Oxford Bulletin of Economics and Statistics

ORIGINAL ARTICLE OPEN ACCESS

Bayesian Estimation of the Normal Location Model: A Non-Standard Approach

Giuseppe De Luca[1] | Jan R. Magnus[2] | Franco Peracchi[3]

[1]Università degli Studi di Palermo, Palermo, Italy | [2]Vrije Universiteit Amsterdam and Tinbergen Institute, Amsterdam, the Netherlands | [3]Università degli Studi di Roma Tor Vergata, Rome, Italy

Correspondence: Giuseppe De Luca (giuseppe.deluca@unipa.it)

Received: 8 March 2024 | **Revised:** 4 February 2025 | **Accepted:** 24 February 2025

Funding: This work was supported by MUR, Italy, PRIN (PRJ-1446 and PRJ-1547); Joint Usage/Research Centre for Behavioural Economics at ISER, Osaka University; and Università degli Studi di Palermo FFR 2023-2024.

Keywords: asymptotic distribution | Bayesian shrinkage estimators | finite-sample distribution | model averaging | normal location model | uniform $\sqrt{n}$-consistency

ABSTRACT

We consider the estimation of the location parameter θ in the normal location model and study the sampling properties of shrinkage estimators derived from a non-standard Bayesian approach that places the prior on a scaled version of θ, interpreted as the "population t-ratio." We show that the finite-sample distribution of these estimators is not centred at θ and is generally non-normal. In the asymptotic theory, we prove uniform $\sqrt{n}$-consistency of our estimators and obtain their asymptotic distribution under a general moving-parameter setup that includes both the fixed-parameter and the local-parameter settings as special cases.

JEL Classification: C11, C13, C51, C52

1 | Motivation

We reconsider possibly the simplest of all estimation problems, namely the estimation of a parameter θ based on a single observation x, normally distributed with unknown mean θ and unit variance, that is,

$$x \sim \mathrm{N}(\theta, 1) \tag{1}$$

Model (1) is known as the (univariate) *normal location model* and the problem of estimating the location parameter θ in this model is known as the *normal location problem*. The normal location problem has been studied from various angles; see, for example, Pericchi and Smith [1], Magnus [2, 3], Johnstone and Silverman [4], Kumar and Magnus [5], DasGupta and Johnstone [6], Johnstone [7], and De Luca et al. [8].

We offer two motivations for studying this seemingly trivial problem. First, there is an intrinsic interest in this problem, based on the fact that the obvious solution of estimating θ by the maximum likelihood estimator $\hat{\theta} = x$ is less obvious than one might think. Second, the problem is closely related to the issue of variable selection and model averaging in linear regression models, based on the "equivalence theorem" of Magnus and Durbin [9].

Let us consider both motivations in more detail. The maximum likelihood estimator $\hat{\theta}$, sometimes called the "usual" estimator, is unbiased, has unit variance and, under squared error loss, its risk is equal to its mean squared error $\mathrm{MSE}(\hat{\theta}) = 1$. It is the unique minimax estimator of θ [10, Proposition 8.6] and it is admissible [11, p. 548]. One may therefore wonder why we wish to consider alternative estimators. To answer this question, let us compare

$\hat{\theta}$ to another estimator of θ, namely the "silly" estimator $\check{\theta} = 0$. Since $\text{MSE}(\check{\theta}) = \theta^2$, we find that the silly estimator is "better" than the usual estimator in the mean squared error sense if and only if $|\theta| < 1$. This simple comparison shows that, although no estimator dominates the usual estimator for all values of θ, it is easy to find other estimators with smaller risks in a non-negligible region of the parameter space.

Ideally, we would like to find an admissible estimator which behaves like $\check{\theta}$ for "sufficiently small" values of θ, and like $\hat{\theta}$ for "sufficiently large" values of θ. This intuitive idea leads to the "pretest" estimator

$$\overline{\theta} = \begin{cases} x & \text{if } |x| > c \\ 0 & \text{if } |x| \leq c \end{cases}$$

for some $c \geq 0$ (possibly $c = 1$), which we can also write as

$$\overline{\theta} = \lambda(x)x \qquad \lambda(x) = \begin{cases} 1 & \text{if } |x| > c \\ 0 & \text{if } |x| \leq c \end{cases}$$

The function $\lambda(x)$ may be interpreted either as a shrinkage function that shrinks the usual estimator of θ towards 0, or as a weight function employed to combine the usual and the silly estimators of θ with weights $\lambda(x)$ and $1 - \lambda(x)$ respectively. Unfortunately, $\overline{\theta}$ is inadmissible (it has a discontinuity at $x = \pm c$), and it has other undesirable properties as well; see Judge and Bock [12], Magnus [2], Leeb and Pötscher [13–15], Kabaila and Leeb [16], and Efron [17].

A continuous version of the pretest estimator is obtained when we allow $\lambda(x)$ to increase smoothly with $|x|$. The question then arises how to choose the function $\lambda(x)$. A simple example is

$$\lambda(x) = \frac{x^2}{x^2 + 1} \tag{2}$$

but there are many alternative specifications. Magnus [3] shows that this smooth generalisation of the pretest estimator already produces an estimator of θ with better sampling properties. However, the estimator based on (2) is still inadmissible [18, p. 278].

Although admissibility is a frequentist concept, it is intimately connected to Bayesian ideas. In fact, subject only to mild conditions, any (generalised) Bayes estimator is admissible [11, Section 4.8]. Thus, instead of specifying the function $\lambda(x)$, we may take a Bayesian approach by specifying a prior distribution $\pi(\theta)$ for θ. Under squared error loss, the posterior mean $\tilde{\theta} = \mathbb{E}(\theta|x)$ is our estimator of θ, and we are interested in its *frequentist* properties. For previous studies on the frequentist properties of Bayesian estimators see, for example, Joanes and Peers [19], Carlin and Louis [20, Chapter 4], Johnstone and Silverman [4], Efron [21, 22], and De Luca et al. [8].

Expanding on our second motivation, the (univariate) normal location problem is also essential in the theory of pretest, shrinkage, and model averaging estimators, which can all be viewed as estimating the location parameters in the diagonal form of a Gaussian sequence model [7, Chapter 2]. For example, the pretest estimator corresponds to a two-step procedure where we first "test" whether $|\theta| > c$, and then choose between the usual and the silly estimator based on this preliminary "test". This procedure mimics a common practice in applied regression analysis, where the analyst first performs a t-test and then chooses the model (and hence the estimator) based on the outcome of this preliminary t-test. Other examples include the empirical Bayes thresholding estimator of Johnstone and Silverman [4], the rescaled spike-and-slab variable selection method of Ishwaran and Rao [23], the penalised least-squares estimators analysed by Pötscher and Leeb [24], the weighted-average least squares (WALS) estimator of Magnus et al. [25], the Bayesian model-averaging estimator of Lee and Oh [26], and the wavelet shrinkage estimators discussed in [7, Chapter 7].

Consider, for example, the linear regression model

$$y = X_1\beta_1 + X_2\beta_2 + u, \qquad u \sim \text{N}(0, \sigma^2 I_n)$$

where X_1 $(n \times k_1)$ is a matrix of "focus" variables that are required to be in the model on theoretical or other grounds, while there is doubt whether the "auxiliary" variables in X_2 $(n \times k_2)$ should be in the model or not. The focus regressors are those whose causal effects on y we wish to study, or whose presence in the model is well established from previous studies, while the auxiliary regressors are controls added to avoid omitted variable bias, or non-linear transformations of a set of original regressors. The "focus parameters" in β_1 are the parameters of interest, while the "auxiliary parameters" in β_2 are treated as nuisance parameters. Because of uncertainty on which auxiliary regressors to include in the model, there are 2^{k_2} models to consider. If $\hat{\beta}_{1(i)}$ denotes the estimator of β_1 in model i, then the model-averaging estimator of β_1 is

$$\tilde{\beta}_1 = \sum_{i=1}^{2^{k_2}} \omega_i \hat{\beta}_{1(i)} \tag{3}$$

where the ω_i are data-dependent model weights. When the auxiliary regressors in X_2 can be orthogonalised using some preliminary data transformation (e.g., principal components, semi-orthogonal transformations of X_2, or wavelet transformations), the equivalence theorem of Magnus and Durbin [9] shows that the choice of the ω_i in (3) is equivalent to the choice of $\lambda(x)$ for a shrinkage estimator of the location parameter θ in model (1). In particular, the theory developed in this article is directly applicable to the WALS estimator; see Magnus and De Luca [27] for a review of the WALS approach and De Luca et al. [8, 28, 29] for additional results. More generally, our paper contributes to the literature by developing Bayesian estimators of θ that are based on a transparent notion of prior ignorance and enjoy desirable sampling properties.

