Electronic Journal of Statistics
Vol. 11 (2017) 3196–3225
ISSN: 1935-7524
DOI: 10.1214/17-EJS1316

Adaptive posterior contraction rates for the horseshoe

Stéphanie van der Pas*

Leiden University
e-mail: svdpas@math.leidenuniv.nl

Botond Szabó*,†

Leiden University and Budapest University of Technology and Economics
e-mail: b.t.szabo@math.leidenuniv.nl

and

Aad van der Vaart†

Leiden University
e-mail: avdvaart@math.leidenuniv.nl

Abstract: We investigate the frequentist properties of Bayesian procedures for estimation based on the horseshoe prior in the sparse multivariate normal means model. Previous theoretical results assumed that the sparsity level, that is, the number of signals, was known. We drop this assumption and characterize the behavior of the maximum marginal likelihood estimator (MMLE) of a key parameter of the horseshoe prior. We prove that the MMLE is an effective estimator of the sparsity level, in the sense that it leads to (near) minimax optimal estimation of the underlying mean vector generating the data. Besides this empirical Bayes procedure, we consider the hierarchical Bayes method of putting a prior on the unknown sparsity level as well. We show that both Bayesian techniques lead to rate-adaptive optimal posterior contraction, which implies that the horseshoe posterior is a good candidate for generating rate-adaptive credible sets.

MSC 2010 subject classifications: Primary 62G15; secondary 62F15.
Keywords and phrases: Horseshoe, sparsity, nearly black vectors, normal means problem, adaptive inference, frequentist Bayes.

Received February 2017.

1. Introduction

The rise of big datasets with few signals, such as gene expression data and astronomical images, has given an impulse to the study of sparse models. The sequence model, or sparse normal means problem, is well studied. In this model,

*Research supported by the Netherlands Organization for Scientific Research.
†The research leading to these results has received funding from the European Research Council under ERC Grant Agreement 320637.

a random vector $Y^n = (Y_1, \dots, Y_n)$ with values in $\mathbb{R}^n$ is observed, and each single observation Y_i is the sum of a fixed mean $\theta_{0,i}$ and standard normal noise ε_i:

$$Y_i = \theta_{0,i} + \varepsilon_i, \quad i = 1, \dots, n. \tag{1.1}$$

We perform inference on the mean vector $\theta_0 = (\theta_{0,1}, \dots, \theta_{0,n})$, and assume it to be sparse in the nearly black sense, meaning that all except an unknown number $p_n = \sum_{i=1}^n \mathbf{1}\{\theta_{0,i} \neq 0\}$ of the means are zero. We assume that p_n increases with n, but not as fast as n: $p_n \to \infty$ and $p_n/n \to 0$ as n tends to infinity.

Many methods to recover θ_0 have been suggested. Those most directly related to this work are [32, 21, 10, 9, 19, 17, 20, 15, 6, 3, 2, 27]. In the present paper we study the Bayesian method based on the *horseshoe prior* [8, 7, 30, 25, 26]. Under this prior the coordinates $\theta_1, \dots, \theta_n$ are (given τ) an i.i.d. sample from a scale mixture of normals with a half-Cauchy prior on the variance, as follows. Given a "global hyperparameter" τ,

$$\begin{aligned} \theta_i \,|\, \lambda_i, \tau &\sim \mathcal{N}(0, \lambda_i^2 \tau^2), \\ \lambda_i &\sim C^+(0,1), \qquad i = 1, \dots, n. \end{aligned} \tag{1.2}$$

In the Bayesian model the observations Y_i follow (1.1) with θ_0 taken equal to θ. The *posterior distribution* is then as usual obtained as the conditional distribution of θ given Y^n. For a given value of τ, possibly determined by an empirical Bayes method, aspects of the posterior distribution of θ, such as its mean and variance, can be computed with the help of analytic formulas and numerical integration [25, 26, 35]. It is also possible to equip τ with a hyper prior, and follow a hierarchical, full Bayes approach. Several MCMC samplers and software packages are available for computation of the posterior distribution [29, 22, 16, 33, 18].

The horseshoe posterior has performed well in simulations [8, 7, 25, 24, 3, 1, 23]. Theoretical investigation in [35] shows that the parameter τ can, up to a logarithmic factor, be interpreted as the fraction of nonzero parameters θ_i. In particular, if τ is chosen to be at most of the order $(p_n/n)\sqrt{\log n/p_n}$, then the horseshoe posterior contracts to the true parameter at the (near) minimax rate of recovery for quadratic loss over sparse models [35]. While motivated by these good properties of the horseshoe prior, we also believe that the results obtained in the present paper give insight in the performance of Bayesian procedures for sparsity in general.

In the present paper we make three novel contributions. First and second we establish the contraction rates of the posterior distributions of θ in the hierarchical, full Bayes case and in the general empirical Bayes case. Third we study the particular empirical Bayes method of estimating τ by the method of maximum Bayesian marginal likelihood.

As the parameter τ can be viewed as measuring sparsity, the first two contributions are both focused on adaptation to the number p_n of nonzero means, which is unlikely to be known in practice. The hierarchical and empirical Bayes methods studied here are shown to have similar performance, both in theory and in a small simulation study, and appear to outperform the ad-hoc estimator introduced in [35]. The horseshoe posterior attains similar contraction rates as the spike-and-slab priors, as obtained in [21, 10, 9], and two-component mixtures, as in [27]. We obtain these results under general conditions on the hyper prior on τ, and for general empirical Bayes methods.

The conditions for the empirical Bayes method are met in particular by the maximum marginal likelihood estimator (MMLE). This is the maximum likelihood estimator of τ under the assumption that the "prior" (1.2) is part of the data-generating model, leaving only τ as a parameter. The MMLE is a natural estimator and is easy to compute. It turns out that the "MMLE plug-in posterior distribution" closely mimics the hierarchical Bayes posterior distribution, as has been observed in other settings [31, 28]. Besides practical benefit, this correspondence provides a theoretical tool to analyze the hierarchical Bayes method, which need not rely on testing arguments (as in [13, 14, 36]).

In the Bayesian framework the spread of the posterior distribution over the parameter space is used as an indication of the error in estimation. For instance, a set of prescribed posterior probability around the center of the posterior distribution (a credible set) is often used in the same way as a confidence region for the parameter. In the follow-up paper [34], we investigate the coverage properties and sizes of the adaptive credible balls and marginal credible intervals.

The paper is organized as follows. We first introduce the MMLE in Section 2. Next we present contraction rates in Section 3, for general empirical and hierarchical Bayes approaches, and specifically for the MMLE. We illustrate the results in Section 4. We conclude with appendices containing all proofs not given in the main text.

1.1. Notation

We use $\Pi(\cdot \mid Y^n, \tau)$ for the posterior distribution of θ relative to the prior (1.2) given fixed τ, and $\Pi(\cdot \mid Y^n)$ for the posterior distribution in the hierarchical setup where τ has received a prior. The empirical Bayes "plug-in posterior" is the first object with a data-based variable $\widehat{\tau}_n$ substituted for τ. In order to stress that this does not entail conditioning on $\widehat{\tau}_n$, we also write $\Pi_\tau(\cdot \mid Y^n)$ for $\Pi(\cdot \mid Y^n, \tau)$, and then $\Pi_{\widehat{\tau}_n}(\cdot \mid Y^n)$ is the empirical Bayes (or plug-in) posterior distribution.

The density of the standard normal distribution is denoted by φ. Furthermore, $\ell_0[p] = \{\theta \in \mathbb{R}^n : \sum_{i=1}^n \mathbf{1}\{\theta_i \neq 0\} \leq p\}$ denotes the class of nearly black vectors, and we abbreviate

$$\zeta_\tau = \sqrt{2\log(1/\tau)}, \qquad \tau_n(p) = (p/n)\sqrt{\log(n/p)}, \qquad \tau_n = \tau_n(p_n).$$

2. Maximum marginal likelihood estimator

In this Section we define the MMLE and compare it to a naive empirical Bayes estimator previously suggested in [35]. In Section 3.1, we show that the MMLE is close to the "optimal" value $\tau_n(p_n) = (p_n/n)\sqrt{\log(n/p_n)}$ with high probability, and leads to posterior contraction at the near-minimax rate.

The marginal prior density of a parameter θ_i in the model (1.2) is given by

$$g_\tau(\theta) = \int_0^\infty \varphi\left(\frac{\theta}{\lambda\tau}\right)\frac{1}{\lambda\tau}\frac{2}{\pi(1+\lambda^2)}\,d\lambda. \tag{2.1}$$

In the Bayesian model the observations Y_i are distributed according to the convolution of this density and the standard normal density. The MMLE is the maximum likelihood estimator of τ in this latter model, given by

$$\widehat{\tau}_M = \operatorname*{argmax}_{\tau\in[1/n,1]} \prod_{i=1}^n \int_{-\infty}^\infty \varphi(y_i-\theta)g_\tau(\theta)\,d\theta. \tag{2.2}$$

The restriction of the MMLE to the interval $[1/n, 1]$ can be motivated by the interpretation of τ as the level of sparsity, as in [35], which makes the interval correspond to assuming that at least one and at most all parameters are nonzero. The lower bound of $1/n$ has the additional advantage of preventing computational issues that arise when τ is very small ([35, 11]). We found the observation in [11] that an empirical Bayes approach cannot replace a hierarchical Bayes one, because the estimate of τ tends to be too small, too general. In both our theoretical study as in our simulation results the restriction that the MMLE be at least $1/n$ prevents a collapse to zero. Our simulations, presented in Section 4, also give no reason to believe that the hierarchical Bayes method is inherently better than empirical Bayes. Indeed, in our studies they behave very similarly (depending on the prior on τ), and we do not see consequences of the parameter "not being strongly identified by the data" as reported in [23].

The MMLE requires one-dimensional maximization and is thus easily computed. The behavior of the quantity to be maximized in (2.2) and the MMLE itself is illustrated in Figure 1. A function for computation is available in the R package 'horseshoe' ([33]).

An interpretation of τ as the fraction of nonzero coordinates motivates another estimator ([35]), which is based on a count of the number of observations that exceed the "universal threshold" $\sqrt{2\log n}$:

$$\widehat{\tau}_S(c_1,c_2) = \max\left\{\frac{\sum_{i=1}^n \mathbf{1}\{|y_i| \geq \sqrt{c_1\log n}\}}{c_2 n}, \frac{1}{n}\right\}, \tag{2.3}$$

where c_1 and c_2 are positive constants. If $c_2 > 1$ and ($c_1 > 2$ or $c_1 = 2$ and $p_n \gtrsim \log n$), then the plug-in posterior distribution with the *simple estimator* $\widehat{\tau}_S(c_1,c_2)$ contracts at the near square minimax rate $p_n\log n$ (see [35], Section 4). This also follows from Theorem 3.2 in the present paper, as $\widehat{\tau}_S(c_1,c_2)$ satisfies Condition 1 below.

In [35], it was observed that the simple estimator is prone to underestimation of the sparsity level if signals are smaller than the universal threshold. This is corroborated by the numerical study presented in Figure 2. The figure shows

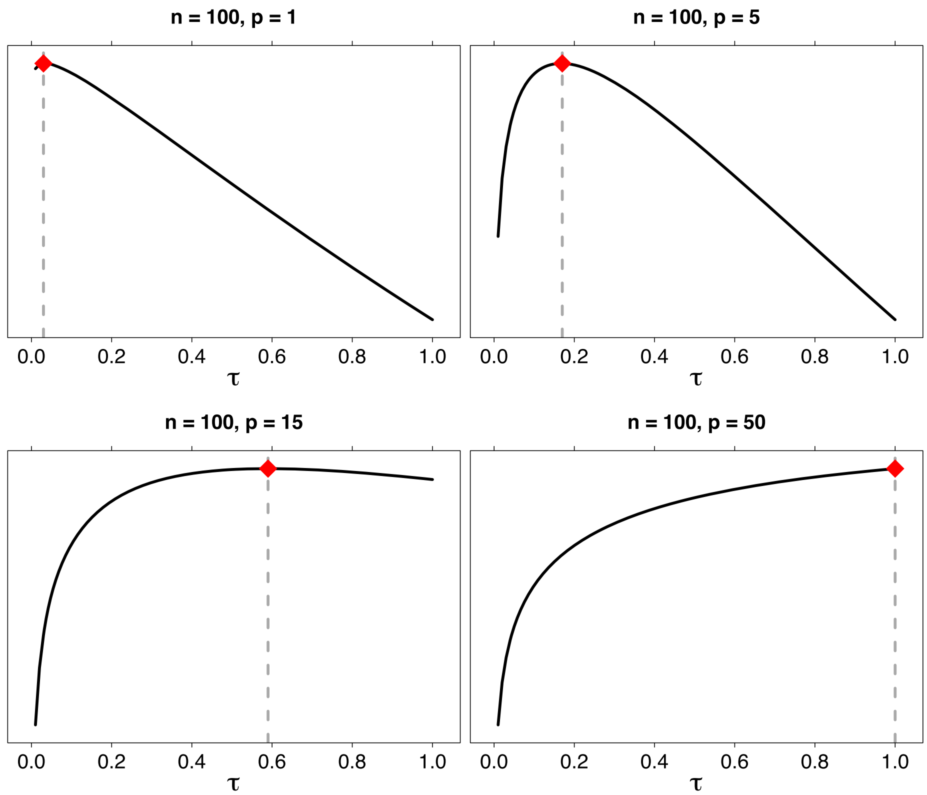

Fig 1. *Logarithm of the quantity to be maximized in* (2.2). *The red dot indicates the location of the MMLE. Each plot was made using a single simulated data set consisting of 100 observations each. From left to right, top to bottom, there are 1, 5, 15 or 40 means equal to 10; the remaining means are equal to zero.*

approximations to the expected values of $\widehat{\tau}_S$ and $\widehat{\tau}_M$ when θ_0 is a vector of length $n = 100$, with p_n coordinates drawn from a $\mathcal{N}(A,1)$ distribution, with $A \in \{1,4,7\}$, and the remaining coordinates drawn from a $\mathcal{N}(0,1/4)$ distribution. For this sample size the "universal threshold" $\sqrt{2\log n}$ is approximately 3, and thus signals with $A=1$ should be difficult to detect, whereas those with $A=7$ should be easy; those with $A=4$ represent a boundary case.

