Classical p-values and the Bayesian posterior probability that the hypothesis is approximately true

Brendan Kline

Abstract. This paper relates p-values for the hypothesis that $\theta = c$ to the Bayesian posterior probability that the hypothesis is approximately true, in the sense that $\theta \in [c - \epsilon, c + \epsilon]$ for a selected $\epsilon > 0$. In a setup with a continuous prior for θ, the results show that a larger (respectively, smaller) p-value does not necessarily correspond to a higher (respectively, lower) probability that θ is close to c. Therefore, the results suggest caution about common ways of using p-values, specifically the use of small p-values as a key standard in empirical research.

Keywords: frequentist, hypothesis, posterior, testing

JEL classification: C11, C12, C18

1. Introduction

This paper is about p-values for the hypothesis $\theta = c$, where θ is a scalar parameter and c is a known constant (e.g., often $c = 0$). An important question is how to use/interpret p-values in empirical practice to draw conclusions about θ. For example, one common usage of p-values in practice concerns the case where θ is a treatment effect, like a regression coefficient. In that case, in practice to establish/publish an empirical finding that the treatment has an effect on an outcome, it is considered important to find a small p-value for the hypothesis that the treatment effect is zero. This is a running example, but the results hold regardless of the meaning of θ.

University of Texas at Austin
E-mail address: brendan.kline@austin.utexas.edu.
Date: September 2021. Thanks to Elie Tamer, and audiences at Cornell University and the Midwest Econometrics Group for helpful comments and discussion. This paper looks at different questions based on some of the ideas that originally appeared in "Bayesian conclusions from classical p-values" by the same author, first distributed in October 2018. Any errors are mine.

The results of Edwards, Lindman, and Savage (1963), Berger and Delampady (1987), Berger and Sellke (1987), and Sellke, Bayarri, and Berger (2001) as representative of a related literature establish that it is possible to "calibrate" a p-value via the corresponding minimum posterior probability that $\theta = c$. In the setup of that literature, there is a positive prior probability that $\theta = c$, and a *class* of priors for θ on $\theta \neq c$. The "minimum" is taken over this class of priors. Although the p-value is smaller than the minimum posterior probability in general, there is a monotone increasing relationship between the p-value and the minimum posterior probability of the hypothesis that $\theta = c$.[1][2] Essentially, this literature implies that only very small p-values are enough to "reject" that $\theta = c$. Conversely, larger p-values correspond to relatively higher minimum posterior probabilities that $\theta = c$. This provides a Bayesian justification for "rejecting" the hypothesis based on a sufficiently small p-value, and "not rejecting" the hypothesis based on larger p-values. Put differently, this provides a justification for using small p-values as a standard for empirical evidence that θ is unlikely to be equal to c.[3] Indeed, the Benjamin, Berger, Johannesson, et al. (2018) proposal to redefine "statistical significance" to be a p-value smaller than 0.005 is based in part on such considerations of how p-values relate to the (minimum) posterior probability that $\theta = c$.

In practice, often the empirical question is not whether $\theta = c$ is literally true but rather whether θ is "close" to c. Specifically, consider "ϵ-approximate" hypotheses of the form $\theta \in [c - \epsilon, c + \epsilon]$ for some specified $\epsilon > 0$ chosen by the econometrician.[4] An ϵ-approximate

[1]There are many "minimum posterior probabilities" proposed in that literature, based on different classes of priors over which the minimum is taken, and monotonicity is a shared property.

[2]Lindley's paradox (e.g., Jeffreys (1939); Lindley (1957)) establishes conditions for the p-value to be less than the posterior probability of the hypothesis for a (uniform) *single* prior on $\theta \neq c$. A review of the literature on minimum posterior probabilities can be found in Held and Ott (2018).

[3]Of course, with classical hypothesis testing without any reference to Bayesian analysis, smaller p-values are interpreted as rejecting the hypothesis, and larger p-values lead to not rejecting the hypothesis and/or the conclusion that there is no empirical evidence against the hypothesis. Along similar lines to the latter case of not rejecting the hypothesis, using a "limited information" posterior that conditions on the dichotomous event of rejecting (or not) the hypothesis $\theta = c$ by classical methods, Abadie (2020) shows the posterior conditional on not rejecting the hypothesis has much higher probability mass at $\theta = c$ compared to the prior, whereas the posterior conditional on rejecting the hypothesis is not too different from the prior.

[4]It does not matter if the ϵ-approximate hypothesis is defined as a closed interval or open interval.

hypothesis is about θ being "close" to c. In the setup with a positive prior probability that $\theta = c$, it can also be justified to use small p-values as a standard for empirical evidence that θ is unlikely to be "close" to c, using the same "calibration" of p-values for the hypothesis $\theta = c$. The reason: when there is positive probability that $\theta = c$ and ϵ is small enough, the posterior probability of $\theta = c$ is approximately the same as the posterior probability of $\theta \in [c - \epsilon, c + \epsilon]$. See Remark 4 for more on this point.

However, particularly in the social sciences, often there is zero prior probability that $\theta = c$. So, in that setting, how does the p-value relate to the posterior probability that θ is close to c? In other words, in that setting, is there a (Bayesian) justification for using small p-values as a key standard for empirical evidence? (Obviously, in that setting it is not interesting to ask about the posterior probability that $\theta = c$, since that is necessarily zero.) This paper relates p-values for the hypothesis that $\theta = c$ to the Bayesian posterior probability of $\theta \in [c - \epsilon, c + \epsilon]$ in a setup with a continuous prior, with zero prior probability that $\theta = c$.

The results allow the econometrician to choose any ϵ. ϵ can be a selected constant or data-dependent. It is possible to interpret ϵ as representing the magnitude of deviations from $\theta = c$ that are viewed as "economically significant." In particular, because ϵ can be data-dependent, in the case that θ reflects a treatment effect, ϵ can be taken to be proportionate to the (estimated) effect of a reference treatment, so that the ϵ-approximate hypothesis relates to the effect of the treatment under study relative to the reference treatment. Similar to other hypothesis testing problems, the scaling of the model parameters is an important consideration; θ, c, and ϵ are all defined in terms of the same scale (units of measurement).

The paper derives a closed-form expression for the posterior probability of the ϵ-approximate hypothesis, as a function of the p-value (and classical estimate of θ). The main results are large sample approximations, which are valid quite generally in parametric and semiparametric models when the Bernstein-von Mises phenomenon holds. There are also finite sample results, based on distributional assumptions on the data generating process and a uniform prior.

The large sample and finite sample results are similar. Then, it is possible to investigate the relationship between the p-value and the posterior probability of the ϵ-approximate hypothesis.

In short, the results of this paper show that a small p-value is neither necessary nor sufficient for there to be a small posterior probability that θ is close to c. In particular, relatively larger p-values can correspond to relatively smaller posterior probabilities that θ is close to c.