We shall consider a small but important variant of model (1), which allows us to study the sampling properties of the proposed Bayesian estimators, both exactly and asymptotically. This variant is explored in Section 2, where we distinguish between two Bayesian approaches to the normal location problem: the standard approach and our proposed novel approach. Section 3 discusses regularity conditions placed on the prior to ensure that the posterior mean resulting from the proposed Bayesian approach satisfies certain desirable properties, while Section 4 specifies four priors which satisfy these conditions. Section 5

studies the finite-sample distribution of the posterior mean. In Section 6 we investigate uniform $\sqrt{n}$-consistency, and in Section 7 we investigate the asymptotic distribution of the posterior mean using a general moving-parameter setup that includes both the fixed-parameter and the local-parameter settings as special cases. Section 8 concludes. Proofs of all propositions are in the Appendix A.

2 | Two Bayesian Approaches

In model (1) there is only one observation, and hence the sample size plays no role and any discussion of asymptotics is meaningless. Things change, however, if we replace x with a sample statistic x_n (e.g., a sample mean) that depends on the sample size n and satisfies

$$x_n \sim \mathrm{N}\left(\theta, \frac{\sigma^2}{n}\right) \tag{4}$$

where $\sigma^2 > 0$ is assumed to be known. The assumption that σ^2 is known is typically motivated by the wish to derive simple and easy-to-interpret approximations that can prove helpful for the asymptotic analysis; see, for example, Pötscher [30], Pötscher and Leeb [24], and DasGupta and Johnstone [6]. The assumption can be relaxed if σ^2 is estimated consistently by an estimator that is independent of x_n; see, for example, Leeb and Pötscher [13]. The normality assumption is employed to derive our Bayesian estimator of θ and its finite-sample distribution, but it will be weakened in the large-sample theory of Sections 6 and 7.

The obvious estimator of θ in model (4) is x_n (the "usual" estimator), which is unbiased and consistent. The mean squared error of the usual estimator is $\mathrm{MSE}(x_n) = \sigma^2/n$, while the silly estimator has $\mathrm{MSE}(0) = \theta^2$. So we prefer the silly estimator (in the MSE sense) if and only if $|\theta| < \sigma/\sqrt{n}$. Defining

$$x_n^* = \frac{x_n}{\sigma/\sqrt{n}}, \qquad \theta_n^* = \frac{\theta}{\sigma/\sqrt{n}}$$

we can write (4) equivalently as

$$x_n^* \sim \mathrm{N}(\theta_n^*, 1) \tag{5}$$

With σ known, the transformed random variable x_n^* is the t-ratio for testing the hypothesis $\theta = 0$, so its mean θ_n^* may be viewed as the "population t-ratio". There is no essential difference between the approach via (4) and the approach via (5). However, if we add a prior, then it *does* make a difference whether we place the prior on θ or on θ_n^*.

The standard Bayesian approach places a prior on θ. This prior does not depend on the sample size, so the resulting posterior mean of θ given x_n is consistent for θ [10, Chapter 10], because, as n increases, the data information becomes increasingly important and dominates the prior information which remains constant.

Since $\mathrm{MSE}(0) < \mathrm{MSE}(x_n)$ if and only if $|\theta_n^*| < 1$ and, more generally, since model selection and model averaging typically depend on diagnostics (such as t-ratios) rather than on parameter estimates, we believe that it makes more sense to place a prior on θ_n^* rather than on θ. This approach, first suggested by Huntsberger [31] and Hjort [32], plays a key role in the Bayesian shrinkage step of the WALS estimator of Magnus et al. [25].

Our approach is related to the rescaled spike-and-slab variable selection method of Ishwaran and Rao [23], where the observed responses are replaced by their $\sqrt{n}$-rescaled values. However, our objectives, and hence our assumptions, are quite different. While Ishwaran and Rao [23] are mainly concerned with the selective shrinkage and oracle properties of the posterior mean, we focus on the sampling properties of the posterior mean as an estimator of θ. Both approaches employ priors that do not depend on the sample size, but while Ishwaran and Rao [23] require that the effect of the prior on the posterior does not vanish as the sample size increases (even if this may lead to inconsistent estimators), we require that the prior on θ_n^* satisfies a minimum of regularity conditions that guarantee uniform $\sqrt{n}$-consistency of our estimator of θ.

The fact that the non-standard Bayesian approach may lead to inconsistent estimators of θ is an important issue. For example, suppose the prior distribution of θ_n^* is $\mathrm{N}(0, \tau^2)$, where τ does not depend on n. Then the posterior distribution of θ_n^* is $\mathrm{N}(\xi\, x_n^*,\ \xi)$, with $\xi = \tau^2/(1+\tau^2)$. Under quadratic loss, the Bayes estimator of θ_n^* is the posterior mean $m_n^* = \mathbb{E}(\theta_n^*|x_n^*) = \xi\, x_n^*$, while $\tilde{\theta}_n = \sigma\, m_n^*/\sqrt{n} = \xi\, x_n$ is the implied estimator of θ. From a frequentist perspective, the bias and variance of $\tilde{\theta}_n$ are $\mathbb{E}(\tilde{\theta}_n) - \theta = (\xi - 1)\theta$ and $\mathbb{V}(\tilde{\theta}_n) = \xi^2\sigma^2/n$, so the variance of $\tilde{\theta}_n$ vanishes as $n \to \infty$ but the bias does not. Hence, with a normal prior for θ_n^*, $\tilde{\theta}_n$ is not a consistent estimator of θ.

We thus need conditions on the prior to ensure that $\tilde{\theta}_n$ enjoys attractive sampling properties as an estimator of θ. Henceforth we shall write the posterior mean of θ_n^* as $m_n^* = m(x_n^*)$ and the posterior variance as $v_n^{*2} = v^2(x_n^*)$, where the functions $m : \mathbb{R} \to \mathbb{R}$ and $v^2 : \mathbb{R} \to \mathbb{R}_+$ have properties that depend on those of the prior density π of θ_n^*.

3 | Regularity Conditions on the Prior

Our first three conditions on the prior π are just mild regularity conditions:

C1. π is symmetric around 0;

C2. π is positive and non-increasing on $(0, \infty)$; and

C3. π is differentiable, except possibly at 0.

Kumar and Magnus [5] show that, under (C1)–(C3), the posterior mean function m satisfies the following properties: $m(-x) = -m(x)$, $m(0) = 0$, $m'(x) > 0$, $0 < m(x) < x$ for $x > 0$, and $m(x) \to \infty$ as $x \to \infty$. Since m is continuous and strictly increasing, its inverse function $\ell : \mathbb{R} \to \mathbb{R}$ exists, is continuous, strictly increasing, and satisfies $\ell(-t) = -\ell(t)$, $\ell(0) = 0$, and $\ell(t) \to \infty$ as $t \to \infty$. Because m is a shrinkage rule, there exists a continuous and symmetric shrinkage function $w(x) = m(x)/x$ for $x \neq 0$, satisfying $w(-x) = w(x)$, $0 < w(x) < 1$, and $w(x) \to v^2(0)$ as $x \to 0$. The latter property follows immediately by l'Hôpital's rule and the Brown–Tweedie formula [1, 33]: $\lim_{x\to 0} m(x)/x = \lim_{x\to 0} m'(x) = v^2(0)$.