The figure shows that in all cases the MMLE (2.2) yields larger estimates of τ than the simple estimator (2.3), and thus leads to less shrinkage. This is expected in light of the results in the following section, which show that the MMLE is of order $\tau_n(p_n)$, whereas the simple estimator is capped at p_n/n. Both estimators appear to be linear in the number of nonzero coordinates of θ_0, with different slopes. When the signals are below the universal threshold, then the simple estimator is unlikely to detect any of them, whereas the MMLE may still pick up some of the signals. We study the consequences of this for the mean square errors in Section 4.

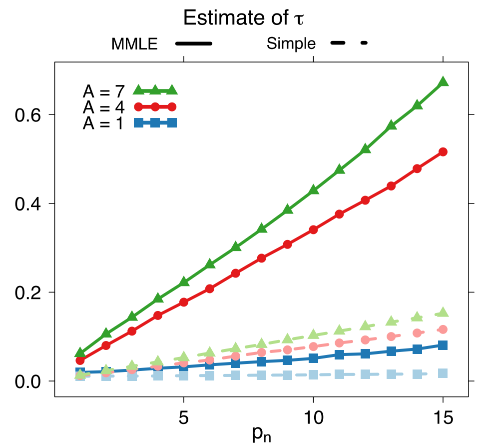

Fig 2. *Approximate expected values of the MMLE* (2.2) *(solid) and the simple estimator* (2.3) *with* $c_1 = 2$ *and* $c_2 = 1$ *(dotted) when* p_n *(horizontal axis) out of* $n = 100$ *parameters are drawn from a* $\mathcal{N}(A, 1)$ *distribution, and the remaining* $(n - p_n)$ *parameters from a* $\mathcal{N}(0, 1/4)$ *distribution. The study was conducted with* $A = 1$ *(■),* $A = 4$ *(●) and* $A = 7$ *(▲). The results as shown are the averages over* $N = 1000$ *replications.*

3. Contraction rates

In this section we establish the rate of contraction of both the empirical Bayes and full Bayes posterior distributions. The *empirical Bayes posterior* is found by replacing τ in the posterior distribution $\Pi(\cdot \mid Y^n, \tau)$ of θ relative to the prior (1.2) with a given τ by a data-based estimator $\widehat{\tau}_n$; we denote this by $\Pi_{\widehat{\tau}_n}(\cdot \mid Y^n)$. The *full Bayes posterior* $\Pi(\cdot \mid Y^n)$ is the ordinary posterior distribution of θ in the model where τ is also equipped with a prior and (1.2) is interpreted as the conditional prior of θ given τ.

The rate of contraction refers to properties of these posterior distributions when the vector Y^n follows a normal distribution on $\mathbb{R}^n$ with mean θ_0 and covariance the identity. We give general conditions on the empirical Bayes estimator $\widehat{\tau}_n$ and the hyper prior on τ that ensure that the square posterior rate of contraction to θ_0 of the resulting posterior distributions is the near minimax rate $p_n \log n$ for estimation of θ_0 relative to the Euclidean norm. We also show that these conditions are met by the MMLE and natural hyper priors on τ.

The minimax rate, the usual criterion for point estimators, has proven to be a useful benchmark for the speed of contraction of posterior distributions as well. The posterior cannot contract faster to the truth than at the minimax rate [13]. The square minimax ℓ_2-rate for the sparse normal means problem is $p_n \log(n/p_n)$ [12]. This is slightly faster (i.e. smaller) than $p_n \log n$, but equivalent if the true parameter vector is not very sparse (if $p_n \leq n^\alpha$, for some $\alpha < 1$, then $(1 - \alpha)p_n \log n \leq p_n \log(n/p_n) \leq p_n \log n$). For adaptive procedures, where the number of nonzero means p_n is unknown, results are usually given in terms of the "near-minimax rate" $p_n \log n$, for example for the spike-and-slab Lasso [27], the Lasso [4], and the horseshoe [35].

3.1. Empirical Bayes

The empirical Bayes posterior distribution achieves the near-minimax contraction rate provided that the estimator $\widehat{\tau}_n$ of τ satisfies the following condition. Let $\tau_n(p) = (p/n)\sqrt{\log(n/p)}$.

Condition 1. There exists a constant $C > 0$ such that $\widehat{\tau}_n \in [1/n, C\tau_n(p_n)]$, with P_{θ_0}-probability tending to one, uniformly in $\theta_0 \in \ell_0[p_n]$.

This condition is weaker than the condition given in [35] for ℓ_2-adaptation of the empirical Bayes posterior mean, which requires asymptotic concentration of $\widehat{\tau}_n$ on the same interval $[1/n, C\tau_n(p_n)]$ but at a rate. In [35] a plug-in value for τ of order $\tau_n(p_n)$ was found to be the largest value of τ for which the posterior distribution contracts at the minimax-rate, and has variance of the same order. Condition 1 can be interpreted as ensuring that $\widehat{\tau}_n$ is of at most this "optimal" order. The lower bound can be interpreted as assuming that there is at least one nonzero mean, which is reasonable in light of the assumption $p_n \to \infty$. In addition, it prevents computational issues, as discussed in Section 2.

A main result of the present paper is that the MMLE satisfies Condition 1.

Theorem 3.1. *The MMLE* (2.2) *satisfies Condition 1.*

Proof. See Appendix A.1. □

A second main result is that under Condition 1 the posterior contracts at the near-minimax rate.

Theorem 3.2. *For any estimator $\widehat{\tau}_n$ of τ that satisfies Condition 1, the empirical Bayes posterior distribution contracts around the true parameter at the near-minimax rate: for any $M_n \to \infty$,*

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathrm{E}_{\theta_0} \Pi_{\widehat{\tau}_n} \Big(\theta : \|\theta_0 - \theta\|_2 \geq M_n \sqrt{p_n \log n} \,|\, Y^n \Big) \to 0.$$

In particular, this is true for $\widehat{\tau}_n$ equal to the MMLE.

Proof. See Appendix B.1. □

3.2. Hierarchical Bayes

The full Bayes posterior distribution contracts at the near minimax rate whenever the prior density π_n on τ satisfies the following two conditions.

Condition 2. The prior density π_n is supported inside $[1/n, 1]$.

Condition 3. Let $t_n = C_u \pi^{3/2}\, \tau_n(p_n)$, with the constant C_u as in Lemma C.7(i). The prior density π_n satisfies

$$\int_{t_n/2}^{t_n} \pi_n(\tau)\, d\tau \gtrsim e^{-cp_n}, \quad \text{for some } c \leq C_u/2.$$

The restriction of the prior distribution to the interval $[1/n, 1]$ can be motivated by the same reasons as discussed under the definition of the MMLE in Section 2. In our simulations (also see [35]) we have also noted that large values produced by for instance a sampler using a half-Cauchy prior, as in the original set-up proposed by [8], were not beneficial to recovery.

This observation seems in agreement with the findings in [23], who also warn against too large values for τ. The authors of the latter paper suggest to choose a prior for τ based on the prior that it induces on the total shrinkage in the posterior mean (called "effective dimension"), making this center around the total number of nonzero parameters. They conclude that "there is no globally optimal prior choice", but suggest as a default a half Cauchy with a small scale determined through an equation involving the expected sparsity.

In the present paper we assume the standard deviation of the noise variables ε_i to be equal to 1, but in practice this will be unknown and it will be natural to scale the prior for θ_i or λ_j in (1.2). Alternatively but equivalently, one may scale the prior for τ, as suggested by [23]. The latter authors also work in the more general linear regression model, rather than the sequence model considered in the present paper. This introduces further complications, through the design matrix X, in particular in $p > n$ situations, where the assumption $X^T X = nI$ of [23] is clearly not tenable. The theory for spike-and-slab priors developed in [9] gives some suggestions for interpretations of effective dimensions of priors, but this deserves deeper investigation in the situation of the horseshoe prior.

As t_n is of the same order as $\tau_n(p_n)$, Condition 3 is similar to Condition 1 in the empirical Bayes case. It requires that there is sufficient prior mass around the "optimal" values of τ. The condition is satisfied by many prior densities, including the usual ones, except in the very sparse case that $p_n \lesssim \log n$, when it requires that π_n is unbounded near zero. Thus, the 'extremely sparse' case, as identified by [5], where $p_n \to s \in (0, \infty]$ and $\log(p_n)/\log(n) \to 0$, is not entirely covered by Condition 3. For this regime we also introduce the following weaker condition, which is still good enough for a contraction rate with additional logarithmic factors.

Condition 4. For t_n as in Condition 3 the prior density π_n satisfies,

$$\int_{t_n/2}^{t_n} \pi_n(\tau)\, d\tau \gtrsim t_n.$$

Example 3.3. The Cauchy distribution on the positive reals, truncated to $[1/n, 1]$, has density $\pi_n(\tau) = (\arctan(1) - \arctan(1/n))^{-1}(1 + \tau^2)^{-1}\mathbf{1}_{\tau\in[1/n,1]}$. This satisfies Condition 2, of course, and Condition 4. It also satisfies the stronger Condition 3 provided $t_n \geq e^{-cp_n}$, i.e. $p_n \geq C \log n$, for a sufficiently large C.

Example 3.4. For the uniform prior on $[1/n, 1]$, with density $\pi_n(\tau) = n/(n - 1)\mathbf{1}_{\tau\in[1/n,1]}$, the same conclusions hold.

Example 3.5. For the prior with density $\pi_n(x) \propto 1/x$ on $[1/n, 1]$, Conditions 2 and 3 hold provided $p_n \gg \operatorname{loglog} n$.

The following lemma is a crucial ingredient of the derivation of the contraction rate. It shows that the posterior distribution of τ will concentrate its mass at most a constant multiple of t_n away from zero. We denote the posterior distribution of τ by the same general symbol $\Pi(\cdot \,|\, Y^n)$.

Lemma 3.6. *If Conditions 2 and 3 hold, then*

$$\inf_{\theta_0 \in \ell_0[p_n]} \mathrm{E}_{\theta_0} \Pi(\tau : \tau \leq 5t_n \,|\, Y^n) \to 1.$$

Furthermore, if only Conditions 2 and 4 hold, then the similar assertion is true but with $5t_n$ *replaced by* $(\log n)t_n$.

Proof. See Appendix B.2. □

We are ready to state the posterior contraction result for the full Bayes posterior.

Theorem 3.7. *If the prior on* τ *satisfies Conditions 2 and 3, then the hierarchical Bayes posterior contracts to the true parameter at the near minimax rate: for any* $M_n \to \infty$,

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathrm{E}_{\theta_0} \Pi(\theta : \|\theta - \theta_0\|_2 \geq M_n \sqrt{p_n \log n} \,|\, Y^n) \to 0.$$

If the prior on τ *satisfies only Conditions 2 and 4, then this is true with* $\sqrt{p_n \log n}$ *replaced by* $\sqrt{p_n} \log n$.

Proof. Using the notation $r_n = \sqrt{p_n \log n}$, we can decompose the left side of the preceding display as

$$\begin{aligned} &\mathrm{E}_{\theta_0} \Big[\int_{\tau \leq 5t_n} + \int_{\tau > 5t_n} \Big] \Pi_\tau(\theta : \|\theta - \theta_0\|_2 \geq M_n r_n \,|\, Y^n) \, \pi(\tau \,|\, Y^n) \, d\tau \\ &\quad \leq \mathrm{E}_{\theta_0} \sup_{\tau \leq 5t_n} \Pi_\tau(\theta : \|\theta - \theta_0\|_2 \geq M_n r_n \,|\, Y^n) + \mathrm{E}_{\theta_0} \Pi(\tau : \tau > 5t_n \,|\, Y^n). \end{aligned}$$

The first term on the right tends to zero by Theorem 3.2, and the second by Lemma 3.6. □

4. Simulation study

We study the relative performances of the empirical Bayes and hierarchical Bayes approaches further through simulation studies, extending the simulation study in [35]. We consider the mean square error (MSE) for empirical Bayes combined with either (i) the simple estimator (with $c_1 = 2, c_2 = 1$) or (ii) the MMLE, and for hierarchical Bayes with (iii) a Cauchy prior on τ, (iv) a Cauchy prior truncated to $[1/n, 1]$ on τ, or (v) a uniform prior on [0,1] on τ.