This has important implications for using and interpreting p-values. As a running example, if θ is a treatment effect (e.g., a regression coefficient), the results can be used to relate a p-value for the hypothesis $\theta = 0$ to the probability that the treatment effect is close to 0. Concretely, close to 0 would mean that $\theta \in [-\epsilon, \epsilon]$ for $\epsilon > 0$ chosen by the econometrician. In this treatment effects setting, the results suggest caution about the practice of using small p-values as a key standard for empirical evidence, like the standard for finding/publishing an effect of a treatment on an outcome.[5] In particular the results show that a larger (respectively, smaller) p-value for the hypothesis that the treatment effect is zero does not necessarily correspond to a higher (respectively, lower) probability that the treatment effect is close to zero. Comparing two studies, a study reporting a larger p-value can actually have a smaller posterior probability of a treatment effect that is close to zero. This suggests caution against using small p-values as a key standard for empirical evidence, as in cutoff rules for determining "significance." This contrasts with the previously discussed results from the literature that do provide justifications for using small p-values as a standard in different settings/setups. The results imply the research community can "miss" (e.g., not publish) treatment effects that are probably *not* close to zero if the research community only "accepts" (or focuses on) treatment effects with small p-values for the hypothesis of zero treatment effect. Formally, "probably not close to zero" refers to the probability of the ϵ-approximate hypothesis. The

[5]It is clear that standards for empirical research are such that a small p-value is important. Some evidence is that Masicampo and Lalande (2012), Leggett, Thomas, Loetscher, and Nicholls (2013), and Brodeur, Lé, Sangnier, and Zylberberg (2016) find high prevalence of p-values around and just below the "significance cutoff" of 0.05.

expressions derived in this paper make it possible to use a p-value to assess the posterior probability of the ϵ-approximate hypothesis. This can be viewed as either a complement or substitute to the standard approach of reporting p-values in empirical research; in particular, it is possible to *retroactively* use these expressions to assess the posterior probability of the ϵ-approximate hypothesis from published research.

1.1. **Notation.** Notation is standard. The $m \times m$ identity matrix is $I_{m \times m}$. The L^p norm of $x \in \mathbb{R}^m$ is $||x||_p$. Let $\delta_{\Sigma,2}(x,y) = ||\Sigma^{-\frac{1}{2}}(x-y)||_2$ for a positive definite Σ. The total variation distance between random variables X and Y is $||X-Y||_{TV}$. For a sequence X_N, $X_N \to^p X$ means converges in probability to X and $X_N \to^d X$ means converges in distribution to X. The indicator $1[E]$ of logical statement E is 1 when E is true and 0 otherwise. As convention, $z1[\text{false}] = 0$; for example, $\frac{x}{y}1[y \neq 0] = 0$ when $y = 0$. A multivariate normal distribution with mean μ and covariance Σ is $\mathcal{N}(\mu, \Sigma)$. A (central) chi-squared distribution with m degrees of freedom is χ^2_m. A noncentral chi-squared distribution with m degrees of freedom and non-centrality parameter λ is $\chi^2_{m,\lambda}$.[6] For a random variable X, the cumulative distribution function is $F_X(\cdot)$, the quantile function is $Q_X(\cdot)$, the complementary cumulative distribution function is $\overline{F}_X(\cdot) \equiv 1 - F_X(\cdot)$, and $X(\mathcal{B}) \equiv P(X \in \mathcal{B})$.

2. Setup

The data is a sample of N i.i.d. observations from P_0, the true data generating process, so the data is $X^{(N)} \equiv \{X_i\}_{i=1}^N$ where $X_i \sim^{iid} P_0$. Let $\psi = (\theta, \gamma_2, \ldots, \gamma_t, \ldots, \gamma_m)$ be the finite-dimensional parameter of the model, with parameter space $\Psi \subseteq \mathbb{R}^m$ for some m. The p-values and posterior probabilities will concern the scalar parameter of interest θ. The true value of ψ is ψ_0. Per Remark 1 about semi-parametric models, there can also be infinite-dimensional nuisance parameters.

[6]This paper uses the parameterization where $E(\chi^2_{m,\lambda}) = m + \lambda$.

There is a classical ("frequentist") estimator $\hat{\psi}_N = (\hat{\theta}_N, \hat{\gamma}_{N,2}, \ldots, \hat{\gamma}_{N,t}, \ldots, \hat{\gamma}_{N,m})$ of ψ, which is the basis for the p-value. Also there is a Bayesian posterior for ψ, denoted $\Pi_{\psi|X^{(N)}}(\cdot)$ so $\Pi_{\psi|X^{(N)}}(A)$ is the posterior probability that $\psi \in A$ conditional on the data. The posterior in this paper conditions on the entire dataset, so this paper relates p-values to a conventional Bayesian analysis based on the entire dataset.

3. Large sample approximation results

The large sample approximation results are based on an assumption about the relationship between the sampling distribution of $\hat{\psi}_N$ and the posterior distribution for ψ.

Assumption 1. *It holds that:*

(1) *The classical estimator $\hat{\psi}_N$ is asymptotically normal, in the sense that $\sqrt{N}(\hat{\psi}_N - \psi_0) \to^d \mathcal{N}(0, \Sigma_0)$ where Σ_0 is nonsingular. There is a consistent estimator $\hat{\Sigma}_N$ of Σ_0.*

(2) *The Bayesian posterior for ψ is asymptotically normal, in the sense that $||\Pi_{\sqrt{N}(\psi - \hat{\psi}_N)|X^{(N)}} - \mathcal{N}(0, \Sigma_0)||_{TV} \to^p 0$.*

The analysis treats Σ_0 as an unknown quantity, with corresponding estimate $\hat{\Sigma}_N$. A possible alternative analysis is the known Σ_0 case. Because Σ_0 typically depends on unknown model parameters, it is unlikely that Σ_0 would be known in empirical practice, so it must be estimated using the data. Further, the known Σ_0 case is not particularly interesting from the perspective of this paper: If Σ_0 were known, then $\hat{\theta}_N$ (implicitly along with N and c) would be enough information to compute the p-value, and therefore reporting a p-value would convey no additional information beyond the information from reporting $\hat{\theta}_N$. But, when Σ_0 is unknown, as in empirical practice, the p-value does contain additional information beyond $\hat{\theta}_N$, which justifies the empirical practice of reporting p-values in addition to $\hat{\theta}_N$. Indeed, a key driving force of the results is that the p-value contains information about Σ_0. This is relevant information since the posterior distribution depends on Σ_0. This condition on the covariance corresponds also to the analogous condition on the posterior used in the finite sample results in Section 5. This provides another intuition for why this condition on the covariance is used; in the finite sample results, it arises because of the unknown (co)variance of the underlying data generating process. Correspondingly, essentially the same condition also holds "in the limit" in the large sample approximation.

Assumption 1 holds as the consequence of theorems commonly known as "Bernstein-von Mises theorem(s)." The following remark quickly summarizes settings where the "Bernstein-von Mises theorem(s)" hold. The results of this section apply to those settings.

Remark 1 (Sufficient conditions for Assumption 1)**.** Assumption 1 holds in these settings:

(1) **Parametric model**: Consider a parametric model P_ψ with finite-dimensional parameter ψ. Then $P_0 = P_{\psi_0}$, where ψ_0 is the true value. Often, $\hat{\psi}_N$ is the maximum likelihood estimate of ψ, and Σ_0 the inverse Fisher information matrix. The prior for ψ must have a continuous and positive density on a neighborhood of ψ_0. Under a few further regularity conditions, the parametric Bernstein-von Mises results (e.g., Le Cam (1986, Chapter 12), Van der Vaart (1998, Chapter 10), Le Cam and Yang (2000, Chapter 8)) imply Assumption 1 is satisfied.