An important requirement for posterior inference is that, when the data information is sufficiently strong, the prior should have bounded influence on $m(x)$ [34]. If this is not the case, then the MSE of m_n^* as an estimator of θ_n^* is not bounded in θ_n^* [33]. Although $m(x)$ is bounded from above by x (for $x > 0$), conditions (C1)–(C3) are not sufficient to imply this additional property.

To see this, consider again the example of a normal prior for θ_n^* and let us introduce the following two functions:

$$g(x) = x - m(x) = (1 - w(x))x$$
$$\psi(\theta) = -\frac{d \ln \pi(\theta)}{d\theta} = -\frac{\pi'(\theta)}{\pi(\theta)}$$

Since $g(x)$ measures the deviation between x and $m(x)$, we call it the "discrepancy function". In our example, $g(x) = (1 - \xi)x$, $w(x) = \xi$, and $\psi(\theta) = \theta/\tau^2$. Hence, the bias of m_n^* is equal to $-g(\theta_n^*)$ and its MSE is equal to $\xi^2 + g(\theta_n^*)^2$, both of which are unbounded in θ_n^* because the function g is unbounded. Since $g(\theta_n^*) = \xi\, \psi(\theta_n^*)$, it follows that the bias and the MSE of m_n^* are unbounded in θ_n^*, because the function ψ is unbounded.

To avoid this problem, we therefore impose the following additional condition:

C4. $\psi(\theta) \to \psi_0$ as $\theta \to \infty$, where $\psi_0 \geq 0$ is some finite constant.

With this additional condition, the limiting behaviour of $g(x)$ is also determined, essentially demonstrating the close relationship between $\psi(\theta)$ and $g(x)$.

Proposition 1. *Under conditions (C1)–(C4), $g(x) \to \psi_0 < \infty$ as $x \to \infty$.*

The proposition implies that the estimator m_n^* of θ_n^* has bounded MSE and is admissible. When $\psi_0 = 0$, it also implies a stronger property, sometimes called "Bayesian robustness", namely that $x - m(x) \to 0$ and $v^2(x) \to 1$ as $x \to \infty$, so that prior information is essentially ignored when x is sufficiently large. For robust priors (i.e., when $\psi_0 = 0$), we also have $w(x) \to 1$ as $x \to \infty$.

In a Bayesian analysis one often needs to formalise the concept of prior ignorance, and a flat (improper) prior is then typically used for computational convenience. In our context, a flat prior fails to capture a notion of prior ignorance which we call "neutrality", namely not knowing whether or not $|\theta_n^*| < 1$, that is, whether or not the silly estimator of θ has a lower MSE than the usual estimator. Our final condition on π ensures that neutrality is satisfied:

C5. π satisfies $\int_{-\infty}^{-1}\pi(\theta)\, d\theta = \int_{-1}^{0}\pi(\theta)\, d\theta = \int_{0}^{1}\pi(\theta)\, d\theta = \int_{1}^{\infty}\pi(\theta)\, d\theta = 1/4$.

Together with condition (C1), condition (C5) implies that the events $|\theta_n^*| < 1$ and $|\theta_n^*| > 1$ are equally likely a priori. In terms of the original parameter θ, this is equivalent to the condition that:

$$\Pr\left(|\theta| < \frac{\sigma}{\sqrt{n}}\right) = \frac{1}{2}$$

from which we see that the prior distribution of θ is asymptotically of the mixed discrete-continuous type with $\Pr(\theta = 0) = 1/2$ and $\Pr(\theta > 0) = \Pr(\theta < 0) = 1/4$. This property links our approach to the spike-and-slab prior originally proposed by Mitchell and Beauchamp [35], which is becoming increasingly popular in the application of Bayesian regularisation methods (see, e.g., [23, 36, 37]).

4 | Specification of the Prior

Many priors satisfy conditions (C1)–(C5). We shall study four. The first two priors are based on the gamma function

$$\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}\, dt \qquad (r > 0)$$

If we substitute $t = c\theta^b$ with $dt = bc\theta^{b-1}\, d\theta$, and let $r = (1 - a)/b$, we obtain the class of reflected gamma-type priors

$$\pi(\theta) = \frac{bc^{(1-a)/b}}{2\Gamma((1-a)/b)}|\theta|^{-a}e^{-c|\theta|^b} \qquad (0 \leq a < 1, b > 0, c > 0)$$

where we have imposed the additional restriction $a \geq 0$, because otherwise the prior will be increasing rather than decreasing for small values of $\theta > 0$. Special cases are the Weibull prior $(a + b = 1)$:

$$\pi(\theta) = \frac{bc}{2}\ |\theta|^{-(1-b)}e^{-c|\theta|^b} \qquad (0 < b \leq 1, c > 0)$$

and the Laplace prior ($a = 0$ and $b = 1$), which is a special case of the Weibull prior:

$$\pi(\theta) = \frac{c}{2}\ e^{-c|\theta|} \qquad (c > 0)$$

The Weibull (and hence also the Laplace) prior satisfies regularity conditions (C1)–(C4). As shown in Kumar and Magnus [5] and Magnus and De Luca [27], a reflected generalised gamma prior is robust if and only if $0 < b < 1$ and is neutral if and only if

$$\int_0^c t^{(1-a-b)/b}e^{-t}\, dt = \frac{\Gamma((1-a)/b)}{2}$$

Thus, the Weibull prior with $b < 1$ is robust, but the Laplace prior is not (although it has bounded influence since $\psi_0 = c > 0$). Neutrality leads to $c = \ln 2$ for both priors. For the Weibull prior, we fix the prior parameter b by the minimax regret criterion for m_n^*, with regret defined as the difference between the MSE of m_n^* and the minimum MSE in estimating θ_n^*. Based on this criterion Magnus and De Luca [27], find that the "optimal" neutral and robust Weibull prior is achieved for $b \approx 0.8876$ and $c = \ln 2$.

The other two priors are based on the beta function

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^1 t^{r-1}(1-t)^{s-1}\, dt \qquad (r > 0,\ s > 0)$$

Substituting $t = (1 + c\theta^b)^{-1}$ with $dt = -bc(1 + c\theta^b)^{-2}\theta^{b-1}d\theta$, and letting $r = 1/a - 1/b$ and $s = 1/b$, we obtain the class of reflected beta-type priors

$$\pi(\theta) = \frac{c^{1/b}b}{2\ B(1/a - 1/b, 1/b)}(1 + c|\theta|^b)^{-1/a} \qquad (0 < a < b, c > 0)$$

special cases of which are the Pareto prior ($b = 1$) :

$$\pi(\theta) = \frac{c(1-a)}{2a}(1 + c|\theta|)^{-1/a} \qquad (0 < a < 1, c > 0)$$

and the Cauchy prior ($a = 1, b = 2, c = 1$) :

$$\pi(\theta) = \frac{1}{\pi}\,\frac{1}{1+\theta^2}$$

where $\pi(\theta)$ denotes the prior function and $\pi \approx 3.14$ denotes Archimedes' constant. The Pareto and Cauchy priors also satisfy conditions (C1)–(C4). A reflected beta-type prior is robust when $b > a > 0$, and hence both the Pareto and the Cauchy priors are robust. The reflected beta-type prior is neutral when

$$\int_{1/(c+1)}^{1} t^{r-1}(1-t)^{s-1}\,dt = \frac{B(r,s)}{2}$$

where $r = 1/a - 1/b$ and $s = 1/b$. Specifically, the Pareto prior is neutral for $c = 2^{a/(1-a)} - 1$, while the Cauchy prior is always neutral. For the Pareto prior, we also fix the prior parameter a by the minimax regret criterion for m_n^*. The "optimal" neutral and robust Pareto prior is achieved for $a \approx 0.0862$ and $c \approx 0.0676$. The maximum regret for m_n^* is equal to 0.4546 under the Weibull prior, 0.4959 under the Pareto prior, 0.5127 under the Laplace prior, and 0.6332 under the Cauchy prior.