We created a ground truth θ_0 of length $n = 400$ with $p_n \in \{20, 200\}$, where each nonzero mean was fixed to $A \in \{1, 2, \ldots, 10\}$. We computed the posterior mean for each of the four procedures, and approximated the MSE by averaging

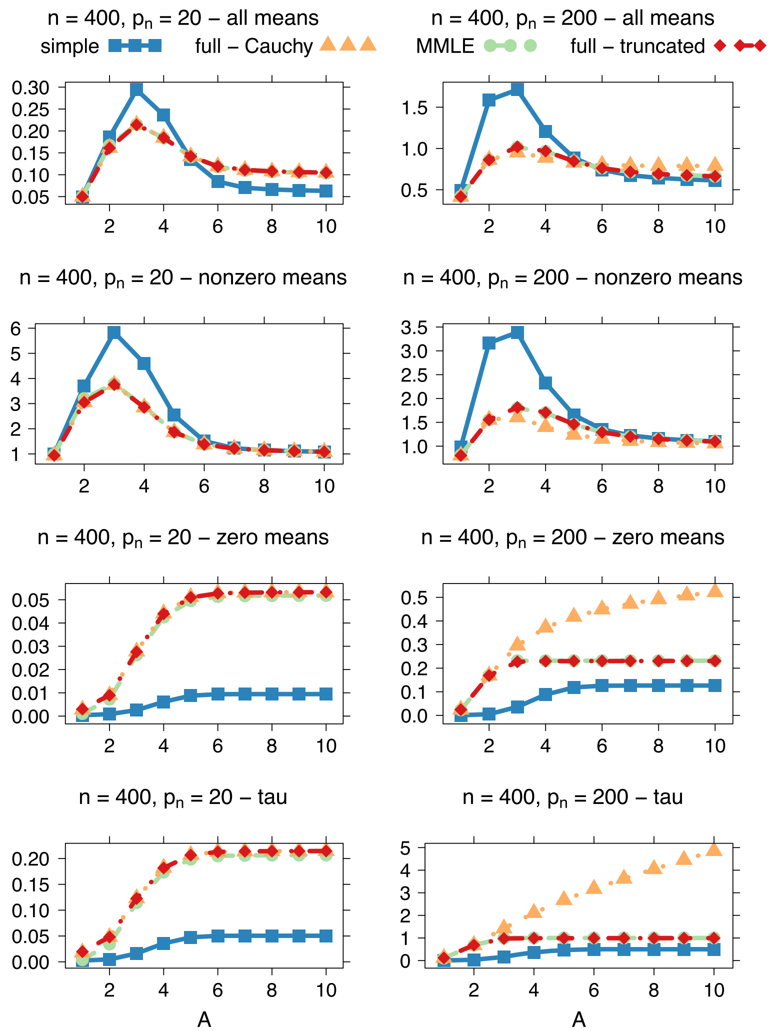

FIG 3. *Mean square error (overall, for the nonzero coordinates, and for the zero coordinates) of the posterior mean corresponding to empirical Bayes with the simple estimator with* $c_1 = 2, c_2 = 1$ *(■) or the MMLE (•) and to hierarchical Bayes with a Cauchy prior on* τ *(▲) or a Cauchy prior truncated to* $[1/n, 1]$ *(♦). The bottom plot shows the average estimated value of* τ *(or the posterior mean in the case of the hierarchical Bayes approaches). The settings are* $n = 400$ *and* $p_n = 20$ *(left) and* $p_n = 200$ *(right); the results are approximations based on averaging over* $N = 100$ *samples for each value of A.*

over $N = 100$ iterations. The results are shown in Figure 3. In addition the figure shows the MSE separately for the nonzero and zero coordinates of θ_0, and the average value (of the posterior mean) of τ. The results for the uniform prior are not plotted, as they are very similar to those for the truncated Cauchy. Full results and standard deviations are given in Appendix D.

The shapes of the curves of the overall MSE for methods (i) and (iii) were discussed in [35]. Values close to the threshold $\sqrt{2\log n} \approx 3.5$ pose the most difficult problem, and hierarchical Bayes with a Cauchy prior performs better below the threshold, while empirical Bayes with the simple estimator performs better above, as the simple estimator is very close to p_n/n in those settings, whereas the values of τ resulting from hierarchical Bayes are much larger.

Four new features stand out in this comparison, with the MMLE and hierarchical Bayes with a truncated Cauchy added in, and the opportunity to study the zero and nonzero means separately. The first is that empirical Bayes with the MMLE and hierarchical Bayes with the Cauchy prior truncated to $[1/n, 1]$ behave very similarly, as was expected from our proofs, in which the comparison of the two methods is fruitfully explored. The second is that the differences between the results for the truncated Cauchy and the uniform prior on τ are negligible, as was expected based on the theoretical results.

Thirdly, while in the most sparse setting ($p_n = 20$), full Bayes with the truncated and non-truncated Cauchy priors yield very similar results, as the mean value of τ does not come close to the 'maximum' of 1 in either approach, the truncated Cauchy (and the MMLE) offer an improvement over the non-truncated Cauchy in the less sparse ($p_n = 200$) setting. The non-truncated Cauchy does lead to lower MSE on the nonzero means close to the threshold, but overestimates the zero means due to the large values of τ. With the MMLE and the truncated Cauchy, the restriction to $[1/n, 1]$ prevents the marginal posterior of τ from concentrating too far away from the 'optimal' values of order $\tau_n(p_n)$, leading to better estimation results for the zero means, and only slightly higher MSE for the nonzero means.

Finally, the lower MSE of the simple estimator for large values of A in case $p_n = 20$ is mostly due to a small improvement in estimating the zero means, compared to the truncated Cauchy and the MMLE. As so many of the parameters are zero, this leads to lower overall MSE. However, close to the threshold, the absolute differences between these methods on the nonzero means can be quite large, and the simple estimator performs worse than all three other methods for these values.

Thus, from an estimation point of view, empirical Bayes with the MMLE or hierarchical Bayes with a truncated Cauchy seem to deliver the best results, only to be outperformed by hierarchical Bayes with a non-truncated Cauchy in a non-sparse setting with all zero means very close to the universal threshold.

Appendix A: Proof of the main result about the MMLE

A.1. Proof of Theorem 3.1

By its definition the MMLE maximizes the logarithm of the marginal likelihood function, which is given by

$$M_\tau(Y^n) = \sum_{i=1}^{n} \log\Big(\int_{-\infty}^{\infty} \varphi(y_i - \theta) g_\tau(\theta) d\theta\Big). \tag{A.1}$$

We split the sum in the indices $I_0 := \{i : \theta_{0,i} = 0\}$ and $I_1 := \{i : \theta_{0,i} \neq 0\}$. By Lemma C.1, with m_τ given by (C.3),

$$\frac{d}{d\tau} M_\tau(Y^n) = \frac{1}{\tau} \sum_{i \in I_0} m_\tau(Y_i) + \frac{1}{\tau} \sum_{i \in I_1} m_\tau(Y_i).$$

By Proposition C.2 the expectations of the terms in the first sum are strictly negative and bounded away from zero for $\tau \geq \varepsilon$, and any given $\varepsilon > 0$. By Lemma C.6 the sum behaves likes its expectation, uniformly in τ. By Lemma C.7 (i) the function m_τ is uniformly bounded by a constant C_u. It follows that for every $\varepsilon > 0$ there exists a constant $C_\varepsilon > 0$ such that, for all $\tau \geq \varepsilon$, and with $p_n = \#(\theta_{0,i} \neq 0)$, the preceding display is bounded above by

$$-\frac{n - p_n}{\tau} C_\varepsilon (1 + o_P(1)) + \frac{p_n}{\tau} C_u.$$

This is negative with probability tending to one as soon as $(n - p_n)/p_n > C_u/C_e$, and in that case the maximum $\widehat{\tau}_M$ of $M_\tau(Y^n)$ is taken on $[1/n, \varepsilon]$. Since this is true for any $\varepsilon > 0$, we conclude that $\widehat{\tau}_M$ tends to zero in probability.

We can now apply Proposition C.2 and Lemma C.3 to obtain the more precise bound on the derivative when $\tau \to 0$ given by

$$\frac{d}{d\tau} M_\tau(Y^n) \leq -\frac{(n - p_n)(2/\pi)^{3/2}}{\zeta_\tau} (1 + o_P(1)) + \frac{p_n}{\tau} C_u. \tag{A.2}$$

This is negative for $\tau/\zeta_\tau \gtrsim p_n/(n - p_n)$, and then $\widehat{\tau}_M$ is situated on the left side of the solution to this equation, or $\widehat{\tau}_M/\zeta_{\widehat{\tau}_M} \lesssim p_n/(n - p_n)$, which implies, that $\widehat{\tau}_M \lesssim \tau_n$, given the assumption that $p_n = o(n)$.

Appendix B: Proofs of the contraction results

Lemma B.1. *For $A > 1$ and every $y \in \mathbb{R}$,*

(i) $|\mathrm{E}(\theta_i \,|\, Y_i = y, \tau) - y| \leq 2\zeta_\tau^{-1}$, *for* $|y| \geq A\zeta_\tau$, *as* $\tau \to 0$.
(ii) $|\mathrm{E}(\theta_i \,|\, Y_i = y, \tau)| \leq |y|$.
(iii) $|\mathrm{E}(\theta_i \,|\, Y_i = y, \tau)| \leq \tau |y| e^{y^2/2}$, *as* $\tau \to 0$.
(iv) $|\operatorname{var}(\theta_i \,|\, Y_i = y, \tau) - 1| \leq \zeta_\tau^{-2}$, *for* $|y| \geq A\zeta_\tau$, *as* $\tau \to 0$.
(v) $\operatorname{var}(\theta_i \,|\, Y_i = y, \tau) \leq 1 + y^2$,
(vi) $\operatorname{var}(\theta_i \,|\, Y_i = y, \tau) \lesssim \tau e^{y^2/2}(y^{-2} \wedge 1)$, *as* $\tau \to 0$.

Proof. Inequalities (iii) and (v) come from Lemma A.2 and Lemma A.4 in [35], while (ii), (iv) and (vi) are implicit in the proofs of Theorems 3.1 and 3.2 (twice) in [35], and (i) with the bound ζ_τ instead of ζ_τ^{-1} is their (17). Alternatively, the posterior mean and variance in these assertions are given in (B.1) and (B.2). Then (ii) and (iv) are immediate from the fact that $0 \leq I_{3/2} \leq I_{1/2} \leq I_{-1/2}$, while (iii) and (vi) follow by bounding $I_{-1/2}$ below by a multiple of $1/\tau$ and $I_{3/2} \leq I_{1/2}$ above by $(1 \wedge y^{-2})e^{y^2/2}$, using Lemmas C.9 and C.10. Assertions (i) and (iv) follow from expanding $I_{-1/2}$ and $I_{1/2}$ and $I_{3/2}$, again using Lemmas C.9 and C.10. ∎

For the proof of Theorem 3.2, we use the following observations. The posterior density of θ_i given $(Y_i = y, \tau)$ is (for fixed τ) an exponential family with density

$$\theta \mapsto \frac{\varphi(y-\theta)g_\tau(\theta)}{\psi_\tau(y)} = c_\tau(y)e^{\theta y}g_\tau(\theta)e^{-\theta^2/2},$$

where g_τ is the posterior density of θ given in (2.1), and ψ_τ is the Bayesian marginal density of Y_i, given in (C.2), and the norming constant is given by

$$c_\tau(y) = \frac{\varphi(y)}{\psi_\tau(y)} = \frac{\pi}{\tau I_{-1/2}(y)},$$

for the function $I_{-1/2}(y)$ defined in (C.1). The cumulant moment generating function $z \mapsto \log \mathrm{E}(e^{z\theta_i} \,|\, Y_i = y, \tau)$ of the family is given by $z \mapsto \log\bigl(c_\tau(y)/c_\tau(y+z)\bigr)$, which is $z \mapsto \log I_{-1/2}(y+z)$ plus an additive constant independent of z. We conclude that the first, second and fourth cumulants are given by

$$\begin{aligned}\hat\theta_i(\tau) = \mathrm{E}(\theta_i \,|\, Y_i = y, \tau) &= \frac{d}{dy}\log I_{-1/2}(y),\\ \mathrm{var}(\theta_i \,|\, Y_i = y, \tau) &= \frac{d^2}{dy^2}\log I_{-1/2}(y), \qquad \text{(B.1)}\\ \mathrm{E}\bigl[\bigl(\theta_i - \hat\theta_i(\tau)\bigr)^4 \,|\, Y_i = y, \tau\bigr] - 3\,\mathrm{var}(\theta_i \,|\, Y_i = y, \tau)^2 &= \frac{d^4}{dy^4}\log I_{-1/2}(y).\end{aligned}$$

The derivatives at the right side can be computed by repeatedly using the product and sum rule together with the identity $I_k'(y) = yI_{k+1}(y)$, for I_k as in (C.1). In addition, since $(\log h)'' = h''/h - (h'/h)^2$, for any function h, and $I_{-1/2}'(y) = yI_{1/2}(y)$ and $I_{-1/2}''(y) = y^2 I_{3/2}(y) + I_{1/2}(y)$, we have

$$\mathrm{var}(\theta_i \,|\, Y_i = y, \tau) = y^2\Bigl[\frac{I_{3/2}}{I_{-1/2}} - \Bigl(\frac{I_{1/2}}{I_{-1/2}}\Bigr)^2\Bigr](y) + \frac{I_{1/2}}{I_{-1/2}}(y). \qquad \text{(B.2)}$$