(2) **Semi-parametric model**: Consider a semi-parametric model $P_{\psi,\eta}$ with finite-dimensional parameter ψ and infinite-dimensional parameter η. Then $P_0 = P_{\psi_0,\eta_0}$, where (ψ_0, η_0) is the true value. Often, $\hat{\psi}_N$ can be taken to be an asymptotically linear and efficient estimator of ψ, and Σ_0 the inverse efficient Fisher information matrix. The marginal prior for ψ must have a continuous and positive density on a neighborhood of ψ_0. Under a few further regularity conditions, the semi-parametric Bernstein von-Mises results (e.g., Shen (2002), Bickel and Kleijn (2012), Castillo (2012), Castillo and Rousseau (2015)) imply Assumption 1 is satisfied.

Per Remark 1, a main condition is that the prior for ψ is continuous, so there is zero prior probability that $\theta = c$ is literally true. The results in this section are approximations to posterior probabilities in large samples, for any prior compatible with Assumption 1.

Remark 1 is not a comprehensive review. There are Bernstein-von Mises results that cover certain nonparametric models (e.g., Castillo and Nickl (2013)), and important specific models, including limited information and moment condition models (e.g., Kwan (1999), Kim (2002), and Chib, Shin, and Simoni (2018)), linear and partially linear regressions (e.g., Bickel and Kleijn (2012) and Norets (2015)) proportional hazard models (e.g., Kim (2006)), and models with quasi-posteriors (e.g., Chernozhukov and Hong (2003)). Assumption 1 may not hold if the model is misspecified as in Kleijn and Van der Vaart (2012) and Müller (2013).

Remark 2 (Parameter of interest is a function of ψ)**.** Suppose Assumption 1 holds, but the parameter of interest is $\tilde{\psi} \in \mathbb{R}^1$ with $\tilde{\psi} = f(\psi)$ for some known function $f(\cdot) : \mathbb{R}^m \to \mathbb{R}^1$ that is continuously differentiable at ψ_0. Suppose the $(1 \times m)$-matrix of first derivatives of $f(\cdot)$ is $F(\cdot)$, and $F(\psi_0)$ is non-zero. The delta theorem (e.g., Van der Vaart (1998, Chapter 3), Wasserman (2004)) implies that Assumption 1 holds for $\tilde{\psi}$, with $\hat{\psi}_N$ replaced by $f(\hat{\psi}_N)$, ψ_0 replaced by $f(\psi_0)$, ψ replaced by $f(\psi)$, and Σ_0 replaced by $F(\psi_0)\Sigma_0 F(\psi_0)'$. Therefore, the results apply to a parameter of interest that is a continuously differentiable function of a parameter that satisfies Assumption 1.

The Wald test is the classical method for testing the hypothesis $\theta = c$. The Wald test statistic is

$$W_N = N(\hat{\theta}_N - c)'(\hat{\Sigma}_{N,11})^{-1}(\hat{\theta}_N - c).$$

By Assumption 1, if the hypothesis $\theta = c$ is true, W_N converges in distribution to χ^2_1. Therefore, the p-value for the hypothesis that $\theta = c$ is

$$p_N = P(\chi^2_1 > W_N) = \overline{F}_{\chi^2_1}(W_N).$$

Theorem 1. *Under Assumption 1, and when the p-value is not 1,*[7] *for any $\epsilon > 0$ chosen by the econometrician (possibly data-dependent per Remark 5), the Bayesian posterior probability of the ϵ-approximate hypothesis that $\theta \in [c-\epsilon, c+\epsilon]$ is approximately $F_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}\left(\frac{\epsilon^2}{|\hat{\theta}_N - c|^2} Q_{\chi^2_1}(1-p_N)\right)$ in large samples. Formally,*

$$\left(\Pi\left(\theta \in [c-\epsilon, c+\epsilon] | X^{(N)}\right) - F_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}\left(\frac{\epsilon^2}{|\hat{\theta}_N - c|^2} Q_{\chi^2_1}(1-p_N)\right)\right) 1[p_N < 1] \to^p 0.$$

All proofs are collected in Appendix A, part of the main paper. Theorem 1 relates the p-value to the posterior probability of the ϵ-approximate hypothesis. Figure 1 displays the contour plot of the posterior probability of the ϵ-approximate hypothesis as a function of the p-value and $\frac{|\hat{\theta}_N - c|}{\epsilon}$. The latter quantity depends on ϵ. To explain by example, if $\frac{|\hat{\theta}_N - c|}{\epsilon} = 1.5$ and the p-value is 0.05, Figure 1 indicates the posterior probability of the ϵ-approximate hypothesis is 0.2562. Note that because Σ_0 is unknown (and therefore must be estimated), $\hat{\theta}_N$ (and c and ϵ) does not uniquely pin down the p-value.

Shortly, Section 4 discusses the main implications of this result for using p-values in practice. Before getting there, it is worthwhile to discuss a few intermediate implications of this result.

One implication of this result is that it shows the posterior probability of the ϵ-approximate hypothesis depends on $\hat{\theta}_N$. The p-value by itself is not determinative of the posterior probability of the ϵ-approximate hypothesis. This contrasts with the fact that the p-value by itself is used to decide on rejecting or not the hypothesis that $\theta = c$ by classical standards, or other "cutoff rules" for significance like Benjamin, Berger, Johannesson, et al. (2018), and the fact that the minimum posterior probability of the hypothesis that $\theta = c$ (with positive prior probability of the hypothesis) of Edwards, Lindman, and Savage (1963), Berger and Delampady (1987), Berger and Sellke (1987), and Sellke, Bayarri, and Berger (2001) depends only on the p-value.

[7] $p_N = 1$ exactly happens with asymptotic probability 0, and finite sample probability 0 with a continuous sampling distribution, so this "exclusion" is irrelevant formally in the asymptotic approximation and also for all practical purposes. Relatedly, $p_N = 0$ exactly cannot happen.