Figure 1 plots the posterior mean and discrepancy functions of the four priors after imposing neutrality and minimax regret optimality. For large θ_n^*, the Laplace prior has thin tails that decay exponentially, while the Cauchy prior has much thicker tails that decay at the rate $(\theta_n^*)^{-2}$. The Weibull and Pareto priors are in between the Laplace and Cauchy priors. This implies that, as x increases, the Cauchy posterior mean function $m(x)$ curves back towards the 45° line much faster than the other three priors. Turning to the discrepancy functions in the right panel, we see that the Laplace prior is not robust since its discrepancy function converges to the prior parameter $c = \ln 2$. The Weibull and Pareto priors are robust since $g(x) \to 0$ as $x \to \infty$, but convergence is slow, especially for the Weibull prior. In contrast, the Cauchy discrepancy function converges to zero very quickly.

Figure 2 plots the shrinkage and the posterior variance functions resulting from our four neutral priors. For small x, the Cauchy prior implies slightly more shrinkage than the other three priors but, for all of them, $w(x) \to v^2(0)$ as $x \to 0$ and $v^2(x) \to 1$ as $x \to \infty$. In agreement with Mitchell [38], the Laplace posterior variance is non-decreasing and converges to 1 from below as $x \to \infty$. Since $v^2(x) = m'(x)$, the Laplace posterior mean is convex for $x > 0$. The Weibull, Pareto, and Cauchy posterior variances reach a maximum greater than 1, before approaching 1 from above as $x \to \infty$. Moreover, for $x > 0$, their shrinkage functions are non-decreasing because $v^2(x) \geq w(x)$. In the next three sections, we show that the properties of the functions m, ℓ, g, v^2, and w play a crucial role in determining the finite-sample and asymptotic properties of our shrinkage estimator of θ.

5 | Finite-Sample Properties

In this section, we study the sampling properties of the shrinkage estimator $\tilde{\theta}_n$ of θ, given by $\tilde{\theta}_n = (\sigma/\sqrt{n})m_n^*$, as developed in Section 2. Extending the results of De Luca et al. [8] on the bias and variance of m_n^*, we establish the finite-sample distribution of $t_n = \sqrt{n}(\tilde{\theta}_n - \theta)/\sigma = m_n^* - \theta_n^*$ under model (5).

Proposition 2. *Let* $x_n^* \sim \mathrm{N}(\theta_n^*, 1)$. *Then,*

$$t_n = \frac{\tilde{\theta}_n - \theta}{\sigma/\sqrt{n}} = w(\theta_n^* + z_n)z_n - \big(1 - w(\theta_n^* + z_n)\big)\theta_n^*$$

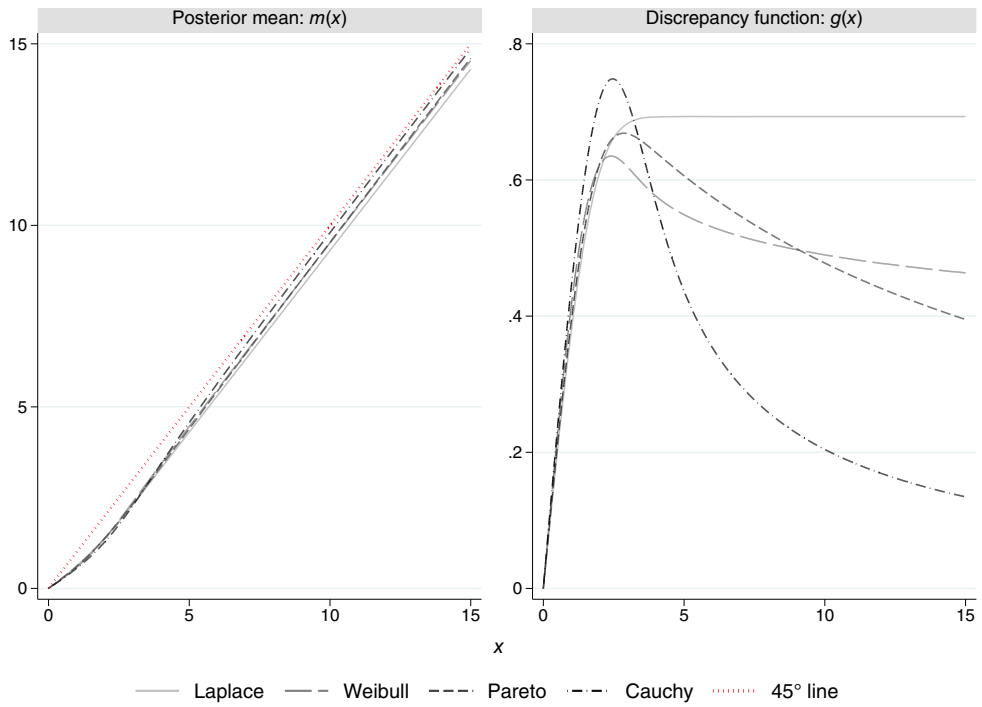

FIGURE 1 | Posterior mean and discrepancy functions under four neutral priors. [Colour figure can be viewed at wileyonlinelibrary.com]

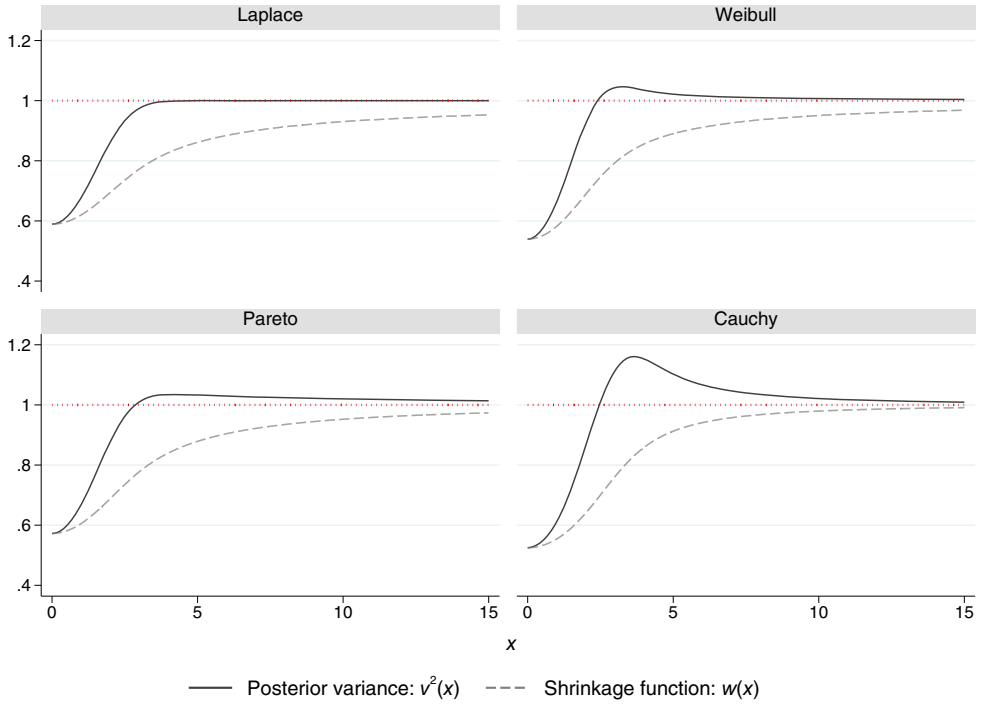

FIGURE 2 | Posterior variance and shrinkage functions under four neutral priors. [Colour figure can be viewed at wileyonlinelibrary.com]

where w is the shrinkage function and $z_n \sim \mathrm{N}(0,1)$. Under conditions $(C1)$–$(C3)$, the density of t_n at a point t is given by

$$f(t;\theta_n^*) = \frac{\phi\big(\ell(t+\theta_n^*) - \theta_n^*\big)}{v^2(\ell(t+\theta_n^*))}$$

where v^2 is the posterior variance function, ℓ is the inverse of the posterior mean function, and ϕ is the standard normal density.

