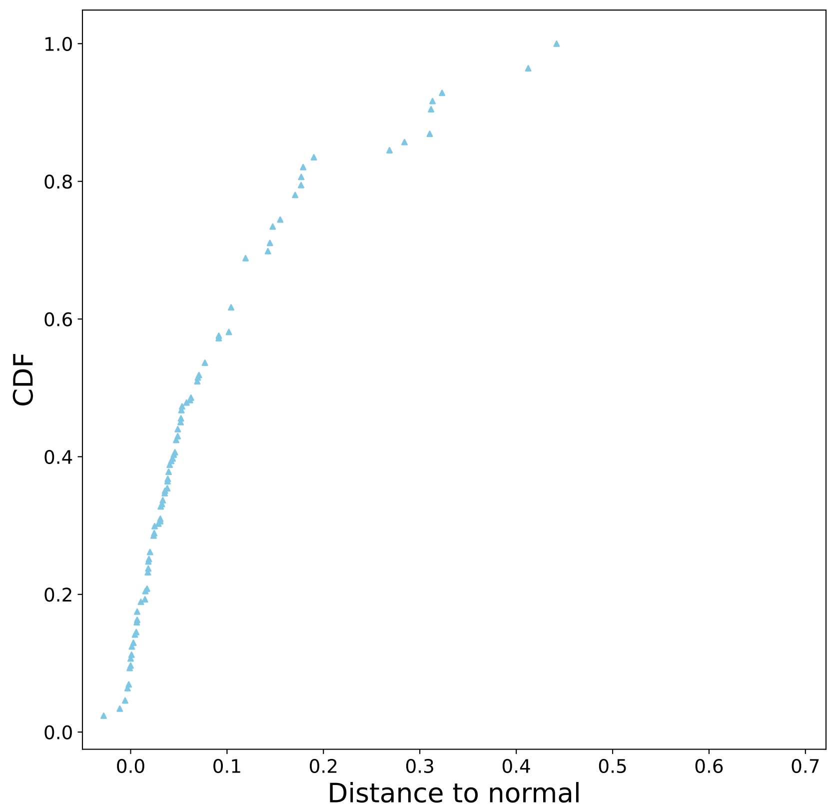
Notes: The plot is a scatterplot. The unit of analysis is an object of interest, with the area of each point inversely proportional to the total number of objects of interest in the same paper as the given object. We include objects of interest for which we were able to compute a Bayes bootstrap by adapting the authors' original bootstrap code. The y-axis reports the signed Kolmogorov distance between the distribution of a set of nonparametric bootstrap replicates and the default normal report, whose mean is given by the point estimate and whose standard deviation is given by the bootstrap standard error. The x-axis reports the signed Kolmogorov distance between the distribution of a set of Bayes bootstrap replicates and the default normal report. In both cases, the number of replicates is equal to the number used in the authors' original bootstrap. For all objects of interest, the displayed 45-degree line passes through a rectangle formed as the Cartesian product of a 95% confidence interval for the signed Kolmogorov distance between the nonparametric bootstrap distribution and the default normal report, and a 95% confidence interval for the signed Kolmogorov distance between the Bayes bootstrap distribution and the default normal report, where these confidence intervals are constructed based on 95% uniform (DKW) bands for the bootstrap distributions.

Appendix Figure 2: Distribution of Signed Kolmogorov Distance to Normal Report, Using Quantiles