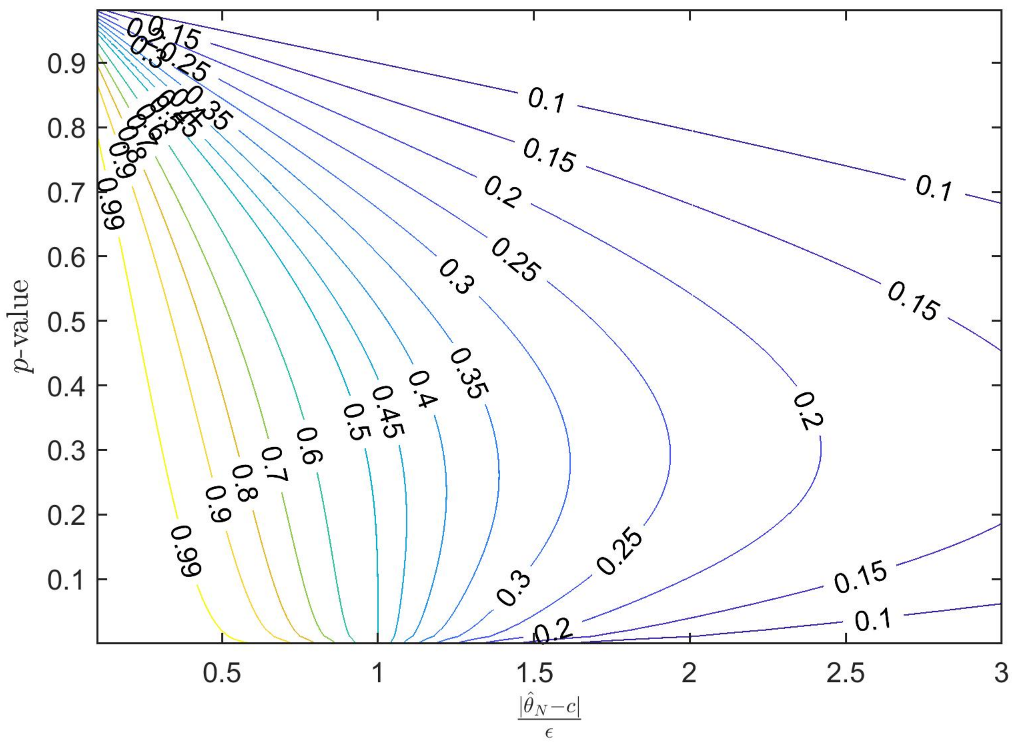

FIGURE 1. Posterior probability of the ϵ-approximate hypothesis that $\theta \in [\theta - \epsilon, \theta + \epsilon]$ for given p-value and $\frac{|\hat{\theta}_N - c|}{\epsilon}$.

Another implication of this result concerns the relationship between the posterior probability of the ϵ-approximate hypothesis and the p-value, even for given $\hat{\theta}_N$.

First, consider the case that the classical estimate provides "evidence against" the ϵ-approximate hypothesis, in the sense that $\hat{\theta}_N \notin [c - \epsilon, c + \epsilon]$, equivalent to $\frac{|\hat{\theta}_N - c|}{\epsilon} > 1$. Corollary 1 shows in that case that there is a non-monotone relationship between the p-value and the posterior probability of the ϵ-approximate hypothesis, for given $\hat{\theta}_N$. This can be seen in Figure 1. Both small p-values and large p-values correspond to a small probability of the ϵ-approximate hypothesis. This contrasts with the fact that small p-values lead to rejecting the hypothesis that $\theta = c$ and large p-values do not, whether by classical standards or other "cutoff rules" for significance, and the fact that there is a monotone relationship between the

p-value and the minimum posterior probability of the hypothesis that $\theta = c$. A large p-value can mean that the ϵ-approximate hypothesis is unlikely to be true, because values of the parameter other than the ϵ-approximate hypothesis values are more likely. Essentially, this effect is driven by what the p-value says about the variance of the posterior, which in turn affects the posterior probability of the ϵ-approximate hypothesis.

Corollary 1. *Under the same conditions as Theorem 1, for fixed $\hat{\theta}_N$ with $\frac{|\hat{\theta}_N - c|}{\epsilon} > 1$: The asymptotic approximation to the posterior probability that $\theta \in [c-\epsilon, c+\epsilon]$ is maximal (as a function of p_N) when $p_N = p_{N,max}$, where $p_{N,max} = \overline{F}_{\chi_1^2}\left(\frac{1}{2\sqrt{\tilde{\epsilon}_N}} \log\left(\frac{-\sqrt{\tilde{\epsilon}_N}-1}{\sqrt{\tilde{\epsilon}_N}-1}\right)\right)$ with $\tilde{\epsilon}_N = \frac{\epsilon^2}{|\hat{\theta}_N - c|^2}$. The asymptotic approximation to the posterior probability of the ϵ-approximate hypothesis is an increasing function of p_N when $p_N \in (0, p_{N,max})$ and a decreasing function of p_N when $p_N \in (p_{N,max}, 1)$. As $p_N \to 0$ or $p_N \to 1$, the asymptotic approximation to the posterior probability of the ϵ-approximate hypothesis limits to 0.*

Second, now consider the case that the classical estimate provides "evidence for" the ϵ-approximate hypothesis, in the sense that $\hat{\theta}_N \in [c-\epsilon, c+\epsilon]$, equivalent to $\frac{|\hat{\theta}_N - c|}{\epsilon} \leq 1$.[8] Corollary 2 shows in that case that the probability of the ϵ-approximate hypothesis is greatest when the p-value is smallest, for given $\hat{\theta}_N$. This can be easily seen in Figure 1. Obviously, this is the exact opposite of common practice, where small p-values are used to reject the hypothesis.

Corollary 2. *Under the same conditions as Theorem 1, for fixed $\hat{\theta}_N$ with $0 < \frac{|\hat{\theta}_N - c|}{\epsilon} \leq 1$: The asymptotic approximation to the posterior probability that $\theta \in [c-\epsilon, c+\epsilon]$ is a decreasing function of p_N. As $p_N \to 1$, the asymptotic approximation to the posterior probability of the ϵ-approximate hypothesis limits to 0. As $p_N \to 0$, the asymptotic approximation to the*

[8]Again, per Theorem 1, the (asymptotically) probability 0 - and thus irrelevant for practical purposes - event that $\hat{\theta}_N = c$ exactly is excluded from the results/discussion.

posterior probability of the ϵ-approximate hypothesis limits[9] *to* 1 *unless* $\frac{|\hat{\theta}_N - c|}{\epsilon} = 1$ *in which case the limit is* $\frac{1}{2}$.

4. Discussion of the results

In sum, the results suggest caution against the use of small p-values as a key standard in empirical research, including via p-value cutoff rules for "significance." The results of this paper show that a small p-value is neither necessary nor sufficient for there to be a small posterior probability that θ is close to c. Small p-values *can be* evidence that θ is probably *not close* to c, but they can also be evidence that θ is probably *close* to c. And larger p-values ("not rejecting") can *also* be evidence that θ is probably not close to c.

To illustrate by a numerical example, suppose θ is the effect of a treatment on an outcome, like a regression coefficient, and the p-value is for the hypothesis $\theta = 0$. For this example, a "close to zero" effect is any effect in $[-0.1, 0.1]$, so $\epsilon = 0.1$. Nothing changes if θ and ϵ are scaled by the same positive constant.

Consider these specifications for the p-value and classical estimate of the effect: $(p = 0.04, \hat{\theta} = 0.12)$ and $(p = 0.11, \hat{\theta} = 0.14)$. The first specification has a smaller p-value, which has the familiar implications: it implies the first specification has a smaller minimum posterior probability that the effect is zero, and it implies rejecting the hypothesis that the effect is zero by classical standards for the first specification. The p-value for the second specification does not achieve "statistical significance" and so it might be difficult to "sell"/publish such a result. It might be said that the second specification shows a "lack of evidence" against the hypothesis that the treatment effect is (close to) zero. Yet, the first specification implies a 36.6% posterior probability of an effect that is close to zero, greater than the 32.1% implied by the second specification. There is a similar analysis for innumerable other specifications, like $(p = 0.04, \hat{\theta} = 0.20)$ and $(p = 0.11, \hat{\theta} = 0.26)$ for example, where the posterior probabilities of

[9]This is a bit difficult to see in Figure 1 since the posterior probability is "close" to 1 only for *extremely* small p_N, for some $\frac{|\hat{\theta}_N - c|}{\epsilon}$.

an effect that is close to zero are respectively 15.1% and 14.9%. There are innumerable other such comparisons, with the footnote giving a few more examples.[10] In these comparisons, the specification with the larger p-value has a smaller (or equal) probability of the treatment effect being close to zero. Of course, in other comparisons, the specification with the larger p-value has a larger probability of the treatment effect being close to zero.