Let us consider again the case of a $\mathrm{N}(0,\tau^2)$ prior for θ_n^*, where $m_n^* = \xi\, x_n^*$, $\ell(t+\theta_n^*) = (t+\theta_n^*)/\xi$, and $v^2(\ell(t+\theta_n^*)) = \xi$. Proposition 2 then implies that

$$f(t;\theta_n^*) = \frac{1}{\sqrt{2\pi\xi}} \exp -\frac{1}{2}\left(\frac{t-(\xi-1)\theta_n^*)}{\xi}\right)^2$$

that is, $t_n \sim \mathrm{N}((\xi-1)\theta_n^*, \xi^2)$. As in Section 2, we have $\mathbb{E}(t_n) = \mathbb{E}(m_n^* - \theta_n^*) = (\xi-1)\theta_n^*$, so the bias of our estimator of θ is $\mathbb{E}(\tilde{\theta}_n - \theta) = (\xi-1)(\sigma\theta_n^*/\sqrt{n}) = (\xi-1)\theta$. The result on the sampling variance of t_n is also in agreement with the "surprising" findings of Efron [22] and De Luca et al. [8, Proposition 2], namely that, for any positive and bounded prior density, the posterior *variance* represents a first-order approximation to the sampling *standard deviation* (not the sampling variance) of m_n^*. For the normal prior, the posterior mean is linear and hence the approximation is exact.

More generally, Proposition 2 shows that, except in the case of a normal prior, the finite-sample distribution of t_n is the same as that of a non-linear function of $z_n \sim \mathrm{N}(0,1)$, and is therefore non-normal in general. This distribution depends on the sample size n only through θ_n^*, and the continuity of ℓ and v^2 ensures that the density f is continuous in both its arguments.

At $\theta_n^* = 0$, the finite-sample distribution of t_n is the same as the distribution of $w(z_n)z_n$, which does not depend on n and is symmetric around the origin. All its odd moments thus vanish. Further, since $\mathbb{E}[(w(z_n))^{2p} z_n^{2p}] < \mathbb{E}(z_n^{2p})$, all its even moments are smaller than the corresponding moments of the standard normal distribution. Setting $p=1$ shows that, at $\theta_n^* = 0$, our shrinkage estimator is unbiased and more efficient than the usual estimator x_n^*.

Figure 3 plots the finite-sample densities of t_n under our four neutral priors for five values of θ_n^*. These densities are all non-normal but, unlike many post-selection estimators (see, e.g., Figure 2 in [14]) and many thresholding estimators (see, e.g., Figures 1–3 in [24]), they are all unimodal and smooth. This is because Bayesian shrinkage estimators are smooth functions of the data. At $\theta_n^* = 0$, all densities are symmetric and centred at the origin. They are also more concentrated than the standard normal density, and leptokurtic. For small non-zero values of θ_n^*, there is a bias which has opposite sign as θ_n^*, while the skewness has the same sign as θ_n^*. For large enough values of θ_n^*, all densities tend to have a normal shape but they are not necessarily centred at 0 because of the bias. In Proposition 4 we shall prove that the finite-sample distribution of t_n indeed converges to a normal distribution as $\theta_n^* \to \infty$.

Figure 4 plots the bias, variance, skewness, and excess kurtosis of m_n^* under our four neutral priors. As θ_n^* increases, the bias of m_n^* converges to zero much faster for the Cauchy and Pareto priors than for the Weibull prior, which is closely related to the behaviour of the discrepancy function shown in the right panel of Figure 1. When θ_n^* is small, the Cauchy posterior mean has a relatively small variance. However, as θ_n^* increases, its variance first reaches a peak of 1.24 at $\theta_n^* = 4.33$ and then approaches 1 monotonically. Although there is no uniform improvement in the MSE, the Cauchy prior may be preferred to the other priors because of its greater robustness. However, robustness is not the only criterion, and the Pareto prior offers, in our view, the

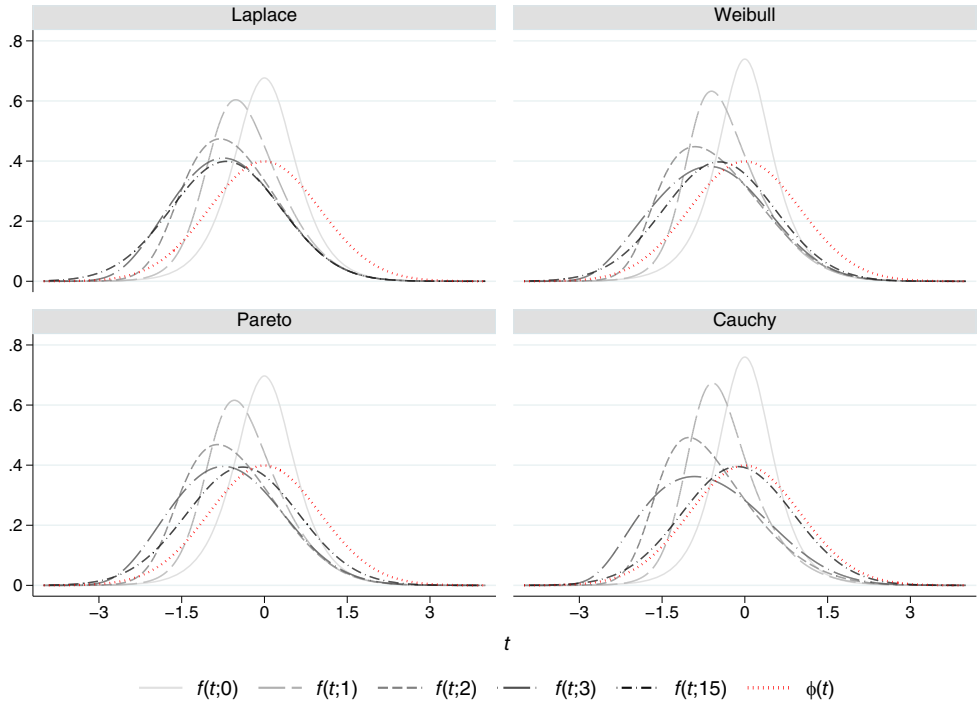

FIGURE 3 | Density of $t_n = \sqrt{n}(\tilde{\theta}_n - \theta)/\sigma$ under four neutral priors and five values of θ_n^*. [Colour figure can be viewed at wileyonlinelibrary.com]

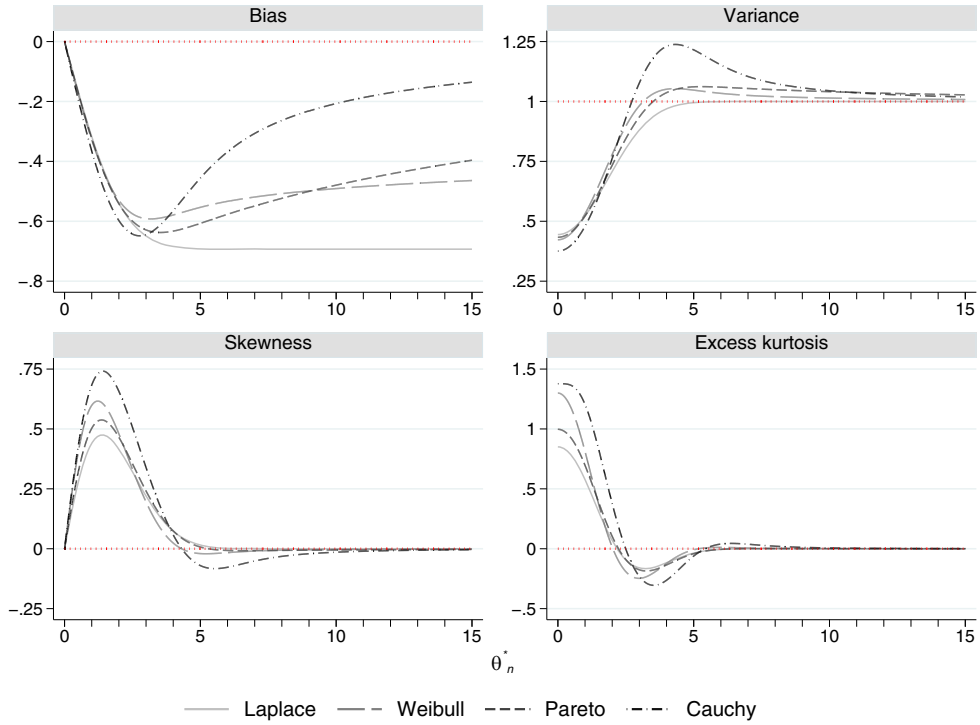

FIGURE 4 | Bias, variance, skewness and excess kurtosis of m_n^* under four neutral priors. [Colour figure can be viewed at wileyonlinelibrary.com]

best compromise between robustness and minimax regret. The skewness reaches a positive maximum around $\theta_n^* = 1.4$ and the peak is largest for the Cauchy prior. As θ_n^* increases, the skewness converges to 0 monotonically for the Laplace prior, while for the other priors, it first attains a negative minimum around $\theta_n^* = 5$ before approaching 0 from below. In contrast, the excess kurtosis attains a global maximum at $\theta_n^* = 0$, the peak being again largest for the Cauchy prior. As θ_n^* increases, the excess kurtosis first decreases, and attains a negative minimum between $\theta_n^* = 3$ and $\theta_n^* = 4$ for all priors, and then approaches 0, although not necessarily monotonically and not necessarily from below.

