NBER WORKING PAPER SERIES

COMMUNICATING SCIENTIFIC UNCERTAINTY VIA APPROXIMATE POSTERIORS

Isaiah Andrews
Jesse M. Shapiro

Working Paper 32038
http://www.nber.org/papers/w32038

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2024, Revised September 2025

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 Giuseppe Cavaliere, Clément de Chaisemartin, Luca Fanelli, Anna Mikusheva, Jon Roth, and conference and seminar participants at the Princeton Day of Statistics, Harvard, MIT, the University of Tokyo, USC, UCLA, the IAAE, Yale, and the Reinhardt Lecture for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w32038

Communicating Scientific Uncertainty via Approximate Posteriors
Isaiah Andrews and Jesse M. Shapiro
NBER Working Paper No. 32038
January 2024, Revised September 2025
JEL No. C18, C44, D81

ABSTRACT

We cast the problem of communicating scientific uncertainty as one of reporting a posterior distribution on an unknown parameter to an audience of Bayesian decision-makers. We establish novel bounds on the audience's regret when the analyst reports an approximation to a posterior that the audience treats as exact. Under a palatable restriction on the audience's decision problems, the bounds take an especially convenient form. Under a further restriction on the audience's priors, a bootstrap distribution can be used as a stand-in posterior. We propose a practical recipe for checking whether a conventional statistical report (say, a normal parameterized by a point estimate and standard error) is a good approximation, and for improving the report if it is not. We illustrate our proposals using the articles in the 2021 *American Economic Review* that use a bootstrap for inference.

Isaiah Andrews
Massachusetts Institute of Technology
Department of Economics
and NBER
iandrews@mit.edu

Jesse M. Shapiro
Harvard University
Department of Economics
and NBER
jesse_shapiro@fas.harvard.edu

A data appendix is available at http://www.nber.org/data-appendix/w32038
A Python package is available at https://github.com/JMSLab/BootstrapReport
A web app is available at https://jmslab.shinyapps.io/BootstrapReport

1 Introduction

Communicating uncertainty about one's conclusions is a fundamental task of scientists. Within empirical economics a conventional approach is to report a statistic, often a standard error, alongside the point estimate of a target parameter (Athey and Imbens, 2023). An important question is whether, and when, this convention accurately conveys the uncertainty about the target parameter.

In this paper we propose to approach the problem of communicating uncertainty about a target parameter from a Bayesian perspective. Theoretically, we argue that this perspective leads to a natural notion of what is a "good" summary of the uncertainty about the target parameter. Practically, we argue that this perspective leads to a widely applicable toolkit for diagnosing and addressing failures of the conventional report.

We begin with the theoretical question of what constitutes a good summary of uncertainty for a Bayesian audience. We formulate this question in a model of statistical communication following Andrews and Shapiro (2021). An analyst makes a report about a target parameter to an audience of agents. Each agent faces a decision problem with an associated loss function, and holds a prior in a density ratio neighborhood of some *central prior*. Different agents may face different decision problems and hold different priors, reflecting heterogeneity in the uses of scientific knowledge. Nevertheless, well-known results imply that reporting the *central posterior* that results from updating the central prior is as good—in terms of the loss it induces—as reporting the full data.

We ask what harm is done if, instead of reporting the central posterior, the analyst reports an approximation to it, and the audience treats the approximation as if it is exact. For an agent with a given prior and loss, we measure the harm by the agent's *posterior regret*: the expected loss from the action they will take based on the approximation, minus the expected loss from the best action they could take given the data. We establish a fundamental bound on the posterior regret. The bound has an intuitive economic structure, and is sharp when the agents' decision problems are sufficiently rich.

We turn next to the question of when to accept a given *conventional report*, such as a point estimate and standard error. We allow that there is a benefit to reporting a conventional approximation to the central posterior, for example because of the convention's simplicity and familiarity. Our theoretical results suggest a natural *procedure* for the analyst

to follow in this case. Compute the conventional report, the central posterior, and the implied regret bound. If the regret bound is smaller than the benefit of adhering to the convention, make the conventional report. If the regret bound is larger than the benefit of adhering to the convention, report the central posterior. This procedure ensures that the audience's regret is no greater than the benefit of adhering to the convention.

To make this procedure a practical one, we need three ingredients: a conventional report, a class of audience decision problems, and a central posterior. For the conventional report, we focus on the practice of reporting a point estimate and standard error, which we model as reporting a normal approximation to the central posterior. We motivate this convention by recalling that Bernstein-von Mises (BvM) theory establishes a range of conditions on a prior and parameter under which the corresponding posterior is well approximated by a normal distribution in large samples. If the central posterior for the target parameter satisfies the conditions for a BvM theorem, then well-known results imply a variety of senses in which a normal distribution parameterized by the point estimate and standard error is a good approximation to the central posterior.

For the class of audience decision problems, we focus on the case where agents' loss functions and density ratio functions are bounded and of bounded increasing and decreasing variation. We argue that this class is sufficiently inclusive for use in practice. Under the proposed assumption, we prove that the bound on regret is itself bounded, both above and below, by scalar multiples of the signed Kolmogorov (SK) metric between the central posterior and the conventional approximation. The SK metric measures the distance between two distributions on the real line by adding together the magnitudes of the largest positive and negative vertical distances between their CDFs. Conveniently, the SK metric can therefore be read directly off of a p-p plot.

For the central posterior, the analyst may have a subjective prior whose implied posterior can serve in this role. To accommodate the remaining (and, we suspect, many) situations where this is not the case, we extend our results to show that if the analyst uses a *surrogate central posterior* that is close to the central posterior, the bounds we derive continue to apply up to a slack term that depends on the difference between the surrogate central posterior and the true central posterior.

We propose to use the Bayesian bootstrap distribution (Rubin 1981, Gasparini 1995, Chamberlain and Imbens 2003) as a default surrogate central posterior. We characterize

two situations in which this is reasonable. One situation is where the central prior is a Dirichlet process, in which case existing results show that the Bayes bootstrap distribution is the limit of a sequence of posteriors as the informativeness of the prior becomes low. We adapt these results to show conditions under which the Bayes bootstrap distribution is close the central posterior in the relevant metric. The other situation is where the target parameter is a functional of some other parameter, for example the CDF of the data or a vector of regression coefficients, that obeys a BvM theorem under both the bootstrap distribution and the central prior. In that case, we show that the Bayes bootstrap distribution serves as a surrogate with high probability as the sample grows large. Under these same conditions, we show that the Bayes bootstrap distribution is similar to that of other weighted bootstraps, such as the nonparametric bootstrap, so that these other bootstraps can reasonably be used when more convenient.

We thus arrive at a practical recipe for researchers in empirical economics. Calculate the point estimate and standard error. Sample from a bootstrap distribution. Make a p-p plot comparing the two distributions. If the distributions are close in SK metric, make the conventional report. If not, report the bootstrap distribution directly, for example with a histogram or CDF. An accompanying Python package and web app called `BootstrapReport` facilitate adoption.

We illustrate practicality by applying the recipe to the universe of articles in the 2021 *American Economic Review* which bootstrap some target parameter, 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 some target parameters, the conventional report approximates the bootstrap distribution well. For others, it does not. The quality of the approximation can differ meaningfully even across different parameters targeted in the same article.

This paper makes two main contributions. The main theoretical contribution is a set of novel bounds on the posterior regret from reporting an approximate central posterior to an audience of decision-makers who believe they have received an exact report.[1] In addition to those we explore here, these bounds may have other uses. For example, in situations where the analyst performs an explicitly Bayesian analysis, and wishes to provide a low-

[1] Our focus on communicating a useful summary of data to an audience of decision-makers is related to research on omniprediction (e.g., Gopalan et al. 2022) and sequential calibration (e.g., Noarov et al. 2025), though our setting, problem statement and, consequently, proposed solutions, are different.

dimensional summary of the posterior distribution for a target parameter (e.g., a posterior mean and standard deviation), our results can be used to bound the harm from doing so.[2]

The main practical contribution is a procedure for evaluating the quality of a normal (or other) approximation to the uncertainty in an estimated parameter. Previous work in the frequentist tradition (e.g., Beran, 1997; Zhan, 2018; Angelini et al., 2022, 2024) has proposed using a bootstrap to diagnose failures of conventional asymptotic approximations.[3] In contrast to this work, and to much other previous work on uncertainty quantification in economics, our approach neither aims for, nor achieves, frequentist guarantees such as size control.[4] Instead, our approach focuses on limiting the potential for bad decision-making by Bayesian agents treating the reported approximation as a central posterior.

There are two main reasons we believe the Bayesian communication perspective we adopt here is useful for empirical economics. First, the description of agents who treat the scientist's report as a summary of a revised belief (posterior) about the target parameter matches well with how we and, we suspect, others, consume empirical research. Second, the use of a Bayesian perspective allows us to make certain statements without reference to asymptotic or other approximations. In particular, if the bootstrap distribution differs meaningfully from the default report, our results establish that there exists *some* agent who would be meaningfully better off basing their decision-making directly on the bootstrap distribution.

Wide applicability of our procedure does require substantive assumptions. When these assumptions are reasonable, our procedure affords a recipe for quantifying uncertainty about a target parameter without assuming approximate normality of the sampling distribution of the estimator, and without specializing the recipe for particular use cases. This is in contrast to the large frequentist literature on failures of the conventional normal approximation, which requires specialized treatments depending on the nature of the target

[2]Cohen and Einav (2007), for example, summarize many marginal posterior distributions with two statistics, the posterior mean and standard deviation. Smets and Wouters (2007) summarize with four statistics, the mode, mean, and 5th and 95th quantiles. In each case, if we think of these statistics as parameterizing some family of approximating distributions, our bound can be used directly to evaluate the quality of the resulting approximation to the sampled posterior.

[3]See also Cavaliere et al. (2025). Previous work in the frequentist tradition also recommends reporting richer bootstrap information than the standard error (see, e.g., Efron 1982, Hall 1992, and Hahn and Liao 2021). Wang (2025) suggests diagnosing approximation failures for GMM via a comparison to a quasi-Bayes posterior distribution.

[4]Indeed, for the frequentist goal of describing the sampling distribution of the target parameter, results in Singh (1981) and Weng (1989) suggest that the nonparametric bootstrap is more accurate than the Bayes bootstrap that we recommend as the default surrogate for the central posterior (see also Hall 1992).

parameter and the reason for the failure of the conventional normal approximation, and which aims at goals, such as size control, that differ from those we pursue here.

The remainder of the paper is organized as follows. Section 2 introduces our communication framework and our main theoretical result bounding the regret from reporting an approximate posterior. Section 3 introduces the conventional report and the procedure to control regret. Section 4 introduces the class of audience decision problems, and the theoretical results, that justify the use of the SK metric. Section 5 introduces the idea of a surrogate central posterior and develops the use of the bootstrap in this role. Section 6 presents the findings from our census of articles in the 2021 *American Economic Review* that use the bootstrap. A running example throughout the article serves to illustrate the main ideas and limitations. An Appendix contains proofs of all results stated in the main text. A Supplemental Appendix contains additional theoretical results, illustrations for the running example, and findings from the bootstrap census.

2 A Bound on Regret from Approximate Posterior 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 fundamental bound on the regret from reporting an approximate posterior belief on the target parameter. We develop practical implications in later sections.

2.1 Communicating a Posterior Report

Consider an analyst who observes data $X \in \mathcal{X}$. The analyst reports some function of the data to an audience of agents, who each must take an action $a \in \mathcal{A}$.[5] The consequences of each agent's decision depend on the value of an unknown *target parameter* $\theta \in \Theta$. Specifically, each agent is endowed with a loss function $L : \mathcal{A} \times \Theta \to [0,\lambda]$ in some class $\mathcal{L}$, so that an agent with loss function $L \in \mathcal{L}$ who takes action $a \in \mathcal{A}$ realizes loss $0 \leq L(a,\theta) \leq \lambda < \infty$ when the target parameter's true value is $\theta \in \Theta$. The assumption that the loss is bounded by λ implies that the decisions taken based on the research findings have bounded stakes.

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

Each agent is further endowed with a prior $\pi \in \Delta(\Theta \times \mathcal{X})$ that is dominated by some *central prior* π^*. We can describe the agent's prior with a marginal prior $\pi(\theta) \in \Delta(\Theta)$ on the target parameter, and a conditional prior $\pi(X|\theta) \in \Delta(\mathcal{X})$ on the data given the target parameter. We assume that each agent's conditional prior given θ agrees with that of the central prior, so that $\pi(X|\theta) = \pi^*(X|\theta)$ for all $\theta \in \Theta$.[6] All prior disagreement in the audience therefore concerns the marginal prior, and we can compactly describe any prior π by a density ratio function $w(\theta) = \frac{d\pi}{d\pi^*}(\theta)$. We let $\mathcal{W}$ denote a class of such density ratio functions, $w \in \mathcal{W}$. We assume that the class $\mathcal{W}$ includes $w(\theta) = 1$, which corresponds to an agent whose prior agrees with the central prior.

Example. (Get out the vote.) Our running example follows Nickerson et al. (2006). The analyst observes the results of an experiment in which $\left(N^T, N^C\right)$ citizens are chosen at random to either receive (T) or not receive (C) a phone call encouraging them to vote. The data consist of the number $X = \left(X^T, X^C\right)$ of citizens in each group who vote. All agents believe that the data are independently binomially distributed according to $X^T | N^T, \beta^T \sim B\left(N^T, \beta^T\right)$ and $X^C | N^C, \beta^C \sim B\left(N^C, \beta^C\right)$ where $\beta^T, \beta^C \in [0,1]$ are probabilities.

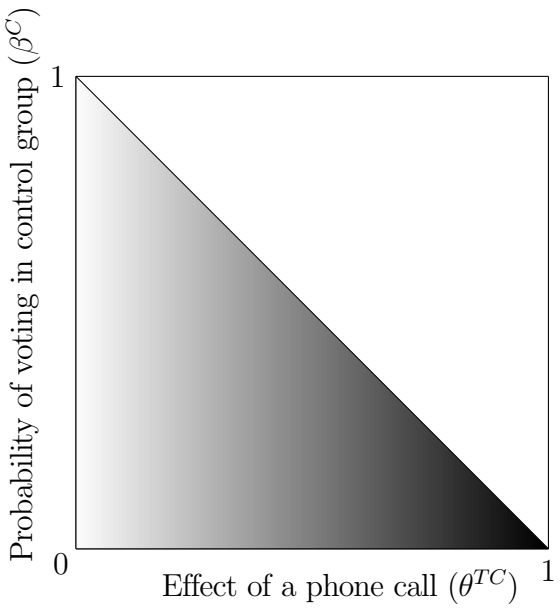
The agents in the audience are campaign managers for local candidates who must decide what share $a \in [0,1] = \mathcal{A}$ of a fixed budget to devote to phone calls instead of other campaign activities. For a given campaign manager, the loss $L(a, \theta) \in [0,1]$ is the probability that their candidate is defeated.

The loss depends on the effectiveness of a phone call, which is encoded in the probabilities $\left(\beta^T, \beta^C\right)$. If the target parameter θ is maximally rich, $\theta = \left(\beta^T, \beta^C\right)$, then the assumption that all disagreement concerns the marginal prior is without loss. If instead the parameter of interest is more restrictive, then our assumptions on the form of disagreement are more substantive.

To illustrate, suppose that the loss depends on $\theta^{TC} = \beta^T - \beta^C \in [-1,1] = \Theta$, the effect of a phone call on the probability of voting. Suppose further that, under the central prior π^*, we have that $\theta^{TC} \sim U[\underline{\theta}, 1]$, for $\underline{\theta} > 0$ a small value, and $\beta^C | \theta^{TC} \sim U\left[0, 1 - \theta^{TC}\right]$, which together suffice to also describe $\beta^T | \theta^{TC}$ and therefore, given the binomial likelihood, the distribution of the data X. Figure 1 depicts the density implied by the central prior.

[6] Here and throughout, we use $\pi(\theta)$ and $\pi(X|\theta)$ as shorthand for the measures $\pi_\theta(B)$ for $B \subseteq \Theta$ and $\pi_{X|\theta}(B|\theta)$ for $B \subseteq \mathcal{X}$, respectively. Other similar notation may be read analogously, and our assumption that each agent's conditional prior agrees with that of the central prior may be read as $\pi_{X|\theta}(B|\theta) = \pi^*_{X|\theta}(B|\theta)$ for all measurable sets $B \subseteq \mathcal{X}$.

Figure 1: Central prior in the example



Notes: The plot shows the density of the joint central prior on the probability β^C of voting for citizens in the control group, and the effect θ^{TC} of a phone call on the probability of voting.

Under the assumption that $\pi(X|\theta) = \pi^*(X|\theta)$, audience members all agree that $\beta^C|\theta^{TC} \sim U\left[0,1-\theta^{TC}\right]$, but may disagree that $\theta^{TC} \sim U[\underline{\theta},1]$. Notice that, under all such prior distributions, we have that $\beta^T > \beta^C a.s.$, i.e., that the intervention increases turnout.

Certain forms of disagreement can be incorporated directly into the loss function, rather than the prior. Suppose, for example, that a particular campaign manager believes that the experiment is not externally valid, so that θ^{TC} overstates the true causal effect of a phone call by a factor of $\eta \geq 1$. We can write such a manager's loss as $L\left(a,\theta^{TC}\right) = \tilde{L}\left(a,\theta^{TC}/\eta\right)$ for $\tilde{L}$ a loss that is parameterized by the true causal effect θ^{TC}/η. $\triangle$

What remains is to describe the analyst's report. The analyst's information consists of the data X. If an agent with prior π were to directly observe the data X, their posterior belief about the target parameter would be $\pi(\theta|X)$. It follows that, from the agent's standpoint, no report can improve upon a report of $\pi(\theta|X)$. Importantly, the same is also true of a report that consists of the *central posterior* $\pi^*(\theta|X)$. The reason is that each agent's posterior is just a reweighting of the central posterior, i.e., $\pi(\theta|X) \propto w(\theta)\pi^*(\theta|X)$.[7]

[7]Here and throughout, we use $w(\theta)\pi^*(\theta|X)$ to denote the measure defined by $\mu(A) = \int_A w(\theta) d\pi^*(\theta|X)$

This is a classical justification for reporting posterior distributions (see, e.g., Raiffa and Schlaifer 1961, Hildreth 1963, and Geweke 1997).

A leading possibility is, then, that the analyst reports $\pi^*(\theta|X)$. For reasons we discuss below, the analyst may instead prefer to report some other distribution $\hat{\pi}^*(\theta|X)$, for example as an approximation to $\pi^*(\theta|X)$. We assume that $\hat{\pi}^*(\theta|X)$ is absolutely continuous with respect to the central prior π^*, and denote by $\Delta_{\pi^*}(\Theta)$ the set of such distributions. We next measure the harm to the agents if $\hat{\pi}^*(\theta|X) \neq \pi^*(\theta|X)$.

2.2 Regret from an Approximate Posterior Report

We assume that each agent treats the analyst's report as if it is a report of the central posterior $\pi^*(\theta|X)$. Each agent therefore forms their perceived posterior belief $\hat{\pi}(\theta|X)$, with $\hat{\pi}(\theta|X) \propto w(\theta)\hat{\pi}^*(\theta|X)$, by reweighting the report. Each agent then seeks to minimize their perceived posterior expected loss $\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[L(a,\theta)]$.[8] A perceived optimal action is then

$$\hat{a} \in \operatorname*{argmin}_{a \in \mathcal{A}} \mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[L(a,\theta)].$$

For concreteness, we proceed here as if an optimal action exists. The proofs in the Appendix cover the more general case where an optimal action may not exist, in which case the bounds we derive below continue to hold up to an arbitrarily small slack term.