This illustrates why the results suggest caution about the practice of using small p-values as a key standard for empirical research. A larger (respectively, smaller) p-value does not necessarily correspond to a higher (respectively, lower) probability that θ is close to c. Comparing two studies, a study reporting a larger p-value can actually have a smaller posterior probability of a treatment effect that is close to 0. In particular, this shows that the research community can "miss" (e.g., not publish) treatment effects that are probably *not* close to zero if the research community only "accepts" (or focuses on) treatment effects with small p-values for the hypothesis of zero treatment effect. In particular, a larger p-value ("not rejecting") does not mean that there is a lack of evidence against the hypothesis that the treatment effect is (close to) zero.

Remark 3 (Confidence intervals)**.** This paper has focused on p-values, but because of the ability to convert between p-values for $\theta = c$ and confidence intervals for θ (for given $\hat{\theta}_N$), the results and discussions also apply to the relationship between confidence intervals and posterior probabilities of ϵ-approximate hypotheses. To illustrate the same issues arise, consider these specifications of 95% confidence intervals for a treatment effect: $[0.0657, 0.1704]$, $[0.0324, 0.2386]$, $[0.0180, 0.2681]$, $[-0.0193, 0.3425]$, $[-0.0648, 0.4230]$. Clearly, these are increasingly large (and nested) 95% confidence intervals. In order, the associated p-values

[10] Consider these comparisons where the first specification has the smaller p-value, which implies rejecting the hypothesis that the effect is zero by classical standards and a smaller minimum posterior probability that the effect is zero. Compare $(p = 0.04, \hat{\theta} = 0.16)$ and $(p = 0.11, \hat{\theta} = 0.24)$: the respective posterior probabilities that the treatment effect is close to zero are 22.0% and 16.4%. Compare $(p = 0.04, \hat{\theta} = 0.12)$ and $(p = 0.90, \hat{\theta} = 0.12)$: the respective posterior probabilities that the treatment effect is close to zero are 36.6% and 8.3%. Compare $(p = 0.0001, \hat{\theta} = 0.08)$ and $(p = 0.11, \hat{\theta} = 0.12)$: the respective posterior probabilities that the treatment effect is close to zero are 83.5% and 39.3%.

for the hypothesis of zero treatment effect range from 0.00001 to 0.15. By the application of common empirical standards, essentially the same used above for p-values, the first specification easily meets the standard for statistical significance and rejecting the hypothesis that the treatment effect is zero, whereas the last specification does not achieve statistical significance and so it might be comparatively difficult to "sell"/publish as evidence of an effect. Because of the failure to reject the hypothesis of zero treatment effect, it might be said that the last specification shows "lack of evidence" against the hypothesis that the treatment effect is zero. What would be the corresponding posterior probabilities that the treatment effect is close to zero, which as before is defined for this example to be an effect in $[-0.1, 0.1]$? Actually, the posterior probabilities that the treatment effect is close to zero are exactly the same for these specifications: 25%. This reflects exactly the same issues discussed previously about p-values, because of the relationship between p-values and confidence intervals.

Remark 4 (Other approaches)**.** This remark is about other approaches. Another case is when there is positive prior probability of $\theta = c$. As discussed in the Introduction, this case seems rare in empirical work in the social sciences. Nevertheless, this case can be analyzed. In this case, the posterior probability of $\theta \in [c-\epsilon, c+\epsilon]$ with ϵ small is approximately the same as the posterior probability of the hypothesis $\theta = c$ (e.g., Berger and Sellke (1987, Section 4)). Therefore, results for the minimum posterior probability of the hypothesis $\theta = c$ and of the hypothesis $\theta \in [c-\epsilon, c+\epsilon]$ with ϵ small are basically the same in this setup. In fact, this is part of the motivation for the literature on minimum posterior probabilities to study the (minimum) posterior probability of $\theta = c$, understanding it can be an approximation to the (minimum) posterior probability that $\theta \in [c-\epsilon, c+\epsilon]$. See for example the discussion in Berger and Sellke (1987). As mentioned in the Introduction, this means that the use of small p-values as a key standard, including p-value cutoff rules for significance like the one suggested by Benjamin, Berger, Johannesson, et al. (2018), can be justified from a Bayesian perspective even if the empirical question is whether θ is close to c, when there is positive prior probability that $\theta = c$.

This paper studies the p-value for the hypothesis $\theta = c$, precisely because that is what is used in empirical practice, but one could in principle conduct a classical hypothesis test of the ϵ-approximate hypothesis (when ϵ is not data-dependent). In the appendix, Abadie (2020) studies the informativeness of rejecting or not rejecting the hypothesis $\theta \in [c - \epsilon, c + \epsilon]$ by such a classical hypothesis test. In large samples, Abadie (2020)'s results imply the posterior probability of the ϵ-approximate hypothesis is 1 conditional on not rejecting and is 0 conditional on rejecting.

Remark 5 (Data-dependent ϵ)**.** ϵ can be data-dependent. For example, in a regression setting (with $c = 0$), ϵ can be a multiple λ of the estimated coefficient $\hat{\beta}$ on a "reference" explanatory variable. The ϵ-approximate hypothesis would be $\theta \in [-\lambda|\hat{\beta}|, \lambda|\hat{\beta}|]$, meaning the effect of one explanatory variable is "close to 0" if it is sufficiently closer to 0 compared to the effect of the "reference" explanatory variable. More specifically, consider as an example the case that θ reflects the treatment effect of a novel treatment, when the outcome is known already in the literature to be impacted by other "reference" treatments. Consider a regression model that includes both the novel treatment and one such "reference" treatment. ϵ can be selected to be proportionate to the estimated treatment effect of the reference treatment, so the ϵ-approximate hypothesis concerns whether or not the novel treatment has an effect that is at least a pre-specified percentage of the effect of the reference treatment. If so, the novel treatment can be said to have an "economically significant" effect, specifically relative to the effect of this reference treatment. Other choices of ϵ are allowed, including fixed constants as in the numerical example above.

5. Finite sample results

A similar analysis can be done in finite samples, with additional distributional assumptions on the data generating process. $t_d(\mu, \Sigma)$ is a t-distribution with location μ, scale Σ, and degrees of freedom d. Thus $t_d(0,1) \equiv t_d$ is the standard t-distribution with d degrees of freedom. $\mathcal{F}_{d_1,d_2}$ is an F-distribution with degrees of freedom d_1 and d_2. Define $G_a(v,w) \equiv F_{t_a}\left(\sqrt{Q_{\mathcal{F}_{1,a}}(1-v)}\,(-1+w)\right) - F_{t_a}\left(\sqrt{Q_{\mathcal{F}_{1,a}}(1-v)}\,(-1-w)\right)$ for $a \in \mathbb{N}$ and $v \in (0,1)$.