6 | Consistency

We next investigate the sampling properties of our shrinkage estimator of θ as $n \to \infty$, relaxing the assumption that x_n is normally distributed. In the current section, we establish (uniform) consistency, and in the next section the asymptotic distribution. Both properties will be proved under the mild assumption that $\sqrt{n}(x_n - \theta)/\sigma \xrightarrow{d} \mathrm{N}(0,1)$.

Formally, $\tilde{\theta}_n$ is (pointwise) $\sqrt{n}$-consistent for θ if, given any $\theta \in \mathbb{R}$ and $\varepsilon > 0$, there exist M and N (depending on both θ and ε) such that

$$n > N \;\Rightarrow\; \Pr[n^{1/2}|\tilde{\theta}_n - \theta| \geq M] \leq \varepsilon$$

implying that $\lim_{n\to\infty} \Pr[n^{\delta}|\tilde{\theta}_n - \theta| \geq \varepsilon] = 0$ for any $\theta \in \mathbb{R}$, $\varepsilon > 0$, and $0 \leq \delta < 1/2$. If the same N and M work equally well for every θ, the consistency is said to be uniform. That is, $\tilde{\theta}_n$ is *uniformly* $\sqrt{n}$-consistent for θ if, for any $\varepsilon > 0$, there exist M and N such that

$$n > N \;\Rightarrow\; \sup_{\theta\in\mathbb{R}} \Pr[n^{1/2}|\tilde{\theta}_n - \theta| \geq M] \leq \varepsilon$$

Proposition 3. *Suppose that* $\sqrt{n}(x_n - \theta)/\sigma \xrightarrow{d} \mathrm{N}(0,1)$ *and that the prior density of* θ_n^* *satisfies conditions* (C1)–(C4). *Then* $\tilde{\theta}_n$ *is uniformly* $\sqrt{n}$*-consistent for* θ, *and hence*

$$\lim_{n\to\infty} \sup_{\theta\in\mathbb{R}} \Pr[n^{\delta}|\tilde{\theta}_n - \theta| \geq \varepsilon] = 0$$

for every $0 \leq \delta < 1/2$ *and every* $\varepsilon > 0$.

The proof of this result exploits the fact that under (C1)–(C4), $g(x_n^*)$ is bounded, and so $|\tilde{\theta}_n - x_n|$ is bounded by a deterministic sequence proportional to $\sigma/\sqrt{n}$. This property highlights the difference between our Proposition 3 and [23, Theorem 2], who prove consistency of the posterior mean from a rescaled spike-and-slab model, where the variance inflation parameter λ_n satisfies the restriction $\lambda_n \to \infty$ and $\lambda_n/n \to 0$ as $n \to \infty$. Our Proposition 3 does not impose this restriction; instead it requires that the prior density of θ_n^* satisfies conditions (C1)–(C4).

Unlike Proposition 2, condition (C4) rules out a conjugate normal prior for θ_n^* because the corresponding discrepancy function would be unbounded (see Section 3). The uniform convergence rate of $\tilde{\theta}_n$ is similar to the convergence rate of conservative post-selection estimators, in contrast to consistent post-selection and thresholding estimators which typically achieve a slower convergence rate that depends on the underlying tuning parameters [24].

7 | Asymptotic Distribution

When establishing the asymptotic distribution of $\tilde{\theta}_n$, we allow explicitly for the possibility that θ depends on n. To structure this dependence, we set $\theta_n = n^{-\delta}\theta_0$ with $\delta \geq 0$ and $\theta_0 \in \mathbb{R}$. The most important special cases are $\delta = 0$ and $\delta = 1/2$. When $\delta = 0$ we have $\theta_n = \theta_0$ (the "fixed-parameter" case), and when $\delta = 1/2$ we have $\theta_n = n^{-1/2}\theta_0$ (the "local-parameter" case). For $\theta_0 = 0$ we have $\theta_n^* = 0$, but for $\theta_0 \neq 0$ our parameter of interest becomes

$$\theta_n^* = \frac{n^{1/2}\theta_n}{\sigma} = \frac{n^{1/2-\delta}\theta_0}{\sigma} \to \begin{cases} \pm\infty & \text{if } 0 \leq \delta < 1/2 \\ \theta_0/\sigma & \text{if } \delta = 1/2 \\ 0 & \text{if } \delta > 1/2 \end{cases}$$

as $n \to \infty$. What matters for the asymptotic distribution of $\tilde{\theta}_n$ is whether θ_n^* diverges or converges to a finite constant, in which case its speed of convergence also matters.

Proposition 4. *Suppose the conditions of Proposition 3 hold and let* $t_n = \sqrt{n}(\tilde{\theta}_n - \theta_n)/\sigma$, *where* $\theta_n = n^{-\delta}\theta_0$ *with* $\delta \geq 0$ *and* $\theta_0 \in \mathbb{R}$. *Let* w *be the shrinkage function and* $z \sim \mathrm{N}(0,1)$. *Then,* $t_n \xrightarrow{d} t$, *where the distribution of* t *depends on* δ *and* θ_0 *as follows*:

a. *If* $0 \leq \delta < 1/2$, *then*

$$t = \begin{cases} w(z)z & \text{if } \theta_0 = 0 \\ z - \mathrm{sign}(\theta_0)\,\psi_0 & \text{if } \theta_0 \neq 0 \end{cases}$$

where ψ_0 *is defined in condition* (C4). *If the prior is robust then* $\psi_0 = 0$.

b. *If* $\delta = 1/2$, *then* $t = w(\zeta + z)z - [1 - w(\zeta + z)]\zeta$ *with density*

$$f(t;\zeta) = \frac{\phi(\ell(t+\zeta) - \zeta)}{v^2(\ell(t+\zeta))}$$

where $\zeta = \theta_0/\sigma$ *and the functions* ϕ, v^2, *and* ℓ *are defined as in Proposition 2.*

c. *If* $\delta > 1/2$, *then* $t = w(z)z$ *with density* $f(t) = \phi(\ell(t))/v^2(\ell(t))$.

In contrast to Proposition 2, where condition (C4) is not needed, we need this condition in Proposition 4 to obtain the distribution of t when $0 \leq \delta < 1/2$ and $\theta_0 \neq 0$. At $\theta_0 = 0$, the value of δ does not matter and the distribution of t_n always converges to the distribution of $t = w(z)z$. When $\theta_0 \neq 0$ there are two relevant intervals for δ, namely $0 \leq \delta < 1/2$ and $\delta \geq 1/2$, though in practice, the most important values are $\delta = 0$ and $\delta = 1/2$, which we have called the fixed-parameter and the local-parameter case, respectively.