If the analyst's report differs from the central posterior, an agent's perceived optimal action $\hat{a}$ may differ from their true posterior optimal action conditional on the data. We define an agent's *posterior regret* to be the increase in expected loss from taking the perceived optimal action $\hat{a}$ rather than the true optimal action. Formally, given data $X \in \mathcal{X}$, an agent with loss $L \in \mathcal{L}$ and density ratio $w \in \mathcal{W}$ has regret from the report $\hat{\pi}^*(\theta|X)$ given by

$$R(\hat{\pi}^*(\theta|X); X, L, w, \pi^*) = \mathrm{E}_{w(\theta)\pi^*(\theta|X)}[L(\hat{a},\theta)] - \min_{a \in \mathcal{A}} \mathrm{E}_{w(\theta)\pi^*(\theta|X)}[L(a,\theta)],$$

where, importantly, both expectations are taken with respect to the agent's true posterior

for measurable $A \subseteq \Theta$.

[8] Here and throughout, for any measure μ on Θ and any function $f : \Theta \to \mathbb{R}$, we let the operator $\mathrm{E}_\mu[f(\theta)] = \frac{\int_\Theta f(\theta) d\mu(\theta)}{\int_\Theta d\mu(\theta)}$ denote the expectation of f with respect to the normalized measure $\mu / \int_\Theta d\mu(\theta)$. If μ is a probability measure, so that $\int_\Theta d\mu(\theta) = 1$, then $\mathrm{E}_\mu[\cdot]$ corresponds to the usual expectation.

$\pi(\theta|X) \propto w(\theta)\pi^*(\theta|X)$.[9]

2.3 A Fundamental Bound on Regret

Our goal is to bound $R(\hat{\pi}^*(\theta|X);X,L,w,\pi^*)$. As notation, for any class $\mathcal{F}$ of functions $\mathcal{Y}\times\Theta\to\mathbb{R}$, we let $\mathcal{F}(\mathcal{Y})$ denote the class of functions formed by fixing a value of $y\in\mathcal{Y}$ and a function $f\in\mathcal{F}$, i.e., functions of the form $f(y,\cdot)$, and we let $f(\mathcal{Y})\subseteq\mathcal{F}(\mathcal{Y})$ denote the set of such functions formed using a particular $f\in\mathcal{F}$. For ease of exposition, we suppose that the set of pairs of expectations of differences of pairs of functions in $\mathcal{L}(\mathcal{A})$ under the true and perceived posteriors is compact.[10] If this condition fails, as we discuss in the Appendix the sharpness result that we state below continues to hold up to an arbitrarily small slack term.

We then have the following result.

Theorem 1. (Fundamental bound on regret) *For any report $\hat{\pi}^*(\theta|X)\in\Delta_{\pi^*}(\Theta)$, data $X\in\mathcal{X}$, loss $L\in\mathcal{L}$, and density ratio $w\in\mathcal{W}$, the regret $R(\hat{\pi}^*(\theta|X);X,L,w,\pi^*)$ is bounded above by*

$$\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)=\sup_{\ell_1,\ell_0\in\mathcal{L}(\mathcal{A}),w\in\mathcal{W}} \mathrm{E}_{w(\theta)\pi^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]$$
$$\text{s.t. } \mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]\leq 0.$$

Moreover, there exist $\overline{w}\in\mathcal{W}$ and $\overline{L}:\mathcal{A}\times\Theta\to\mathbb{R}$ such that $\overline{L}(\mathcal{A})\subseteq\mathcal{L}(\mathcal{A})$ and $R\big(\hat{\pi}^*(\theta|X);X,\overline{L},\overline{w},\pi^*\big)=\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$.

Intuitively, the regret bound $\overline{R}$ measures the difference between the approximation $\hat{\pi}^*(\theta|X)$ and the true central posterior $\pi^*(\theta|X)$, an interpretation that we return to later. To motivate the form of the bound, observe that if we fix an action $a\in\mathcal{A}$, any loss function $L\in\mathcal{L}$ implies a function $\ell(\theta)=L(a,\theta)$ only of the target parameter. Consider an agent choosing between actions a_0 and a_1, with corresponding $\ell_0,\ell_1\in\mathcal{L}(\mathcal{A})$. The worst case for this agent arises when ℓ_1 has a weakly lower perceived posterior expectation than ℓ_0, but a much larger true posterior expectation, since in this case the agent is willing to make

[9] If multiple optimal actions $\hat{a}$ exist, we consider the selection among them that yields the largest regret, i.e., the selection that maximizes the value of $\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[L(\hat{a},\theta)]$.

[10] That is, we assume that the set

$$\big\{\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)],\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]:\ell_1,\ell_0\in\mathcal{L}(\mathcal{A}),w\in\mathcal{W}\big\}$$

is compact.

a choice yielding a large regret. The proof of Theorem 1 shows that characterizing the worst-case two-action decision problem is sufficient to bound the regret over all decision problems. The theorem also states that the bound is sharp in the sense that, if the audience is sufficiently rich, then there exists some agent whose regret achieves the bound.

The bound $\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$ can be practical to work with in some cases. For example, when $\mathcal{L}(\mathcal{A})$ is convex and $w(\theta)=1$, the bound solves a convex programming problem. When $|\Theta|<\infty$ the program is finite-dimensional. When $|\Theta|=\infty$ the program is infinite-dimensional, but approximation methods exist under suitable restrictions on $\mathcal{L}(\mathcal{A})$ (see, e.g., Devolder et al. 2010). When $w(\theta)$ is non-constant, the program becomes non-convex but may remain tractable in some situations. As we seek a method that is generally computationally light, we prefer to offer a bound that can be read directly from the distributions $\hat{\pi}^*(\theta|X)$ and $\pi^*(\theta|X)$. We later derive such a bound based on an integral probability metric.

Remark 1. Theorem 1 fixes a particular realization of the data and imagines that agents treat the analyst's report $\hat{\pi}^*(\theta|X)$ as if it is a report of the central posterior $\pi^*(\theta|X)$. We may instead imagine that, as in Andrews and Shapiro (2021), the analyst commits in advance to a reporting procedure and each agent updates according to Bayes' rule with full knowledge of the procedure. In that case, Supplemental Appendix A shows that each agent's expected regret is bounded above by the agent's expectation of the bound derived in Theorem 1, though this bound is not sharp.

3 Controlling the Regret from a Conventional Report

Within empirical economics, it is conventional to report a point estimate and standard error (Athey and Imbens, 2023). Scientific conventions are often imperfect but they do have value, for example because of their simplicity and familiarity (see, e.g., the discussion in Benjamin et al. 2018). Here, we consider the situation of an analyst who values adhering to a convention but wishes to limit the harm from a poor approximation. We show how the analyst can use the bound in Theorem 1 to balance these goals. Lastly, we formalize the convention of reporting a point estimate and standard error as a normal (Gaussian) approximation to the central posterior, and discuss how this convention can be justified via Bernstein-von Mises theory.

3.1 A Procedure for Controlling Regret

Consider the situation of an analyst who must choose between reporting the central posterior $\pi^*(\theta|X)$ and reporting some particular approximation $\pi^0(\theta|X) \in \Delta_{\pi^*}(\Theta)$ prescribed by convention. There is an intrinsic value of $\kappa \geq 0$ to reporting the conventional approximation, in the sense that, all else equal, each agent would be willing to accept an additional loss of $\kappa \geq 0$ to receive the conventional report.

We continue to assume that agents treat any report as if it is a report of $\pi^*(\theta|X)$, and update accordingly. This may be because agents trust that $\pi^0(\theta|X)$ approximates $\pi^*(\theta|X)$, or because agents find it difficult to compute the posterior $\pi(\theta|\pi^0(\theta|X))$ that results from updating their beliefs directly from the conventional report.

The analyst seeks a procedure that, for any $X \in \mathcal{X}$, satisfies the following criteria:

(C1) The analyst reports $\pi^0(\theta|X)$ only when the regret from doing so is no more than κ for all agents.

(C2) The analyst reports $\pi^*(\theta|X)$ only when the regret from reporting $\pi^0(\theta|X)$ exceeds κ for some agent.

A procedure that satisfies Criterion (C1) avoids using the conventional report when doing so imposes, on some agent, regret that exceeds the intrinsic value κ of adhering to the convention. A procedure that satisfies Criterion (C2) conserves the conventional report when doing so does not impose a large regret on any agent.

Consider the following procedure for the analyst.

Procedure 1 Controlling regret from a conventional report

- **Calculate**
  - central posterior $\pi^*(\theta|X)$
  - conventional approximation $\pi^0(\theta|X)$
  - regret bound $\overline{R}(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$
- **Choose** threshold τ
- **Report** threshold τ and
  - $\hat{\pi}^*(\theta|X)=\pi^0(\theta|X)$ if $\overline{R}(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)\leq\tau$
  - $\hat{\pi}^*(\theta|X)=\pi^*(\theta|X)$ otherwise

We then have the following corollary of Theorem 1, taking $\overline{L}$ to be the loss function defined in the statement of the theorem.

Corollary 1. *If $\tau\leq\kappa$, then Procedure 1 satisfies Criterion (C1). If $\tau\geq\kappa$ and $\overline{L}\in\mathcal{L}$, then Procedure 1 satisfies Criterion (C2). If $\tau=\kappa$ and $\overline{L}\in\mathcal{L}$, then any procedure for choosing between reports $\pi^*(\theta|X)$ and $\pi^0(\theta|X)$ that satisfies Criteria (C1) and (C2) is equivalent to Procedure 1.*

Corollary 1 states that for τ sufficiently small, Procedure 1 satisfies Criterion (C1). Corollary 1 further states that, for τ sufficiently large, Procedure 1 satisfies Criterion (C2) if the class of losses is sufficiently rich, in the sense that it contains the loss $\overline{L}$ that attains the bound in Theorem 1. When both conditions hold, so $\tau=\kappa$, Procedure 1 is the unique procedure for choosing between reporting the central posterior and reporting the conventional approximation that satisfies both criteria.

Implementing Procedure 1 requires specifying a conventional approximation. We focus on a conventional normal approximation, which can be justified by Bernstein-von Mises (BvM) theory.

3.2 Justification for Conventional Normal Reports

One way to interpret the common convention of reporting a point estimate and standard error is as a normal approximation to the central posterior, that is,

$$\pi^0(\theta|X) = \pi^N(\theta|X) = N\Big(\hat{\theta}(X), \widehat{\Sigma}(X)\Big)$$

where $\hat{\theta}(X)$ is a point estimate, $\widehat{\Sigma}(X)$ is an estimated variance, and Θ is a finite-dimensional Euclidean space. One way to justify such a convention is via Bernstein-von Mises (BvM) theory, which provides a range of conditions on the prior and target parameter under which $\pi^N(\theta|X) \approx \pi^*(\theta|X)$ in large samples.[11] Such approximations appear reasonable in many economic settings.

Example. (Get out the vote, continued.) The first row of plots in Figure 2 shows five marginal prior distributions on θ^{TC}. The middle plot in this row corresponds to the central prior π^*, which is uniform. The left plots correspond to audience priors that are "pessimistic" about the intervention in the sense that they put more density on lower values of the effect θ^{TC} than does the central prior. The right plots correspond to audience priors that are "optimistic" about the intervention in the sense that they put more density on higher values of the effect θ^{TC} than does the central prior.

The second row of plots shows, for each prior distribution in the first row, a histogram of MCMC draws from the corresponding posterior distribution, conditional on the realized data X reported in Nickerson et al. (2006, Table 3), pooled across congressional districts. The middle plot in this row corresponds to the central posterior $\pi^*\big(\theta^{TC}|X\big)$. The left and right plots in this row correspond to the audience posteriors $\pi\big(\theta^{TC}|X\big)$. The audience posteriors are more similar to the central posterior than the audience priors are to the central prior. But, the audience posteriors reflect the pessimism and optimism of the corresponding priors.

In the second row, we overlay on each histogram the corresponding perceived posterior $\hat{\pi}\big(\theta^{TC}|X\big)$ formed by reweighting the normal approximation $\pi^N\big(\theta^{TC}|X\big)$ that is parameterized by the maximum likelihood point estimate $\hat{\theta}^{TC}(X)$ and estimated variance $\widehat{\Sigma}^{TC}(X)$, and truncated to $\theta^{TC} \in [\underline{\theta}, 1]$. For the central posterior, the normal approximation applies directly, $\hat{\pi}\big(\theta^{TC}|X\big) = \pi^N\big(\theta^{TC}|X\big)$. For the other posteriors, the normal approximation

[11] See, for example, Theorems 12.1 and 12.8 of Ghosal and van der Vaart (2017).

must be reweighted according to $\hat{\pi}\left(\theta^{TC}|X\right) \propto w(\theta)\pi^N\left(\theta^{TC}|X\right)$. In all cases, the perceived posterior is a close visual match to the true posterior.

An alternative way for agents to update would be to treat a report of $\hat{\theta}^{TC}(X)$, $\widehat{\Sigma}^{TC}(X)$ as describing a quadratic approximation to the log-likelihood for θ^{TC} rather than as a normal approximation to the posterior for θ^{TC}. Using a point estimate and standard error as the basis for a quadratic approximation to a log-likelihood can be justified by large-sample approximations similar to those underlying BvM theory (see, e.g., Geyer 2013). If the agents treat the log-likelihood as quadratic with mode $\hat{\theta}^{TC}(X)$ and Hessian $\widehat{\Sigma}^{TC}(X)$, they arrive at posteriors identical to the perceived posteriors appearing in the second row of plots in Figure 2. Interpreting the conventional normal report as an approximation to the likelihood therefore offers an alternative justification for the model of agents' belief updating that we encode in our definition of regret. $\triangle$

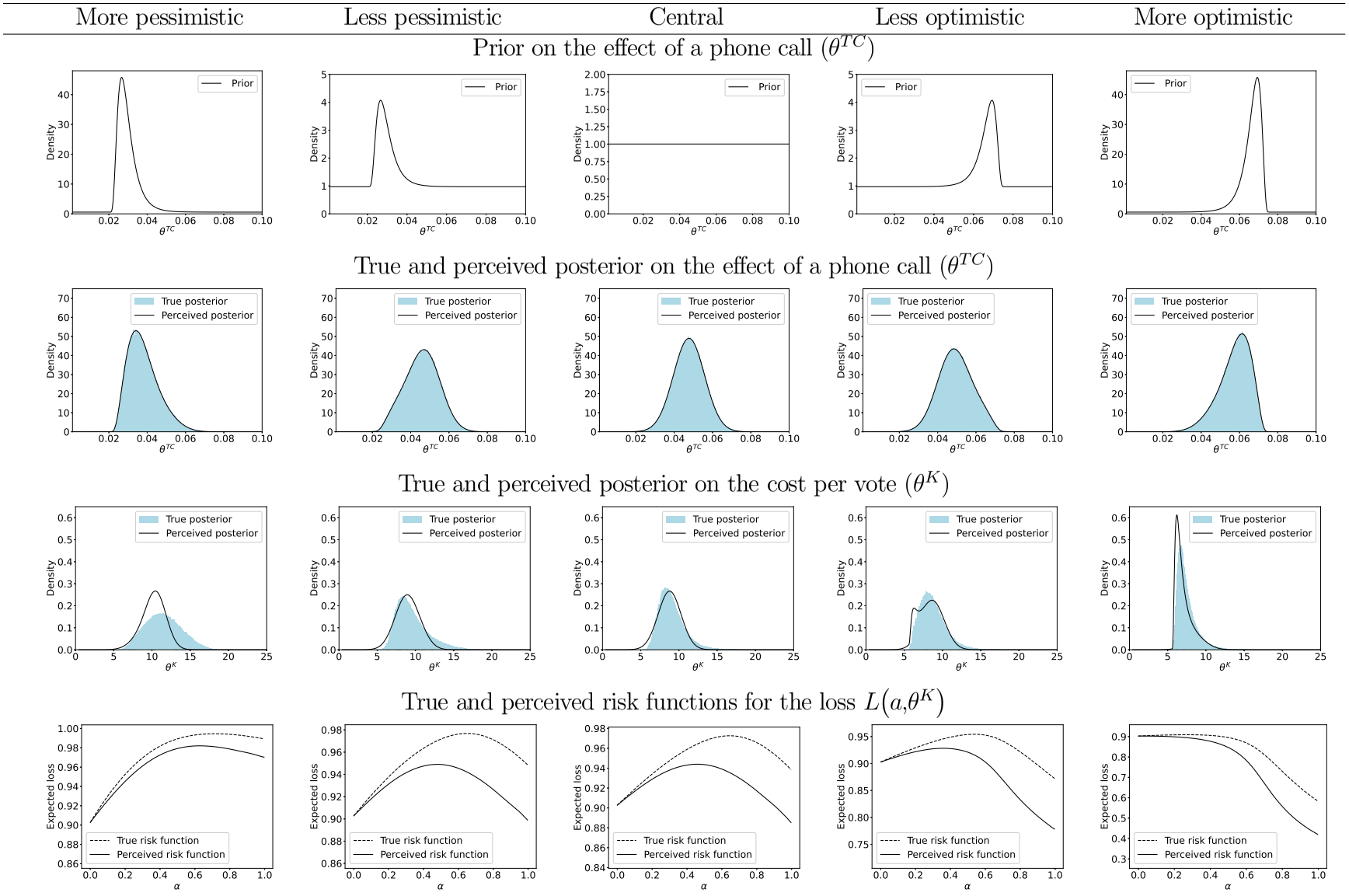
3.3 Failures of Conventional Normal Reports

While there are many economic settings where the approximation $\pi^N(\theta|X) \approx \pi^*(\theta|X)$ appears reasonable, there are others where it does not. Such situations are an important motivation for Procedure 1.

Example. (Get out the vote, continued.) Nickerson et al. (2006, Table 6) report the cost per vote obtained via phone calls. We can write this parameter as $\theta^K = K/\left(\beta^T - \beta^C\right)$ where $K > 0$ is the cost per phone call and where we recall that β^T and β^C are the probabilities of voting in the treatment and control groups, respectively. Previously we took the target parameter to be $\theta^{TC} = \left(\beta^T - \beta^C\right)$, but the target parameter $\theta^K = K/\left(\beta^T - \beta^C\right)$ may be more directly useful to local campaign managers interested in optimally allocating their budgets.

The third row of plots in Figure 2 shows, for each prior distribution on θ^{TC} in the first row, a histogram of MCMC draws from the corresponding posterior distribution $\pi\left(\theta^K|X\right)$. We overlay on each histogram the corresponding perceived posterior $\hat{\pi}\left(\theta^K|X\right)$ formed by reweighting the normal approximation $\pi^N\left(\theta^K|X\right)$ that is parameterized by the maximum likelihood point estimate $\hat{\theta}^K(X)$ and the delta-method estimated variance $\widehat{\Sigma}^K(X)$, and truncated to $\theta^K \in \left[K, \frac{K}{\underline{\theta}}\right]$. In all cases, the perceived posterior is visually distinct from the true posterior.

Figure 2: Example prior distributions, posterior distributions, and decision problems



Notes: The first row shows the central prior, and four possible audience priors, on the effect θ^{TC} of a phone call on the probability of voting. The histograms in the second row describe the corresponding true posteriors on θ^{TC} given the observed data, approximated by MCMC. The histograms in the third row describe the corresponding true posteriors on θ^{K} given the observed data, approximated by MCMC. We overlay on each histogram the density of the perceived posterior formed by reweighting the normal distribution parameterized by the maximum likelihood point estimate and (delta method) standard error. The fourth row describes the expected value of a particular loss function with respect to the true and perceived posteriors.