Assumption 2. *It holds that:*

1. *The classical sampling distribution of $\frac{\hat{\theta}_N - \theta_0}{\hat{\sigma}_N}$ has distribution $t_{N-d}(0,1)$, i.e., a standard t-distribution with $N-d$ degrees of freedom, for some estimated scale factor $\hat{\sigma}_N > 0$ and integer d such that $N - d > 2$.*
2. *The Bayesian posterior $\theta | X^{(N)}$ has distribution $t_{N-d}(\hat{\theta}_N, \hat{\sigma}_N^2)$.*

An example of Assumption 2 is when $X_i \sim^{i.i.d.} \mathcal{N}(\theta, \sigma^2)$ with θ and σ^2 unknown. Then, $d = 1$, $\hat{\theta}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$, and $\hat{\sigma}_N^2 = \frac{1}{N}\left(\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\theta}_N)^2\right)$. The prior is uniform on $(\theta, \log\sigma)$. See Gelman, Carlin, Stern, and Rubin (2003, Section 3.2). Another example is when θ is a linear regression coefficient with homoskedastic, normally-distributed unobservables. Then, d is the number of explanatory variables (counting the intercept), $\hat{\theta}_N$ is the OLS estimate, and $\hat{\sigma}_N^2$ is the estimate of the variance of the OLS estimate. The prior is a uniform prior. See Gelman, Carlin, Stern, and Rubin (2003, Section 14.2) or Lancaster (2004, Section 3.3.3). As with the large sample approximation, and the discussion of Assumption 1, note again the analysis concerns the case of unknown variance (or scale) with corresponding estimate $\hat{\sigma}_N^2$.

The classical test statistic is $W_N = \left(\frac{\hat{\theta}_N - c}{\hat{\sigma}_N}\right)^2$. By Assumption 2, if the hypothesis $\theta = c$ is true, W_N has distribution $\mathcal{F}_{1,N-d}$. Therefore, the p-value is $p_N = \overline{F}_{\mathcal{F}_{1,N-d}}(W_N)$.

Theorem 2. *Under Assumption 2, and when the p-value is not* 1,[11] *for any* $\epsilon > 0$ *chosen by the econometrician (possibly data-dependent), the Bayesian posterior probability of the* ϵ*-approximate hypothesis that* $\theta \in [c-\epsilon, c+\epsilon]$ *is* $G_{N-d}\left(p_N, \frac{\epsilon}{|\hat{\theta}_N - c|}\right)$.

Lemma 1. $G_a(v,w) \to F_{\chi^2_{1,Q_{\chi^2_1}(1-v)}}\left(w^2 Q_{\chi^2_1}(1-v)\right)$ *as* $a \to \infty$.

As $N \to \infty$, Theorem 2 limits to Theorem 1 by applying Lemma 1. In fact, Theorem 2 and Theorem 1 are practically the same numerically when $N - d \geq 1000$, and also when $N - d \geq 300$ and $p_N \geq 0.001$, among other cases.[12] Therefore, practically the same discussion of the asymptotic approximation applies also to the finite sample result.

6. Conclusions

This paper derives closed-form expressions for the posterior probability of the ϵ-approximate hypothesis. The properties are discussed, which contradict some common practices about using p-values. In the running example of treatment effects, a quick implication of the results is that larger p-values can correspond to lower posterior probabilities that the treatment effect is close to zero, potentially lower than from a smaller p-value. This shows that the research community can "miss" (e.g., not publish) treatment effects that are probably *not* close to zero if the research community only "accepts" (or focuses on) treatment effects with small p-values for the hypothesis of zero treatment effect. In particular, a larger p-value ("not rejecting") does not mean that there is a lack of evidence against the hypothesis that the treatment effect is (close to) zero. On the other hand, it is possible to *retroactively* use the expressions derived in this paper to assess the posterior probability of these ϵ-approximate hypotheses from published research, thereby assessing the evidence about whether the treatment effect is (close to) zero.

[11] Similar to the asymptotic approximation, $p_N = 1$ happens with probability 0 when $\hat{\theta}_N$ has a continuous sampling distribution and $p_N = 0$ cannot happen in this setup.

[12] This claim is based on the maximal absolute difference between the posterior probability of the ϵ-approximate hypothesis per Theorem 2 and the asymptotic approximation, where the maximum is across values of p_N, ϵ, and $\hat{\theta}_N$. In the stated cases, the maximal absolute difference is less than 0.003.

Appendix A. Technical appendix

Lemma 2. *Under Assumption 1,* $||\Pi_{\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}} - \mathcal{N}(0,\hat{\Sigma}_N)||_{TV} \to^p 0$ *and* $||\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}} - \mathcal{N}(0,I_{m\times m})||_{TV} \to^p 0$.

Proof of Lemma 2. $||\mathcal{N}(0,\hat{\Sigma}_N)-\mathcal{N}(0,\Sigma_0)||_{TV} \to^p 0$ since total variation distance is bounded above by the square root of Kullback-Leibler divergence (e.g., DasGupta (2008, Chapter 2)) and $\hat{\Sigma}_N \to^p \Sigma_0$ by Assumption 1. Therefore, $||\Pi_{\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}} - \mathcal{N}(0,\hat{\Sigma}_N)||_{TV} \to^p 0$.

Let $h_N(\cdot):\mathbb{R}^m \to \mathbb{R}^m$ be $h_N(w) = \hat{\Sigma}_N^{-\frac{1}{2}}w$. Since $h_N(\cdot)$ is continuous, $h_N^{-1}(\mathcal{B})$ is Borel for any Borel $\mathcal{B}$. Therefore, $|\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}}(\mathcal{B}) - \mathcal{N}(0,I_{m\times m})(\mathcal{B})| = |\Pi_{\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}}(h_N^{-1}(\mathcal{B})) - \mathcal{N}(0,\hat{\Sigma}_N)(h_N^{-1}(\mathcal{B}))|$. Because $||\Pi_{\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}} - \mathcal{N}(0,\hat{\Sigma}_N)||_{TV} \to^p 0$, $||\Pi_{\hat{\Sigma}_N^{-\frac{1}{2}}\sqrt{N}(\psi-\hat{\psi}_N)|X^{(N)}} - \mathcal{N}(0,I_{m\times m})||_{TV} \to^p 0$. □

For the hypothesis $\tilde{\psi}=c$, where $\tilde{\psi}\in\mathbb{R}^{\tilde{m}}$ is a sub-vector of components of ψ (e.g., $\tilde{\psi}=\theta$ or $\tilde{\psi}=(\theta,\gamma_2)$), and $\tilde{\Sigma}$ is the corresponding sub-matrix of Σ, the test statistic is

$$\tilde{W}_N = N(\hat{\tilde{\psi}}_N - c)'\hat{\tilde{\Sigma}}_N^{-1}(\hat{\tilde{\psi}}_N - c).$$

The p-value is

$$\tilde{p}_N = P(\chi^2_{\tilde{m}} > \tilde{W}_N) = \overline{F}_{\chi^2_{\tilde{m}}}(\tilde{W}_N).$$