Part (a) of Proposition 4 shows that, when $\delta = 0$, the asymptotic distribution of $\tilde{\theta}_n$ is normal (standard normal if the prior is robust) for all $\theta_0 \neq 0$, but not for $\theta_0 = 0$. Thus, for θ_0 not equal

but close to zero, the asymptotic approximation could misrepresent important features of the finite-sample distribution of $\tilde{\theta}_n$. Moreover, as for the classical Hodges estimator (see, e.g., [10, pp. 109–110]), the fixed-parameter case gives the misleading impression that $\tilde{\theta}_n$ is superefficient at the origin, as it is asymptotically equivalent to x_n when $\theta_0 \neq 0$ but asymptotically more efficient when $\theta_0 = 0$. This over-optimistic conclusion reflects the lack of uniformity in the convergence of the finite-sample distribution to the asymptotic distribution. Related results are discussed in detail in Leeb and Pötscher [14, 15, 39], Pötscher [30], and Pötscher and Leeb [24].

Part (b) of Proposition 4 shows that, when $\delta = 1/2$, the asymptotic density of $t_n = \sqrt{n}(\tilde{\theta}_n - \theta_n)/\sigma$ is the same as the finite-sample density in Proposition 2, which justifies using the local-parameter framework to study the sampling properties of $\tilde{\theta}_n$. In a neighbourhood of $\zeta = 0$, the asymptotic distribution of t_n is characterised by sizeable departures from normality, both in terms of skewness and excess kurtosis. This is also the region of the parameter space where $\tilde{\theta}_n$ has a smaller asymptotic risk than the usual estimator x_n, although the risk improvements are not uniform. However, for sufficiently large values of ζ, the asymptotic distribution of t_n is normal.

Although one can construct consistent estimators of the finite-sample distribution function of $\sqrt{n}(\tilde{\theta}_n - \theta)$ for any given value of θ, these estimators are not uniformly close to the true distribution function. In fact [39, Lemmas 3.1 and 3.5], show that no uniformly consistent estimator of the distribution of shrinkage-type estimators can exist. This "impossibility result" is a general feature of a large class of estimators of θ that includes post-selection, thresholding, and model-averaging estimators. In our case, it means that one should be careful with inference based on the estimated distribution of $\tilde{\theta}_n$, as there is no guarantee that it will be close to the true distribution.

8 | Conclusions

We have investigated the sampling properties of a Bayesian estimator of θ in the normal location model $x_n \sim \mathrm{N}(\theta, \sigma^2/n)$ with σ^2 known. Unlike the standard Bayesian approach, which places a prior on θ that does not depend on the sample size n, we placed a prior on the "population t-ratio" $\theta_n^* = \sqrt{n}\theta/\sigma$. This non-standard approach is motivated by the fact that model selection and model averaging estimators typically depend on diagnostics such as t-ratios. Moreover, since MSE comparisons between alternative estimators of θ depend crucially on θ_n^*, it leads to a transparent notion of prior ignorance about θ_n^* that is formalised in the neutrality condition (C5).

Results on the large-sample properties of standard Bayesian estimators of θ do not directly extend to our approach. In addition to the mild regularity conditions (C1)–(C3), condition (C4) requires the prior on θ_n^* to have bounded influence on the posterior mean function. This rules out a conjugate normal prior for θ_n^* and is related to the notion of Bayesian robustness.

Our main results are as follows. First, we extend earlier results on the first two sampling moments of our estimator $\tilde{\theta}_n = n^{-1/2}\sigma\mathbb{E}(\theta_n^*|x_n)$ and derive its finite-sample distribution under the assumption that σ is known. We show that this distribution displays sizeable departures from normality in terms of skewness and excess kurtosis. Second, the choice of prior affects both the speed at which the estimation bias tends to zero and the speed at which the sampling variance tends to one. This bias-precision trade-off depends on the thickness of the prior tails. Third, $\tilde{\theta}_n$ is a uniformly $\sqrt{n}$-consistent estimator of θ under mild conditions on x_n. Finally, we derive the asymptotic distribution of $\tilde{\theta}_n$ under a moving-parameter setup that encompasses both the fixed-parameter and the local-parameter settings.

Acknowledgements

We are grateful to Aad van der Vaart and Xinyu Zhang for helpful discussions, and to two referees for constructive comments. Earlier versions of this paper were presented at ICEEE 2023, the University of Palermo, and the University of Rome Tor Vergata. Open access publishing facilitated by Universita degli Studi di Palermo, as part of the Wiley - CRUI-CARE agreement.

References

1. L. R. Pericchi and A. F. M. Smith, "Exact and Approximate Posterior Moments for a Normal Location Parameter," *Journal of the Royal Statistical Society, Series B* 54 (1992): 793–804.

2. J. R. Magnus, "The Traditional Pretest Estimator," *Theory of Probability and its Applications* 44 (1999): 293–308.

3. J. R. Magnus, "Estimation of the Mean of a Univariate Normal Distribution With Known Variance," *Econometrics Journal* 5 (2002): 225–236.

4. I. M. Johnstone and B. W. Silverman, "Needles and Straw in Haystacks: Empirical Bayes Estimates of Possibly Sparse Sequences," *Annals of Statistics* 32 (2004): 1594–1649.

5. K. Kumar and J. R. Magnus, "A Characterization of Bayesian Robustness for a Normal Location Parameter," *Sankhyā: The Indian Journal of Statistics* 75 (2013): 216–237.

6. A. DasGupta and I. Johnstone, "Asymptotic Risk and Bayes Risk of Thresholding and Superefficient Estimates and Optimal Thresholding," in *Contemporary Developments in Statistical Theory, Springer Proceedings in Mathematics & Statistics*, vol. 68, ed. S. Lahiri, A. Schick, A. SenGupta, and T. Sriram (Springer, 2014), 40–67.

7. I. M. Johnstone, "Gaussian Estimation: Sequence and Wavelet Models," in *Draft version* (2019), statweb.stanford.edu.

8. G. De Luca, J. R. Magnus, and F. Peracchi, "Sampling Properties of the Bayesian Posterior Mean With an Application to WALS Estimation," *Journal of Econometrics* 230 (2022): 299–317.

9. J. R. Magnus and J. Durbin, "Estimation of Regression Coefficients of Interest When Other Regression Coefficients Are of no Interest," *Econometrica* 67 (1999): 639–643.

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

11. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. (Springer, 1985).

12. G. G. Judge and M. E. Bock, "Biased Estimation," in *Handbook of Econometrics*, vol. 1, ed. Z. Griliches and M. D. Intriligator (North-Holland, 1983), 599–649.

13. H. Leeb and B. M. Pötscher, "The Finite-Sample Distribution of Post-Model-Selection Estimators and Uniform Versus Non-uniform Approximations," *Econometric Theory* 19 (2003): 100–142.

14. H. Leeb and B. M. Pötscher, "Model Selection and Inference: Facts and Fiction," *Econometric Theory* 21 (2005): 21–59.

15. H. Leeb and B. M. Pötscher, "Can One Estimate the Conditional Distribution of Post-Model-Selection Estimators?," *Annals of Statistics* 34 (2006a): 2554–2591.

16. P. Kabaila and H. Leeb, "On the Large-Sample Minimal Coverage Probability of Confidence Intervals After Model Selection," *Journal of the American Statistical Association* 101 (2006): 619–629.

17. B. Efron, "Estimation and Accuracy After Model Selection," *Journal of the American Statistical Association* 109 (2014): 991–1007.

18. W. E. Strawderman and A. Cohen, "Admissibility of Estimators of the Mean Vector of a Multivariate Normal Distribution With Quadratic Loss," *Annals of Mathematical Statistics* 42 (1971): 270–296.

19. D. N. Joanes and H. W. Peers, "On the Sampling Properties of Bayesian Point Estimators," *Journal of the American Statistical Association* 69 (1974): 560–564.

20. B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. (Chapman and Hall/CRC, 2000).

21. B. Efron, "Bayesian Inference and the Parametric Bootstrap," *Annals of Applied Statistics* 6 (2012): 1971–1997.

22. B. Efron, "Frequentist Accuracy of Bayesian Estimates," *Journal of the Royal Statistical Society, Series B* 77 (2015): 617–646.

23. H. Ishwaran and J. S. Rao, "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *Annals of Statistics* 33 (2005): 730–773.

24. B. M. Pötscher and H. Leeb, "On the Distribution of Penalized Maximum Likelihood Estimators: The LASSO, SCAD, and Thresholding," *Journal of Multivariate Analysis* 100 (2009): 2065–2082.