B.1. Proof of Theorem 3.2

Proof. Set $r_n = \sqrt{p_n \log n}$ and $\tau_n = \tau_n(p_n)$. By Condition 1 and the triangle inequality,

$$\begin{aligned}&\mathrm{E}_{\theta_0}\Pi_{\hat\tau_n}\Bigl(\theta : \|\theta_0 - \theta\|_2 \ge M_n r_n \,|\, Y^n\Bigr)\\ &\quad\le \mathrm{E}_{\theta_0}\mathbf{1}_{\hat\tau_n \in [1/n, C\tau_n]}\Pi_{\hat\tau_n}\Bigl(\theta : \|\theta_0 - \hat\theta(\hat\tau_n)\|_2 + \|\theta - \hat\theta(\hat\tau_n)\|_2 \ge M_n r_n \,|\, Y^n\Bigr)\\ &\qquad + o(1)\\ &\quad\le \mathrm{E}_{\theta_0}\sup_{\tau \in [1/n, C\tau_n]}\Pi_\tau\Bigl(\theta : \|\theta_0 - \hat\theta(\tau)\|_2 + \|\theta - \hat\theta(\tau)\|_2 \ge M_n r_n \,|\, Y^n\Bigr) + o(1).\end{aligned}$$

Hence, in view of Chebyshev's inequality, it is sufficient to show that, with $\mathrm{var}(\theta \,|\, Y^n, \tau) = \mathrm{E}\bigl(\|\theta - \hat\theta(\tau)\|^2 \,|\, Y^n, \tau\bigr)$,

$$P_{\theta_0}\Big(\sup_{\tau\in[1/n,C\tau_n]}\|\theta_0-\hat\theta(\tau)\|_2\geq (M_n/2)r_n\Big)=o(1), \tag{B.3}$$

$$P_{\theta_0}\Big(\sup_{\tau\in[1/n,C\tau_n]}\mathrm{var}(\theta\,|\,Y^n,\tau)\geq M_n r_n^2\Big)=o(1). \tag{B.4}$$

To prove (B.3) we first use Lemma B.1(i)+(ii) to see that $|\hat\theta_i(\tau)|\lesssim\zeta_\tau$ and next the triangle inequality to see that $|\hat\theta_i(\tau)-\theta_{0,i}|\lesssim\zeta_\tau+|Y_i-\theta_{0,i}|$, as $\tau\to 0$. This shows that

$$\mathrm{E}_{\theta_{0,i}}\sup_{\tau\in[1/n,\tau_n]}(\theta_{0,i}-\hat\theta_i(\tau))^2\lesssim\sup_{\tau\geq 1/n}\zeta_\tau^2+\mathrm{var}_{\theta_{0,i}}Y_i\lesssim\log n. \tag{B.5}$$

Second we use Lemma B.1 (iii) and (ii) to see that $|\hat\theta_i(\tau)|$ is bounded above by $\tau|Y_i|e^{Y_i^2/2}$ if $|Y_i|\leq\zeta_{\tau_n}$ and bounded above by $|Y_i|$ otherwise, so that

$$\mathrm{E}_0\sup_{\tau\in[1/n,C\tau_n]}|\hat\theta_i(\tau)|^2\lesssim\int_0^{\zeta_{\tau_n}}(C\tau_n)^2y^2e^{y^2}\varphi(y)\,dy+\int_{\zeta_{\tau_n}}^\infty y^2\varphi(y)\,dy\lesssim\tau_n\zeta_{\tau_n}.$$

Applying the upper bound (B.5) for the p_n non-zero coordinates $\theta_{0,i}$, and the upper bound in the last display for the zero parameters, we find that

$$\mathrm{E}_{\theta_0}\sup_{\tau\in[1/n,C\tau_n]}\|\theta_0-\hat\theta(\tau)\|_2^2\lesssim p_n\log n+(n-p_n)\tau_n\zeta_{\tau_n}\lesssim p_n\log n.$$

Next an application of Markov's inequality leads to (B.3).

The proof of (B.4) is similar. For the nonzero $\theta_{0,i}$ we use the fact that $\mathrm{var}(\theta_i\,|\,Y_i,\tau)\leq 1+\zeta_\tau^2\lesssim\log n$, by Lemma B.1 (iv) and (v), while for the zero $\theta_{0,i}$ we use that $\mathrm{var}(\theta_i\,|\,Y_i,\tau)$ is bounded above by $\tau e^{Y_i^2/2}$ for $|Y_i|\leq\zeta_{\tau_n}$ and bounded above by $1+Y_i^2$ otherwise, by Lemma B.1 (vi) and (v). For the two cases of parameter values this gives bounds for $\mathrm{E}_{\theta_{0,i}}\sup_{\tau\in[1/n,C\tau_n]}\mathrm{var}(\theta_i\,|\,Y_i,\tau)$ of the same form as the bounds for the square bias, resulting in the overall bound $p_n\log n+(n-p_n)\tau_n\zeta_{\tau_n}\lesssim p_n\log n$ for the sum of these variances. An application of Markov's inequality gives (B.4). □

B.2. Proof of Lemma 3.6

The number t_n defined in Condition 4 is the (approximate) solution to the equation $p_nC_u/\tau=C_e(n-p)/(2\zeta_\tau)$, for $C_e=(\pi/2)^{3/2}$. By the decomposition (A.2), with P_{θ_0}-probability tending to one,

$$\frac{\partial}{\partial\tau}M_\tau(Y^n)<\begin{cases}p_nC_u/(t_n/2), & \text{if } t_n/2\leq\tau\leq t_n,\\ 0 & \text{if } \tau>t_n,\\ -p_nC_u/(2t_n), & \text{if } \tau\geq 2t_n.\end{cases}$$

Therefore, for $M_\tau(Y^n)$ defined in (A.1), $\tau_{\min}=\mathrm{argmin}_{\tau\in[t_n/2,t_n]}M_\tau(Y^n)$, and $\tau\geq 2t_n$,

$$M_\tau(Y^n) - M_{\tau_{\min}}(Y^n) = \Big[\int_{\tau_{\min}}^{t_n} + \int_{t_n}^{2t_n} + \int_{2t_n}^{\tau}\Big] \frac{\partial}{\partial s} M_s(Y^n)\, ds$$
$$\le (t_n/2)p_n C_u/(t_n/2) + 0 - (\tau - 2t_n)p_n C_u/(2t_n)$$
$$= -(\tau - 4t_n)p_n C_u/(2t_n),$$

where the right hand side is further bounded from above by $-\tau p_n C_u/(10t_n)$ for $\tau \ge 5t_n$. Since $\pi(\tau \,|\, Y^n) \propto \pi(\tau)e^{M_\tau(Y^n)}$ by Bayes's formula, with P_{θ_0}-probability tending to one, for $c_n \ge 5$

$$\Pi(\tau \ge c_n t_n \,|\, Y^n) \le \frac{\int_{\tau \ge c_n t_n} e^{M_{\tau_{\min}}(Y^n) - \tau p_n C_u/(10t_n)}\pi(\tau)\, d\tau}{\int_{\tau \in [t_n/2, t_n]} e^{M_{\tau_{\min}}(Y^n)}\pi(\tau)\, d\tau}$$
$$\lesssim \frac{e^{-c_n p_n C_u/10}}{\int_{\tau \in [t_n/2, t_n]} \pi(\tau)\, d\tau}.$$

Under Condition 3 this tends to zero if $c_n \ge 5$. Under the weaker Condition 4 this is certainly true for $c_n \ge \log n$.

Appendix C: Lemmas supporting the MMLE results

For $k \in \{-1/2, 1/2, 3/2\}$ define a function $I_k : \mathbb{R} \to \mathbb{R}$ by

$$I_k(y) := \int_0^1 z^k \frac{1}{\tau^2 + (1-\tau^2)z} e^{y^2 z/2}\, dz. \tag{C.1}$$

The Bayesian marginal density of Y_i given τ is the convolution $\psi_\tau := \varphi * g_\tau$ of the standard normal density and the prior density of g_τ, given in (2.1). The latter is a half-Cauchy mixture of normal densities $\varphi_{\tau\lambda}$ with mean zero and standard deviation $\tau\lambda$. By Fubini's theorem it follows that ψ_τ is a half-Cauchy mixture of the densities $\varphi * \varphi_{\tau\lambda}$. In other words

$$\psi_\tau(y) = \int_0^\infty \frac{e^{-\frac{1}{2}y^2/(1+\tau^2\lambda^2)}}{\sqrt{1+\tau^2\lambda^2}\sqrt{2\pi}} \frac{2}{1+\lambda^2}\frac{1}{\pi}\, d\lambda = \int_0^1 \frac{e^{-\frac{1}{2}y^2(1-z)}}{\sqrt{2\pi}\pi} \frac{\tau z^{-1/2}}{\tau^2(1-z)+z}\, dz$$
$$= \frac{\tau}{\pi} I_{-1/2}(y)\varphi(y), \tag{C.2}$$

where the second step follows by the substitution $1 - z = (1+\tau^2\lambda^2)^{-1}$ and some algebra. Note that $I_{-1/2}$ depends on τ, but this has been suppressed from the notation I_k.

Set

$$m_\tau(y) = y^2 \frac{I_{1/2}(y) - I_{3/2}(y)}{I_{-1/2}(y)} - \frac{I_{1/2}(y)}{I_{-1/2}(y)}. \tag{C.3}$$

Lemma C.1. *The derivative of the log-likelihood function takes the form*

$$\frac{d}{d\tau} M_\tau(y^n) = \frac{1}{\tau}\sum_{j=1}^n m_\tau(y_j).$$

Proof. From (C.2) we infer that, with a dot denoting the partial derivative with respect to τ,

$$\frac{\dot\psi_\tau}{\psi_\tau}=\frac{1}{\tau}+\frac{\dot I_{-1/2}}{I_{-1/2}}=\frac{I_{-1/2}+\tau\dot I_{-1/2}}{\tau I_{-1/2}}=\frac{\int_0^1\frac{e^{y^2z/2}}{\sqrt{z}N(z)^2}[N(z)-2\tau^2(1-z)]\,dz}{\tau I_{-1/2}},$$

where $N(z)=\tau^2(1-z)+z=\tau^2+(1-\tau^2)z$. By integration by parts,

$$y^2(I_{1/2}-I_{3/2})(y)=\int_0^1\frac{\sqrt{z}(1-z)}{N(z)}y^2e^{y^2z/2}\,dz=-2\int_0^1 e^{y^2z/2}\,d\Big[\frac{\sqrt{z}(1-z)}{N(z)}\Big].$$

Substituting the right hand side in formula (C.3), we readily see by some algebra that τ^{-1} times the latter formula reduces to the right side of the preceding display. ◻

Proposition C.2. *Let $Y\sim N(\theta,1)$. Then $\sup_{\tau\in[\varepsilon,1]}\mathrm{E}_0m_\tau(Y)<0$ for every $\varepsilon>0$, and as $\tau\to 0$,*

$$\mathrm{E}_\theta m_\tau(Y)=\begin{cases}-\frac{2^{3/2}}{\pi^{3/2}}\frac{\tau}{\zeta_\tau}\big(1+o(1)\big), & |\theta|=o(\zeta_\tau^{-2}),\\ o(\tau^{1/16}\zeta_\tau^{-1}), & |\theta|\le\zeta_\tau/4.\end{cases}\tag{C.4}$$

Proof. Let κ_τ be the solution to the equation $e^{y^2/2}/(y^2/2)=1/\tau$, that is

$$e^{\kappa_\tau^2/2}=\frac{1}{\tau}\,\kappa_\tau^2/2,\qquad \kappa_\tau\sim\zeta_\tau+\frac{2\log\zeta_\tau}{\zeta_\tau},\qquad \zeta_\tau=\sqrt{2\log(1/\tau)}.$$

We split the integral over $(0,\infty)$ into the three parts $(0,\zeta_\tau)$, (ζ_τ,κ_τ), and (κ_τ,∞), where we shall see that the last two parts give negligible contributions.

By Lemma C.7(vi) and (vii), if $|\theta|\kappa_\tau=O(1)$,

$$\int_{|y|\ge\kappa_\tau}m_\tau(y)\varphi(y-\theta)\,dy\lesssim\int_{z\ge\kappa_\tau-|\theta|}\varphi(z)\,dz\lesssim\frac{e^{-(\kappa_\tau-\theta)^2/2}}{\kappa_\tau-\theta}\lesssim\frac{e^{-\kappa_\tau^2/2}}{\kappa_\tau},$$

$$\int_{\zeta_\tau\le|y|\le\kappa_\tau}m_\tau(y)\varphi(y-\theta)\,dy\lesssim\int_{\zeta_\tau\le|y|\le\kappa_\tau}\frac{\tau e^{y^2/2-(y-\theta)^2/2}}{y^2}\,dy\lesssim\frac{\tau(\kappa_\tau-\zeta_\tau)}{\zeta_\tau^2}.$$

By the definition of κ_τ, both terms are of smaller order than τ/ζ_τ.