The difference between the reweighted normal approximation and the agent's posterior opens up the possibility of poor decision-making. To illustrate this possibility, we imagine a local campaign manager with a fixed budget $M>0$, a share $a\in[0,1]$ of which is allocated to phone calls. The remaining funds are spent on a last-minute rally the night before the election, which generates a return of Y votes per dollar, where Y is uncertain and drawn according to a known distribution with CDF $G(\cdot)$. The opponent is ahead by V votes, and the loss is the probability of defeat. We can therefore write the loss as

$$L\big(a,\theta^K\big)=G\left(\frac{V/M-a/\theta^K}{1-a}\right).$$

For any prior π, we can define both the true risk function $\mathrm{E}_{\pi(\theta^K|X)}\big[L\big(a,\theta^K\big)\big]$ and the perceived risk function obtained by replacing $\pi\big(\theta^K|X\big)$ with a reweighting of $\pi^N\big(\theta^K|X\big)$. The fourth row of plots in Figure 2 shows, for an illustrative parameterization of the loss, the true and perceived risk functions associated with each example prior. Under both the true and perceived posteriors, the optimal action for the more pessimistic agent is to devote no resources to phone calls, $a=\hat{a}=0$, and the optimal action for the more optimistic agent is to devote all resources to phone calls, $a=\hat{a}=1$. Because their optimal actions are identical under the true and perceived posteriors, these agents experience no regret from the conventional normal report. By contrast, under the true posterior, the optimal action for the agents with the less pessimistic and central priors is to devote no resources to phone calls, $a=0$, whereas these agents' optimal action under the perceived posterior is to devote all resources to phone calls, $\hat{a}=1$. These agents experience regret of, respectively, 0.046 and 0.036, from the conventional normal report.

Procedure 1 can avoid this regret. We have learned that there is an agent who experiences regret of at least 0.046 from the conventional report. If this regret exceeds the threshold τ, then Procedure 1 prescribes reporting $\pi^*\big(\theta^K|X\big)$ rather than $\pi^N\big(\theta^K|X\big)$. $\triangle$

The example shows, in a particular setting, how Procedure 1 can help the analyst to control the regret from reporting a poor approximation to the central posterior. To enable practitioners to apply Procedure 1 in a wide range of economic settings, we would ideally like to avoid the need to calculate the (possibly infinite-dimensional) program in Theorem 1. We next propose a more convenient bound that avoids that computation.

4 A Convenient Bound Using the SK Metric

To facilitate adoption of Procedure 1, we would like a more convenient bound that avoids the need to solve the (possibly infinite-dimensional) program in Theorem 1. In this section we first derive such a bound using an integral probability metric that depends on the classes of losses $\mathcal{L}$ and density ratio functions $\mathcal{W}$ in the audience. We then propose default choices for these classes that we argue are reasonable for many economic settings, and show that these choices lead to an especially convenient procedure.

4.1 A Bound in Terms of an Integral Probability Metric

For probability measures $\mu,\nu \in \Delta(\Theta)$ and a class $\mathcal{F}$ of functions $f:\Theta\to\mathbb{R}$ that are absolutely integrable w.r.t. μ,ν, let

$$d_{\mathcal{F}}(\mu,\nu) = \sup_{f\in\mathcal{F}} |\mathrm{E}_\mu[f(\theta)] - \mathrm{E}_\nu[f(\theta)]|$$

denote the largest difference in the expected values that the two measures assign to any function in $\mathcal{F}$. When $|\mathrm{E}_\mu[f(\theta)]-\mathrm{E}_\nu[f(\theta)]|\neq 0$ for some $f\in\mathcal{F}$ whenever $\mu\neq\nu$, $d_{\mathcal{F}}(\cdot,\cdot)$ is an integral probability metric (IPM).[12] IPMs play a role in many areas of probability theory. For example, if $\mathcal{F}$ is the set of all functions $f:\Theta\to[0,1]$, then $d_{\mathcal{F}}(\mu,\nu)=d_{TV}(\mu,\nu)$ is the total variation distance, also defined as the largest difference in measure assigned to any set.[13] To concisely describe other function classes, for scalars $\underline{\phi}\leq\overline{\phi}$ we let $\underline{\phi}\leq\mathcal{F}\leq\overline{\phi}$ denote that $f:\Theta\to\left[\underline{\phi},\overline{\phi}\right]$ for all $f\in\mathcal{F}$. Further, for function classes $\mathcal{G}$ and $\mathcal{F}$ we let $\mathcal{G}\cdot\mathcal{F}$ denote the class formed by the product of $g\in\mathcal{G}$ and $f\in\mathcal{F}$, i.e., functions of the form $g(\cdot)f(\cdot)$.

We now state a bound on regret that takes the form of an IPM. This bound will be the basis of our practical recommendations.

Proposition 1. (Convenient bound on regret) *For any report $\hat{\pi}^*(\theta|X)\in\Delta_{\pi^*}(\Theta)$, data $X\in\mathcal{X}$, loss class $0\leq\mathcal{L}\leq\lambda$, scalar $\omega\geq 1$, and density ratio class $\omega^{-1}\leq\mathcal{W}\leq\omega$, the bound $\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$ is itself bounded above by $2\omega\cdot d_{\mathcal{W}\cdot\mathcal{L}(\mathcal{A})}(\hat{\pi}^*(\theta|X),\pi^*(\theta|X))$. If, further, $\mathcal{L}(\mathcal{A})$ is centrosymmetric around $\frac{\lambda}{2}$ and $\mathcal{L}(\mathcal{A})$ contains all constant functions with values in $[0,\lambda]$, then $\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$ is bounded below by $d_{\mathcal{L}(\mathcal{A})}(\hat{\pi}^*(\theta|X),\pi^*(\theta|X))$.*

[12] Otherwise, it is a pseudometric, as it always satisfies symmetry, nonnegativity, and the triangle inequality.

[13] That is, $d_{TV}(\mu,\nu)=\sup_{\Theta'\subseteq\Theta}|\mu(\Theta')-\nu(\Theta')|$.

Proposition 1 states that, if the density ratios in the audience are bounded, the fundamental regret bound $\overline{R}$ is no larger than a scalar multiple of the IPM for the function class $\mathcal{W}\cdot\mathcal{L}(\mathcal{A})$ that depends on products of the density ratios $w\in\mathcal{W}$ and losses $L(\mathcal{A})\in\mathcal{L}(\mathcal{A})$ in the audience. The scalar multiple depends, in turn, on the bound ω on the density ratio neighborhood. Proposition 1 further states that, under suitable conditions, the bound $\overline{R}$ is bounded below by the IPM for the function class $\mathcal{L}(\mathcal{A})$ that depends only on the class of losses in the audience. When the bound $\overline{R}$ is sharp, it follows directly that some agent has regret at least as large as the IPM for $\mathcal{L}(\mathcal{A})$.

4.2 Losses and Density Ratios of Bounded Signed Variation

To use Proposition 1 in practice, we must select a class of losses $\mathcal{L}$ and density ratios $\mathcal{W}$. In some settings these choices may be clear from the communication problem faced by the analyst. For the remaining settings where this is not the case, we recommend default choices.

For our recommended default we consider the case where θ is scalar, $\Theta\subseteq\mathbb{R}$, and take $\mathcal{L}$ and $\mathcal{W}$ to consist of bounded functions of bounded signed variation. For any function $f:\Theta\to\mathbb{R}$, let

$$V(f)=\max\left\{\sup_{K\in\mathbb{N},t_0\leq\ldots\leq t_K\in\Theta}\sum_{k=1}^{K}(f(t_k)-f(t_{k-1}))_+,\sup_{K\in\mathbb{N},t_0\leq\ldots\leq t_K\in\Theta}\sum_{k=1}^{K}(f(t_{k-1})-f(t_k))_+\right\}$$

denote the signed variation of the function, where we take $(x)_+=\max\{x,0\}$ to be the positive part of $x\in\mathbb{R}$. The signed variation is akin to the usual notion of total variation, but takes separate account of increasing and decreasing regions. For any class of functions $\mathcal{F}$, we let $V(\mathcal{F})$ denote the largest signed variation of all functions in the class, i.e., $V(\mathcal{F})=\sup_{f\in\mathcal{F}}V(f)$.

We suppose losses in $\mathcal{L}$, which we recall are bounded by λ, also have signed variation bounded by λ, $V(\mathcal{L})\leq\lambda$. Bounded signed variation seems like a palatable restriction. Any loss function with $L(a,\theta)\in[0,\lambda]$ that is, for each $a\in\mathcal{A}$, either quasi-concave or quasi-convex in θ must have signed variation in θ bounded by λ. This admits a wide range of losses under which, for instance, there is a best or worst value $\theta(a)$ for each $a\in\mathcal{A}$, and the loss increases or decreases with distance from this value. It also admits monotone losses under which, for each $a\in\mathcal{A}$, the agent always prefers either higher or lower values of θ.

We likewise suppose that the density ratio functions in $\mathcal{W}$ are bounded, $\omega^{-1}\leq\mathcal{W}\leq\omega$,

and that they have bounded signed variation, $V(\mathcal{W}) \leq (\omega - \omega^{-1})$. This class of density ratio functions is again permissive. Any density ratio function with $w(\theta) \in [\omega^{-1}, \omega]$ that is quasi-concave or quasi-convex in θ has signed variation bounded by $\omega - \omega^{-1}$. This admits density ratios that are monotone in the distance from some peak $\overline{\theta}$ or trough $\underline{\theta}$, such that the agent finds values near to $\overline{\theta}$ or far from $\underline{\theta}$ more likely than under the central prior. It also admits any monotone density ratio function such that the agent finds higher or lower values of θ more likely than under the central prior.

What bounded signed variation does rule out are losses and density ratio functions that oscillate repeatedly as θ changes, for instance functions that are highly cyclical in θ, or that depend sensitively on the parity (even or odd) of the digits in θ.

Example. (Get out the vote, continued.) Recall that the loss is $L\left(a, \theta^K\right) = G\left(\frac{V/M - a/\theta^K}{1-a}\right)$, which is bounded by 1. Because for any $G(\cdot)$ the loss is monotone in θ^K for any a, this loss also has signed variation bounded by 1. For the signed variation of $L\left(a', \theta^K\right)$ to be greater than one for some $a' \in [0,1]$ would require that $L\left(a', \theta^K\right)$ is non-monotone in θ^K, for example because a more effective phone campaign induces more turnout from the opponent's voters. And, it would require that $L\left(a', \theta^K\right)$ is not single-peaked in θ^K, so that, over the range of θ^K, the loss switches multiple times from increasing to decreasing in θ^K. Turning to the density ratios $w(\theta)$, because the density ratios implied by the priors in Figure 2 are quasi-concave, their signed variation is bounded above by $\omega - \omega^{-1}$. For the agent experiencing greatest regret, this bound is 3.8. $\triangle$

4.3 Equivalence to the Signed Kolmogorov Metric

The class of functions of bounded signed variation pairs with a convenient IPM. We first define this metric and then show its connection to the class of functions of bounded signed variation.

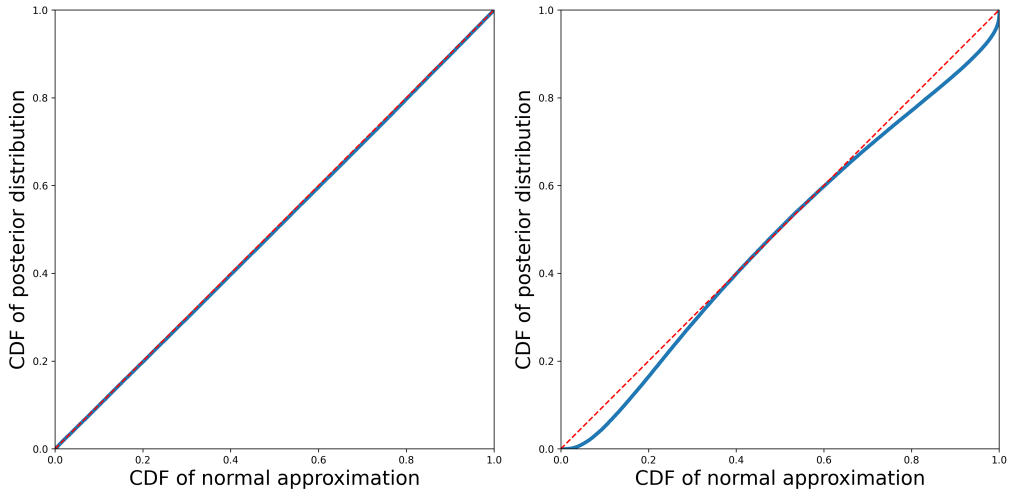
Definition 1. For measures $\mu, \nu \in \Delta(\mathbb{R})$, the *signed Kolmogorov (SK) metric* is

$$d_{SK}(\mu, \nu) = \sup_{t \in \Theta} (\mu(\theta \leq t) - \nu(\theta \leq t))_+ + \sup_{t \in \Theta} (\nu(\theta \leq t) - \mu(\theta \leq t))_+$$

where $\mu(\theta \leq t) = \mu(\{\theta \in \Theta : \theta \leq t\})$.

Figure 3: Example p-p plots

(a) Target parameter is the effect of a phone call (b) Target parameter is the cost per vote



Notes:
Panel (a) shows a p-p plot comparing the central posterior $\pi^*\left(\theta^{TC}|X\right)$ to the normal approximation $\pi^N\left(\theta^{TC}|X\right)$ for the effect θ^{TC} of a phone call. Panel (b) shows a p-p plot comparing the central posterior $\pi^*\left(\theta^K|X\right)$ to the normal approximation $\pi^N\left(\theta^K|X\right)$ for the cost per vote θ^K. In both panels, the dashed line is the 45-degree line.

The signed Kolmogorov (SK) distance between two measures is found by adding together the largest positive and negative vertical distances between their respective CDFs.[14] The SK metric can be read off of a p-p plot, which, for two measures μ,ν, is a plot of the CDF of one measure against the CDF of the other, i.e., of $\{\mu(\theta\leq t),\nu(\theta\leq t)\}_{t\in\Theta}$.

Example. (Get out the vote, continued.) Figure 3a shows a p-p plot comparing the central posterior $\pi^*\left(\theta^{TC}|X\right)$ to the normal approximation $\pi^N\left(\theta^{TC}|X\right)$ for the effect θ^{TC} of a phone call. If the plot were on the 45-degree line, we would have $d_{SK}\left(\pi^*\left(\theta^{TC}|X\right),\pi^N\left(\theta^{TC}|X\right)\right)=0$. In fact the SK metric is small, at approximately 0.005.

Figure 3b shows a p-p plot comparing the central posterior $\pi^*\left(\theta^K|X\right)$ to the normal approximation $\pi\left(\theta^K|X\right)$ for the cost per vote θ^K. Here, the SK metric is larger, at approximately 0.052. $\triangle$

The SK metric is equivalent to the IPM generated by the class of non-negative functions with value and signed variation both bounded by one. The following lemma shows this

[14]This metric is distinct from what Filion (2015) terms the "signed Kolmogorov-Smirnov test."

by extending a result of Müller (1997) for the Kolmogorov metric.

Lemma 1. *If $\mathcal{F}$ is the set of all functions $f:\mathbb{R}\to[0,1]$ with $V(f)\leq 1$, then $d_{\mathcal{F}}(\mu,\nu)=d_{SK}(\mu,\nu)$ for any probability measures $\mu,\nu\in\Delta(\mathbb{R})$.*

When the classes of losses $\mathcal{L}$ and density ratio functions $\mathcal{W}$ are bounded and of bounded signed variation, Lemma 1 immediately implies a connection between the SK metric and the IPMs that bound the regret in Proposition 1.

Corollary 2. *Fix any probability measures $\mu,\nu\in\Delta(\Theta)$. If $0\leq\mathcal{L}\leq\lambda$ and $V(\mathcal{L})\leq\lambda$, then $d_{\mathcal{L}(\mathcal{A})}(\mu,\nu)\leq\lambda d_{SK}(\mu,\nu)$, with equality when $\mathcal{L}$ is otherwise unrestricted. If, further, $\omega^{-1}\leq\mathcal{W}\leq\omega$ and $V(\mathcal{W})\leq(\omega-\omega^{-1})$, then $d_{\mathcal{W}\cdot\mathcal{L}(\mathcal{A})}(\mu,\nu)\leq\lambda(2\omega-\omega^{-1})d_{SK}(\mu,\nu)$.*

Corollary 2 states that if the function classes $\mathcal{L}$ and $\mathcal{W}$ take the form we propose in Section 4.2, then the IPMs appearing in the bounds in Proposition 1 can be replaced with ones proportional to the SK metric. We now make use of this fact to specialize Procedure 1.

4.4 Procedure Based on the SK Metric

We propose the following specialization of Procedure 1 for the case where $\Theta\subseteq\mathbb{R}$.

Procedure 2 Controlling regret with the SK metric

- **Calculate**
  - SK metric $d_{SK}(\pi^0(\theta|X),\pi^*(\theta|X))$
- **Report** threshold τ and
  - $\hat{\pi}^*(\theta|X)=\pi^0(\theta|X)$ if $d_{SK}\leq\tau$
  - $\hat{\pi}^*(\theta|X)=\pi^*(\theta|X)$ otherwise

All other details follow Procedure 1.

We then have the following corollary of Proposition 1 and Lemma 1.

Corollary 3. *Suppose that $\mathcal{L}$ and $\mathcal{W}$ satisfy the conditions of Corollary 2. Then if $\tau\leq\frac{\kappa}{2\lambda\omega(2\omega-\omega^{-1})}$, Procedure 2 satisfies Criterion (C1). Moreover, whenever Procedure 2 recommends reporting $\pi^*(\theta|X)$, if $\mathcal{L}$ is maximally rich then there exists some agent in the audience whose regret from reporting $\pi^0(\theta|X)$ is at least $\lambda\tau$.*

Because the upper bounds in Proposition 1 and Corollary 2 are not generally sharp, Procedure 2 will not generally satisfy both Criterion (C1) and Criterion (C2) for the same

value of κ. However because Proposition 1 establishes a lower bound on regret, if Procedure 2 recommends abandoning the conventional report, and the class of losses is sufficiently rich, then there is *some* agent whose regret from the conventional report is at least $\lambda\tau$.

Example. (Get out the vote, continued.) A loss function $L\big(a,\theta^K\big)$ maps the budget share $a\in[0,1]$ and cost per vote $\theta^K>0$ to the probability of defeat. A loss of 1 is certain defeat and a loss of 0 is certain victory. If all loss functions have signed variation bounded by one and the class $\mathcal{L}$ is otherwise unrestricted, then making the conventional report $\pi^N\big(\theta^K|X\big)$ about the cost per vote θ^K implies a willingness to increase some agent's probability of defeat by 0.052, relative to their posterior optimal action, in order to preserve the scientific convention. $\triangle$

5 A Bootstrap Surrogate for the Central Posterior

To facilitate adoption of Procedure 2, we would ideally like a default choice of central posterior $\pi^*(\theta|X)$ that is both convenient and reasonable for many economic settings. Because no one choice of posterior is likely exactly right for many different settings, we instead aim for a posterior that serves as a good approximation, in an appropriate sense, to posteriors from many reasonable priors. In this section we first formalize the idea of such a surrogate posterior. We then propose to use a bootstrap distribution as a default surrogate posterior, and we describe two classes of posteriors for which its surrogacy is justified.

5.1 Use of a Surrogate Central Prior

Procedure 2 requires that the analyst calculate the central posterior. Suppose that the analyst has access instead to a *surrogate central posterior* $\pi^S(\theta|X)$ that is close to the central posterior in the sense that $d_{SK}\big(\pi^S(\theta|X),\pi^*(\theta|X)\big)\leq\delta$ for some small $\delta\in[0,1]$. The triangle inequality implies that

$$\big|d_{SK}\big(\pi^0(\theta|X),\pi^S(\theta|X)\big)-d_{SK}\big(\pi^0(\theta|X),\pi^*(\theta|X)\big)\big|\leq\delta.$$

Therefore, if the surrogate is far from the conventional report, then so is the central posterior. Moreover, under the conditions of Corollary 2, $d_{\mathcal{W}\cdot\mathcal{L}(\mathcal{A})}\big(\pi^S(\theta|X),\pi^*(\theta|X)\big)\leq\lambda(2\omega-\omega^{-1})\delta$, so that by Proposition 1 reporting the surrogate controls regret even when the conventional report is not a good approximation to the central posterior.

We can therefore modify Procedure 2 as follows.

Procedure 3 Controlling regret with a surrogate central posterior

- **Calculate**
  - surrogate posterior $\pi^S(\theta|X)$
  - SK metric $d_{SK}\big(\pi^0(\theta|X),\pi^S(\theta|X)\big)$
- **Report** threshold τ and
  - $\hat{\pi}^*(\theta|X)=\pi^0(\theta|X)$ if $d_{SK}\leq\tau$
  - $\hat{\pi}^*(\theta|X)=\pi^S(\theta|X)$ otherwise

All other details follow Procedure 1.