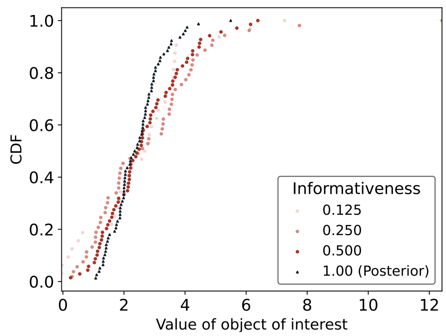
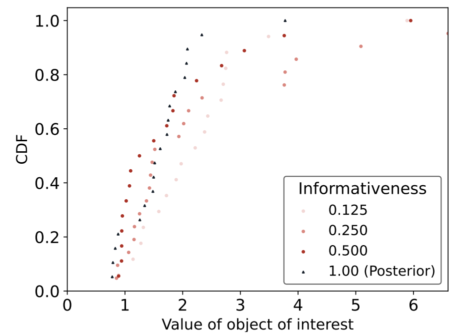
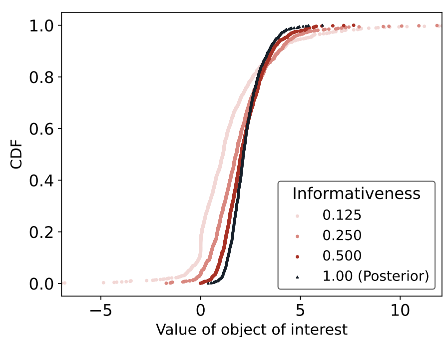
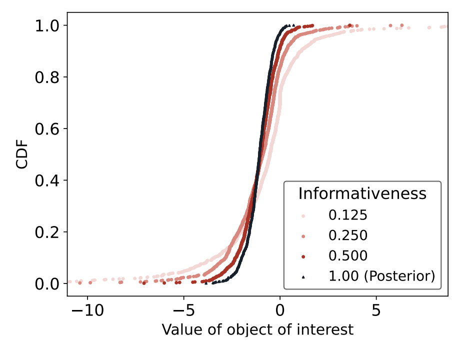
Notes: The plot is a weighted empirical CDF. The unit of analysis is an object of interest and, for each paper, we weight each object of interest by the inverse of the number of objects of interest associated with the paper. For each object of interest we calculate the signed Kolmogorov distance between the distribution of bootstrap replicates and the quantile normal report, whose mean is given by the point estimate and whose standard deviation is taken to match the difference between the 97.5th and 2.5th quantiles of the empirical distribution of the replicates. We also calculate the signed Kolmogorov distance between the distribution of the bootstrap replicates and the closest normal report, whose mean and standard deviation are chosen to minimize SK distance. The plot shows the weighted empirical CDF of each of these two distances across the objects of interest in our census.

Appendix Figure 3: Distribution of Total Variation Distance to Default Normal Report



Notes: The plot is a weighted empirical CDF. The unit of analysis is an object of interest and, for each paper, we weight each object of interest by the inverse of the number of objects of interest associated with the paper. For each object of interest we estimate the total variation distance between the bootstrap distribution and the default normal report, whose mean is given by the point estimate and whose standard deviation is given by the bootstrap standard error. The plot shows the weighted empirical CDF of this distance across the objects of interest in our census. To estimate the total variation distance, we smooth the empirical distribution of the J bootstrap replicates using a kernel density smoother with Gaussian kernel. To choose the bandwidth of the kernel, we take J draws from the default normal report, smooth the draws using a kernel density smoother with Gaussian kernel, and choose the smoother's bandwidth to minimize the total variation distance between the smoothed distribution of the J draws from the default normal report and the exact distribution of the default normal report. We then take the total variation distance between the smoothed distribution of the bootstrap replicates and the exact distribution of the default normal report, subtracting the total variation distance between the smoothed normal draws and the exact distribution of the default normal report as a bias correction.

Appendix Figure 4: Illustration of the Class of Reference Priors

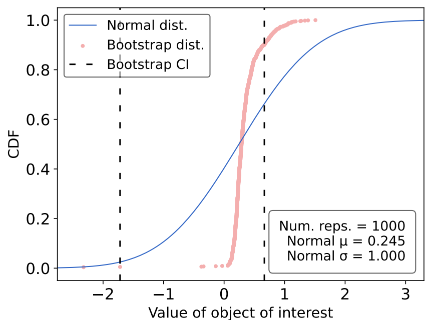
Paper	**Least SK distance**	**Greatest SK distance**
3		
9		