Lemma 3. *Under Assumption 1,*

$$\sup_{0<q<1}\left| q - \Pi\left(\delta_{\hat{\tilde{\Sigma}}_N,2}(\tilde{\psi},c) \le \sqrt{\frac{Q_{\chi^2_{\tilde{m},Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)}}(q)}{N}} \mid X^{(N)}\right)\right| \to^p 0.$$

Proof of Lemma 3. $||\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)||_2^2 = \sqrt{N}(\tilde{\psi}-c)'\hat{\tilde{\Sigma}}_N^{-1}\sqrt{N}(\tilde{\psi}-c) = \sqrt{N}(\tilde{\psi}-\hat{\tilde{\psi}}_N+\hat{\tilde{\psi}}_N-c)'\hat{\tilde{\Sigma}}_N^{-1}\sqrt{N}(\tilde{\psi}-\hat{\tilde{\psi}}_N+\hat{\tilde{\psi}}_N-c)$. Let $Z\sim\mathcal{N}(0,I_{\tilde{m}\times\tilde{m}})$. For any Borel $\mathcal{B}\subseteq\mathbb{R}^{\tilde{m}}$, $\Pi_{\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)|X^{(N)}}(\mathcal{B}) = \Pi_{\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-\hat{\tilde{\psi}}_N+\hat{\tilde{\psi}}_N-c)|X^{(N)}}(\mathcal{B}) = \Pi_{\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-\hat{\tilde{\psi}}_N)|X^{(N)}}(\mathcal{B}-\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\tilde{\psi}}_N-c))$. By Assumption 1 and Lemma 2, $|\Pi_{\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-\hat{\tilde{\psi}}_N)|X^{(N)}}(\mathcal{B}-\hat{\tilde{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\tilde{\psi}}_N-c)) - (Z+$

$\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))(\mathcal{B})| = |\Pi_{\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-\hat{\hat{\psi}}_N)|X^{(N)}}(\mathcal{B} - \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c)) - Z(\mathcal{B} - \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))| \to^p 0$. Therefore, $||\Pi_{\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)|X^{(N)}} - (Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))||_{TV} \to^p 0$.

Let $h(\cdot) : \mathbb{R}^{\tilde{m}} \to \mathbb{R}$ be given by $h(z) = z'z$. Since $h(\cdot)$ is continuous, $h^{-1}(\mathcal{B}) \subseteq \mathbb{R}^{\tilde{m}}$ is Borel for any Borel $\mathcal{B}$. Therefore, $|\Pi_{||\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)||_2^2|X^{(N)}}(\mathcal{B}) - (Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))'(Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))(\mathcal{B})| = |\Pi_{\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)|X^{(N)}}(h^{-1}(\mathcal{B})) - (Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))(h^{-1}(\mathcal{B}))|$. Because $||\Pi_{\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)|X^{(N)}} - (Z+\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))||_{TV} \to^p 0$ by the above, $||\Pi_{||\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)||_2^2|X^{(N)}} - (Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))'(Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))||_{TV} \to^p 0$.

$C_N = (Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c))'(Z + \hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\hat{\hat{\psi}}_N - c)) \sim \chi^2_{\tilde{m},\tilde{W}_N}$. Therefore, $\sup_{0<q<1}|\Pi_{||\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)||_2^2|X^{(N)}}((-\infty, Q_{\chi^2_{\tilde{m},Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)}}(q)]) - C_N((-\infty, Q_{\chi^2_{\tilde{m},Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)}}(q)])| \to^p 0$. By definition, $\tilde{W}_N = Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)$, so $C_N((-\infty, Q_{\chi^2_{\tilde{m},Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)}}(q)]) = C_N((-\infty, Q_{\chi^2_{\tilde{m},\tilde{W}_N}}(q)]) = q$. The result follows since by definition $\Pi(\delta^2_{\hat{\hat{\Sigma}}_N,2}(\tilde{\psi}, c) \le \frac{Q_{\chi^2_{\tilde{m},Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)}}(q)}{N} \mid X^{(N)}) = \Pi_{||\hat{\hat{\Sigma}}_N^{-\frac{1}{2}}\sqrt{N}(\tilde{\psi}-c)||_2^2|X^{(N)}}((-\infty, Q_{\chi^2_{\tilde{m},Q_{\chi^2_{\tilde{m}}}(1-\tilde{p}_N)}}(q)])$. □

Lemma 4. *Under Assumption 1,*

$$\sup_{0<q<1}\left|\left(q - \Pi\left(|\theta - c| \le \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}(q)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c| \mid X^{(N)}\right)\right)1[p_N < 1]\right| = o_p(1).$$

Proof of Lemma 4. $Q_{\chi^2_1}(1-p_N) = W_N = N(\hat{\Sigma}_{N,11})^{-1}(\hat{\theta}_N - c)^2$, so $(\hat{\Sigma}_{N,11})^{-\frac{1}{2}} = \frac{\sqrt{Q_{\chi^2_1}(1-p_N)}}{\sqrt{N}|\hat{\theta}_N - c|}$ when $|\hat{\theta}_N - c| \neq 0$ which is equivalent to $p_N < 1$. When $|\hat{\theta}_N - c| \neq 0$, $\delta_{\hat{\Sigma}_{N,11},2}(\theta, c) = (\hat{\Sigma}_{N,11})^{-\frac{1}{2}}|\theta - c| = \frac{\sqrt{Q_{\chi^2_1}(1-p_N)}}{\sqrt{N}|\hat{\theta}_N - c|}|\theta - c|$. The result follows from Lemma 3 by substitution. □

Proof of Theorem 1. By simplification, $\Pi\left(\theta \in [c-\epsilon, c+\epsilon]|X^{(N)}\right) = \Pi\left(|\theta - c| \le \sqrt{\frac{Q_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}\left(F_{\chi^2_{1,Q_{\chi^2_1}(1-p_N)}}\left(\frac{\epsilon^2}{|\hat{\theta}_N - c|^2}Q_{\chi^2_1}(1-p_N)\right)\right)}{Q_{\chi^2_1}(1-p_N)}}|\hat{\theta}_N - c||X^{(N)}\right)$. Therefore, the result follows from Lemma 4. □

Proof of Corollary 1 and Corollary 2. By Johnson, Kotz, and Balakrishnan (1995, page 441), $F_{\chi^2_{1,b}}(ab)$ where $a = \frac{\epsilon^2}{|\hat{\theta}_N - c|^2}$ and $b = Q_{\chi^2_1}(1-p_N)$ can be written $\Phi(\sqrt{b}(\sqrt{a}-1)) - \Phi(-\sqrt{b}(\sqrt{a}+1))$. The derivative with respect to b is $\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(ab - 2\sqrt{a}b + b))(\sqrt{a} - 1)\frac{1}{2}b^{-\frac{1}{2}} + \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(ab + 2\sqrt{a}b + b))(\sqrt{a}+1)\frac{1}{2}b^{-\frac{1}{2}}$. The sign of the derivative is the same as the sign of $\exp(2\sqrt{a}b)(\sqrt{a}-1) + (\sqrt{a}+1)$. If $a \geq 1$, the derivative is obviously positive. The first part of Corollary 2 follows, since b is a decreasing function of p_N. If $a < 1$, the derivative is negative (resp. positive, zero) exactly when b is greater than (resp. less than, equal to) $\frac{1}{2\sqrt{a}}\log\left(\frac{-\sqrt{a}-1}{\sqrt{a}-1}\right)$. The first part of Corollary 1 follows, since this means the posterior probability is a decreasing (resp., increasing) function of p_N when p_N is greater than (resp., less than) $1 - F_{\chi^2_1}\left(\frac{1}{2\sqrt{a}}\log\left(\frac{-\sqrt{a}-1}{\sqrt{a}-1}\right)\right)$. The limits in these Corollaries are apparent from the representation from Johnson, Kotz, and Balakrishnan (1995, page 441) that implies that $F_{\chi^2_{1,t^2}}(s^2t^2) = \Phi(t(s-1)) - \Phi(t(-s-1))$. □