We then have the following Corollary of Proposition 1 and Lemma 1.

Corollary 4. *Suppose that $\mathcal{L}$ and $\mathcal{W}$ satisfy the conditions of Corollary 2 and that $d_{SK}\big(\pi^S(\theta|X),\hat{\pi}^*(\theta|X)\big)\leq\delta$. Then Procedure 3 satisfies Criterion (C1) for $\tau\leq\frac{\kappa}{2\lambda\omega(2\omega-\omega^{-1})}-\delta$. Moreover, whenever Procedure 3 recommends reporting $\pi^S(\theta|X)$, if $\mathcal{L}$ is maximally rich then there exists some agent in the audience whose regret from reporting $\pi^0(\theta|X)$ is at least $\lambda(\tau-\delta)$.*

5.2 Two Classes of Central Priors with a Bootstrap Surrogate Posterior

We now describe two classes of central priors for which a bootstrap can serve as a surrogate posterior. To do this, we specialize to the case where $X=(X_1,...,X_n)\in\mathcal{X}_0^n=\mathcal{X}$ consists of $n\in\mathbb{N}$ i.i.d. draws from some (possibly unknown) probability distribution $P\in\Delta(\mathcal{X}_0)$. We further suppose that $\theta=\theta(P)$ for $\theta(\cdot)$ a known functional, so that θ is identified from the distribution of the data, as with a point-identified parameter, pseudotrue value from a GMM procedure, or bound on an identified set. A prior π specifies a distribution for P, which in turn implies a distribution for θ. We let $P_n\in\Delta(\mathcal{X}_0)$ denote the empirical distribution of the sample.

5.2.1 Dirichlet Process Central Priors

Suppose that $\pi^*_\alpha(P)=DP(\alpha,Q)$ where $DP(\cdot,\cdot)$ denotes a Dirichlet process (e.g., Ghosal and van der Vaart 2017, Chapter 4), $\alpha>0$ is a scalar, and $Q\in\Delta(\mathcal{X}_0)$ is a distribution for a single observation X_i. The parameter α controls the informativeness of the prior, with a smaller value corresponding to a less informative prior. The parameter Q controls the central tendency of the prior, and in particular corresponds to the marginal distribution of a single observation X_i under the prior.

Example. (Get out the vote, continued.) An observation $X_i \in \{0,1\}^2$ encodes both whether citizen i votes and whether citizen i is assigned to receive a phone call. A measure Q describes a contingency table for these two events. Any given sample X implies a sample contingency table P_n. A prior is a distribution over contingency tables.

A prior of the form $\pi^*_\alpha(P) = DP(\alpha,Q)$ implies a posterior of the form

$$\pi^*_\alpha(P|X) = DP\left(\alpha+n, \frac{\alpha}{\alpha+n}Q + \frac{n}{\alpha+n}P_n\right).$$

The centering measure of the posterior distribution is a weighted average of the prior contingency table Q and the sample contingency table P_n. As the informativeness of the prior becomes large, $\alpha \to \infty$, the posterior centers on the prior contingency table. As the informativeness of the prior becomes small, $\alpha \to 0$, the posterior centers on the sample contingency table.

Just as the informativeness of the prior is controlled by the parameter α, the informativeness of the sample is controlled by the sample size n. As the informativeness of the sample becomes large, $n \to \infty$, the posterior centers on the sample contingency table, and with increasing precision, so that draws from the posterior become increasingly concentrated on the observed distribution of the data. As the informativeness of the sample becomes small, $n \to 0$, the posterior converges to the prior distribution. As the informativeness of the prior also becomes small, $\alpha \to 0$, the prior distribution remains centered on the prior contingency table, but becomes less concentrated, so that draws from the prior typically place almost all mass in one cell (e.g., almost all citizens are treated and vote), with the location of that cell drawn randomly from the prior contingency table. $\triangle$

The *Bayes bootstrap distribution* is defined as $\pi^B(P|X) = DP(n,P_n)$. The Bayes bootstrap distribution can be conveniently sampled by calculating the functional $\theta(\cdot)$ on data X whose observations have been reweighted by a standard Dirichlet (Rubin 1981). Well-known properties of this bootstrap, together with conjugacy properties of Dirichlet processes, imply that the Bayes bootstrap distribution is a surrogate posterior for a Dirichlet process central prior with small informativeness α.

Lemma 2. *If $\pi^*_\alpha(P) = DP(\alpha,Q)$ for some Q, $\theta(P)$ is continuous with respect to convergence in distribution almost everywhere in the support of $\pi^B(P|X)$, and either (i) $\pi^B(\theta|X)$ is*

a continuous distribution or (ii) the range of $\theta(P)$ is countable, then

$$\lim_{\alpha\to 0} d_{SK}\left(\pi^B(\theta|X),\pi^*_\alpha(\theta|X)\right)=0.$$

Lemma 2 states that, under continuity conditions, the Bayes bootstrap distribution for θ is close, in SK metric, to the central posterior $\pi^*_\alpha(\theta|X)$ when the corresponding central prior $\pi^*_\alpha(P)$ is not very informative.

Lemma 2 justifies Procedure 3 under Dirichlet process central priors $\pi^*_\alpha(P)=DP(\alpha,Q)$ with small α. We think this justification is reasonable for situations where audience members have limited prior information about the target parameter, and so are happy to adopt a nonparametric prior with low informativeness. For other situations, we turn to an alternative justification for the use of $\pi^B(P|X)$ as a surrogate.

5.2.2 Central Posteriors that Agree on a BvM Parameter

Suppose that $\theta(P)=\theta(\beta(P))$ for known functions $\beta:\Delta(\mathcal{X}_0)\to\mathcal{B}$ and $\theta:\mathcal{B}\to\mathbb{R}$, where $\mathcal{B}$ may be infinite-dimensional.[15] On its own this restriction has no content, as we can always take $\beta(P)=P$ or $\beta(P)=\theta(P)$.

Example. (Get out the vote, continued.) The parameter $\beta=\left(\beta^T,\beta^C\right)$ describes the probability of voting with and without a phone call. The target parameters $\theta^{TC}(\beta)=\beta^T-\beta^C$ and $\theta^K(\beta)=K/\left(\beta^T-\beta^C\right)$ can both be written as functions of β. $\triangle$

If the central posterior $\pi^*(\beta|X)$ and Bayes bootstrap distribution $\pi^B(\beta|X)$ are close in a metric which implies closeness of $\pi^*(\theta|X)$ and $\pi^B(\theta|X)$ in SK, then the Bayes bootstrap distribution can serve as a surrogate. To state this formally, for a class of functions $\mathcal{F}$ let $\mathcal{F}\circ\theta$ denote the class of functions of the form $f\circ\theta=f(\theta(\cdot))$ for $f\in\mathcal{F}$.

Lemma 3. *If $d_{\mathcal{G}}\left(\pi^B(\beta|X),\pi^*(\beta|X)\right)\leq\delta$ for a class of functions $\mathcal{G}$, and $\mathcal{F}\circ\theta\subseteq\mathcal{G}$ for $\mathcal{F}$ the class of functions with value and signed variation bounded by one, i.e., the maximal class satisfying $0\leq\mathcal{F}\leq 1$ and $V(\mathcal{F})\leq 1$, it follows that $d_{SK}\left(\pi^B(\theta|X),\pi^*(\theta|X)\right)\leq\delta$.*

Agreement on an underlying parameter β can, in turn, be justified by BvM theory. BvM theory provides conditions on the central prior π^*, parameters β and the function

[15] Formally, $\theta(P)=\theta_\beta(\beta(P))$, where $\theta:\Delta(\mathcal{X}_0)\to\mathbb{R}$ and $\theta_\beta:\mathcal{B}\to\mathbb{R}$. We suppress the distinction between $\theta(\cdot)$ and $\theta_\beta(\cdot)$ in the main text to ease notation.

class $\mathcal{G}$ under which $\pi^*(\beta|X)$ approaches the conventional normal approximation in the metric $d_{\mathcal{G}}$ as n grows large. When β is finite-dimensional, standard results require only that $\mathcal{G}$ is bounded (see, e.g., van der Vaart 1998, Chapter 10; Ghosal and van der Vaart 2017, Chapter 12.3). When β is infinite-dimensional, existing results further require that $\mathcal{G}$ have a bounded Lipschitz constant (see, e.g., Castillo and Nickl 2014).

Example. (Get out the vote, continued.) For $\beta=\left(\beta^T,\beta^C\right)$, the left plot in Supplemental Appendix Figure 1 shows that the central posterior $\pi^*(\beta|X)$ is visually similar to the conventional normal approximation $N\left(\hat{\beta}(X),\hat{\Sigma}_{\hat{\beta}}(X)\right)$. $\triangle$

Importantly, BvM theory can also be applied to the Bayes bootstrap distribution. Supplemental Appendix C shows a variety of conditions under which the hypotheses of Lemma 3 hold with probability tending to one as n grows large.

Example. (Get out the vote, continued.) The right plot in Supplemental Appendix Figure 1 shows that the Bayes bootstrap distribution $\pi^B(\beta|X)$ is visually similar to the conventional normal approximation $N\left(\hat{\beta}(X),\hat{\Sigma}_{\hat{\beta}}(X)\right)$. $\triangle$

5.3 Procedure with a Bootstrap Surrogate

The preceding arguments suggest the following refinement to Procedure 3.

Procedure 4 Controlling regret with Bayes bootstrap and a conventional normal report

---

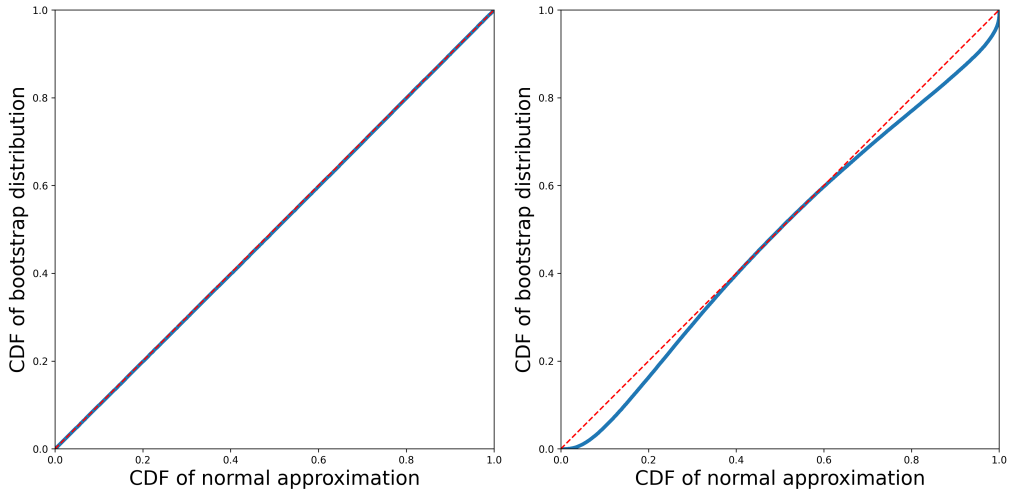
- **Calculate**
  - Bayes bootstrap distribution $\pi^B(\theta|X)$
  - Conventional normal report $\pi^0(\theta|X)=\pi^N(\theta|X)$
  - SK metric $d_{SK}\left(\pi^N(\theta|X),\pi^B(\theta|X)\right)$
- **Report** threshold τ and
  - $\hat{\pi}^*(\theta|X)=\pi^N(\theta|X)$ if $d_{SK}\leq\tau$
  - $\hat{\pi}^*(\theta|X)=\pi^B(\theta|X)$ otherwise

All other details follow Procedure 1.

Corollary 5. *Suppose that $\mathcal{L}$ and $\mathcal{W}$ satisfy the conditions of Corollary 2 and that either (i) $\pi^*_\alpha(P)=DP(\alpha,Q)$ for α such that $d_{SK}\left(\pi^B(\theta|X),\pi^*_\alpha(\theta|X)\right)\leq\delta$ or*

Figure 4: Comparing the Bayes bootstrap distribution to the normal approximation

(a) Target parameter is the effect of a phone call (b) Target parameter is the cost per vote



Notes:
Panel (a) shows a p-p plot comparing the Bayes bootstrap distribution $\pi^B\left(\theta^{TC}|X\right)$ to the normal approximation $\pi^N\left(\theta^{TC}|X\right)$ for the effect θ^{TC} of a phone call. Panel (b) shows a p-p plot comparing the Bayes bootstrap distribution $\pi^B\left(\theta^K|X\right)$ to the normal approximation $\pi^N\left(\theta^K|X\right)$ for the cost per vote θ^K. In both panels, the dashed line is the 45-degree line.

(ii) $d_{\mathcal{F}\circ\theta}\left(\pi^B(\beta|X),\pi^*(\beta|X)\right)\leq\delta$ *for* $\mathcal{F}$ *the maximal class satisfying* $0\leq\mathcal{F}\leq 1$ *and* $V(\mathcal{F})\leq 1$. *Then Procedure 4 satisfies Criterion (C1) for* $\tau\leq\frac{\kappa}{2\lambda\omega(2\omega-\omega^{-1})}-\delta$. *Moreover, whenever Procedure 4 recommends reporting* $\pi^B(\theta|X)$, *if* $\mathcal{L}$ *is maximally rich then there exists some agent in the audience whose regret from reporting* $\pi^N(\theta|X)$ *is at least* $\lambda(\tau-\delta)$.

Example. (Get out the vote, continued.) Figure 4a shows a p-p plot comparing the Bayes bootstrap distribution $\pi^B\left(\theta^{TC}|X\right)$ to the normal approximation $\pi^N\left(\theta^{TC}|X\right)$ for the effect of the intervention. The SK metric is small, at approximately 0.003. Figure 4b shows a p-p plot comparing the Bayes bootstrap distribution $\pi^B\left(\theta^K|X\right)$ to the normal approximation $\pi^N\left(\theta^K|X\right)$ for the cost per vote. The SK metric is larger, at approximately 0.052. The conclusions from Figure 4, which uses the bootstrap as a surrogate, are therefore similar to those from Figure 3, which uses the central posterior.

Notice that here the Bayes bootstrap distribution is a surrogate for the central posterior but not necessarily for all agents' posteriors. Nevertheless, Procedure 4 allows that agents may reweight the Bayes bootstrap distribution as if it were the central posterior. To us, central priors of the form envisioned in Lemma 3 seem easier to contemplate, and hence

to reweight, than those of the form envisioned in Lemma 2. We therefore think that the justification via Lemma 3 is more appealing for situations where the audience has substantive beliefs about the parameter of interest. $\triangle$

Remark 2. We recommend using the Bayes bootstrap distribution because it admits a direct interpretation following Lemma 2. Weng (1989) shows a sense in which the nonparametric bootstrap approximates the posterior distribution corresponding to a Dirichlet process prior in a large sample, albeit not as accurately as does the Bayes bootstrap. Supplemental Appendix D shows that, under the same BvM conditions that can make the Bayes bootstrap a useful surrogate for the central posterior, the Bayes bootstrap distribution will be close to that of other weighted bootstraps, including the nonparametric bootstrap. The analyst may therefore alternatively substitute the nonparametric bootstrap as a surrogate.

6 Application to a Bootstrap Census

Procedure 4 is applicable to a wide range of economic settings. To demonstrate the procedure's applicability, we applied it to the papers in the 2021 *American Economic Review* that use a bootstrap (Andrews and Shapiro, 2025). We now describe our methods and findings.

6.1 Census of Papers Using a Bootstrap

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. Supplemental 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.

For each paper, we attempted to reproduce the bootstrap replicates for all objects of interest 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 papers, with only 1 paper for which we were unable to obtain the replicates.

Consistent with the wide applicability of Procedure 4, 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.

Our main calculations will use the replicates that we obtained from the replication materials or from the authors. Given these replicates, applying Procedure 4 is trivial and can be done in a web app.

Although our theoretical development focuses on the Bayes bootstrap, only 1 paper in the census actually uses the Bayes bootstrap. Among the others, the most popular form of bootstrap is the nonparametric bootstrap (10 articles including block bootstraps), followed by various forms of parametric bootstrap (3 articles).[16] As we discuss in Section 5, there are reasons to expect different bootstrap procedures to yield similar distributions in large samples. For a subset of articles using a nonparametric bootstrap, we were able to use the published replication code and data to compute a Bayes bootstrap. In no case do we reject that the SK distance to the default normal differs between the two bootstrap distributions; see Supplemental Appendix Figure 2 for details.

Applying Procedure 4 requires specifying a conventional report. For our main analysis we use the normal approximation $\pi^0(\theta|X) = \pi^N(\theta|X) = N\left(\hat{\theta}(X), \widehat{\Sigma}(X)\right)$ implied by the reported point estimate and the bootstrap standard error, defined as the standard deviation of the bootstrap distribution. All of the papers in the census report point estimates for all objects of interest, and 10 out of 14 report bootstrap standard errors. The remaining 4 out of 14 papers report bootstrap confidence intervals. To better capture such situations, Supplemental Appendix Figure 3 shows results where we take the variance of the default normal to match the difference between the 97.5th and 2.5th percentiles of the bootstrap distribution, rather than its standard deviation.

[16] We include in the category of parametric bootstraps procedures that treat some statistics as exactly normally distributed. 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.

6.2 Findings from the Bootstrap Census

Figure 5 shows a series of p-p plots comparing the CDF of bootstrap replicates, i.e., samples from $\pi^B(\theta|X)$, to the conventional normal report $\pi^N(\theta|X)$. Each plot reports the maximum positive and negative vertical distance between the CDF of the bootstrap replicates and of the conventional report. The sum of these distances is an estimator of $SK\big(\pi^B(\theta|X),\pi^N(\theta|X)\big)$ that is consistent in the number of bootstrap replicates. For each paper in our census, we depict two p-p plots, corresponding to the objects with the smallest and largest estimated SK metric among the objects of interest in the paper. The figure shows that, as in our running example, the estimated SK metric can differ meaningfully even across objects of interest reported in the same paper.

Across all of the objects of interest in our census, we find that the estimated SK metric is greater than 0.1 in 72 percent of cases, and is greater than 0.2 in 32 percent of cases. In these cases, if the Bayes bootstrap is a good surrogate for the central posterior, then under the conditions of Corollary 4, the value of the conventional report must represent a share of at least 0.1 or 0.2, respectively, of the worst-case loss to justify reporting the conventional normal report. Supplemental Appendix Figure 3 shows the full distribution of the estimated SK metric across the objects of interest in our census.