Because $e^{y^2/2}/y^2$ is increasing for large y and reaches the value τ^{-1}/ζ_τ^2 at $y=\zeta_\tau$, Lemma C.9 gives that $I_{-1/2}(y)=\pi\tau^{-1}(1+O(1/\zeta_\tau^2))$ uniformly in y in the interval $(0,\zeta_\tau)$. Therefore

$$\int_{|y|\le\zeta_\tau}m_\tau(y)\varphi(y-\theta)\,dy=\int_0^{\zeta_\tau}\frac{y^2I_{1/2}(y)-y^2I_{3/2}(y)-I_{1/2}(y)}{\tau^{-1}\pi}\varphi(y)\,dy+R_\tau,$$

where the remainder R_τ is bounded in absolute value by $\int_0^{\zeta_\tau}|y^2(I_{1/2}-I_{3/2})(y)-I_{1/2}(y)|\varphi(y)\,dy$ times $\sup_{0\le y\le\zeta_\tau}\big|\varphi(y-\theta)/(I_{-1/2}(y)\varphi(y))-1/(\tau^{-1}\pi)\big|$, which is bounded above by $\tau\big(\zeta_\tau^{-2}+e^{|\theta|\zeta_\tau-\theta^2/2}-1\big)=o(\tau\zeta_\tau^{-1})$, for $|\theta|=o(\zeta_\tau^{-2})$. By

Lemma C.10 the integrand in the integral is bounded above by a constant for y near 0 and by a multiple of y^{-2} otherwise, and hence the integral remains bounded. Thus the remainder R_τ is negligible. By Fubini's theorem the integral in the preceding display can be rewritten

$$\frac{\tau}{\pi}\int_0^1 \frac{\sqrt{z}}{\tau^2+(1-\tau^2)z}\int_0^{\zeta_\tau}\big[y^2(1-z)-1\big]\frac{e^{-y^2(1-z)/2}}{\sqrt{2\pi}}\,dy\,dz$$
$$=-\frac{\tau}{\pi}\int_0^1 \frac{\sqrt{z}}{\tau^2+(1-\tau^2)z}\int_{\zeta_\tau}^{\infty}\big[y^2(1-z)-1\big]\frac{e^{-y^2(1-z)/2}}{\sqrt{2\pi}}\,dy\,dz$$

by the fact that the inner integral vanishes when computed over the interval $(0,\infty)$ rather than $(0,\zeta_\tau)$. Since $\int_y^\infty[(va)^2-1]\varphi(va)\,dv=y\varphi(ya)$, it follows that the right side is equal to

$$-\frac{\tau}{\pi}\int_0^1 \frac{\sqrt{z}}{\tau^2+(1-\tau^2)z}\frac{\zeta_\tau\, e^{-\zeta_\tau^2(1-z)/2}}{\sqrt{2\pi}}\,dz.$$

We split the integral in the ranges $(0,1/2)$ and $(1/2,1)$. For z in the first range we have $1-z\geq 1/2$, whence the contribution of this range is bounded in absolute value by

$$\frac{\zeta_\tau\tau}{\pi\sqrt{2\pi}}e^{-\zeta_\tau^2/4}\int_0^{1/2}\frac{\sqrt{z}}{(1-\tau^2)z}\,dz=O(\zeta_\tau\tau e^{-\zeta_\tau^2/4}).$$

Uniformly in z in the range $(1/2,1)$ we have $\tau^2+(1-\tau^2)z\sim z$, and the corresponding contribution is

$$-\frac{\tau}{\pi}\int_{1/2}^1\frac{1}{\sqrt{z}}\frac{\zeta_\tau\, e^{-\zeta_\tau^2(1-z)/2}}{\sqrt{2\pi}}\,dz=-\frac{\tau}{\pi\zeta_\tau\sqrt{2\pi}}\int_0^{\zeta_\tau^2/2}\frac{1}{\sqrt{1-u/\zeta_\tau^2}}e^{-u/2}\,du.$$

by the substitution $\zeta_\tau^2(1-z)=u$. The integral tends to $\int_0^\infty e^{-u/2}\,du=2$, and hence the expression is asymptotic to half the expression as claimed.

The second statement follows by the same estimates, where now we use that $e^{|\theta|2\zeta_\tau-\theta^2/2}\leq\tau^{-15/16}$, if $|\theta|\leq\zeta_\tau/4$.

Since $\mathrm{E}_0 m_\tau(Y)\sim -c\tau/\zeta_\tau$ for a positive constant c, as $\tau\downarrow 0$, the continuous function $\tau\mapsto \mathrm{E}_0 m_\tau(Y)$ is certainly negative if $\tau>0$ and τ is close to zero. To see that it is bounded away from zero as τ moves away from 0, we computed $\mathrm{E}_0 m_\tau(Y)$ via numerical integration. The result is shown in Figure 4. ☐

Lemma C.3. *For any $\varepsilon_\tau\downarrow 0$ and uniformly in $I_0\subseteq\{i:|\theta_{0,i}|\leq\zeta_\tau^{-1}\}$ with $|I_0|\gtrsim n$,*

$$\sup_{1/n\leq\tau\leq\varepsilon_\tau}\frac{1}{|I_0|}\Big|\sum_{i\in I_0}m_\tau(Y_i)\frac{\zeta_\tau}{\tau}-\sum_{i\in I_0}\mathrm{E}_{\theta_0}m_\tau(Y_i)\frac{\zeta_\tau}{\tau}\Big|\overset{P_{\theta_0}}{\to}0.$$

Similarly, uniformly in $I_1\subseteq\{i:|\theta_{0,i}|\leq\zeta_\tau/4\}$,

$$\sup_{1/n\leq\tau\leq\varepsilon_\tau}\frac{1}{|I_1|}\Big|\sum_{i\in I_1}m_\tau(Y_i)\frac{\zeta_\tau}{\tau^{1/32}}-\sum_{i\in I_1}\mathrm{E}_{\theta_0}m_\tau(Y_i)\frac{\zeta_\tau}{\tau^{1/32}}\Big|\overset{P_{\theta_0}}{\to}0.$$

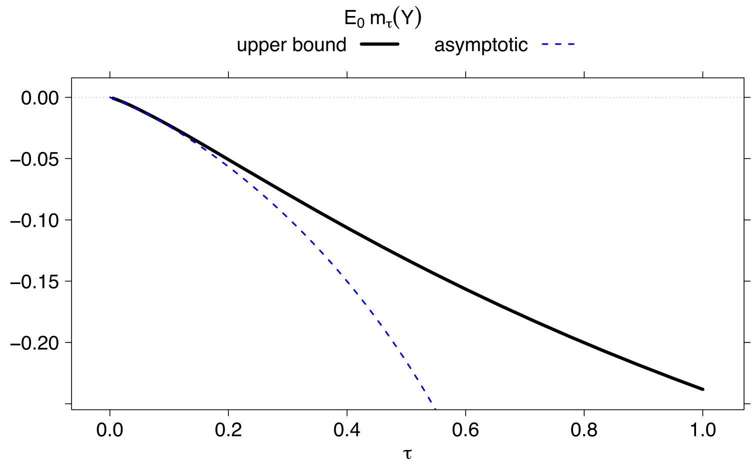

FIG 4. *Upper bound on* $\mathrm{E}_0 m_\tau(Y)$ *as computed with the R* `integrate()` *routine (solid line). The upper bound* $m_\tau(y) \le y^2$ *was used for* $|y| > 500$ *for numerical stability. The dashed line shows the asymptotic value* (C.4).

Proof. Write $G_n(\tau) = |I_0|^{-1} \sum_{i \in I_0} m_\tau(Y_i)(\zeta_\tau)/\tau$. In view of Corollary 2.2.5 of [37] (applied with $\psi(x) = x^2$) it is sufficient to show that $\mathrm{var}_{\theta_0} G_n(\tau) \to 0$ for some τ, and

$$\int_0^{\mathrm{diam}_n} \sqrt{N(\varepsilon, [1/n, 1], d_n)}\, d\varepsilon = o(1), \tag{C.5}$$

where d_n is the intrinsic metric defined by its square $d_n^2(\tau_1, \tau_2) = \mathrm{var}_{\theta_0}\big(G_n(\tau_1) - G_n(\tau_2)\big)$, diam_n is the diameter of the interval $[1/n, 1]$ with respect to the metric d_n, and $N(\varepsilon, A, d_n)$ is the covering number of the set A with ε radius balls with respect to the metric d_n.

If $|\theta_{0,i}| \le \zeta_\tau^{-1}$, then in view of Lemma C.5, as $\tau \to 0$,

$$\mathrm{var}_{\theta_0} G_n(\tau) \le \frac{1}{|I_0|} \mathrm{E}_{\theta_0}\big(m_\tau(Y)\zeta_\tau/\tau\big)^2 = o(\tau^{-1}/|I_0|).$$

This tends to zero, as $\tau n \ge 1$ by assumption. Combining this with the triangle inequality we also see that the diameter diam_n tends to 0.

Next we deal with the entropy. The metric d_n is up to a constant equal to the square root of the left side of (C.6). By Lemma C.4 it satisfies

$$d_n(\tau_1, \tau_2) \lesssim |I_0|^{-1/2} |\tau_2/\tau_1 - 1| \tau_1^{-1/2}.$$

To compute the covering number of the interval $[1/n, 1]$, we cover this by dyadic blocks $[2^i/n, 2^{i+1}/n]$, for $i = 0, 1, 2, ..., \log_2 n$. On the ith block the distance $d_n(\tau_1, \tau_2)$ is bounded above by a multiple of $n|\tau_1 - \tau_2|/2^{3i/2}$. We conclude that the ith block can be covered by a multiple of $\varepsilon^{-1} 2^{-i/2}$ balls of radius ε. Therefore the whole interval $[1/n, 1]$ can be covered by a multiple of $\varepsilon^{-1}\sum_i 2^{-i/2} \lesssim \varepsilon^{-1}$ balls of radius ε. Hence the integral of the entropy is bounded by

$$\int_0^{\mathrm{diam}_n} \sqrt{N(\varepsilon, [1/n, 1], d_n)}\, d\varepsilon \lesssim \int_0^{\mathrm{diam}_n} \varepsilon^{-1/2}\, d\varepsilon.$$

This tends to zero as diam_n tends to zero.

The second assertion of the lemma follows similarly, where we use the second parts of Lemmas C.5 and C.4. □

Lemma C.4. *Let $Y \sim N(\theta, 1)$. For $|\theta| \lesssim \zeta_\tau^{-1}$ and $0 < \tau_1 < \tau_2 \leq 1/2$,*

$$\mathrm{E}_\theta \left(\frac{\zeta_{\tau_1}}{\tau_1} m_{\tau_1}(Y) - \frac{\zeta_{\tau_2}}{\tau_2} m_{\tau_2}(Y) \right)^2 \lesssim (\tau_2 - \tau_1)^2 \tau_1^{-3}. \tag{C.6}$$

Furthermore, for $|\theta| \leq \zeta_\tau/4$, and $\varepsilon = 1/16$ and $0 < \tau_1 < \tau_2 \leq 1/2$,

$$\mathrm{E}_\theta \left(\frac{\zeta_{\tau_1}}{\tau_1^\varepsilon} m_{\tau_1}(Y) - \frac{\zeta_{\tau_2}}{\tau_2^\varepsilon} m_{\tau_2}(Y) \right)^2 \lesssim (\tau_2 - \tau_1)^2 \tau_1^{-2-\varepsilon}.$$

Proof. In view of Lemma C.11 the left side of (C.6) is bounded above by, for $\dot{m}_\tau$ denoting the partial derivative of m_τ with respect to τ,

$$\begin{aligned}
&(\tau_1 - \tau_2)^2 \sup_{\tau \in [\tau_1, \tau_2]} \mathrm{E}_\theta \Big(\frac{\zeta_\tau}{\tau} \dot{m}_\tau(Y) - \frac{\zeta_\tau + \zeta_\tau^{-1}}{\tau^2} m_\tau(Y) \Big)^2 \\
&\quad \leq (\tau_1 - \tau_2)^2 \Big[2 \sup_{\tau \in [\tau_1, \tau_2]} \mathrm{E}_\theta \Big(\frac{\zeta_\tau}{\tau} \dot{m}_\tau(Y) \Big)^2 + 2 \sup_{\tau \in [\tau_1, \tau_2]} \mathrm{E}_\theta \Big(\frac{\zeta_\tau + \zeta_\tau^{-1}}{\tau^2} m_\tau(Y) \Big)^2 \Big].
\end{aligned}$$

By Lemma C.5 the second expected value on the right hand side is bounded from above by a multiple of $\sup_{\tau \in [\tau_1, \tau_2]} \tau^{-3} \lesssim \tau_1^{-3}$.

