Electronic Journal of Statistics
Vol. 8 (2014) 2585–2618
ISSN: 1935-7524
DOI: 10.1214/14-EJS962

The horseshoe estimator: Posterior concentration around nearly black vectors

S. L. van der Pas*

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

B. J. K. Kleijn

Korteweg-de Vries Institute for Mathematics, University of Amsterdam
e-mail: b.kleijn@uva.nl

and

A. W. van der Vaart†

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

Abstract: We consider the horseshoe estimator due to Carvalho, Polson and Scott (2010) for the multivariate normal mean model in the situation that the mean vector is sparse in the nearly black sense. We assume the frequentist framework where the data is generated according to a fixed mean vector. We show that if the number of nonzero parameters of the mean vector is known, the horseshoe estimator attains the minimax ℓ_2 risk, possibly up to a multiplicative constant. We provide conditions under which the horseshoe estimator combined with an empirical Bayes estimate of the number of nonzero means still yields the minimax risk. We furthermore prove an upper bound on the rate of contraction of the posterior distribution around the horseshoe estimator, and a lower bound on the posterior variance. These bounds indicate that the posterior distribution of the horseshoe prior may be more informative than that of other one-component priors, including the Lasso.

MSC 2010 subject classifications: 62F15, 62F10.

Keywords and phrases: Sparsity, horseshoe prior, worst case risk, Bayesian inference, empirical Bayes, posterior contraction, normal means model.

Received February 2014.

*Research supported by Netherlands Organization for Scientific Research NWO.

†The research leading to these results has received funding from the European Research Council under ERC Grant Agreement 320637.

1. Introduction

We consider the normal means problem, where we observe a vector $Y \in \mathbb{R}^n$, $Y = (Y_1, \ldots, Y_n)$, such that

$$Y_i = \theta_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

for independent normal random variables ε_i with mean zero and variance σ^2. The vector $\theta = (\theta_1, \ldots, \theta_n)$ is assumed to be sparse, in the 'nearly black' sense that the number of nonzero means

$$p_n := \#\{i : \theta_i \neq 0\}$$

is $o(n)$ as $n \to \infty$. A natural Bayesian approach to recovering θ would be to induce sparsity through a 'spike and slab' prior (Mitchell and Beauchamp, 1988), which consists of a mixture of a Dirac measure at zero and a (heavy-tailed) continuous distribution. Johnstone and Silverman (2004) analyzed an empirical Bayes version of this approach, where the mixing weight is obtained by marginal maximum likelihood. In the frequentist set-up that the data is generated according to a fixed mean vector, they showed that the empirical Bayes coordinatewise posterior median attains the minimax rate, in ℓ_q norm, $q \in (0, 2]$, for mean vectors that are either nearly black or of bounded ℓ_p norm, $p \in (0, 2]$. Castillo and Van der Vaart (2012) analyzed a fully Bayesian version, where the proportion of nonzero coefficients is modelled by a prior distribution. They identified combinations of priors on this proportion and on the nonzero coefficients (the 'slab') that yield posterior distributions concentrating around the underlying mean vector at the minimax rate in ℓ_q norm, $q \in (0, 2]$, for mean vectors that are nearly black, and in ℓ_q norm, $q \in (0, 2)$ for mean vectors of bounded weak ℓ_p norm, $p \in (0, q)$. Other work on empirical Bayes approaches to the two-group model includes (Efron, 2008; Jiang and Zhang, 2009; Yuan and Lin, 2005).

As a full Bayesian approach with a mixture of a Dirac and a continuous component may require exploration of a model space of size 2^n, implementation on large datasets is currently impractical, although Castillo and Van der Vaart (2012) present an algorithm which can compute several aspects of the posterior in polynomial time, provided sufficient memory can be allocated. Several authors, including (Armagan, Dunson and Lee, 2013; Griffin and Brown, 2010), have proposed one-component priors, which model the spike at zero by a peak in the prior density at this point. For most of these proposals, theoretical justification in terms of minimax risk rates or posterior contraction rates is lacking. The Lasso estimator (Tibshirani, 1996), which arises as the MAP estimator after placing a Laplace prior with common parameter on each θ_i, is an exception. It attains close to the minimax risk rate in ℓ_q, $q \in [1, 2]$ (Bickel, Ritov and Tsybakov (2009)). It has however been recently shown that the corresponding full posterior distribution contracts at a much slower rate than the mode (Castillo, Schmidt-Hieber and Van der Vaart, 2014). This is undesirable, because this implies that the posterior distribution cannot provide an adequate measure of uncertainty in the estimate.