Figure 5: Comparison of Bootstrap Distribution to Default Normal Report

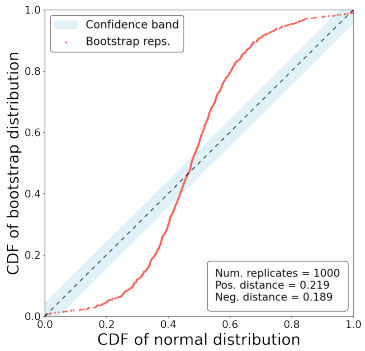
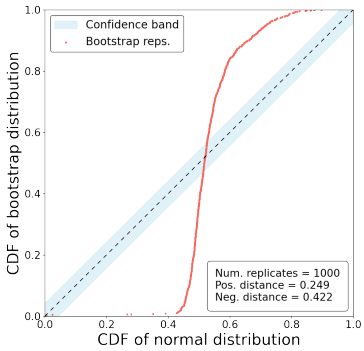
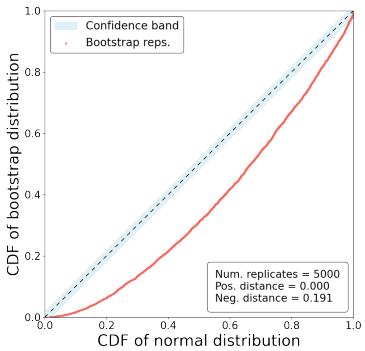
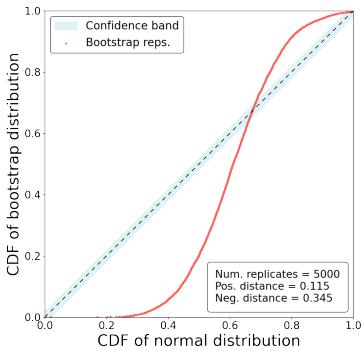
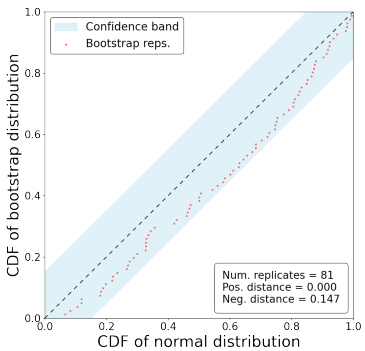
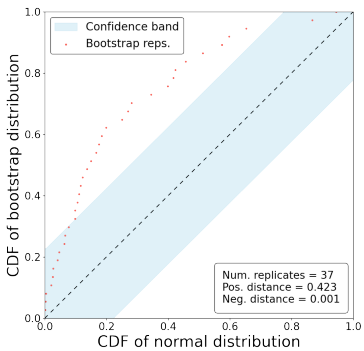
Paper	Object with... least SK distance	Object with... greatest SK distance
1		
2		
3		

Figure 5 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

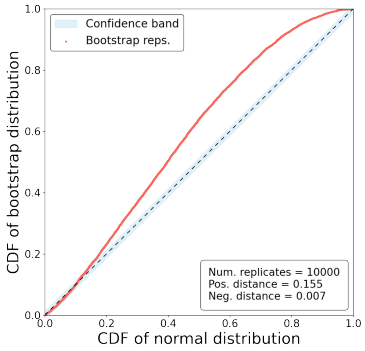
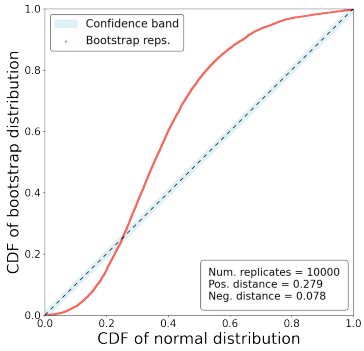
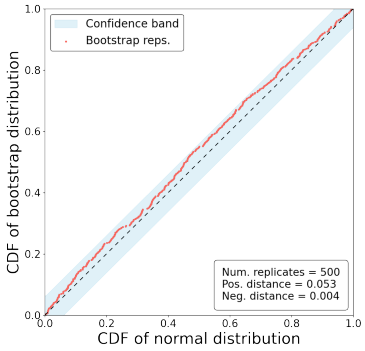
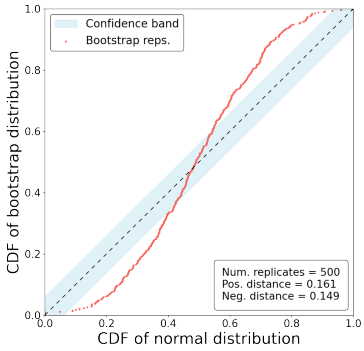
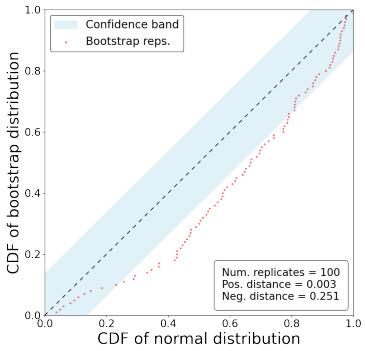
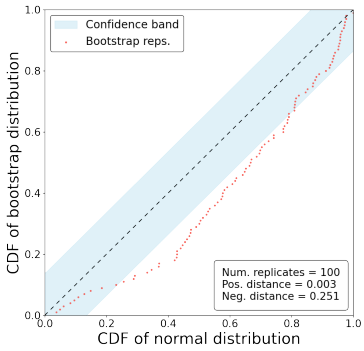
Paper	**Object with...**	
	least SK distance	**greatest SK distance**
4		
5		
6		

Figure 5 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

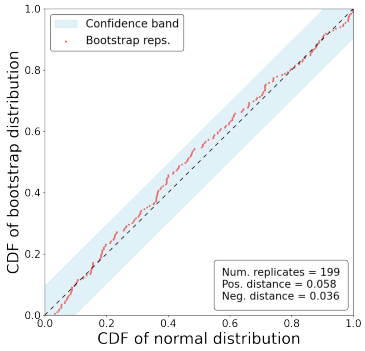
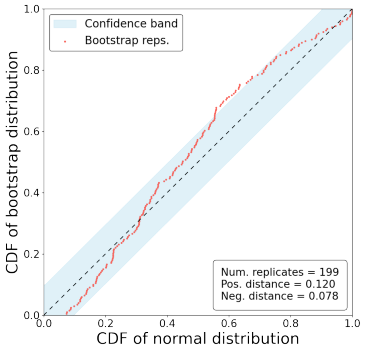
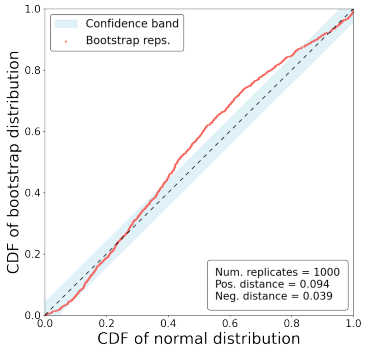
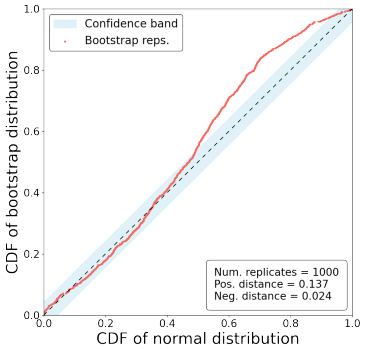
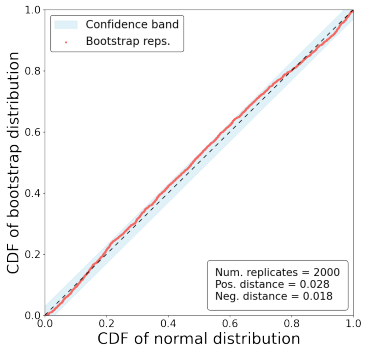
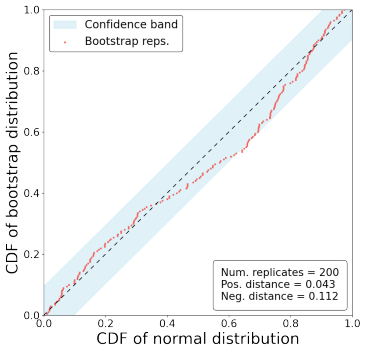
Paper	Object with... least SK distance	Object with... greatest SK distance
7		
8		
9		

Figure 5 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

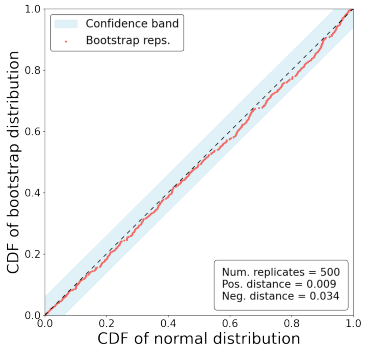
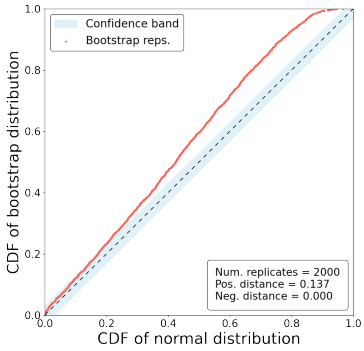
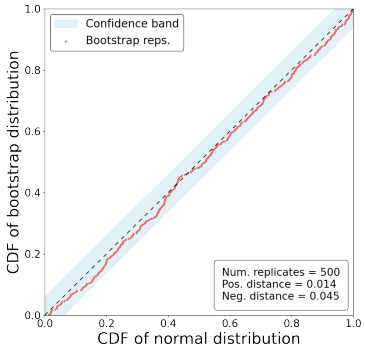
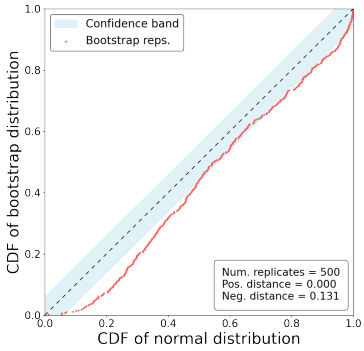
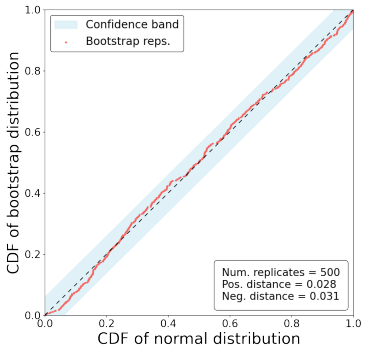
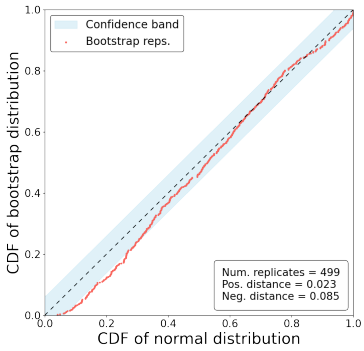
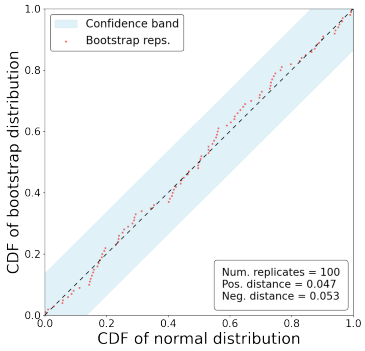
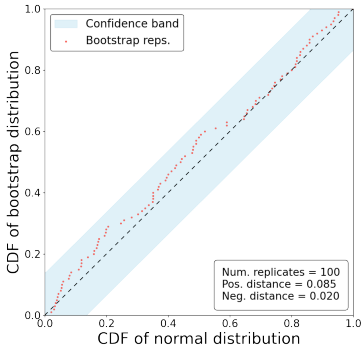
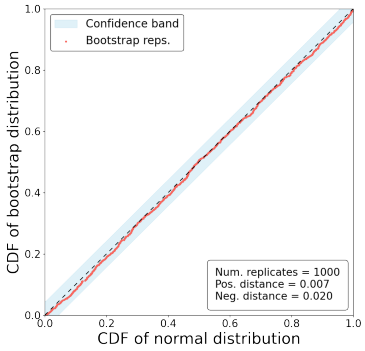
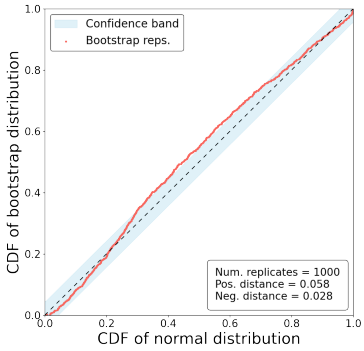
Paper	**Object with...**	
	least SK distance	**greatest SK distance**
10		
11		
12		

Figure 5 (cont'd): Comparison of Bootstrap Distribution to Default Normal Report

Paper	Object with... least SK distance	Object with... 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 SK distance.

Appendix: Proofs

To extend Theorem 1 to the case where an optimal action may not exist, we need to define regret for this case. For $\varepsilon > 0$ define the set of ε-optimal actions under (π,L), $\mathcal{A}_\varepsilon(\pi,L) = \{\hat{a} \in \mathcal{A} : E_\pi[L(\hat{a},\theta)] \leq \inf_{a\in\mathcal{A}} E_\pi[L(a,\theta)] + \varepsilon\}$. We consider the worst-case regret over ε-optimal actions as $\varepsilon \to 0$,

$$R(\hat{\pi}^*(\theta|X);X,L,w,\pi^*) =$$

$$\lim_{\varepsilon\to 0}\left(\sup\left\{\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[L(\hat{a},\theta)-L(a,\theta)]:\ \begin{array}{c}\hat{a}\in\mathcal{A}_\varepsilon(w(\cdot)\hat{\pi}^*(\cdot|X),L), a\in\mathcal{A},\\ \mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[L(\hat{a},\theta)-L(a,\theta)]\leq 0\end{array}\right\}\right).$$

Theorem 2. *For R as defined above,*

$$\sup_{w\in\mathcal{W}}\sup_{L\in\mathcal{L}} R(\hat{\pi}^*(\theta|X);X,L,w,\pi^*) \leq \overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$$

for $\overline{R}$ as defined in Theorem 1.

Proof of Theorem 2 For any $L\in\mathcal{L}$ and $w\in\mathcal{W}$, consider any pair of actions $\hat{a}$ and a such that the agent weakly prefers $\hat{a}$ to a under $\hat{\pi}$, $\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[L(\hat{a},\theta)-L(a,\theta)]\leq 0$. Note that since $L(a,\theta),L(\hat{a},\theta)\in\mathcal{L}(\mathcal{A})$, it is immediate that $\mathrm{E}_{\pi(\theta|X)}[L(\hat{a},\theta)-L(a,\theta)]$ is bounded above by

$$\sup_{\ell_1,\ell_0\in\mathcal{L}(\mathcal{A}),w\in\mathcal{W}} \mathrm{E}_{w(\theta)\pi^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)] \text{ subject to: } \mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]\leq 0.$$

Hence, we see that for any $\varepsilon>0$ and any $\hat{a}\in\mathcal{A}_\varepsilon(\hat{\pi},L), a\in\mathcal{A}$ such that $\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[L(\hat{a},\theta)-L(a,\theta)]\leq 0$,

$$\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[L(\hat{a},\theta)-L(a,\theta)]\leq \overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*(\theta|X)),$$

from which the result is immediate. □

Proof of Theorem 1 The first part of the result is immediate from Theorem 2. For the second part of the result, note that for any ℓ_1, $\ell_0\in\mathcal{L}(\mathcal{A})$, $w\in\mathcal{W}$ such that

$$\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]\leq 0, \tag{1}$$

if we consider $\mathcal{A}=\{0,1\}$ and the loss

$$L(a,\theta)=\begin{cases}\ell_1(\theta) & \text{if } a=1\\ \ell_0(\theta) & \text{if } a=0\end{cases},$$

then the agent weakly prefers $\hat{a}=1$ based on their perceived posterior, but will incur regret $\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]$ from this action. Hence, if we take the worst case over ℓ_1, $\ell_0\in\mathcal{L}(\mathcal{A})$, $w\in\mathcal{W}$ satisfying (1), we obtain risk $\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$, where the conditions of the theorem ensure that this worst-case is attained at some $\overline{\ell}_1$, $\overline{\ell}_0$, $\overline{w}$, and we may take

$$\overline{L}=\begin{cases}\overline{\ell}_1(\theta) & \text{if } a=1\\ \overline{\ell}_0(\theta) & \text{if } a=0\end{cases}.$$

Note that $\overline{L}(\mathcal{A})\subseteq\mathcal{L}(\mathcal{A})$, but we may have $\overline{L}\notin\mathcal{L}$. If, contrary to our assumptions, the set

$$\left\{\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)],\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]:\ell_1,\ell_0\in\mathcal{L}(\mathcal{A}),w\in\mathcal{W}\right\}$$

is not compact, we can select ℓ_1, $\ell_0\in\mathcal{L}(\mathcal{A})$, $w\in\mathcal{W}$ to come arbitrarily close to the worst-case regret $\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)$. □

Proof of Corollary 1 If $\tau\leq\kappa$, Procedure 1 reports $\pi^0(\theta|X)$ only if $\overline{R}(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)\leq\kappa$. Criterion (C1) is thus satisfied by Theorem 1. If $\tau\geq\kappa$ and $\overline{L}\in\mathcal{L}$, Procedure 1 reports $\pi^*(\theta|X)$ only if $\overline{R}(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)>\kappa$. Criterion (C2) is thus satisfied by Theorem 1. By Theorem 1, if $\overline{L}\in\mathcal{L}$ any procedure satisfying Criteria (C1) and (C2) must report $\pi^0(\theta|X)$ if and only if $\overline{R}(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)\leq\kappa$, and thus is equivalent to Procedure 1 with $\tau=\kappa$. □

Proof of Proposition 1 Since $w(\theta)\geq\omega^{-1}$ by assumption, for any $w\in\mathcal{W}$ and any ℓ_1, $\ell_0\in\mathcal{L}(\mathcal{A})$, $\int w(\theta)d\pi^*(\theta|X)\geq\omega^{-1}$ and

$$\mathrm{E}_{w(\theta)\pi^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]\leq\omega\int(\ell_1(\theta)-\ell_0(\theta))w(\theta)d\pi^*(\theta|X).$$

Note, next, that $\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]\leq 0$ implies that

$$\omega\int(\ell_1(\theta)-\ell_0(\theta))w(\theta)d\pi^*(\theta|X)\leq$$

$$\omega\left(\int(\ell_1(\theta)-\ell_0(\theta))w(\theta)d\pi^*(\theta|X)-\int(\ell_1(\theta)-\ell_0(\theta))w(\theta)d\hat{\pi}^*(\theta|X)\right)\leq$$

$$\leq 2\omega\cdot\sup_{f\in\mathcal{W}\cdot\mathcal{L}(\mathcal{A})}\left|\int f(\theta)d\pi^*(\theta|X)-\int f(\theta)d\hat{\pi}^*(\theta|X)\right|=2\omega\cdot d_{\mathcal{W}\cdot\mathcal{L}(\mathcal{A})}(\hat{\pi}^*(\theta|X),\pi^*(\theta|X)).$$

Since this holds for all $w\in\mathcal{W}$ and any ℓ_1, $\ell_0\in\mathcal{L}(\mathcal{A})$ with $\mathrm{E}_{w(\theta)\hat{\pi}^*(\theta|X)}[\ell_1(\theta)-\ell_0(\theta)]\leq 0$, the first part of the result is immediate from Theorem 1.

For the second part of the result, centrosymmetry implies that

$$\sup_{f\in\mathcal{L}(\mathcal{A})}\left|\int f(\theta)d\pi^*(\theta|X)-\int f(\theta)d\hat{\pi}^*(\theta|X)\right|=\sup_{f\in\mathcal{L}(\mathcal{A})}\left(\int f(\theta)d\pi^*(\theta|X)-\int f(\theta)d\hat{\pi}^*(\theta|X)\right).$$

Consider $\tilde{f}$ which attains the preceding supremum (which exists by our compactness assumptions), $\mathcal{A}=\{0,1\}$, and the loss

$$L(a,\theta)=\begin{cases}\tilde{f}(\theta) & \text{if } a=1\\ \int\tilde{f}(\theta)d\hat{\pi}^*(\theta|X) & \text{if } a=0\end{cases}.$$

By construction, the agent is indifferent between the two actions based on their perceived posterior, but choosing $\hat{a}=1$ yields a regret of

$$\int\tilde{f}(\theta)d\pi^*(\theta|X)-\int\tilde{f}(\theta)d\hat{\pi}^*(\theta|X)\geq d_{\mathcal{L}(\mathcal{A})}(\pi^*(\theta|X),\hat{\pi}^*(\theta|X)).$$

The result is immediate. □

Proof of Lemma 1 We first show that for $\mathcal{I}$ the set of indicator functions for (open, closed, and half-open) intervals in $\mathbb{R}$, $d_{\mathcal{I}}(\mu,\nu)=d_{SK}(\mu,\nu)$.