To handle the first expected value, we note that the partial derivative of I_k with respect to τ is given by $\dot{I}_k = 2\tau(J_{k+1} - J_k)$, for

$$J_k(y) = \int_0^1 \frac{z^k}{(\tau^2 + (1 - \tau^2)z)^2} e^{y^2 z/2} dz. \tag{C.7}$$

Therefore, by (C.3),

$$\begin{aligned}
\dot{m}_\tau(y) &= (y^2 - 1) \frac{\dot{I}_{1/2}}{I_{-1/2}}(y) - y^2 \frac{\dot{I}_{3/2}}{I_{-1/2}}(y) - \frac{\dot{I}_{1/2}}{I_{-1/2}}(y) m_\tau(y) \\
&= 2\tau \Big[(y^2 - 1) \frac{J_{3/2} - J_{1/2}}{I_{-1/2}}(y) - y^2 \frac{J_{5/2} - J_{3/2}}{I_{-1/2}}(y) - \frac{J_{1/2} - J_{-1/2}}{I_{-1/2}}(y) m_\tau(y) \Big].
\end{aligned}$$

Since $J_k \leq I_{k-1}/(1 - \tau^2)$ and $J_k \leq I_k/\tau^2$, and $k \mapsto I_k$ and $k \mapsto J_k$ are decreasing and nonnegative, we have that

$$0 \leq \frac{J_{3/2} - J_{5/2}}{I_{-1/2}} \leq \frac{J_{1/2} - J_{3/2}}{I_{-1/2}} \leq \frac{J_{1/2}}{I_{-1/2}} \leq 4,$$

$$0 \le \frac{J_{-1/2} - J_{1/2}}{I_{-1/2}} \le \frac{J_{-1/2}}{I_{-1/2}} \le \frac{1}{\tau^2}. \tag{C.8}$$

By combining the preceding two displays we conclude

$$\mathrm{E}_\theta \dot{m}_\tau^2(Y) \lesssim \tau^2\Big[1 + \mathrm{E}_\theta Y^4 + \frac{1}{\tau^4}\mathrm{E}_\theta m_\tau^2(Y)\Big]. \tag{C.9}$$

Here $\mathrm{E}_\theta Y^4$ is bounded and $\mathrm{E}_\theta m_\tau^2(Y)$ is bounded above by $\tau\zeta_\tau^{-2}$ by Lemma C.5. It follows that $(\zeta_\tau/\tau)^2\mathrm{E}_\theta \dot{m}_\tau^2(Y)$ is bounded by a multiple of $\tau^{-3} \le \tau_1^{-3}$.

For the proof of the second assertion of the lemma, when $|\theta| \le \zeta_\tau/4$, we argue similarly, but now must bound,

$$(\tau_1 - \tau_2)^2\Big[2 \sup_{\tau\in[\tau_1,\tau_2]} \mathrm{E}_\theta\Big(\frac{\zeta_\tau}{\tau^\varepsilon}\dot{m}_\tau(Y)\Big)^2 + 2 \sup_{\tau\in[\tau_1,\tau_2]} \mathrm{E}_\theta\Big(\frac{\varepsilon\zeta_\tau + \zeta_\tau^{-1}}{\tau^{1+\varepsilon}} m_\tau(Y)\Big)^2\Big].$$

The same arguments as before apply, now using the second bound from Lemma C.5. ∎

Lemma C.5. *Let $Y \sim N(\theta, 1)$. Then, as $\tau \to 0$,*

$$\mathrm{E}_\theta m_\tau^2(Y) = \begin{cases} o(\tau\zeta_\tau^{-2}), & |\theta| \lesssim \zeta_\tau^{-1}, \\ o(\tau^{1/16}\zeta_\tau^{-2}), & |\theta| \le \zeta_\tau/4. \end{cases}$$

Proof. By Lemma C.7 (i), (vi) and (vii) we have, if $|\theta|\zeta_\tau \lesssim 1$,

$$\int_{|y|\ge\kappa_\tau} m_\tau^2(y)\varphi(y-\theta)\,dy \lesssim \int_{|z|\ge\kappa_\tau-\theta}^\infty \varphi(z)\,dz \lesssim e^{-(\kappa_\tau-\theta)^2/2}(\kappa_\tau-\theta)^{-1} \lesssim \tau\zeta_\tau^{-3},$$

$$\int_{\zeta_\tau\le|y|\le\kappa_\tau} m_\tau^2(y)\varphi(y-\theta)\,dy \lesssim \int_{\zeta_\tau}^{\kappa_\tau} \tau y^{-2} e^{y^2/2-(y-\theta)^2/2}\,dy = \tau(\kappa_\tau - \zeta_\tau)\zeta_\tau^{-2},$$

$$\int_{|y|\le\zeta_\tau} m_\tau^2(y)\varphi(y-\theta)\,dy \lesssim \tau^2 \int_0^{\zeta_\tau} (y^{-4}\wedge 1)e^{y^2/2}e^{\theta\zeta_\tau-\theta^2/2}\,dy \lesssim \tau\zeta_\tau^{-4}.$$

All three expressions on the right are $o(\tau\zeta_\tau^{-2})$.

The second assertion of the lemma follows by the same inequalities, together with the inequalities $e^{-(\kappa_\tau-\theta)^2/2} \le \tau^{-9/32}$ and $e^{|\theta|2\zeta_\tau-\theta^2/2} \le \tau^{-15/16}$, if $|\theta| \le \zeta_\tau/4$. ∎

Lemma C.6. *If the cardinality of $I_0 := \{i : \theta_{0,i} = 0\}$ tends to infinity, then*

$$\sup_{1/n\le\tau\le1} \frac{1}{|I_0|}\Big|\sum_{i\in I_0} m_\tau(Y_i) - \sum_{i\in I_0} \mathrm{E}_{\theta_0} m_\tau(Y_i)\Big| \overset{P_{\theta_0}}{\to} 0.$$

Proof. By Lemma C.7(i) we have that $\mathrm{E}_0 m_\tau^2(Y_i) \lesssim 1$ uniformly in τ and by the proof of Lemma C.4 $\mathrm{E}_0(m_{\tau_1} - m_{\tau_2})^2(Y_i) \lesssim |\tau_1 - \tau_2|^2/\tau_1$, uniformly in $0 < \tau_1 < \tau_2 \le 1$. The first shows that the marginal variances of the process $G_n(\tau) := |I_0|^{-1}\sum_{i\in I_0} m_\tau(Y_i)$ tend to zero as $|I_0| \to \infty$. The second allows to control the entropy integral of the process and complete the proof, in the same way as the proof of Lemma C.3. ∎

Lemma C.7. *The function $y \mapsto m_\tau(y)$ is symmetric about 0 and nondecreasing on $[0,\infty)$ with*

(i) $-1 \le m_\tau(y) \le C_u$, *for all $y \in \mathbb{R}$ and all $\tau \in [0,1]$, and some $C_u < \infty$.*
(ii) $m_\tau(0) = -(2\tau/\pi)(1+o(1))$, *as $\tau \to 0$.*
(iii) $m_\tau(\zeta_\tau) = 2/(\pi\zeta_\tau^2)(1+o(1))$, *as $\tau \to 0$.*
(iv) $m_\tau(\kappa_\tau) = 1/(\pi+1)/(1+o(1))$, *as $\tau \to 0$.*
(v) $\sup_{y \ge A\zeta_\tau} |m_\tau(y) - 1| = O(\zeta_\tau^{-2})$, *as $\tau \to 0$, for every $A > 1$.*
(vi) $m_\tau(y) \sim \tau e^{y^2/2}/(\pi y^2/2 + \tau e^{y^2/2})$, *as $\tau \to 0$, uniformly in $|y| \ge 1/\varepsilon_\tau$, for any $\varepsilon_\tau \downarrow 0$.*
(vii) $|m_\tau(y)| \lesssim \tau e^{y^2/2}(y^{-2} \wedge 1)$, *as $\tau \to 0$, for every y.*

Proof. As seen in the proof of Lemma C.1 the function m_τ can be written

$$m_\tau(y) = 1 + \tau \frac{\dot{I}_{-1/2}}{I_{-1/2}}(y) = 1 + 2\tau^2 \int_0^1 \frac{z-1}{\tau^2 + (1-\tau^2)z} g_y(z)\, dz,$$

for $z \mapsto g_y(z)$ the probability density function on $[0,1]$ with $g_y(z) \propto e^{y^2/2} z^{-1/2}/(\tau^2 + (1-\tau^2)z)$. If y increases, then the probability distribution increases stochastically, and hence so does the expectation of the increasing function $z \mapsto (z-1)/(\tau^2 + (1-\tau^2)z)$. (More precisely, note that g_{y_2}/g_{y_1} is increasing if $y_2 > y_1$ and apply Lemma C.12.)

(i). The inequality $m_\tau(y) \ge -1$ is immediate from the definition of (C.3) of m_τ and the fact that $I_{3/2} \le I_{1/2} \le I_{-1/2}$. For the upper bound it suffices to show that both $\sup_y m_\tau(y)$ remains bounded as $\tau \to 0$ and that $\sup_y \sup_{\tau \ge \delta} m_\tau(y) < \infty$ for every $\delta > 0$.

The first follows from the monotonicity and (v).

For the proof of the second we note that if $\tau \ge \delta > 0$, then $\delta^2 \le \tau^2 + (1-\tau^2)z \le 1$, for every $z \in [0,1]$, so that the denominators in the integrands of $I_{-1/2}, I_{1/2}, I_{3/2}$ are uniformly bounded away from zero and infinity and hence

$$m_\tau(y) \le y^2 \frac{I_{1/2}(y) - I_{3/2}(y)}{I_{-1/2}(y)} \le \frac{1}{\delta^2} \frac{y^2 \int_0^1 \sqrt{z}(1-z) e^{y^2 z/2}\, dz}{\int_0^1 z^{-1/2} e^{y^2 z/2}\, dz}.$$

After changing variables $zy^2/2 = v$, the numerator and denominator take the forms of the integrals in the second and first assertions of Lemma C.8, except that the range of integration is $(0, y^2/2)$ rather than $(1, y)$. In view of the lemma the quotient approaches 1 as $y \to \infty$. For y in a bounded interval the leading factor y^2 is bounded, while the integral in the numerator is smaller than the integral in the denominator, as $z(1-z) \le z \le z^{-1/2}$, for $z \in [0,1]$.

Assertions (ii)-(v) are consequences of the representation (C.3), Lemmas C.9 and C.10 and the fact that $I_{1/2}(0) = \int_0^1 z^{-1/2} dz\big(1 + O(\tau^2)\big) \to 2$.

Assertions (vi) and (vii) are immediate from Lemmas C.9 and C.10. ∎

C.1. Technical lemmas

Lemma C.8. *For any k, as $y \to \infty$,*

$$\int_1^y u^k e^u \, du = y^k e^y \big(1 - k/y + O(1/y^2)\big).$$

Consequently, as $y \to \infty$,

$$\int_1^y u^k e^u \, du - \frac{1}{y} \int_1^y u^{k+1} e^u \, du = y^{k-1} e^y \big(1 + O(1/y)\big).$$

Proof. By integrating by parts twice, the first integral is seen to be equal to

$$y^k e^y - e - k y^{k-1} e^y + ke + R,$$

where R satisfies

$$\begin{aligned} |R| &= |k(k-1)| \int_1^y u^{k-2} e^u \, du \\ &\le |k(k-1)| \int_1^{y/2} (1 \vee (y/2)^{k-2}) e^u \, du \\ &\qquad + |k(k-1)| \int_{y/2}^y ((y/2)^{k-2} \vee y^{k-2}) e^u \, du \\ &\lesssim |k(k-1)| \Big[(1 \vee y^{k-2}) e^{y/2} + y^{k-2} e^y \Big]. \end{aligned}$$

The second assertion follows by applying the first one twice. □

Lemma C.9. *There exist functions R_τ with $\sup_y |R_\tau(y)| = O(\sqrt{\tau})$ as $\tau \downarrow 0$, such that*

$$I_{-1/2}(y) = \Big(\frac{\pi}{\tau} + \sqrt{y^2/2} \int_1^{y^2/2} \frac{1}{v^{3/2}} e^v \, dv \Big) \big(1 + R_\tau(y)\big).$$

Furthermore, given $\varepsilon_\tau \to 0$ there exist functions S_τ with $\sup_{y \ge 1/\varepsilon_\tau} |S_\tau(y)| = O(\sqrt{\tau} + \varepsilon_\tau^2)$, such that, as $\tau \downarrow 0$,

$$I_{-1/2}(y) = \Big(\frac{\pi}{\tau} + \frac{e^{y^2/2}}{y^2/2} \Big) \big(1 + S_\tau(y)\big).$$

Proof. For the proof of the first assertion we separately consider the ranges $|y| \le 2\zeta_\tau$ and $|y| > 2\zeta_\tau$. For $|y| \le 2\zeta_\tau$ we split the integral in the definition of $I_{-1/2}$ over the intervals $(0, \tau)$, $(\tau, (2/y^2) \wedge 1)$ and $((2/y^2) \wedge 1, 1)$, where we consider the third interval empty if $y^2/2 \le 1$. Making the changes of coordinates $z = u\tau^2$ in the first integral, and $(y^2/2)z = v$ in the second and third integrals, we see that

$$I_{-1/2}(y) = \frac{1}{\tau} \int_0^{1/\tau} \frac{1}{\sqrt{u}} \frac{1}{1 + (1-\tau^2)u} e^{y^2 \tau^2 u/2} \, du$$

$$+\sqrt{y^2/2}\Big[\int_{y^2\tau/2}^{y^2/2\wedge 1}+\int_{y^2/2\wedge 1}^{y^2/2}\Big]\frac{1}{\sqrt{v}}\frac{1}{\tau^2y^2/2+(1-\tau^2)v}e^v\,dv$$