In general one would use a posterior distribution both for recovery and for uncertainty quantification. For the first, a measure of centre, such as a median or mode, suffices. For the second, one typically employs a credible set, which is defined as a central set of prescribed posterior probability. For realistic uncertainty quantification it is necessary that the posterior contracts to its center at the same rate as the posterior median or mode approaches the true parameter.

In this paper we study the posterior distribution resulting from the horseshoe prior, which is a one-component prior, introduced in (Carvalho, Polson and Scott, 2010, 2009) and expanded upon in (Scott, 2011; Polson and Scott, 2012a,b). It combines a pole at zero with Cauchy-like tails. The corresponding estimator does not face the computational issues of the point mass mixture models. Carvalho, Polson and Scott (2010) already showed good behaviour of the horseshoe estimator in terms of Kullback-Leibler risk when the true mean is zero. Datta and Ghosh (2013) proved some optimality properties of a multiple testing rule induced by the horseshoe estimator. In this paper, we prove that the horseshoe estimator achieves the minimax quadratic risk, possibly up to a multiplicative constant. We furthermore prove that the posterior variance is of the order of the minimax risk, and thus the posterior contracts at the minimax rate around the underlying mean vector. These results are proven under the assumption that the number p_n of nonzero parameters is known. However, we also provide conditions under which the horseshoe estimator combined with an empirical Bayes estimator still attains the minimax rate, when p_n is unknown.

This paper is organized as follows. In Section 2, the horseshoe prior is described and a summary of simulation results is given. The main results, that the horseshoe estimator attains the minimax squared error risk (up to a multiplicative constant) and that the posterior distribution contracts around the truth at the minimax rate, are stated in Section 3. Conditions on an empirical Bayes estimator of the key parameter τ such that the minimax ℓ_2 risk will still be obtained are given in Section 4. The behaviour of such an empirical Bayes estimate is compared to a full Bayesian version in a numerical study in Section 5. Section 6 contains some concluding remarks. The proofs of the main results and supporting lemmas are in the appendix.

1.1. Notation

We write $A_n \asymp B_n$ to denote $0 < \lim_{n\to\infty} \inf \frac{A_n}{B_n} \leq \lim_{n\to\infty} \sup \frac{A_n}{B_n} < \infty$ and $A_n \lesssim B_n$ to denote that there exists a positive constant c independent of n such that $A_n \leq cB_n$. $A \vee B$ is the maximum of A and B, and $A \wedge B$ the minimum of A and B. The standard normal density and cumulative distribution are denoted by ϕ and Φ and we set $\Phi^c = 1 - \Phi$. The norm $\|\cdot\|$ will be the ℓ_2 norm and the class of nearly black vectors will be denoted by $\ell_0[p_n] := \{\theta \in \mathbb{R}^n : \#(1 \leq i \leq n : \theta_i \neq 0) \leq p_n\}$.

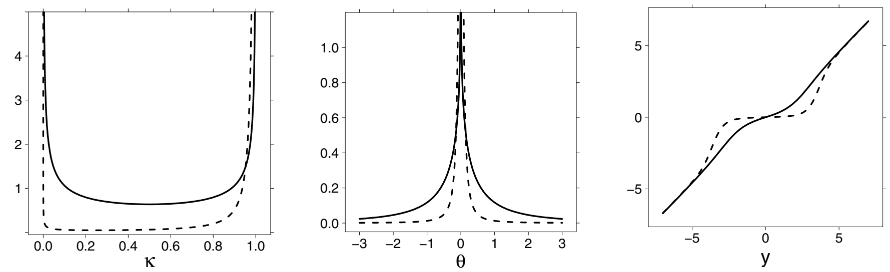

Fig 1. *The effect of decreasing τ on the priors on κ (left) and θ (middle) and the posterior mean $T_\tau(y)$ (right). The solid line corresponds to $\tau = 1$, the dashed line to $\tau = 0.05$. Decreasing τ results in a higher prior probability of shrinking the observations towards zero.*

2. The horseshoe prior

In this section, we give an overview of some known properties of the horseshoe estimator which will be relevant to the remainder of our discussion. The horseshoe prior for a parameter θ modelling an observation $Y \sim \mathcal{N}(\theta, \sigma^2 I_n)$ is defined hierarchically (Carvalho, Polson and Scott, 2010):

$$\theta_i \mid \lambda_i, \tau \sim \mathcal{N}(0, \sigma^2\tau^2\lambda_i^2), \quad \lambda_i \sim C^+(0,1),$$

for $i = 1, \ldots, n$, where $C^+(0,1)$ is a standard half-Cauchy distribution. The parameter τ is assumed to be fixed in this paper, rendering the θ_i independent *a priori*. The corresponding density p_τ increases logarithmically around zero, while its tails decay quadratically. The posterior density of θ_i given λ_i and τ is normal with mean $(1-\kappa_i)y_i$, where $\kappa_i = \frac{1}{1+\tau^2\lambda_i^2}$. Hence, by Fubini's theorem:

$$\mathbb{E}[\theta_i \mid y_i, \tau] = (1 - \mathbb{E}[\kappa_i \mid y_i, \tau])y_i.$$