To see that this is the case, note that since SK is a metric, $|SK(\mu,\nu)|=SK(\mu,\nu)$. Note that for all $t_\mu,t_\nu\in\mathbb{R}$,

$$|\mu(\theta\leq t_\mu)-\mu(\theta\leq t_\nu)-\nu(\theta\leq t_\mu)+\nu(\theta\leq t_\nu)|=$$

$$|\mu(\theta\in(t_\mu\wedge t_\nu,t_\mu\vee t_\nu])-\nu(\theta\in(t_\mu\wedge t_\nu,t_\mu\vee t_\nu])|,$$

where $a\wedge b=\min\{a,b\}$ and $a\vee b=\max\{a,b\}$. Hence, we can rewrite d_{SK} as the IPM

generated by indicator functions for half-open intervals

$$d_{SK}(\mu,\nu)=\sup_{l,u}\left|\int 1\{t\in(l,u]\}d\mu(t)-\int 1\{t\in(l,u]\}d\nu(t)\right|.$$

By definition $d_{\mathcal{I}}(\mu,\nu)\geq d_{SK}(\mu,\nu)$, so to prove the result it suffices to show that we cannot have a strict inequality.

To this end, note that for any $-\infty\leq a\leq b\leq\infty$ and any measure $\mu\in\Delta(\mathbb{R})$,

$$\lim_{\varepsilon\downarrow 0}\int 1\{t\in(a-\varepsilon,b]\}d\mu(t)=\int 1\{t\in[a,b]\}d\mu(t),$$

while for any $-\infty\leq a<b\leq\infty$

$$\lim_{\varepsilon\downarrow 0}\int 1\{t\in(a,b-\varepsilon]\}d\mu(t)=\int 1\{t\in(a,b)\}d\mu(t),$$

$$\lim_{\varepsilon\downarrow 0}\int 1\{t\in(a-\varepsilon,b-\varepsilon]\}d\mu(t)=\int 1\{t\in[a,b)\}d\mu(t).$$

Since the same argument applies when integrating over ν, it is immediate that $d_{\mathcal{I}}(\mu,\nu)=d_{SK}(\mu,\nu)$. The result of Lemma 1 thus follows provided $d_{\mathcal{I}}(\mu,\nu)=d_{\mathcal{F}}(\mu,\nu)$, as we prove in the following lemma.

Lemma 4. *For $\mathcal{I}$ the set of indicator functions for (open, closed, and half-open) intervals in $\mathbb{R}$, and $\mathcal{F}$ the set of all functions $f:\mathbb{R}\to[0,1]$ with $V(f)\leq 1$, we have $d_{\mathcal{I}}(\mu,\nu)=d_{\mathcal{F}}$ for all measures μ,ν.*

The proof of Lemma 4 is broadly similar to arguments in Müller (1997), and is provided in the Supplemental Appendix. □

Proof of Corollary 2 Under the assumptions of the corollary, note that for $\ell\in\mathcal{L}(\mathcal{A})$ and $w\in\mathcal{W}$, $0\leq\ell(\theta)w(\theta)\leq\lambda\omega$. Note further that for $\theta'\leq\theta$,

$$(\ell(\theta)w(\theta)-\ell(\theta')w(\theta'))_+=(\ell(\theta)w(\theta)-\ell(\theta')w(\theta)+\ell(\theta')w(\theta)-\ell(\theta')w(\theta'))_+\leq$$

$$\ell(\theta')(w(\theta)-w(\theta'))_+ + w(\theta)(\ell(\theta)-\ell(\theta'))_+.$$

Hence, for an increasing sequence $t_0,...,t_K$,

$$\sum_{k=1}^{K}(\ell(t_k)w(t_k)-\ell(t_{k-1})w(t_{k-1}))_+ \leq \sum_{k=1}^{K}\ell(t_k)(w(t_k)-w(t_{k-1}))_+ + \sum_{k=1}^{K}w(t_k)(\ell(t_{k-1})-\ell(t_k))_+$$

$$\leq \omega\sum_{k=1}^{K}(\ell(t_{k-1})-\ell(t_k))_+ + \lambda\sum_{k=1}^{K}(w(t_k)-w(t_{k-1}))_+ \leq \omega\lambda+\lambda\left(\omega-\omega^{-1}\right).$$

Since $\omega \geq 1$, it follows that $0 \leq \mathcal{W}\cdot\mathcal{L}(\mathcal{A}) \leq \lambda(2\omega-\omega^{-1})$ and $V(\mathcal{W}\cdot\mathcal{L}(\mathcal{A})) \leq \lambda(2\omega-\omega^{-1})$, from which the result is immediate noting that for the first part of the result we can take $\omega=1$. □

Proof of Corollary 3 By Corollary 2,

$$d_{\mathcal{W}\cdot\mathcal{L}(\mathcal{A})}(\hat{\pi}^*(\theta|X),\pi^*(\theta|X)) \leq \lambda\left(2\omega-\omega^{-1}\right)d_{SK}(\hat{\pi}^*(\theta|X),\pi^*(\theta|X)).$$

By the first part of Proposition 1, we thus know that

$$\overline{R}(\hat{\pi}^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*) \leq 2\omega\lambda\left(2\omega-\omega^{-1}\right)d_{SK}(\hat{\pi}^*(\theta|X),\pi^*(\theta|X)).$$

Procedure 2 reports $\pi^0(\theta|X)$ only when $d_{SK}(\pi^0(\theta|X),\pi^*(\theta|X)) \leq \tau$, and hence when $\overline{R}(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*) \leq 2\omega\lambda(2\omega-\omega^{-1})\tau$, so the first part of the result follows from Corollary 1.

For the second part of the result, note that Procedure 2 reports $\pi^*(\theta|X)$ only when $d_{SK}(\pi^0(\theta|X),\pi^*(\theta|X)) > \tau$. By our assumption that $\mathcal{L}$ is maximally rich, there exists an agent with the loss used to construct the lower bound in the proof of Proposition 1. Hence, there exists an agent with regret at least

$$d_{\mathcal{L}(\mathcal{A})}\left(\pi^0(\theta|X),\pi^*(\theta|X)\right) = \lambda d_{SK}\left(\pi^0(\theta|X),\pi^*(\theta|X)\right) > \lambda\tau,$$

which implies the result. □

Proof of Corollary 4 Procedure 3 reports π^0 only when $d_{SK}\left(\pi^0(\theta|X),\pi^S(\theta|X)\right) \leq \tau$. By the triangle inequality, this implies that $d_{SK}(\pi^0(\theta|X),\pi^*(\theta|X)) \leq \tau+\delta$, and thus by the argument in the proof of Corollary 3 above that

$$\overline{R}\left(\pi^0(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*\right) \leq 2\omega\lambda\left(2\omega-\omega^{-1}\right)(\tau+\delta),$$

so the first part of the result follows from Proposition 1.

Procedure 3 reports π^S only when $d_{SK}\left(\pi^0(\theta|X),\pi^S(\theta|X)\right) > \tau$, in which case the triangle inequality implies that $d_{SK}(\pi^0(\theta|X),\pi^*(\theta|X)) > \tau - \delta$. By our assumption that $\mathcal{L}$ is maximally rich, there exists an agent with the loss used to construct the lower bound in the proof of Proposition 1. Hence, there exists an agent with regret at least

$$d_{\mathcal{L}(\mathcal{A})}\left(\pi^0(\theta|X),\pi^*(\theta|X)\right) = \lambda d_{SK}\left(\pi^0(\theta|X),\pi^*(\theta|X)\right) > \lambda(\tau-\delta),$$

which implies the result. □

Proof of Lemma 2 By Theorem 4.6 in Ghosal and van der Vaart (2017), for any $Q \in \Delta(\mathcal{X}_0)$ and $\pi^*_\alpha = DP(\alpha,Q)$, $\pi^*_\alpha(P|X) = DP\left(\alpha+n, \frac{\alpha}{\alpha+n}Q + \frac{n}{\alpha+n}P_n\right)$. By Theorem 4.16 of Ghosal and van der Vaart (2017), $\pi^*_\alpha(P|X) \to_d \pi^B(P|X) = DP(n,P_n)$ as $\alpha \to 0$, where $\to_d$ denotes convergence in distribution. By assumption θ is continuous in P (with respect to the topology of weak convergence) almost everywhere on the support of $\pi^B(P|X)$. Hence, by the continuous mapping theorem (e.g., Theorem 18.11 of van der Vaart, 1998), $\pi^*_\alpha(\theta|X) \to_d \pi^B(\theta|X)$. By the definition of convergence in distribution, it follows that the distribution function of $\pi^*_\alpha(\theta|X)$ converges to that of $\pi^B(\theta|X)$ at every continuity point of the latter distribution function. If $\pi^B(\theta|X)$ is continuous then all values of θ are continuity points for the distribution function of $\pi^B(\theta|X)$. If instead the range of $\theta(P)$ is countable then for any point t in the range there is a value $\varepsilon(t) > 0$ such that $t+\varepsilon(t)$ is smaller than any point in the range above t, and is trivially a continuity point of $\pi^B(\theta|X)$. Hence, convergence in distribution implies that the distribution function of $\pi^*_\alpha(\theta|X)$ at $t+\varepsilon(t)$ converges to that of $\pi^B(\theta|X)$, and thus that the same holds at t. For both cases, convergence in distribution thus implies point-wise convergence of distribution functions. However, point-wise convergence of distribution functions implies uniform convergence (by, e.g., Lemma 2.11 of van der Vaart, 1998), and uniform convergence of distribution functions implies convergence in d_{SK}. □

Proof of Lemma 3 By assumption $f(\theta(\cdot)) \in \mathcal{G}$ for all $f \in \mathcal{F}$, so by Lemma 1,

$$d_{SK}\left(\pi^B(\theta|X),\pi^*(\theta|X)\right) = \sup_{f\in\mathcal{F}} \left|\mathrm{E}_{\pi^B(\theta|X)}[f(\theta)] - \mathrm{E}_{\pi^*(\theta|X)}[f(\theta)]\right| =$$

$$\sup_{f\in\mathcal{F}} \left|\mathrm{E}_{\pi^B(\beta|X)}[f(\theta(\beta))] - \mathrm{E}_{\pi^*(\beta|X)}[f(\theta(\beta))]\right| \leq \sup_{g\in\mathcal{G}} \left|\mathrm{E}_{\pi^B(\beta|X)}[g(\beta)] - \mathrm{E}_{\pi^*(\beta|X)}[g(\beta)]\right|,$$

from which the result is immediate. □

Proof of Corollary 5 Follows immediately from Lemmas 2 and 3 and Corollary 4. □

References

Andrews, Isaiah and Jesse Shapiro (2025): "Bootstrap Census," Zenodo, https://doi.org/10.5281/zenodo.17085552.

Andrews, Isaiah and Jesse M. Shapiro (2021): "A Model of Scientific Communication," *Econometrica*, 89 (5), 2117–2142.

Angelini, Giovanni, Giuseppe Cavaliere, and Luca Fanelli (2022): "Bootstrap Inference and Diagnostics in State Space Models: With Applications to Dynamic Macro Models," *Journal of Applied Econometrics*, 37 (1), 3–22.

——— (2024): "An Identification and Testing Strategy for Proxy-SVARs with Weak Proxies," *Journal of Econometrics*, 238 (2), 105604.

Athey, Susan and Guido 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, Berkeley Initiative for Transparency in the Social Sciences Keynote Lecture; video accessed in December 2023.

Benjamin, Daniel J, James O Berger, Magnus Johannesson, Brian A Nosek, E-J Wagenmakers, Richard Berk, Kenneth A Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, et al. (2018): "Redefine Statistical Significance," *Nature Human Behaviour*, 2 (1), 6–10.

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

Castillo, Ismaël and Richard Nickl (2014): "On the Bernstein–von Mises Phenomenon for Nonparametric Bayes Procedures," *Annals of Statistics*, 42 (5), 1941–1969.

Cavaliere, Giuseppe, Luca Fanelli, and Iliyan Georgiev (2025): "Bootstrap Diagnostic Tests," *arXiv econ.EM*, 2509.01351, https://arxiv.org/abs/2509.01351; accessed on 03 September 2025.

Chamberlain, Gary and Guido W Imbens (2003): "Nonparametric Applications of Bayesian Inference," *Journal of Business and Economic Statistics*, 21 (1), 12–18.

Cohen, Alma and Liran Einav (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97 (3), 745–788.

Devolder, Olivier, François Glineur, and Yurii Nesterov (2010): "Solving Infinite-Dimensional Optimization Problems by Polynomial Approximation," in *Recent Advances in Optimization and its Applications in Engineering: The 14th Belgian-French-German Conference on Optimization*, Springer, 31–40.

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

Filion, Guillaume J (2015): "The Signed Kolmogorov-Smirnov Test: Why it Should Not be Used," *GigaScience*, 4 (1).

Gasparini, Mauro (1995): "Exact Multivariate Bayesian Bootstrap Distributions of Moments," *Annals of Statistics*, 23 (3), 762–768.

Geweke, John (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,, Cambridge: Cambridge University Press, vol. 3, 128–165.

Geyer, Charles J (2013): "Asymptotics of Maximum Likelihood Without the LLN or CLT or Sample Size Going to Infinity," in *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*, Institute of Mathematical Statistics, vol. 10, 1–25.

Ghosal, Subhashis and Aad van der Vaart (2017): *Fundamentals of Nonparametric Bayesian Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

Gopalan, Parikshit, Adam Tauman Kalai, Omer Reingold, Vatsal Sharan, and Udi Wieder (2022): "Omnipredictors," in *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, ed. by Mark Braverman, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, vol. 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 79:1–79:21.

Hahn, Jinyong and Zhipeng Liao (2021): "Bootstrap Standard Error Estimates and Inference," *Econometrica*, 89 (4), 1963–1977.

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

Hildreth, Clifford (1963): "Bayesian Statisticians and Remote Clients," *Econometrica*, 31 (3), 422–438.

Müller, Alfred (1997): "Integral Probability Metrics and their Generating Classes of Functions," *Advances in Applied Probability*, 29 (2), 429–443.

Nickerson, David W, Ryan D Friedrichs, and David C King (2006): "Partisan Mobilization Campaigns in the Field: Results from a Statewide Turnout Experiment in Michigan," *Political Research Quarterly*, 59 (1), 85–97.

Noarov, Georgy, Ramya Ramalingam, Aaron Roth, and Stephan Xie (2025): "High-Dimensional Prediction for Sequential Decision Making," *Proceedings of the 42nd International Conference on Machine Learning*, PMLR 267.

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

Rubin, Donald B. (1981): "The Bayesian Bootstrap," *Annals of Statistics*, 9 (1), 130–134.

Singh, Kesar (1981): "On the Asymptotic Accuracy of Efron's Bootstrap," *Annals of Statistics*, 1187–1195.

Smets, Frank and Rafael Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97 (3), 586–606.

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

Wang, Andrew (2025): "A General Diagnostic of the Normal Approximation in GMM Models," *Econometrics Journal*, 28 (2), 261–275.

Weng, Chung-Sing (1989): "On a Second-Order Asymptotic Property of the Bayesian Bootstrap Mean," *Annals of Statistics*, 705–710.

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

Supplement to "Communicating Scientific Uncertainty via Approximate Posteriors"

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

Abstract

This supplement contains theoretical results, data description, and plots to accompany the material in Andrews and Shapiro (2025), "Communicating Scientific Uncertainty via Approximate Posteriors."

A Ex-Ante Regret Bounds for Fully Bayesian Agents

Our main results assume that agents treat the analyst's reported distribution $\hat{\pi}^*(\theta|X)$ as if it were the central posterior. If instead the analyst publicly commits to take $\hat{\pi}^*(\theta|X)=c(X)$ for some function $c:\mathcal{X}\to\Delta(\Theta)$ and the agent takes their Bayes-optimal action based on $c(X)$, then the ex-ante expected regret for each agent is bounded above by the agent's ex-ante expectation of our regret bound.

Specifically, if the agent takes the optimal action based on their posterior belief after observing $c(X)$, then as in Andrews and Shapiro (2021) their expected loss is $\mathrm{E}_\pi[\inf_a \mathrm{E}_\pi[L(a,\theta)|c(X)]]$. By construction this is bounded above by $\mathrm{E}_\pi[L(b(c(X)),\theta)]$ for any function $b:\Delta(\Theta)\to\mathcal{A}$. However, the actions $\hat{a}$ studied in the main text are one such function, so

$$\mathrm{E}_\pi\left[\inf_a \mathrm{E}_\pi[L(a,\theta)|c(X)]\right] \leq \mathrm{E}_\pi[L(\hat{a},\theta)],$$

which immediately implies a bound on the expected regret under the agent's prior

$$\mathrm{E}_\pi\left[\inf_a \mathrm{E}_\pi[L(a,\theta)|c(X)] - \inf_a \mathrm{E}_\pi[L(a,\theta)|X]\right] \leq \mathrm{E}_\pi[R(\hat{\pi}^*(\theta|X);X,L,w,\pi^*)].$$

Theorem 1 implies that the right hand side is, in turn, bounded above by $\mathrm{E}_\pi\left[\overline{R}(\pi^*(\theta|X);X,\mathcal{L},\mathcal{W},\pi^*)\right]$. Unlike for our main results, however, this bound is not in general tight. For instance if c is constant irrespective of the data, then the value of $c(X)$ is irrelevant for the actions, and

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

regret, of a Bayesian agent, since this agent knows that $c(X)$ is not informative about X (or θ) and so ignores it. By contrast, an agent who acts based on their perceived posterior, as we assume in the main text, will take different actions (and thus incur different regret) for different choices of constant c.

B Proof of Lemma 4 from Main Appendix

Recall that we aim to show that for $\mathcal{I}$ the set of indicator functions for (open, closed, and half-open) intervals in $\mathbb{R}$, and $\mathcal{F}$ the set of all functions $f:\mathbb{R}\to[0,1]$ with $V(f)\leq 1$, we have $d_{\mathcal{I}}(\mu,\nu)=d_{\mathcal{F}}$ for all measures μ,ν.

To this end, for $\varepsilon>0$ let us consider a set $[-C,C]$ such that $\mu([-C,C])\geq 1-\frac{\varepsilon}{2}$ and $\nu([-C,C])\geq 1-\frac{\varepsilon}{2}$, and note that for $\mathcal{F}$ the class of functions with range contained in $[0,1]$ and signed variation bounded by one and any $f\in\mathcal{F}$,

$$\left|\left|\int_{-C}^{C} f(t)d\mu(t)-\int_{-C}^{C} f(t)d\nu(t)\right|-\left|\int f(t)d\mu(t)-\int f(t)d\nu(t)\right|\right|<\varepsilon.$$

For $f\in\mathcal{F}$, $t\in[-C,C]$, $f_{[a,b]}$ the restriction of f to $[a,b]$, and

$$IV\left(f_{[a,b]}\right)=\sup_{K\in\mathbb{N},t_0\leq\ldots\leq t_K\in[a,b]}\sum_{k=1}^{K}\left(f(t_k)-f(t_{k-1})\right)_+,$$

$$DV\left(f_{[a,b]}\right)=\sup_{K\in\mathbb{N},t_0\leq\ldots\leq t_K\in[a,b]}\sum_{k=1}^{K}\left(f(t_{k-1})-f(t_k)\right)_+,$$

in the same spirit as Jordan's theorem for functions of bounded variation we can write

$$f(t)=f(-C)+IV\left(f_{[-C,t]}\right)-DV\left(f_{[-C,t]}\right).$$

Hence, $f(t)=f_+(t)+f_-(t)$ for $f_+(t)=IV\left(f_{[-C,t]}\right)$ and $f_-(t)=f(-C)-DV\left(f_{[-C,t]}\right)$, where f_+ is increasing with $f_+(-C)=0$ and $f_+(C)\leq 1$, while f_- is decreasing with $f_-(-C)=f(-C)$ and $f_-(C)\geq f(-C)-1$.