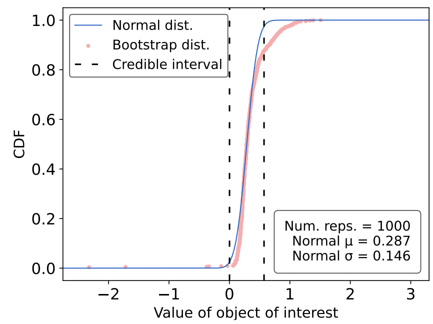
Notes: Each row corresponds to an article in our bootstrap census for which it is possible to compute a Bayes bootstrap by modifying the authors' original bootstrap code. Each plot shows draws from the Dirichlet process distribution $DP(\alpha n, P_n)$ for several values of α in powers of 2, where $\alpha = 1$ corresponds to the Bayes bootstrap distribution and $\alpha \to 0$ corresponds to the limiting prior whose posterior is the Bayes bootstrap distribution. The number of draws is given by the number of bootstrap replicates in the authors' original bootstrap procedure. Each row includes two plots, corresponding to the two objects in Figure 3, with rows following the order in Figure 3. All values are normalized by dividing by the bootstrap standard error. Draws more than 10 bootstrap standard errors away from the mean are excluded from the plots.

Appendix Figure 5: Illustration of Bootstrap Confidence Interval

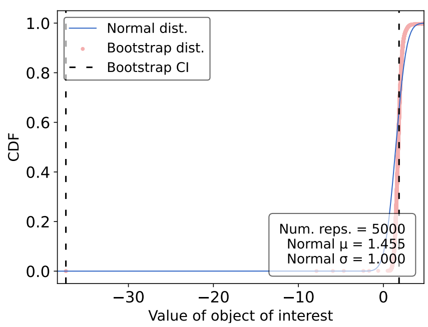
Paper	**Comparing bootstrap distribution to...**	
	default normal report and bias-corrected bootstrap confidence interval	**distance-minimizing normal report and associated credible interval**