For $|y|\leq 2\zeta_\tau$, the exponential in the first integral tends to 1, uniformly in $u\leq 1/\tau$. Since $e^u-1\leq ue^u$, for $u\geq 0$, replacing it by 1 gives an error of at most

$$\frac{1}{\tau}\int_0^{1/\tau}\frac{1}{\sqrt{u}}\frac{e^{y^2\tau/2}y^2\tau^2u}{1+(1-\tau^2)u}\,du\lesssim\frac{1}{\tau}y^2\tau^{3/2}.$$

As $(1-\tau^2)(1+u)\leq 1+(1-\tau^2)u\leq 1+u$, dropping the factor $1-\tau^2$ from the denominator makes a multiplicative error of order $1+O(\tau^2)$. Since $\int_0^\infty u^{-1/2}/(1+u)\,du=\pi$ and $\int_{1/\tau}^\infty u^{-1/2}/(1+u)\,du\lesssim\tau^{1/2}$, the first term gives a contribution of $\pi/\tau+O(\tau^{-1/2})$, uniformly in $|y|\leq 2\zeta_\tau$. In the second integral we bound the factor $\tau^2y^2/2+(1-\tau^2)v$ below by $(1-\tau^2)v$, the exponential e^v above by e and the upper limit of the integral by 1, and next evaluate the integral to be bounded by a constant times $\tau^{-1/2}$. For the third integral we separately consider the cases that $y^2/2\leq 1$ and $y^2/2>1$. In the first case the third integral contributes nothing; the second term (the integral) in the assertion of the lemma is bounded and hence also contributes a negligible amount relative to π/τ. Finally consider the case that $y^2/2>1$. If in the third integral we replace $\tau^2y^2/2+(1-\tau^2)v$ by v, we obtain the second term in the assertion of the lemma. The difference is bounded above by

$$\sqrt{y^2/2}\int_1^{y^2/2}\frac{1}{\sqrt{v}}\frac{\tau^2v+\tau^2y^2}{v(\tau^2y^2/2+(1-\tau^2)v)}e^v\,dv$$
$$\lesssim\tau^2\sqrt{y^2/2}\int_1^{y^2/2}(v^{-3/2}+y^2v^{-5/2})e^v\,dv.$$

This is negligible relative to the integral in the assertion. This concludes the proof of the first assertion of the lemma for the range $|y|\leq 2\zeta_\tau$.

For $|y|$ in the interval $(2\zeta_\tau,\infty)$ we split the integral in the definition of $I_{-1/2}$ into the ranges $[0,1/3]$ and $(1/3,1]$. The contribution of the first range is bounded above by

$$\frac{1}{\tau^2}e^{y^2/6}\int_0^{1/3}z^{-1/2}\,dz\ll\sqrt{\tau}\frac{e^{y^2/2}}{y^2/2},$$

for $|y|\geq 2\zeta_\tau$. This is negligible relative to the integral in the assertion, which expands as $e^{y^2/2}/\sqrt{y^2/2}$, as claimed by the second assertion of the lemma. In the contribution of the second range we use that $z\leq\tau^2+(1-\tau^2)z\leq(1+2\tau^2)z$, for $z\geq 1/3$, and see that this is up to a multiplicative term of order $1+O(\tau^2)$ equal to

$$\int_{1/3}^1 z^{-3/2}e^{y^2z/2}\,dz=\sqrt{y^2/2}\Big[\int_1^{y^2/2}-\int_1^{y^2/6}\Big]v^{-3/2}e^v\,dv.$$

Applying Lemma C.8, we see that the contribution of the second integral is bounded above by a multiple of $(y^2/2)^{-1}e^{y^2/6}$, which is negligible relative to the first.

To prove the second assertion of the lemma we expand the integral in the first assertion with the help of Lemma C.8. ☐

Lemma C.10. *For $k>0$, there exist functions $R_{\tau,k}$ with $\sup_y|R_{\tau,k}(y)| = O(\tau^{2k/(k+1)})$, and for given $\varepsilon_\tau\to 0$ functions $S_{\tau,k}$ with $\sup_{y\geq 1/\varepsilon_\tau}|S_{\tau,k}(y)| = O(\tau^{2k/(2k+1)}+\varepsilon_\tau^2)$, such that, as $\tau\downarrow 0$,*

$$I_k(y) = \frac{1}{(y^2/2)^k}\int_0^{y^2/2} v^{k-1}e^v\,dv\big(1+R_{\tau,k}(y)\big) \lesssim \big(1\wedge y^{-2}\big)e^{y^2/2},$$
$$I_k(y) = \frac{e^{y^2/2}}{y^2/2}\big(1+S_{\tau,k}(y)\big).$$

There also exist functions $\bar R_\tau$ with $\sup_y|\bar R_\tau(y)| = O(\tau^{1/2})$ and $\bar S_\tau$ with $\sup_{y\geq 1/\varepsilon_\tau}|\bar S_\tau(y)| = O(\sqrt{\tau}+\varepsilon_\tau^2)$, such that, as $\tau\downarrow 0$ and $\varepsilon_\tau\to 0$,

$$I_{1/2}(y)-I_{3/2}(y) = \frac{1}{\sqrt{y^2/2}}\int_0^{y^2/2}\frac{1-2v/y^2}{\sqrt{v}}e^v\,dv\big(1+\bar R_\tau(y)\big) \lesssim (1\wedge y^{-4})e^{y^2/2},$$
$$I_{1/2}(y)-I_{3/2}(y) = \frac{e^{y^2/2}}{(y^2/2)^2}\big(1+\bar S_\tau(y)\big).$$

Proof. We split the integral in the definition of I_k over the intervals $[0,\tau^a]$ and $[\tau^a,1]$, for $a=2/(k+1)$. The contribution of the first integral is bounded above by

$$e^{\tau^a y^2/2}\int_0^{\tau^a}\frac{z^k}{(1-\tau^2)z}\,dz \lesssim e^{\tau^a y^2/2}\tau^{ka}.$$

In the second integral we use that $z\leq \tau^2+(1-\tau^2)z\leq(\tau^{2-a}+1-\tau^2)z$, for $z\geq\tau^a$, to see that the integral is $1+O(\tau^{2-a})$ times

$$\int_{\tau^a}^1\frac{z^k}{z}e^{y^2z/2}\,dz \gtrsim e^{\tau^a y^2/2}.$$

Combining these displays, we see that

$$I_k(y) = \int_{\tau^a}^1 z^{k-1}e^{y^2z/2}\,dz(1+O(\tau^{2-a})+O(\tau^{ka})).$$

This remains valid if we enlarge the range of integration to $[0,1]$. The change of coordinates $zy^2/2=v$ completes the proof of the equality in the first assertion.

For the second assertion we expand the integral in the first assertion with the help of the second assertion of Lemma C.8. Note here that for $k>-1$ the integrals in the latter lemma can be taken over $(0,y)$ instead of $(1,y)$, since the difference is a constant.

The inequality in the first assertion is valid for $y \to \infty$, in view of the second assertion, and from the fact that $G(y) := (y^2/2)^{-k} \int_0^{y^2/2} v^{k-1} e^v \, dv$ possesses a finite limit as $y \downarrow 0$ it follows that it is also valid for $y \to 0$. For intermediate y the inequality follows since the continuous function $y \mapsto G(y) e^{-y^2/2}/(y^{-2} \wedge 1)$ is bounded on compacta in $(0, \infty)$.

For the proofs of the assertions concerning $I_{1/2} - I_{3/2}$ we write

$$I_{1/2}(y) - I_{3/2}(y) = \Bigl(\int_0^\tau + \int_\tau^1\Bigr) \frac{\sqrt{z}(1-z)}{\tau^2 + (1-\tau^2)z} e^{y^2 z/2} \, dz.$$

Next we follow the same approach as previously. □

Lemma C.11. *For any stochastic process $(V_\tau : \tau > 0)$ with continuously differentiable sample paths $\tau \mapsto V_\tau$, with derivative written as $\dot{V}_\tau$,*

$$\mathrm{E}(V_{\tau_2} - V_{\tau_1})^2 \le (\tau_2 - \tau_1)^2 \sup_{\tau \in [\tau_1, \tau_2]} \mathrm{E}\dot{V}_\tau^2.$$

Proof. By the Newton-Leibniz formula, the Cauchy-Schwarz inequality, Fubini's theorem and the mean integrated value theorem, for $\tau_2 \ge \tau_1$,

$$\begin{aligned} \mathrm{E}\bigl(V_{\tau_1} - V_{\tau_2}\bigr)^2 &= \mathrm{E}\Bigl(\int_{\tau_1}^{\tau_2} \dot{V}_\tau \, d\tau\Bigr)^2 \le \mathrm{E}(\tau_2 - \tau_1) \int_{\tau_1}^{\tau_2} \dot{V}_\tau^2 \, d\tau \\ &= (\tau_2 - \tau_1) \int_{\tau_1}^{\tau_2} \mathrm{E}\dot{V}_\tau \, d\tau \le (\tau_2 - \tau_1)^2 \sup_{\tau \in [\tau_1, \tau_2]} \mathrm{E}\dot{V}_\tau^2 \, d\tau. \end{aligned}$$

□

Lemma C.12. *If $f_1, f_2 : [0, \infty) \to [0, \infty)$ are probability densities such that f_2/f_1 is monotonely increasing, then, for any monotonely increasing function h,*

$$\mathrm{E}_{f_1} h(X) \le \mathrm{E}_{f_2} h(X).$$

Proof. Define $g = f_2/f_1$. Since $\int_0^\infty f_1(x) dx = \int_0^\infty f_1(x) g(x) \, dx$ and g is monotonely increasing, there exists an $x_0 > 0$ such that $g(x) \le 1$ for $x < x_0$ and $g(x) \ge 1$ for $x > x_0$. Therefore

$$\begin{aligned} 0 &= h(x_0) \int_0^\infty f_1(x)\bigl(g(x) - 1\bigr) \, dx \\ &\le \int_0^{x_0} f_1(x) h(x)\bigl(g(x) - 1\bigr) \, dx + \int_{x_0}^\infty f_1(x) h(x)\bigl(g(x) - 1\bigr) \, dx. \end{aligned}$$

By the definition of g the right side is $\mathrm{E}_{f_2} h(X) - \mathrm{E}_{f_1} h(X)$. □

Appendix D: Additional simulation results

We give more details on the simulation results of Section 4. Per scenario ($p_n = 20$ or $p_n = 200$) we give the MSE and standard deviation (in parentheses) for the four methods described in Section 4, as well as for hierarchical Bayes with the uniform distribution. In addition, we give the average value of τ per method.

TABLE 1
$p_n = 20$, overall MSE and standard deviation over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.05 (0.00)	0.05 (0.00)	0.05 (0.00)	0.05 (0.01)	0.05 (0.00)
2	0.19 (0.01)	0.17 (0.02)	0.16 (0.02)	0.16 (0.02)	0.16 (0.02)
3	0.29 (0.04)	0.21 (0.04)	0.21 (0.04)	0.21 (0.04)	0.21 (0.04)
4	0.24 (0.06)	0.18 (0.04)	0.18 (0.04)	0.18 (0.04)	0.18 (0.04)
5	0.14 (0.05)	0.14 (0.04)	0.14 (0.04)	0.14 (0.04)	0.14 (0.04)
6	0.08 (0.03)	0.12 (0.03)	0.12 (0.03)	0.12 (0.03)	0.12 (0.03)
7	0.07 (0.02)	0.11 (0.03)	0.11 (0.03)	0.11 (0.03)	0.11 (0.03)
8	0.07 (0.02)	0.11 (0.02)	0.11 (0.02)	0.11 (0.02)	0.11 (0.03)
9	0.06 (0.02)	0.10 (0.02)	0.11 (0.02)	0.11 (0.02)	0.11 (0.02)
10	0.06 (0.02)	0.10 (0.02)	0.10 (0.02)	0.10 (0.02)	0.11 (0.02)

TABLE 2
$p_n = 20$, MSE and standard deviation of the nonzero parameters over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.99 (0.03)	0.98 (0.04)	0.94 (0.05)	0.94 (0.06)	0.94 (0.06)
2	3.70 (0.23)	3.20 (0.43)	3.06 (0.35)	3.05 (0.35)	3.06 (0.35)
3	5.84 (0.88)	3.79 (0.79)	3.75 (0.76)	3.75 (0.77)	3.74 (0.76)
4	4.60 (1.22)	2.87 (0.84)	2.85 (0.84)	2.85 (0.83)	2.85 (0.84)
5	2.54 (0.97)	1.87 (0.66)	1.87 (0.66)	1.87 (0.66)	1.87 (0.66)
6	1.51 (0.59)	1.38 (0.47)	1.38 (0.47)	1.38 (0.47)	1.38 (0.47)
7	1.23 (0.39)	1.21 (0.37)	1.21 (0.37)	1.21 (0.37)	1.21 (0.37)
8	1.15 (0.34)	1.15 (0.34)	1.15 (0.34)	1.15 (0.34)	1.15 (0.34)
9	1.11 (0.33)	1.11 (0.33)	1.11 (0.33)	1.11 (0.33)	1.11 (0.33)
10	1.08 (0.32)	1.08 (0.32)	1.08 (0.32)	1.08 (0.32)	1.08 (0.32)