For each $s\in\mathbb{N}$, consider the partition of $[0,1]$ by $\frac{z}{2^s}$ for $z\in Z_s=\{1,\ldots,2^s\}$. Note that the

simple functions

$$f_{s,+}(t)=\sum_{z\in Z_s}\frac{1\{f_+(t)\geq\frac{z}{2^s}\}}{2^s},f_{s,-}(t)=\sum_{z\in Z_s}\frac{1\{-f_-(t)\geq\frac{z}{2^s}+f(-C)-1\}}{2^s}+f(-C)-1$$

differ from f_+ and f_-, respectively, by at most $\frac{1}{2^s}$ in the sup norm. Next, let

$$t_{s,+,z}=\inf\Big\{t:f_+(t)\geq\frac{z}{2^s}\Big\},o_{s,+,z}=1\Big\{f_+(t_{s,+,z})\neq\frac{z}{2^s}\Big\}$$

$$t_{s,-,z}=\sup\Big\{t:-f_-(t)\geq\frac{z}{2^s}+f(-C)-1\Big\},o_{s,-,z}=1\Big\{f_-(t_{s,-,z})\neq\frac{z}{2^s}+f(-C)-1\Big\},$$

and note that

$$f_{s,+}(t)=\frac{1}{2^s}\sum_{z\in Z_s}(o_{s,+,z}1\{t>t_{s,+,z}\}+(1-o_{s,+,z})1\{t\geq t_{s,+,z}\})$$

while

$$f_{s,-}(t)=-\frac{1}{2^s}\sum_{z\in Z_s}(o_{s,-,z}1\{t<t_{s,-,z}\}+(1-o_{s,-,z})1\{t\leq t_{s,-,z}\})+f(-C)-1.$$

Let

$$f_{s,z}(t)=\begin{array}{c}o_{s,+,z}1\{t>t_{s,+,z}\}+(1-o_{s,+,z})1\{t\geq t_{s,+,z}\}+\\ o_{s,-,z}1\{t<t_{s,-,z}\}+(1-o_{s,-,z})1\{t\leq t_{s,-,z}\}-1\{t_{s,+,z}<t_{s,-,z}\}\end{array},$$

and note that we can write $f_{s,z}(t)$ as either $1\{t\in I\}$ or $1\{t\notin I\}$ for an interval I depending on the values of $(t_{s,+,z},o_{s,+,z},t_{s,-,z},o_{s,-,z})$. Moreover,

$$f_s(t)=f_{s,+}(t)+f_{s,-}(t)=\frac{1}{2^s}\sum_{z\in Z_s}(f_{s,z}(t)+1\{t_{s,+,z}<t_{s,-,z}\}+f(-C)-1),$$

so

$$\Big|\int_{-C}^{C}f_s(t)d\mu(t)-\int_{-C}^{C}f_s(t)d\nu(t)\Big|=$$

$$\Big|\frac{1}{2^s}\sum_{z\in Z_s}\Big(\int_{-C}^{C}f_{s,z}(t)d\mu(t)-\int_{-C}^{C}f_{s,z}(t)d\nu(t)\Big)\Big|\leq$$

$$\frac{1}{2^s}\sum_{z\in Z_s}\Big|\int_{-C}^{C}f_{s,z}(t)d\mu(t)-\int_{-C}^{C}f_{s,z}(t)d\nu(t)\Big|\leq$$

$$\sup_I \left| \int_{-C}^{C} 1\{t \in I\} d\mu(t) - \int_{-C}^{C} 1\{t \in I\} d\nu(t) \right| \vee \left| \int_{-C}^{C} 1\{t \notin I\} d\mu(t) - \int_{-C}^{C} 1\{t \notin I\} d\nu(t) \right|. \quad \text{(S1)}$$

Note, however, that since $\mu([-C,C]) \geq 1 - \frac{\varepsilon}{2}$ and $\nu([-C,C]) \geq 1 - \frac{\varepsilon}{2}$,

$$\left| \int_{-C}^{C} 1\{t \in I\} d\mu(t) - \int 1\{t \in I\} d\mu(t) \right| \leq \frac{\varepsilon}{2}$$

and similarly for ν, so

$$\left| \left| \int_{-C}^{C} 1\{t \in I\} d\mu(t) - \int_{-C}^{C} 1\{t \in I\} d\nu(t) \right| - \left| \int 1\{t \in I\} d\mu(t) - \int 1\{t \in I\} d\nu(t) \right| \right| < \varepsilon$$

and the same holds for $1\{t \notin I\}$, so (S1) is bounded above by $d_{\mathcal{I}}(\mu,\nu) + \varepsilon$. Since $f_s \to f$ uniformly in $t \in [-C,C]$ as $s \to \infty$, we thus have that

$$\left| \int_{-C}^{C} f(t) d\mu(t) - \int_{-C}^{C} f(t) d\nu(t) \right| =$$

$$\lim_{s \to \infty} \left| \int_{-C}^{C} f_s(t) d\mu(t) - \int_{-C}^{C} f_s(t) d\nu(t) \right| \leq d_I(\mu,\nu) + \varepsilon.$$

Moreover, since $0 \leq f(t) \leq 1$,

$$\left| \int_{-C}^{C} f(t) d\mu(t) - \int f(t) d\mu(t) \right| \leq \frac{\varepsilon}{2}$$

and similarly for ν, so by the triangle inequality

$$\left| \int f(t) d\mu(t) - \int f(t) d\nu(t) \right| \leq d_I(\mu,\nu) + 2\varepsilon.$$

Since we can repeat this argument for all $\varepsilon > 0$, we obtain that $\left| \int f(t) d\mu(t) - \int f(t) d\nu(t) \right| \leq d_{\mathcal{I}}(\mu,\nu)$ for all $f \in \mathcal{F}$. Conversely, $1\{t \in I\}$ is an element of $\mathcal{F}$ for all I, so

$$\sup_{f \in \mathcal{F}} \left| \int f(t) d\mu(t) - \int f(t) d\nu(t) \right| \leq d_{\mathcal{I}}(\mu,\nu),$$

proving the result.

C Sufficient Conditions for Bernstein-von Mises

This section provides sufficient conditions for the Bernstein-von Mises (BvM) results discussed in Section 5.2 of the paper. As in that section, we assume $X=(X_1,...,X_n)\in\mathcal{X}_0^n=\mathcal{X}$ consists of $n\in\mathbb{N}$ i.i.d. draws from a distribution $P\in\Delta(\mathcal{X}_0)$. Our goal is to provide sufficient conditions on (P,β,θ,π^*) such that for $\mathcal{F}$ the set of functions with range contained in [0,1] and signed variation bounded by one, there exists a class of functions $\mathcal{G}$ with $\mathcal{F}\circ\theta\subseteq\mathcal{G}$ such that for all $\delta>0$, $d_{\mathcal{F}\circ\theta}\big(\pi^B(\beta|X),\pi^*(\beta|X)\big)\leq\delta$ with probability tending to one as $n\to\infty$, allowing us to apply Lemma 3.

Our arguments in this section rely on approximate normality of the posterior $\pi^*(\beta|X)$ as $n\to\infty$. BvM results in the literature typically quantify deviations from normality using IPMs, specifically total variation metric when β is finite-dimensional and bounded Lipschitz metric when β is infinite-dimensional. Consequently, we structure our analysis in this appendix around these two metrics.

C.1 BvM in Total Variation

Let us first consider the case where the central posterior $\pi^*(\beta|X)$ satisfies a BvM Theorem in total variation. For this case, we will limit attention to the case where β corresponds to a finite-dimensional vector of moments, $\beta(P)=E_P[\psi(X_i)]\in\mathcal{B}\subseteq\mathbb{R}^q$. Correspondingly, let $\beta(P_n)=\frac{1}{n}\sum_{i=1}^n\psi(X_i)$ and $\Sigma(P)=\text{Var}_P(\psi(X_i))$ denote the sample mean and population variance, respectively. We assume that as $n\to\infty$, the reference posterior $\pi^*(\beta|X)$ converges to a normal with mean $\beta(P_n)$ and variance $\frac{1}{n}\Sigma(P)$.

Assumption S1. *As* $n\to\infty$

$$d_{TV}\left(\pi^*(\beta|X),N\left(\beta(P_n),\frac{1}{n}\Sigma(P)\right)\right)\to_p 0.$$

A variety of sufficient conditions for Assumption S1 are available in the literature—see for instance Theorem 12.8 in Ghosal and van der Vaart (2017). These results apply to general functionals and do not in general require that $\beta(P)$ correspond to a vector of population moments. However, this additional structure allows us to establish a corresponding result for the Bayes Bootstrap.

Lemma S1. *Provided $\Sigma(P)$ is finite and full-rank, as $n \to \infty$*

$$d_{TV}\left(\pi^B(\beta|X), N\left(\beta(P_n), \frac{1}{n}\Sigma(P)\right)\right) \to_p 0.$$

Recall that if we let $\mathcal{G}$ collect all functions $g:\mathcal{B}\to[0,1]$ then $d_{\mathcal{G}}(\mu,\nu) = d_{TV}(\mu,\nu)$. Since functions in $\mathcal{F}$ have range contained in [0,1] it is immediate that $\mathcal{F}\circ\theta \subseteq \mathcal{G}$ for all θ. Hence, Assumption S1 and Lemma S1 together imply that the conditions of Lemma 3 hold with probability tending to one, without imposing any further restrictions on θ.

Proposition S1. *If Assumption S1 and the conditions of Lemma S1 hold, then for $\mathcal{G}$ the class of all functions $g:\mathcal{B}\to[0,1]$, as $n\to\infty$*

$$d_{\mathcal{G}}\big(\pi^B(\beta|X), \pi^*(\beta|X)\big) \to_p 0,$$

so for any $\delta > 0$ and any $\theta:\mathcal{B}\to\mathbb{R}$ the conditions of Lemma 3 hold with probability tending to one.

C.2 BvM in Bounded Lipschitz

We next consider the case where $\pi^*(\beta|X)$ satisfies a BvM Theorem in bounded Lipschitz metric. Convergence in bounded Lipschitz metric is implied by convergence in total variation, so this condition is weaker than that considered in the previous section, and has been used in infinite-dimensional settings where convergence in total variation is too demanding.

Let $\beta:\Delta(\mathcal{X}_0)\to\mathcal{B}$ be some (possibly infinite-dimensional) transformation of the data distribution and $\mathcal{B}$ a normed vector space. Let $\beta(P_n)$ be the analogous transformation of the empirical distribution. We assume β is "well-behaved" in the sense that $\beta(P_n)$ is asymptotically normal.

Assumption S2. *As $n\to\infty$*

$$\sqrt{n}(\beta(P_n) - \beta(P)) \to_d \mathcal{GP}(0, \Sigma(P))$$

where $\mathcal{GP}(0,\Sigma(P))$ is the mean-zero Gaussian process with covariance function $\Sigma(P)$.

We assume that the posterior distribution $\pi^*(\beta|X)$ matches the limiting distribution in large samples, where the discrepancy is measured in bounded Lipschitz metric, $d_{BL} = d_{\mathcal{F}_{BL}}$, for $\mathcal{F}_{BL}$ is the set of functions with absolute value and Lipschitz constant bounded by one.

Assumption S3. *As* $n \to \infty$

$$d_{BL}\left(\pi^*\left(\sqrt{n}(\beta - \beta(P_n))|X\right), \mathcal{GP}(0, \Sigma(P))\right) \to_p 0.$$

For the case where $\beta(P)$ is the distribution function, sufficient conditions for Assumption S3 are provided in Castillo and Nickl (2014).

In parallel, we will assume that the Bayes bootstrap distribution, appropriately normalized, converges to the same limit.

Assumption S4. *As* $n \to \infty$

$$d_{BL}\left(\pi^B\left(\sqrt{n}(\beta - \beta(P_n))|X\right), \mathcal{GP}(0, \Sigma(P))\right) \to_p 0.$$

Convergence of the bootstrap distribution to normality in bounded Lipschitz metric is the standard definition of bootstrap consistency, and a wide variety of sufficient conditions are available: see for instance Theorems 3.6.13 and 3.9.11 in van der Vaart and Wellner (1996).

We further restrict the function $\theta(\beta)$ to satisfy a mild continuity condition, and require that the distributions $\pi^B(\theta|X)$ and $\pi^*(\theta|X)$ not have point masses in large samples. See Supplemental Appendix D for further discussion of both assumptions.

Assumption S5. *For all* $\epsilon > 0$*, there exists a constant* $K(\epsilon) \in [0, \infty)$*, a sequence of functions* $\theta_{n,\epsilon}$ *with Lipschitz constants* $K(\epsilon)$*, and an open set* $\mathcal{Z}_\epsilon$ *such that for* $Z_\beta \sim Q_\beta = \mathcal{GP}(0, \Sigma(P))$*, some norm* $\|\cdot\|$*, and* c_ϵ *the* $1-\epsilon$ *quantile of* $\|Z_\beta\|$*,*

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

In addition, $\|Z_\beta\|$ *is continuously distributed.*

Assumption S6. *For* $\mathcal{B}_{\varepsilon,n}(\theta) = \left\{\tilde{\theta} : \left|\tilde{\theta} - \theta\right| < \varepsilon/\sqrt{n}\right\}$,

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

Under these assumptions, the conditions of Lemma 3 again hold with probability tending to one.

Proposition S2. *Suppose Assumptions S2-S6 hold. Then for $\mathcal{G}$ equal to $\mathcal{F}\circ\theta$, as $n\to\infty$*

$$d_{\mathcal{G}}\big(\pi^B(\beta|X),\pi^*(\beta|X)\big)\to_p 0,$$

so for any $\delta>0$ the conditions of Lemma 3 hold with probability tending to one.

C.3 Proofs for Appendix C

Proof of Lemma S1 Note that

$$d_{TV}\left(\pi^B(\beta|X),N\left(\beta(P_n),\frac{1}{n}\Sigma(P)\right)\right)=d_{TV}\big(\pi^B\big(\sqrt{n}(\beta-\beta(P_n))|X\big),N(0,\Sigma(P))\big).$$

Hence, to prove the result, it suffices to show that the right hand side tends to zero.

By Theorem 12.2 of Ghosal and van der Vaart (2017), we know that

$$\pi^B\big(\sqrt{n}(\beta-\beta(P_n))|X\big)\to_d N(0,\Sigma(P)),$$

where since convergence in bounded Lipschitz metric is equivalent to convergence in distribution, this immediately implies that

$$d_{BL}\big(\pi^B\big(\sqrt{n}(\beta-\beta(P_n))|X\big),N(0,\Sigma(P))\big)\to_p 0.$$

Note, next, that by Proposition 4.3 in Ghosal and van der Vaart (2017),

$$\mathrm{E}_{\pi^B(\beta|X)}\big[\sqrt{n}(\beta-\beta(P_n))\big]=0,$$

$$\mathrm{Var}_{\pi^B(\beta|X)}\big(\sqrt{n}(\beta-\beta(P_n))\big)=\frac{n}{n+1}\Sigma(P_n)=\frac{n}{n+1}\mathrm{Var}_{P_n}(\psi(X_i)).$$

Hence, the first two moments of $\sqrt{n}(\beta-\beta(P_n))$ correspond to those of $N\big(0,\frac{n}{n+1}\Sigma(P_n)\big)$, while since $\frac{n}{n+1}\Sigma(P_n)\to_p\Sigma(P)$ as $n\to\infty$,

$$d_{BL}\left(N\left(0,\frac{n}{n+1}\Sigma(P_n)\right),N(0,\Sigma(P))\right)\to_p 0,$$

so by the triangle inequality

$$d_{BL}\left(\pi^B\big(\sqrt{n}(\beta-\beta(P_n))|X\big),N\left(0,\frac{n}{n+1}\Sigma(P_n)\right)\right)\to_p 0.$$

Moreover, since $M_n = \sqrt{\frac{n+1}{n}}\Sigma(P_n)^{-\frac{1}{2}} \to_p \Sigma(P)^{-\frac{1}{2}}$,

$$d_{BL}\big(\pi^B\big(M_n\sqrt{n}(\beta-\beta(P_n))|X\big),N(0,I)\big)\to_p 0.$$

Theorem 2.2 of Choudhuri (1998) shows that if the convex hull of $\{\psi(X_i): i\in\{1,...,n\}\}$ has an interior, then the density of $\pi^B(\sqrt{n}(\beta-\beta(P_n))|X)$ is log-concave, and thus so is the density of $\pi^B(M_n\sqrt{n}(\beta-\beta(P_n))|X)$. Since we have assumed $\Sigma(P)$ is full rank, this condition on the convex hull holds with probability tending to one. Proposition 1 of Meckes and Meckes (2014) implies that for log-concave distributions μ and ν with mean zero and identity variance

$$d_{TV}(\mu,\nu)\leq C\sqrt{d_{BL}(\mu,\nu)}$$

for a dimension-dependent constant C. It follows immediately that

$$d_{TV}\left(\pi^B\big(\sqrt{n}(\beta-\beta(P_n))|X\big),N\left(0,\frac{n}{n+1}\Sigma(P_n)\right)\right)=$$

$$d_{TV}\left(\pi^B\big(M_n\sqrt{n}(\beta-\beta(P_n))|X\big),N\left(0,\frac{n}{n+1}\Sigma(P_n)\right)\right)\to_p 0$$

and, since $d_{TV}\big(N\big(0,\frac{n}{n+1}\Sigma(P_n)\big),N(0,\Sigma(P))\big)\to_p 0$, that

$$d_{TV}\big(\pi^B\big(\sqrt{n}(\beta-\beta(P_n))|X\big),N(0,\Sigma(P))\big)\to_p 0$$

as we aimed to show. □

Proof of Proposition S1 By the triangle inequality,

$$d_{\mathcal{G}}\big(\pi^B(\beta|X),\pi^*(\beta|X)\big)=d_{TV}\big(\pi^B(\beta|X),\pi^*(\beta|X)\big)\leq$$

$$d_{TV}\left(\pi^B(\beta|X),N\left(\beta(P_n),\frac{1}{n}\Sigma(P)\right)\right)+d_{TV}\left(\pi^*(\beta|X),N\left(\beta(P_n),\frac{1}{n}\Sigma(P)\right)\right),$$

where the second term converges to zero by Assumption S1, and the first converges to zero by Lemma S1. □

Proof of Proposition S2 The result follows by the same argument used to prove Proposition S3 below. Specifically, Assumptions S2 and S5 imply Assumption S7 below, while Assumptions S3, S4, and S6 imply Assumption S8. Hence, it follows from Proposition S3

that $d_{SK}\big(\pi^B(\theta|X),\pi^*(\theta|X)\big)\to_p 0$, from which the result is immediate. □

D 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 S7. *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:*

(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 $\epsilon>0$, and for $\mathcal{F}_{L,K}$ the set of non-negative functions with Lipschitz constant bounded by K, there exists a constant $K(\epsilon)\in[0,\infty)$, a sequence of functions $\theta_{n,\epsilon}\in\mathcal{F}_{L,K(\epsilon)}$, 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 S7(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.[2] Assumption S7(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 $\hat{\theta}$.[3]

For instance, if θ is the maximum of two means, $\theta(P_{0,n})=\max\big\{\mathrm{E}_{P_{0,n}}[X_{i,1}],\mathrm{E}_{P_{0,n}}[X_{i,2}]\big\}$ where $(X_{i,1},X_{i,2})$ are bounded, Assumption S7 holds if we take $\hat{\beta}_n=\beta_n(P_n)=\sqrt{n}F_{P_n}$ to

[2] 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).

[3] Assumption S7(b) is implied by local Lipschitz conditions such as those considered by Kitagawa et al. (2020).

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 S7 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 S8. *For a bootstrap with distribution $\eta(X)$ and other objects as defined in Assumption S7, we have that for $\mathcal{F}_{BL,K}$ the set of functions with Lipschitz constant and absolute value bounded by K,*

(a)

$$\sup_{h\in BL_1}\left\{\mathrm{E}_{\eta_\beta(X)}\left[h\left(\beta-\hat{\beta}_n\right)\right]-\mathrm{E}_{Q_\beta}\left[h\left(\hat{\beta}_n-\beta_{0,n}\right)\right]\right\}\to_p 0. \tag{S2}$$

(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 S8(a) states a sense in which the bootstrap consistently recovers the asymptotic distribution of the statistics $\hat{\beta}_n$. Importantly, Assumption S8(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 S8(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 S8(b) allows for cases where the bootstrap need not be consistent, such as the examples discussed above.