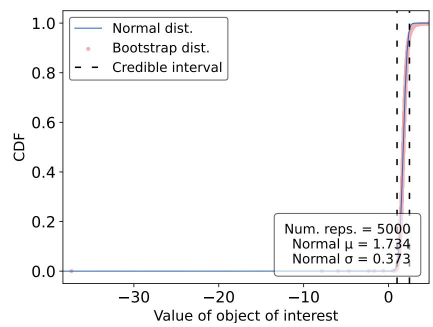
1



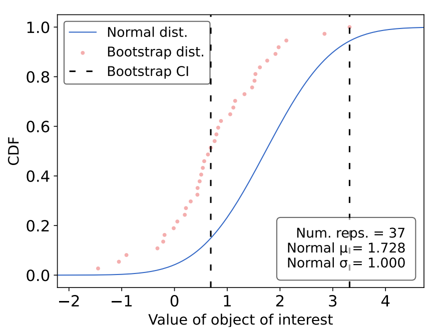


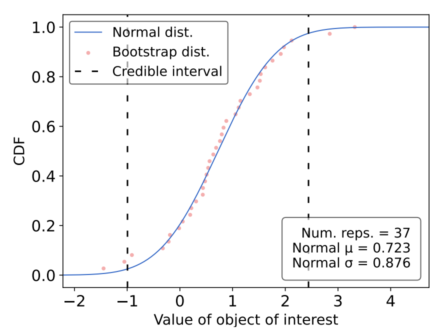
2



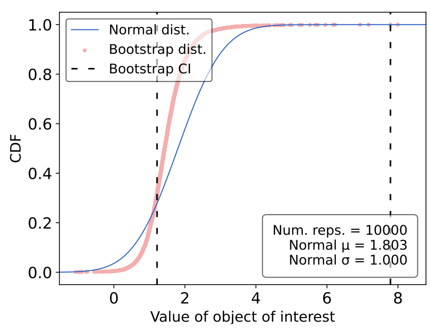
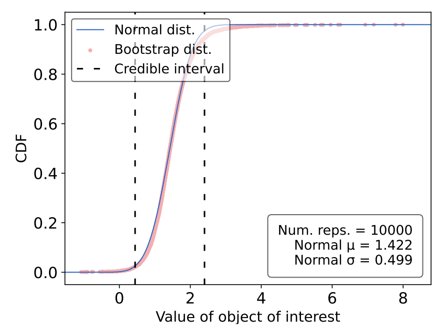
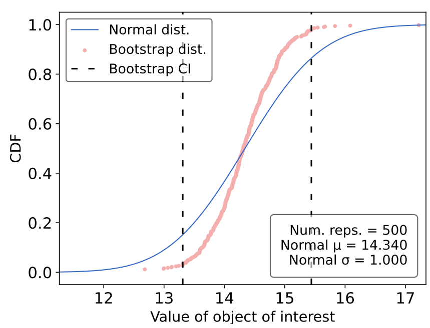
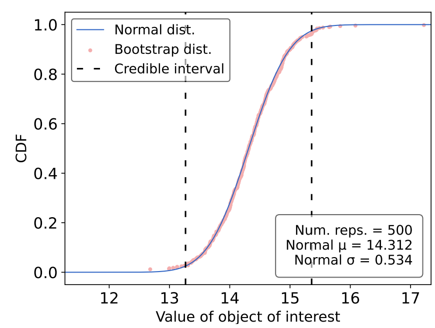
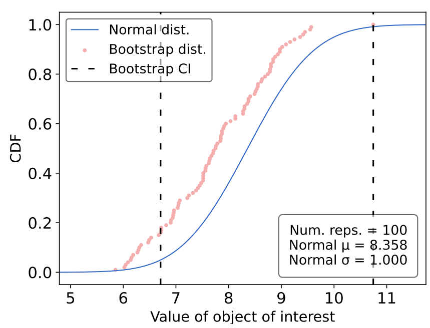
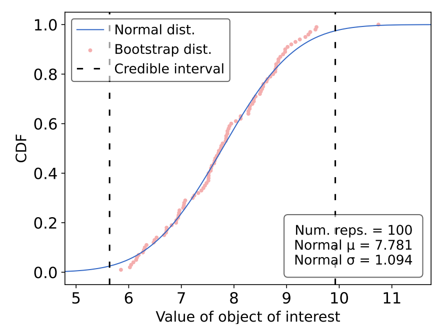


3

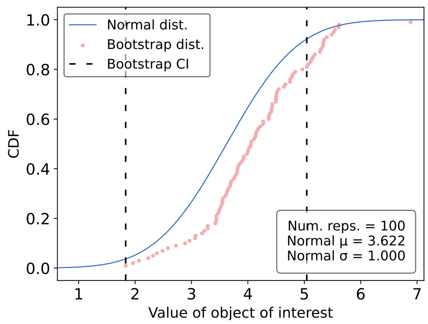
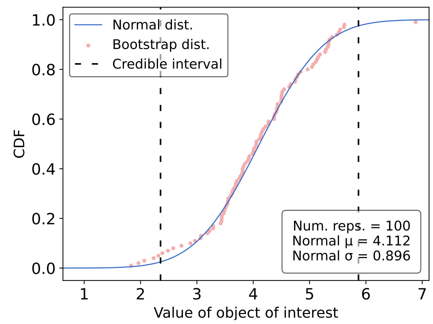
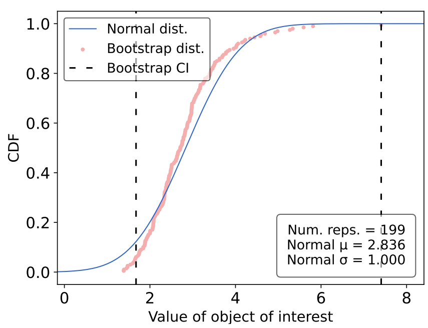
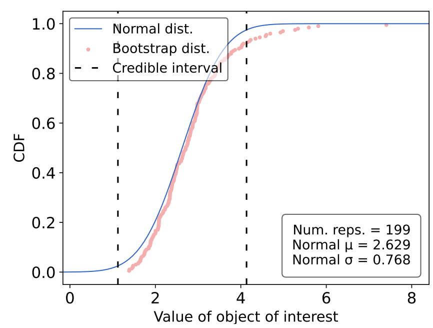
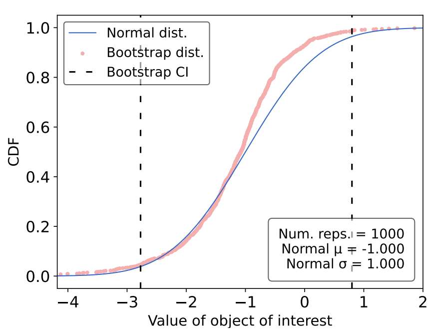
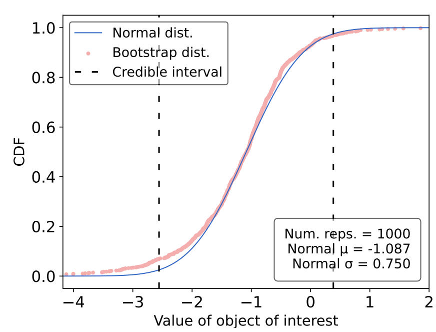