25. J. R. Magnus, O. Powell, and P. Prüfer, "A Comparison of Two Averaging Techniques With an Application to Growth Empirics," *Journal of Econometrics* 154 (2010): 139–153.

26. J. Lee and H.-S. Oh, "Bayesian Regression Based on Principal Components for High-Dimensional Data," *Journal of Multivariate Analysis* 117 (2013): 175–192.

27. J. R. Magnus and G. De Luca, "Weighted-Average Least Squares (WALS): A Survey," *Journal of Economic Surveys* 30 (2016): 117–148.

28. G. De Luca, J. R. Magnus, and F. Peracchi, "Weighted-Average Least Squares Estimation of Generalized Linear Models," *Journal of Econometrics* 204 (2018): 1–17.

29. G. De Luca, J. R. Magnus, and F. Peracchi, "Weighted-Average Least Squares (WALS): Confidence and Prediction Intervals," *Computational Economics* 61 (2023): 1637–1664.

30. B. M. Pötscher, "The Distribution of Model Averaging Estimators and an Impossibility Result Regarding Its Estimation," *IMS Lecture Notes Monograph Series* 52 (2006): 113–129.

31. D. V. Huntsberger, "A Generalization of a Preliminary Testing Procedure for Pooling Data," *Annals of Mathematical Statistics* 26 (1955): 734–743.

32. N. N. Hjort, "Bayes Estimators and Asymptotic Efficiency in Parametric Counting Process Models," *Scandinavian Journal of Statistics* 13 (1986): 63–85.

33. L. D. Brown, "Admissible Estimators, Recurrent Diffusions and Insoluble Boundary Value Problems," *Annals of Mathematical Statistics* 42 (1971): 855–903.

34. B. Sansó and L. R. Pericchi, "Near Ignorance Classes of Log-Concave Priors for the Location Model," *TEST* 1 (1992): 39–46.

35. T. J. Mitchell and J. J. Beauchamp, "'Bayesian Variable Selection in Linear Regression' (With Discussion)," *Journal of the American Statistical Association* 83 (1988): 1023–1036.

36. A. Abadie and M. Kasy, "Choosing Among Regularized Estimators in Empirical Economics: The Risk of Machine Learning," *Review of Economics and Statistics* 101 (2019): 1–20.

37. D. Giannone, M. Lenza, and G. E. Primiceri, "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica* 89 (2021): 2409–2437.

38. F. S. Mitchell, "A Note on Posterior Moments for a Normal Mean With Double-Exponential Prior," *Journal of the Royal Statistical Society, Series B* 56 (1994): 605–610.

39. H. Leeb and B. M. Pötscher, "Performance Limits for Estimators of the Risk or Distribution of Shrinkage-Type Estimators, and Some General Lower Risk-Bound Results," *Econometric Theory* 22 (2006b): 69–97.

Appendix A

Proofs

Proof of Proposition 1. This result is both a special case and an extension of [5, Theorem 4.1]. Kumar and Magnus prove, under conditions (C1)–(C4), that the equivalence

$$\lim_{\theta\to\infty} \psi(\theta) = \psi_0 \;\Leftrightarrow\; \lim_{x\to\infty} g(x) = \psi_0$$

holds for the special case $\psi_0 = 0$. In our case, we need only one direction of this equivalence (special case), but we need it for $\psi_0 \geq 0$ (extension). The extension is straightforward and is obtained by defining a new variable $r(\theta) = \psi(\theta) - \psi_0$. □

Proof of Proposition 2. Write $x_n^* = \theta_n^* + z_n$. Then, since $m_n^* = w(x_n^*)x_n^*$, the first result follows from

$$\sqrt{n}(\tilde{\theta}_n - \theta)/\sigma = m_n^* - \theta_n^* = w(\theta_n^* + z_n)z_n - [1 - w(\theta_n^* + z_n)]\theta_n^*$$

Since $t_n = m_n^* - \theta_n^* = m(x_n^*) - \theta_n^*$ is a continuous one-to-one transformation of x_n^* with inverse function $x_n^* = \ell(t_n + \theta_n^*)$, its density at a point t is

$$f(t;\theta_n^*) = |\ell'(t+\theta_n^*)|\ \phi(\ell(t+\theta_n^*) - \theta_n^*)$$

where $\ell'(t+\theta_n^*) = [m'(x_n^*)]^{-1}$. The second result then follows from the Brown–Tweedie formula, using the fact that $m'(x) = v^2(x) > 0$. □

Proof of Proposition 3. Since the function g is bounded, there exists a finite constant $G > 0$ such that $0 < g(x) \leq G$ for all x. The triangle inequality gives

$$|\tilde{\theta}_n - \theta| \leq |x_n - \theta| + |\tilde{\theta}_n - x_n|$$

from which we obtain

$$\begin{aligned}\Pr\big(n^{1/2}|\tilde{\theta}_n - \theta| \geq M\big) &\leq \Pr\big(n^{1/2}|x_n - \theta| + n^{1/2}|\tilde{\theta}_n - x_n| \geq M\big)\\ &\leq \Pr\big(n^{1/2}|x_n - \theta| \geq M/2\big) + \Pr\big(n^{1/2}|\tilde{\theta}_n - x_n| \geq M/2\big)\\ &\leq \frac{4\sigma^2}{M^2} + \Pr\big(n^{1/2}|\tilde{\theta}_n - x_n| \geq \sigma G\big)\end{aligned}$$

where the first term on the last line uses Chebyshev's inequality, and the second term follows by choosing $M > 2\sigma G$. Since

$$n^{1/2}|\tilde{\theta}_n - x_n| = \sigma|m(x_n^*) - x_n^*| = \sigma|g(x_n^*)| \leq \sigma G$$

almost surely, the result follows by setting $M = n^{1/2-\delta}\varepsilon$ and choosing n sufficiently large, so that $M > 2\sigma G$. □

Proof of Proposition 4. Cases (b) and (c), where $\delta \geq 1/2$, follow directly from Proposition 2 and the assumption that

$$\sqrt{n}(x_n - \theta)/\sigma = x_n^* - \theta_n^* = z_n \xrightarrow{d} z \sim \mathrm{N}(0,1)$$

Note that $\zeta = 0$ in case (c). Let us prove case (a), where $0 \leq \delta < 1/2$. We first rewrite $t_n = \sqrt{n}(\tilde{\theta}_n - \theta_n)/\sigma$ as

$$t_n = m_n^* - \theta_n^* = z_n - g(x_n^*)$$

with $m_n^* = w(x_n^*)x_n^*$, $\theta_n^* = n^{1/2-\delta}\theta_0/\sigma$, and $g(x_n^*) = x_n^* - m_n^*$. Since $z_n \xrightarrow{d} z$, we only need to study the asymptotic behaviour of $g(x_n^*)$. If $\theta_0 > 0$, then $\theta_n^* \to \infty$ as $n \to \infty$. Hence,

$$\Pr\left(x_n^* \geq M\right) = \Pr\left(z_n \geq M - \theta_n^*\right) \to 1$$

for every $M > 0$, which we write as $\text{plim} x_n^* = \infty$. By the (generalised) continuous mapping theorem (see Theorem 18.11 and Example 18.4 in [10]), we then find

$$\text{plim } g(x_n^*) = g(\text{plim } x_n^*) = g(\infty) = \psi_0$$

Similarly, if $\theta_0 < 0$, then $\theta_n^* \to -\infty$ as $n \to \infty$ so that

$$\Pr(x_n^* \leq -M) = \Pr(z_n \leq -M - \theta_n^*) \to 1$$

for every $M > 0$, which we write as $\text{plim} x_n^* = -\infty$. Since m is an odd function, we then find $\text{plim} g(x_n^*) = -\psi_0$. Finally, if $\theta_0 = 0$, then $x_n^* = z_n$ so that $m_n^* = w(z_n)z_n$, $g(x_n^*) = (1 - w(z_n))z_n$, and $t_n = w(z_n)z_n \xrightarrow{d} w(z)z$. □