If Assumption S7 holds for the estimator $\hat{\theta}$ and Assumption S8 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 S3. *If Assumption S7 holds for $\hat{\theta}$ and Assumption S8 holds for bootstraps A*

and B with distributions $\eta^A(X)$ and $\eta^B(X)$, then

$$d_{SK}\big(\eta_\theta^A(X),\eta_\theta^B(X)\big)\to_p 0.$$

Proof of Proposition S3 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 S8(a) implies that

$$\sup_{h\in\mathcal{F}_{BL,1}}\Big\{\mathrm{E}_{\eta_\beta^A(X)}\Big[h\Big(\beta-\hat{\beta}_n\Big)\Big]-\mathrm{E}_{\eta_\beta^B(X)}\Big[h\Big(\beta-\hat{\beta}_n\Big)\Big]\Big\}=$$

$$\sup_{h\in\mathcal{F}_{BL,1}}\Big\{\mathrm{E}\Big[h\Big(\hat{\beta}_n^{*,A}-\hat{\beta}_n\Big)|X\Big]-\mathrm{E}\Big[h\Big(\hat{\beta}_n^{*,B}-\hat{\beta}_n\Big)|X\Big]\Big\}\to_p 0.$$

To prove Proposition S3, we first show that the bootstrap distributions for $\hat{\theta}_n^{*,A}=\theta_n\Big(\hat{\beta}_n^{*,A}\Big)$ and $\hat{\theta}_n^{*,B}=\theta_n\Big(\hat{\beta}_n^{*,B}\Big)$ are equivalent in bounded Lipschitz metric,

$$\sup_{h\in\mathcal{F}_{BL,1}}\Big|\mathrm{E}\Big[h\Big(\hat{\theta}_n^{*,A}\Big)|X\Big]-\mathrm{E}\Big[h\Big(\hat{\theta}_n^{*,A}\Big)|X\Big]\Big|\to_p 0. \tag{S3}$$

To establish the equivalence (S3), note that

$$\sup_{h\in\mathcal{F}_{BL,1}}\Big|\mathrm{E}\Big[h\Big(\hat{\theta}_n^{*,A}\Big)|X\Big]-\mathrm{E}\Big[h\Big(\hat{\theta}_n^{*,B}\Big)|X\Big]\Big|=$$

$$\sup_{h\in\mathcal{F}_{BL,1}}\Big|\mathrm{E}\Big[h\Big(\tilde{\theta}_n\Big(\hat{\beta}_n^{*,A}-\beta_{0,n}\Big)\Big)|X\Big]-\mathrm{E}\Big[h\Big(\tilde{\theta}_n\Big(\hat{\beta}_n^{*,B}-\beta_{0,n}\Big)\Big)|X\Big]\Big|$$

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

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

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

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 S7(a) and S8(a) imply that

$$\begin{pmatrix}\hat{\beta}_n^{*,A}-\hat{\beta}_n\\ \hat{\beta}_n-\beta_{0,n}\end{pmatrix}\to_d\begin{pmatrix}Z_\beta^*\\ Z_\beta\end{pmatrix},\ Z_\beta^*,Z_\beta\overset{i.i.d.}{\sim}F_\beta. \tag{S4}$$

Hence,

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

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

Note, however, that

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

where the first inequality follows from (S4) and the fact that $\|Z_\beta\|$ is continuously distributed, while the second follows from Assumption S7(a), the joint convergence (S4), the Portmanteau Lemma (Lemma 2.2 of van der Vaart 1998), and the fact that $\mathcal{C}_\varepsilon$ is open. We thus have that

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{h\in\mathcal{F}_{BL,1}}\left|\mathrm{E}\left[h\left(\tilde{\theta}_n\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)-h\left(\theta_{n,\epsilon}\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)|X\right]\right|\right]\leq 2\epsilon,$$

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

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{h\in\mathcal{F}_{BL,1}}\left|\mathrm{E}\left[h\left(\theta_n\left(\hat{\beta}_n^{*,A}\right)\right)|X\right]-\mathrm{E}\left[h\left(\theta_n\left(\hat{\beta}_n^{*,B}\right)\right)|X\right]\right|\right]\leq$$

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{h\in\mathcal{F}_{BL,1}}\left|\mathrm{E}\left[h\left(\theta_{n,\epsilon}\left(\hat{\beta}_n^{*,A}-\beta_{0,n}\right)\right)|X\right]-\mathrm{E}\left[h\left(\theta_{n,\epsilon}\left(\hat{\beta}_n^{*,B}-\beta_{0,n}\right)\right)|X\right]\right|\right]+4\epsilon\leq$$

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

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

$$\sup_{h\in\mathcal{F}_{BL,K(\epsilon)}}\left|\mathrm{E}\left[h\left(\hat{\beta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\beta}_n^{*,B}\right)|X\right]\right|=K(\epsilon)\cdot\sup_{h\in\mathcal{F}_{BL,1}}\left|\mathrm{E}\left[h\left(\hat{\beta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\beta}_n^{*,B}\right)|X\right]\right|,$$

so Assumption S8(a) implies that

$$\limsup_{n\to\infty}\mathrm{E}\left[\sup_{h\in\mathcal{F}_{BL,1}}\left|\mathrm{E}\left[h\left(\hat{\theta}_n^{*,A}\right)|X\right]-\mathrm{E}\left[h\left(\hat{\theta}_n^{*,B}\right)|X\right]\right|\right]\leq 4\epsilon.$$

Since we can repeat this argument for all $\epsilon > 0$, we have verified (S3).

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 $h_\upsilon \in \mathcal{F}_{BL,1}$ 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 $\upsilon \in (0,1)$

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

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

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

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

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

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

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

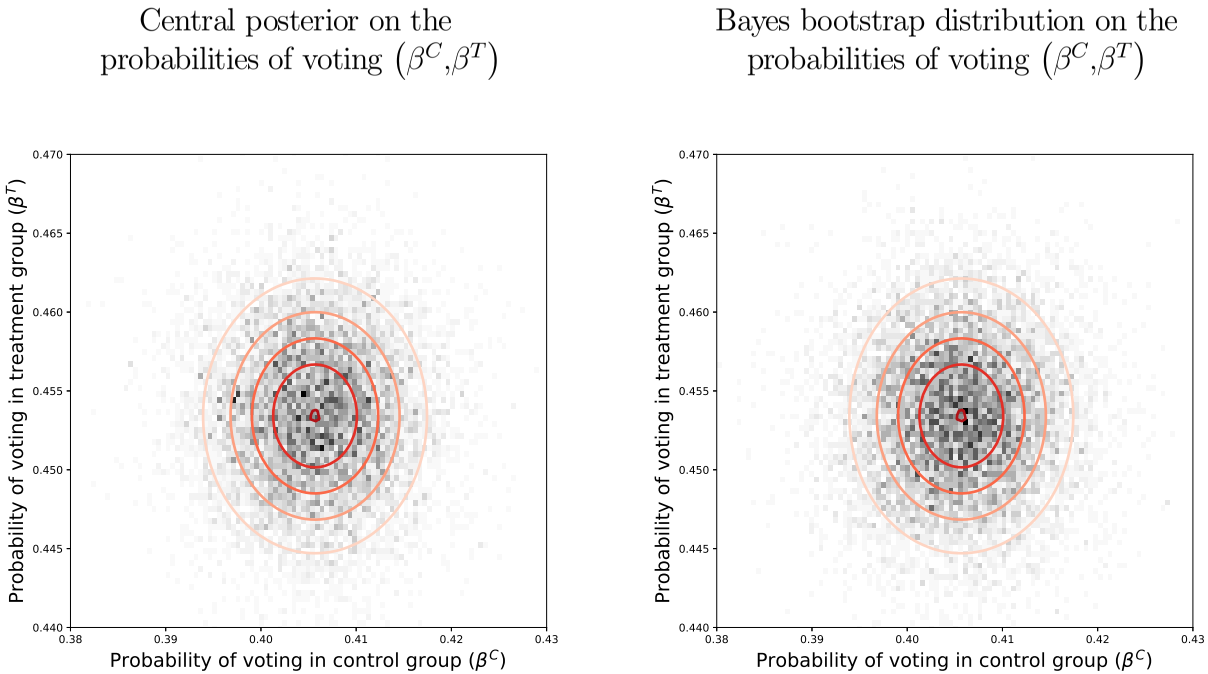
However, we have already established that the first term goes to zero for all $\upsilon \in (0,1)$, while the second and third terms go to zero as $\upsilon \to 0$ by Assumption S8(b). Hence, $\eta_{\theta_n}^A(X)$ and $\eta_{\theta_n}^B(X)$ converge in Kolmogorov metric. Since SK is bounded by twice the Kolmogorov distance

$$SK\left(\eta_\theta^A(X), \eta_\theta^B(X)\right) \leq 2 \sup_{\tilde{\theta}} \left| F_{\eta_\theta^A(X)}\left(\tilde{\theta}\right) - F_{\eta_\theta^B(X)}\left(\tilde{\theta}\right) \right|,$$

the conclusion of the proposition follows immediately. □

E Additional Plots from Running Example

Online Appendix Figure 1: Central prior and posteriors on BvM parameters



Notes: The left plot is a heat plot of samples from the central posterior on (β^C,β^T). The right plot is a heat plot of samples from the Bayes bootstrap distribution on (β^C,β^T). On both heat plots, we overlay a contour plot of the density of the normal approximation parameterized by the maximum likelihood estimate and its associated covariance matrix.

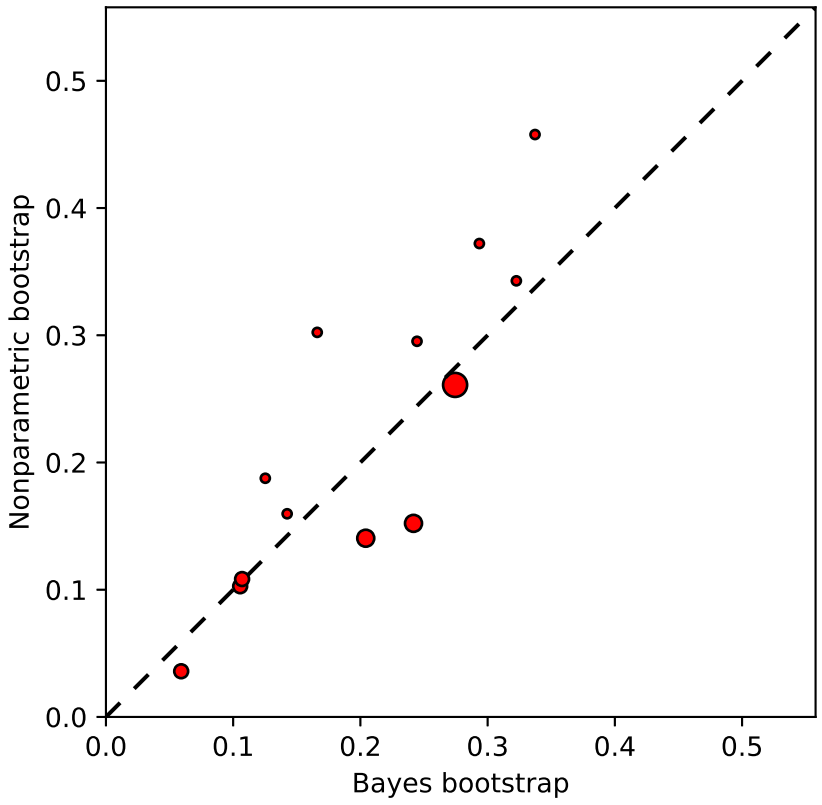
F Additional Findings from Bootstrap Census

Supplemental Appendix Table 1: List of Papers in Bootstrap Census

Citation	Objects	Transmitted?
Abebe et al. (2021b,a)	7	N
Adermon et al. (2021a,b)	2	Y
Bailey et al. (2021b,a)	6	Y
Bourreau et al. (2021b,a)	20	Y
Braguinsky et al. (2021b,a)	2	N
Dinerstein and Smith (2021b,a)	1	N
Finkelstein et al. (2021, 2022)	3	Y
Goodman-Bacon (2021b,a)	2	Y
Känzig (2021b,a)	6	N
Kostøl and Myhre (2021b,a)	5	Y
Mueller et al. (2021, 2020)	3	N
Reimers and Waldfogel (2021b,a)	7	N
Seibold (2021b,a)	4	Y
Weaver (2021b,a)	13	Y

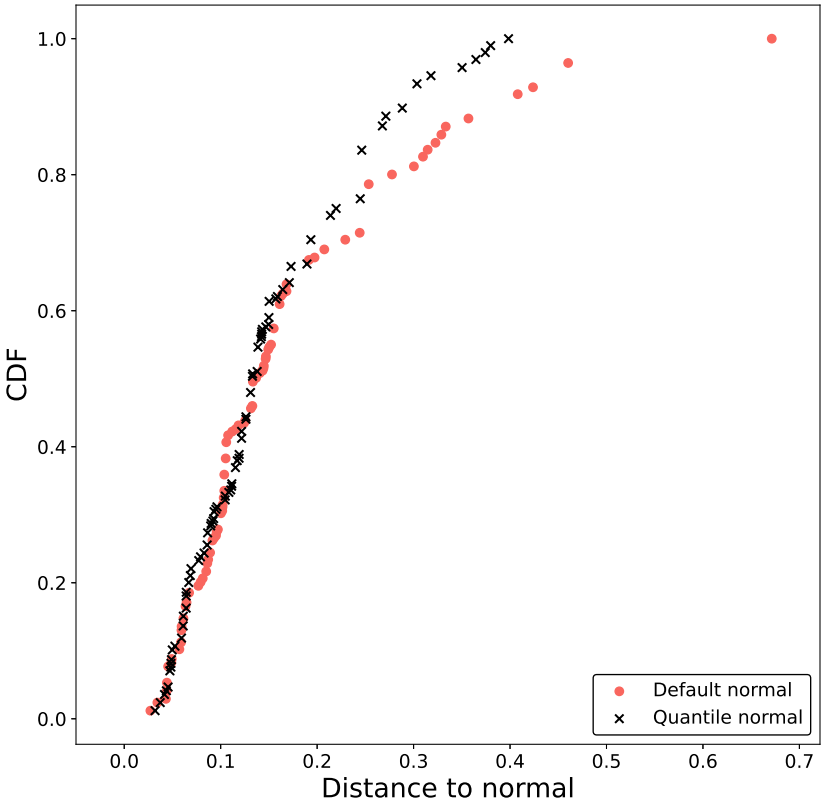
Notes: For each paper in our bootstrap census, the table reports the number of objects of interest for which we obtain replicates and an indicator for whether or not we received a transmission of bootstrap replicates directly from the authors. Papers are in ascending alphabetical order by the first author's last name.

Supplemental Appendix Figure 2: SK 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 SK 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 SK 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 SK distance between the nonparametric bootstrap distribution and the default normal report, and a 95% confidence interval for the SK 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.

Supplemental Appendix Figure 3: Distribution of SK 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 in the paper. For each object of interest we calculate the SK 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 SK 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. The plot shows the weighted empirical CDF of each of these two distances across the objects of interest in our census.

Supplemental Appendix References

Abebe, Girum, A. Stefano Caria, and Esteban Ortiz-Ospina (2021a): "Data and code for: The Selection of Talent: Experimental and Structural Evidence from Ethiopia," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-05-26., https://doi.org/10.3886/E127401V1.

——— (2021b): "The Selection of Talent: Experimental and Structural Evidence from Ethiopia," *American Economic Review*, 111 (6), 1757–1806.

Adermon, Adrian, Mikael Lindahl, and Mårten Palme (2021a): "Dynastic Human Capital, Inequality, and Intergenerational Mobility," *American Economic Review*, 111 (5), 1523–1548.

——— (2021b): "Replication code for: Dynastic Human Capital, Inequality and Intergenerational Mobility," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-04-09., https://doi.org/10.3886/E123761V1.

Andrews, Isaiah and Jesse M. Shapiro (2021): "A Model of Scientific Communication," *Econometrica*, 89 (5), 2117–2142.

Bailey, Martha J., Shuqiao Sun, and Brenden Timpe (2021a): "Data and Code for: Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-11-19., https://doi.org/10.3886/E146361V1.

——— (2021b): "Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency," *American Economic Review*, 111 (12), 3963–4001.

Bourreau, Marc, Yutec Sun, and Frank Verboven (2021a): "Code and Data for: Market Entry, Fighting Brands and Tacit Collusion," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-10-19., https://doi.org/10.3886/E138921V1.

——— (2021b): "Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market," *American Economic Review*, 111 (11), 3459–3499.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson (2021a): "Data and Code for: Product Innovation, Product Diversification, and Firm Growth: Evidence from Japan's Early Industrialization," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-11-19., https://doi.org/10.3886/E145421V1.

——— (2021b): "Product Innovation, Product Diversification, and Firm Growth: Evidence from Japan's Early Industrialization," *American Economic Review*, 111 (12), 3795–3826.

Castillo, Ismaël and Richard Nickl (2014): "On the Bernstein–von Mises Phenomenon for Nonparametric Bayes Procedures," *Annals of Statistics*, 42 (5), 1941–1969.

Choudhuri, Nidhan (1998): "Bayesian Bootstrap Credible Sets for Multidimensional Mean Functional," *Annals of Statistics*, 26 (6), 2104–2127.

Dinerstein, Michael and Troy D. Smith (2021a): "Data and Code for: Quantifying the Supply Response of Private Schools to Public Policies," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-09-28., https://doi.org/10.3886/E141201V1.

——— (2021b): "Quantifying the Supply Response of Private Schools to Public Policies," *American Economic Review*, 111 (10), 3376–3417.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (2021): "Place-Based Drivers of Mortality: Evidence from Migration," *American Economic Review*, 111 (8), 2697–2735.

——— (2022): "Data and Code for: Place-Based Drivers of Mortality: Evidence from Migration," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-08-08., https://doi.org/10.3886/E125381V2.

Ghosal, Subhashis and Aad van der Vaart (2017): *Fundamentals of Nonparametric Bayesian Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press.

Goodman-Bacon, Andrew (2021a): "Data and Code for: The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health and Labor Market Outcomes," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-07-15., https://doi.org/10.3886/E127621V1.

——— (2021b): "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes," *American Economic Review*, 111 (8), 2550–2593.

Känzig, Diego R. (2021a): "Data and Code for: The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-03-19., https://doi.org/10.3886/E122886V1.

——— (2021b): "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements," *American Economic Review*, 111 (4), 1092–1125.

Kitagawa, Toru, José Luis Montiel Olea, Jonathan Payne, and Amilcar Velez (2020): "Posterior Distribution of Nondifferentiable Functions," *Journal of Econometrics*, 217 (1), 161–175.

Kostøl, Andreas R. and Andreas S. Myhre (2021a): "Code for: Labor Supply Responses to Learning the Tax and Benefit Schedule," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-10-19., https://doi.org/10.3886/E144121V1.

——— (2021b): "Labor Supply Responses to Learning the Tax and Benefit Schedule," *American Economic Review*, 111 (11), 3733–3766.

Meckes, Elizabeth S and Mark W Meckes (2014): "On the Equivalence Of Modes of Convergence for Log-concave Measures," in *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2011-2013*, Springer, 385–394.

Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa (2020): "Data and Codes for: Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-12-17., https://doi.org/10.3886/E120501V1.

——— (2021): "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias," *American Economic Review*, 111 (1), 324–363.

Reimers, Imke and Joel Waldfogel (2021a): "Data and Code for: Digitization and Pre-Purchase Information: The Causal and Welfare Effects of Reviews and Crowd Ratings," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-05-24., https://doi.org/10.3886/E130802V1.

——— (2021b): "Digitization and Pre-purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings," *American Economic Review*, 111 (6), 1944–1971.

Seibold, Arthur (2021a): "Data and Code for: Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-03-19., https://doi.org/10.3886/E120903V1.

——— (2021b): "Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities," *American Economic Review*, 111 (4), 1126–1165.

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

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

Weaver, Jeffrey (2021a): "Data for: Jobs for Sale: Corruption and Misallocation in Hiring," Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-09-23., https://doi.org/10.3886/E141881V1.

——— (2021b): "Jobs for Sale: Corruption and Misallocation in Hiring," *American Economic Review*, 111 (10), 3093–3122.