Appendix Figure 5 (continued): Illustration of Bootstrap Confidence Interval

Paper	**Comparing bootstrap distribution to...**	
	default normal report and bias-corrected bootstrap confidence interval	**distance-minimizing normal report and associated credible interval**
4		
5		
6		

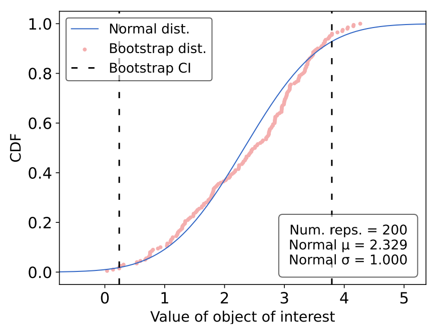
Appendix Figure 5 (continued): Illustration of Bootstrap Confidence Interval

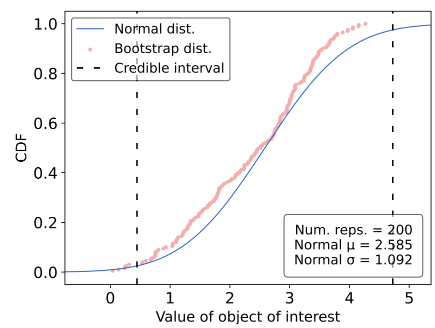
Paper	**Comparing bootstrap distribution to...**	
	default normal report and bias-corrected bootstrap confidence interval	**distance-minimizing normal report and associated credible interval**
7		
8		
9		

Appendix Figure 5 (continued): Illustration of Bootstrap Confidence Interval

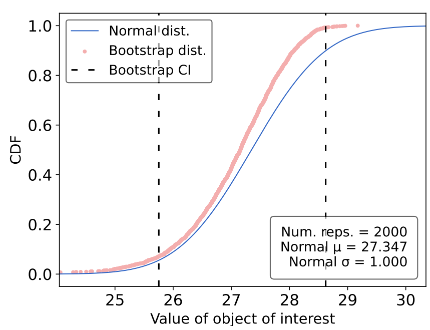
Paper	**Comparing bootstrap distribution to...**	
	default normal report and bias-corrected bootstrap confidence interval	**distance-minimizing normal report and associated credible interval**