TABLE 3
$p_n = 20$, MSE and standard deviation of the parameters equal to zero over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
2	0.00 (0.00)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
3	0.00 (0.00)	0.03 (0.01)	0.03 (0.01)	0.03 (0.01)	0.03 (0.01)
4	0.01 (0.01)	0.04 (0.02)	0.04 (0.02)	0.04 (0.02)	0.04 (0.02)
5	0.01 (0.01)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
6	0.01 (0.01)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
7	0.01 (0.01)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
8	0.01 (0.01)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
9	0.01 (0.01)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
10	0.01 (0.01)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)

Table 4

$p_n = 20$, mean value of τ and standard deviation over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.00 (0.00)	0.00 (0.00)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
2	0.00 (0.00)	0.03 (0.02)	0.05 (0.02)	0.05 (0.02)	0.05 (0.02)
3	0.02 (0.00)	0.11 (0.02)	0.12 (0.02)	0.12 (0.02)	0.12 (0.02)
4	0.04 (0.00)	0.17 (0.02)	0.18 (0.02)	0.18 (0.02)	0.18 (0.02)
5	0.05 (0.00)	0.20 (0.02)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)
6	0.05 (0.00)	0.20 (0.02)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)
7	0.05 (0.00)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)	0.22 (0.02)
8	0.05 (0.00)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)	0.22 (0.02)
9	0.05 (0.00)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)	0.22 (0.02)
10	0.05 (0.00)	0.21 (0.02)	0.21 (0.02)	0.21 (0.02)	0.22 (0.02)

Table 5

$p_n = 200$, overall MSE and standard deviation over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.49 (0.01)	0.42 (0.02)	0.41 (0.02)	0.41 (0.02)	0.41 (0.02)
2	1.59 (0.07)	0.86 (0.05)	0.86 (0.05)	0.86 (0.05)	0.86 (0.05)
3	1.71 (0.14)	1.01 (0.07)	0.95 (0.07)	1.02 (0.07)	1.02 (0.07)
4	1.21 (0.12)	0.97 (0.08)	0.89 (0.07)	0.97 (0.08)	0.97 (0.08)
5	0.89 (0.08)	0.84 (0.07)	0.83 (0.07)	0.84 (0.07)	0.84 (0.07)
6	0.74 (0.07)	0.76 (0.06)	0.80 (0.06)	0.76 (0.06)	0.76 (0.06)
7	0.68 (0.06)	0.72 (0.06)	0.79 (0.06)	0.72 (0.06)	0.72 (0.06)
8	0.64 (0.05)	0.69 (0.06)	0.79 (0.06)	0.69 (0.06)	0.79 (0.06)
9	0.63 (0.05)	0.67 (0.05)	0.79 (0.06)	0.67 (0.05)	0.67 (0.05)
10	0.61 (0.05)	0.66 (0.05)	0.79 (0.06)	0.66 (0.05)	0.66 (0.05)

Table 6

$p_n = 200$, MSE and standard deviation of the nonzero parameters over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.98 (0.01)	0.81 (0.05)	0.81 (0.04)	0.81 (0.04)	0.80 (0.04)
2	3.17 (0.14)	1.55 (0.11)	1.55 (0.11)	1.55 (0.11)	1.55 (0.11)
3	3.39 (0.27)	1.80 (0.14)	1.61 (0.14)	1.81 (0.14)	1.81 (0.14)
4	2.33 (0.24)	1.70 (0.15)	1.40 (0.13)	1.71 (0.15)	1.71 (0.15)
5	1.65 (0.16)	1.46 (0.14)	1.24 (0.11)	1.46 (0.14)	1.46 (0.14)
6	1.35 (0.13)	1.29 (0.12)	1.15 (0.10)	1.29 (0.12)	1.29 (0.12)
7	1.22 (0.11)	1.20 (0.11)	1.11 (0.10)	1.20 (0.11)	1.20 (0.11)
8	1.16 (0.10)	1.15 (0.10)	1.09 (0.09)	1.15 (0.10)	1.15 (0.10)
9	1.12 (0.10)	1.12 (0.10)	1.07 (0.09)	1.12 (0.10)	1.12 (0.10)
10	1.10 (0.09)	1.10 (0.09)	1.06 (0.09)	1.10 (0.09)	1.10 (0.09)

TABLE 7
$p_n = 200$, MSE and standard deviation of the parameters equal to zero over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.00 (0.00)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
2	0.01 (0.01)	0.17 (0.04)	0.17 (0.04)	0.17 (0.04)	017 (0.04)
3	0.04 (0.02)	0.23 (0.04)	0.30 (0.05)	0.23 (0.04)	0.23 (0.04)
4	0.09 (0.03)	0.23 (0.04)	0.37 (0.06)	0.23 (0.04)	0.23 (0.04)
5	0.12 (0.03)	0.23 (0.04)	0.42 (0.06)	0.23 (0.04)	0.23 (0.04)
6	0.13 (0.03)	0.23 (0.04)	0.45 (0.06)	0.23 (0.04)	0.23 (0.04)
7	0.13 (0.03)	0.23 (0.04)	0.47 (0.07)	0.23 (0.04)	0.23 (0.04)
8	0.13 (0.03)	0.23 (0.04)	0.49 (0.07)	0.23 (0.04)	0.23 (0.04)
9	0.13 (0.03)	0.23 (0.04)	0.51 (0.07)	0.23 (0.04)	0.23 (0.04)
10	0.13 (0.03)	0.23 (0.04)	0.52 (0.07)	0.23 (0.04)	0.23 (0.04)

TABLE 8
$p_n = 200$, mean value of τ and standard deviation over 100 simulation repetitions.

A	Simple estimator	MMLE	Cauchy	Truncated Cauchy	Uniform
1	0.00 (0.00)	0.11 (0.04)	0.12 (0.03)	0.12 (0.03)	0.12 (0.03)
2	0.03 (0.01)	0.68 (0.05)	0.69 (0.05)	0.69 (0.05)	0.69 (0.05)
3	0.16 (0.02)	1.00 (0.00)	1.42 (0.06)	0.97 (0.00)	0.98 (0.00)
4	0.35 (0.02)	1.00 (0.00)	2.11 (0.05)	0.99 (0.00)	0.99 (0.00)
5	0.47 (0.01)	1.00 (0.00)	2.68 (0.05)	0.99 (0.00)	0.99 (0.00)
6	0.50 (0.00)	1.00 (0.00)	3.17 (0.05)	0.99 (0.00)	0.99 (0.00)
7	0.50 (0.00)	1.00 (0.00)	3.63 (0.05)	0.99 (0.00)	0.99 (0.00)
8	0.50 (0.00)	1.00 (0.00)	4.05 (0.05)	0.99 (0.00)	0.99 (0.00)
9	0.50 (0.00)	1.00 (0.00)	4.46 (0.06)	0.99 (0.00)	0.99 (0.00)
10	0.50 (0.00)	1.00 (0.00)	4.85 (0.06)	0.99 (0.00)	0.99 (0.00)

References

[1] ARMAGAN, A., DUNSON, D. B., AND LEE, J. Generalized double Pareto shrinkage. *Statistica Sinica 23* (2013), 119–143. MR3076161

[2] BHADRA, A., DATTA, J., POLSON, N. G., AND WILLARD, B. The horseshoe+ estimator of ultra-sparse signals. To appear in Bayesian Analysis., 2015.

[3] BHATTACHARYA, A., PATI, D., PILLAI, N. S., AND DUNSON, D. B. Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association 110*, 512 (2015), 1479–1490. PMID: 27019543. MR3449048

[4] BICKEL, P. J., RITOV, Y., AND TSYBAKOV, A. B. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics 37*, 4 (2009), 1705–1732. MR2533469

[5] Bogdan, M., Chakrabarti, A., Frommlet, F., and Ghosh, J. K. Asymptotic bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist. 39*, 3 (06 2011), 1551–1579. MR2850212

[6] Caron, F., and Doucet, A. Sparse Bayesian nonparametric regression. In *Proceedings of the 25th International Conference on Machine Learning* (New York, NY, USA, 2008), ICML '08, ACM, pp. 88–95.

[7] Carvalho, C. M., Polson, N. G., and Scott, J. G. Handling sparsity via the horseshoe. *Journal of Machine Learning Research, W&CP 5* (2009), 73–80.

[8] Carvalho, C. M., Polson, N. G., and Scott, J. G. The horseshoe estimator for sparse signals. *Biometrika 97*, 2 (2010), 465–480. MR2650751

[9] Castillo, I., Schmidt-Hieber, J., and van der Vaart, A. Bayesian linear regression with sparse priors. *Ann. Statist. 43*, 5 (10 2015), 1986–2018. MR3375874

[10] Castillo, I., and Van der Vaart, A. W. Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist. 40*, 4 (2012), 2069–2101. MR3059077

[11] Datta, J., and Ghosh, J. K. Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Analysis 8*, 1 (2013), 111–132. MR3036256

[12] Donoho, D. L., Johnstone, I. M., Hoch, J. C., and Stern, A. S. Maximum entropy and the nearly black object (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological) 54*, 1 (1992), 41–81. MR1157714

[13] Ghosal, S., Ghosh, J. K., and Van der Vaart, A. W. Convergence rates of posterior distributions. *The Annals of Statistics 28*, 2 (2000), 500–531. MR1790007

[14] Ghosal, S., Lember, J., and van der Vaart, A. Nonparametric Bayesian model selection and averaging. *Electron. J. Stat. 2* (2008), 63–89. MR2386086

[15] Ghosh, P., and Chakrabarti, A. Posterior concentration properties of a general class of shrinkage estimators around nearly black vectors. arXiv:1412.8161v2, 2015.

[16] Gramacy, R. B. *monomvn: Estimation for multivariate normal and Student-t data with monotone missingness*, 2014. R package version 1.9-5.

[17] Griffin, J. E., and Brown, P. J. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis 5*, 1 (2010), 171–188. MR2596440

[18] Hahn, R. P., He, J., and Lopes, H. *fastHorseshoe: The Elliptical Slice Sampler for Bayesian Horseshoe Regression*, 2016. R package version 0.1.0.

[19] Jiang, W., and Zhang, C.-H. General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist. 37*, 4 (08 2009), 1647–1684. MR2533467

[20] Johnson, V. E., and Rossell, D. On the use of non-local prior densities in Bayesian hypothesis tests. *J. R. Stat. Soc. Ser. B Stat. Methodol. 72*, 2 (2010), 143–170. MR2830762

[21] Johnstone, I. M., and Silverman, B. W. Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist. 32*, 4 (2004), 1594–1649. MR2089135

[22] Makalic, E., and Schmidt, D. F. A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*, 1 (Jan 2016), 179–182.

[23] Piironen, J., and Vehtari, A. On the Hyperprior Choice for the Global Shrinkage Parameter in the Horseshoe Prior. *ArXiv e-prints* (Oct. 2016).

[24] Polson, N. G., and Scott, J. G. Shrink globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian Statistics 9*, J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, Eds. Oxford University Press, 2010. MR3204017

[25] Polson, N. G., and Scott, J. G. Good, great or lucky? Screening for firms with sustained superior performance using heavy-tailed priors. *The Annals of Applied Statistics 6*, 1 (2012), 161–185. MR2951533

[26] Polson, N. G., and Scott, J. G. On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis 7*, 4 (2012), 887–902. MR3000018

[27] Ročková, V. Bayesian estimation of sparse signals with a continuous spike-and-slab prior. submitted manuscript, available at `http://stat.wharton.upenn.edu/~vrockova/rockova2015.pdf`, 2015.

[28] Rousseau, J., and Szabo, B. Asymptotic behaviour of the empirical bayes posteriors associated to maximum marginal likelihood estimator. *Ann. Statist. 45*, 2 (04 2017), 833–865. MR3650402

[29] Scott, J. G. Parameter expansion in local-shrinkage models. arXiv:1010.5265, 2010.

[30] Scott, J. G. Bayesian estimation of intensity surfaces on the sphere via needlet shrinkage and selection. *Bayesian Analysis 6*, 2 (2011), 307–328. MR2806246

[31] Szabo, B. T., van der Vaart, A. W., and van Zanten, J. Empirical Bayes scaling of Gaussian priors in the white noise model. *Electron. J. Statist. 7* (2013), 991–1018. MR3044507

[32] Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol. 58*, 1 (1996), 267–288. MR1379242

[33] van der Pas, S., Scott, J., Chakraborty, A., and Bhattacharya, A. *horseshoe: Implementation of the Horseshoe Prior*, 2016. R package version 0.1.0.

[34] van der Pas, S., Szabó, B., and van der Vaart, A. Uncertainty quantification for the horseshoe. *ArXiv e-prints* (July 2016).

[35] van der Pas, S. L., Kleijn, B. J. K., and van der Vaart, A. W. The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron. J. Statist. 8*, 2 (2014), 2585–2618. MR3285877

[36] van der Vaart, A., and van Zanten, H. Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist. 37*, 5B (2009), 2655–2675. MR2541442

[37] van der Vaart, A. W., and Wellner, J. A. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics. MR1385671