Proof of Theorem 2. Follow the strategy of Theorem 1. $W_N = Q_{\mathcal{F}_{1,N-d}}(1-p_N)$ and $\hat{\sigma}_N^{-1} = \frac{\sqrt{Q_{\mathcal{F}_{1,N-d}}(1-p_N)}}{|\hat{\theta}_N - c|}$. The posterior for $\hat{\sigma}_N^{-1}(\theta - c)$ is $t_{N-d}(\hat{\sigma}_N^{-1}(\hat{\theta}_N - c), 1)$. Therefore, $\Pi(|\theta - c| \leq \epsilon | X^{(N)}) = \Pi(-\hat{\sigma}_N^{-1}\epsilon - \hat{\sigma}_N^{-1}(\hat{\theta}_N - c) \leq \hat{\sigma}_N^{-1}(\theta - c) - \hat{\sigma}_N^{-1}(\hat{\theta}_N - c) \leq \hat{\sigma}_N^{-1}\epsilon - \hat{\sigma}_N^{-1}(\hat{\theta}_N - c)|X^{(N)}) = P(-\hat{\sigma}_N^{-1}\epsilon - \hat{\sigma}_N^{-1}(\hat{\theta}_N - c) \leq t_{N-d} \leq \hat{\sigma}_N^{-1}\epsilon - \hat{\sigma}_N^{-1}(\hat{\theta}_N - c)) = P(-\hat{\sigma}_N^{-1}\epsilon - \hat{\sigma}_N^{-1}|\hat{\theta}_N - c| \leq t_{N-d} \leq \hat{\sigma}_N^{-1}\epsilon - \hat{\sigma}_N^{-1}|\hat{\theta}_N - c|)$ by symmetry of the t_{N-d} distribution. The final expression is $G_{N-d}(p_N, \frac{\epsilon}{|\hat{\theta}_N - c|})$ using the expression for $\hat{\sigma}_N^{-1}$. □

Proof of Lemma 1. $\mathcal{F}_{1,a} \to \chi^2_1$ and $t_a \to \mathcal{N}(0,1)$ as $a \to \infty$. Then use the representation for $F_{\chi^2_{1,t^2}}(s^2t^2)$ used in the proof of Corollaries 1 and 2. □

References

Abadie, A. (2020): "Statistical non-significance in empirical economics," *American Economic Review: Insights*, 2(2), 193–208.

Benjamin, D. J., J. O. Berger, M. Johannesson, et al. (2018): "Redefine statistical significance," *Nature Human Behaviour*, 2(1), 6.

Berger, J. O., and M. Delampady (1987): "Testing precise hypotheses," *Statistical Science*, 2(3), 317–335.

Berger, J. O., and T. Sellke (1987): "Testing a point null hypothesis: The irreconcilability of p values and evidence," *Journal of the American Statistical Association*, 82(397), 112–122.

Bickel, P., and B. Kleijn (2012): "The semiparametric Bernstein–von Mises theorem," *The Annals of Statistics*, 40(1), 206–237.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg (2016): "Star wars: The empirics strike back," *American Economic Journal: Applied Economics*, 8(1), 1–32.

Castillo, I. (2012): "A semiparametric Bernstein–von Mises theorem for Gaussian process priors," *Probability Theory and Related Fields*, 152(1-2), 53–99.

Castillo, I., and R. Nickl (2013): "Nonparametric Bernstein–von Mises theorems in Gaussian white noise," *The Annals of Statistics*, 41(4), 1999–2028.

Castillo, I., and J. Rousseau (2015): "A Bernstein–von Mises theorem for smooth functionals in semiparametric models," *The Annals of Statistics*, 43(6), 2353–2383.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

Chib, S., M. Shin, and A. Simoni (2018): "Bayesian estimation and comparison of moment condition models," *Journal of the American Statistical Association*, 113(524), 1–13.

DasGupta, A. (2008): *Asymptotic Theory of Statistics and Probability*. Springer, New York.

Edwards, W., H. Lindman, and L. J. Savage (1963): "Bayesian statistical inference for psychological research," *Psychological Review*, 70(3), 193.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003): *Bayesian Data Analysis*. CRC Press, 2 edn.

Held, L., and M. Ott (2018): "On p-values and Bayes factors," *Annual Review of Statistics and Its Application*, 5, 393–419.

Jeffreys, H. (1939): *Theory of Probability*. The Clarendon Press, Oxford.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995): *Continuous Univariate Distributions*, vol. 2. Wiley, New York, 2 edn.

Kim, J.-Y. (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107(1-2), 175–193.

Kim, Y. (2006): "The Bernstein–von Mises theorem for the proportional hazard model," *The Annals of Statistics*, 34(4), 1678–1700.

Kleijn, B., and A. Van der Vaart (2012): "The Bernstein-von-Mises theorem under misspecification," *Electronic Journal of Statistics*, 6, 354–381.

Kwan, Y. K. (1999): "Asymptotic Bayesian analysis based on a limited information estimator," *Journal of Econometrics*, 88(1), 99–121.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blackwell Oxford.

Le Cam, L. (1986): *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

Le Cam, L., and G. L. Yang (2000): *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York, 2 edn.

Leggett, N. C., N. A. Thomas, T. Loetscher, and M. E. Nicholls (2013): "The life of p: "Just significant" results are on the rise," *The Quarterly Journal of Experimental Psychology*, 66(12), 2303–2309.

Lindley, D. V. (1957): "A statistical paradox," *Biometrika*, 44(1/2), 187–192.

Masicampo, E., and D. R. Lalande (2012): "A peculiar prevalence of p values just below. 05," *The Quarterly Journal of Experimental Psychology*, 65(11), 2271–2279.

Müller, U. K. (2013): "Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix," *Econometrica*, 81(5), 1805–1849.

Norets, A. (2015): "Bayesian regression with nonparametric heteroskedasticity," *Journal of Econometrics*, 185(2), 409–419.

Sellke, T., M. Bayarri, and J. O. Berger (2001): "Calibration of p values for testing precise null hypotheses," *The American Statistician*, 55(1), 62–71.

Shen, X. (2002): "Asymptotic normality of semiparametric and nonparametric posterior distributions," *Journal of the American Statistical Association*, 97(457), 222–235.

Van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.

WASSERMAN, L. (2004): *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York.