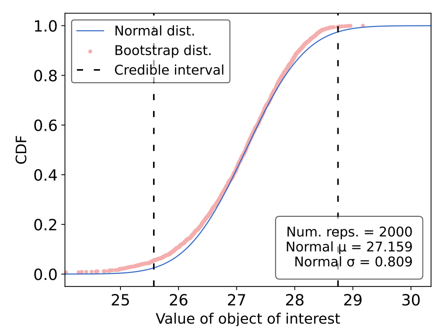
10



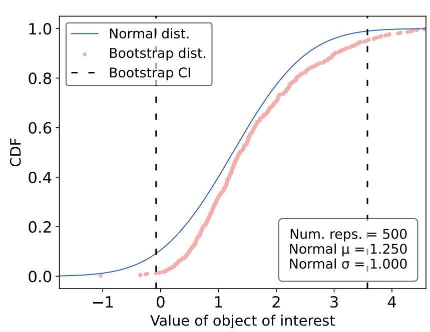


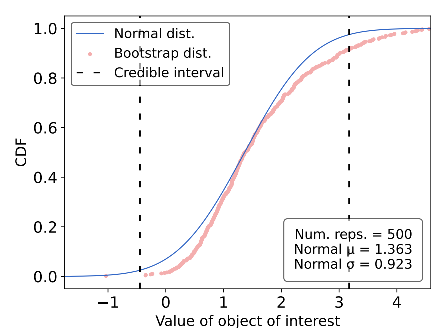
11



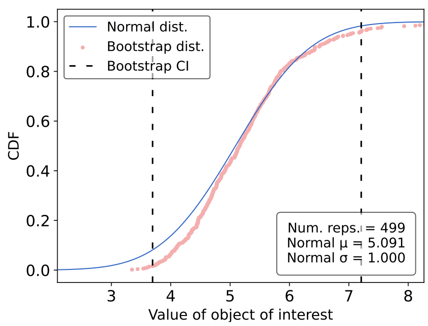
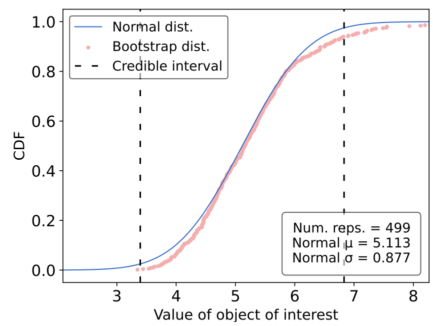
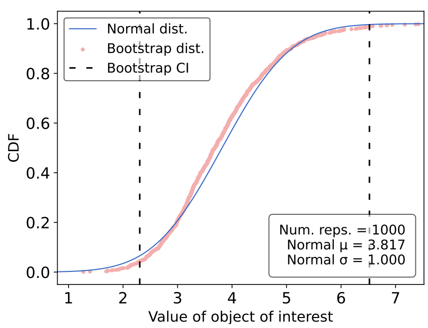
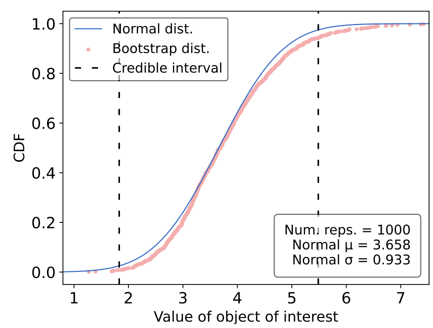


12





Appendix Figure 5 (continued): Illustration of Bootstrap Confidence Interval

Paper	**Comparing bootstrap distribution to...**	
	default **normal report and bias-corrected bootstrap confidence interval**	**distance-minimizing normal report and associated credible interval**
13		
14		

Notes: Each row corresponds to an article in our bootstrap census and focuses on the object of interest with the greatest signed Kolmogorov distance as defined in Figure 3. The first column compares the empirical CDF of the bootstrap replicates to the CDF of the default normal report, whose mean is given by the point estimate and whose standard deviation is given by the bootstrap standard error. In the first column, the vertical dashed lines depict the endpoints of a bias-corrected 95% confidence interval (CI). The second column compares the empirical CDF of the bootstrap replicates to the CDF of the distance-minimizing normal report, whose mean and standard deviation are chosen to minimize the signed Kolmogorov distance to the distribution of the bootstrap replicates. In the second column, the vertical dashed lines depict the endpoints of a centered 95% credible interval based on the distance-minimizing normal report. Each plot legend reports the number of replicates and the mean and standard deviation of the normal. All values are normalized by dividing by the standard deviation of the bootstrap replicates. Rows (papers) are in descending order according to their greatest signed Kolmogorov distance.