The posterior mean $\mathbb{E}[\theta \mid y, \tau]$ will be referred to as the horseshoe estimator and denoted by $T_\tau(y)$. The horseshoe prior takes its name from the prior on κ_i, which is given by:

$$p_\tau(\kappa_i) = \frac{\tau}{\pi}\frac{1}{1-(1-\tau^2)\kappa_i}(1-\kappa_i)^{-\frac{1}{2}}\kappa_i^{-\frac{1}{2}}.$$

If $\tau = 1$, this reduces to a Be$(\frac{1}{2}, \frac{1}{2})$ distribution, which looks like a horseshoe. As illustrated in Figure 1, decreasing τ skews the prior distribution on κ_i towards one, corresponding to more mass near zero in the prior on θ_i and a stronger shrinkage effect in $T_\tau(y)$.

The posterior mean can be expressed as:

$$T_\tau(y_i) = y_i\left(1 - \frac{2\Phi_1\left(\frac{1}{2}, 1, \frac{5}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}{3\Phi_1\left(\frac{1}{2}, 1, \frac{3}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}\right) = y_i\frac{\int_0^1 z^{\frac{1}{2}}\frac{1}{\tau^2+(1-\tau^2)z}e^{\frac{y_i^2}{2\sigma^2}z}dz}{\int_0^1 z^{-\frac{1}{2}}\frac{1}{\tau^2+(1-\tau^2)z}e^{\frac{y_i^2}{2\sigma^2}z}dz}, \tag{1}$$

where $\Phi_1(\alpha, \beta, \gamma; x, y)$ denotes the degenerate hypergeometric function of two variables (Gradshteyn and Ryzhik, 1965).

An unanswered question so far has been how τ should be chosen. Intuitively, τ should be small if the mean vector is very sparse, as the horseshoe prior will then place more of its mass near zero. By approximating the posterior distribution of τ^2 given $\kappa = (\kappa_1, \dots, \kappa_n)$ in case a prior on τ is used, Carvalho, Polson and Scott (2010) show that if most observations are shrunk near zero, τ will be very small with high probability. They suggest a half-Cauchy prior on τ. Datta and Ghosh (2013) implemented this prior on τ and their plots of posterior draws for τ at various sparsity levels indicate the expected relationship between τ and the sparsity level: the posterior distribution of τ tends to concentrate around smaller values when the underlying mean vector is sparser. As will be discussed further in the next section, the value $\tau = \frac{p_n}{n}$ (up to a log factor) is optimal in terms of mean square error and posterior contraction rates.

In case τ is estimated empirically, as will be considered in Section 4, the horseshoe estimator can be computed by plugging this estimate into expression (1), thereby avoiding the use of MCMC. Other aspects of the posterior, such as the posterior variance, can be computed using such a plug-in procedure as well. Polson and Scott (2012a) and Polson and Scott (2012b) consider computation of the horseshoe estimator based on the representation in terms of degenerate hypergeometric functions, as these can be efficiently computed using converging series of confluent hypergeometric functions. They report unproblematic computations for τ^2 between $\frac{1}{1000}$ and 1000. A second option is to apply a quadrature routine to the integral representation in (1). As the continuity and symmetry of $T_\tau(y)$ in y can be taken advantage of when computing the horseshoe estimator for a large number of observations, the complexity of these computations mostly depends on the value of τ. Both approaches will be slower for smaller values of τ. Hence, if we use the (estimated) sparsity level $\frac{p_n}{n}$ (up to a log factor) for τ, the computation of the horseshoe estimator will be slower if there are fewer nonzero parameters. As noted by Scott (2010), problems arise in Gibbs sampling precisely when τ is small as well. Hence care needs to be taken with any computational approach if $\frac{p_n}{n}$ is suspected to be very close to zero.

The performance of the horseshoe prior, with additional priors on τ and σ^2, in various simulation studies has been very promising. Carvalho, Polson and Scott (2010) simulated sparse data where the nonzero components were drawn from a Student-t density and found that the horseshoe estimator systematically beat the MLE, the double-exponential (DE) and normal-exponential-gamma (NEG) priors, and the empirical Bayes model due to Johnstone and Silverman (2004) in terms of square error loss. Only when the signal was neither sparse nor heavy-tailed did the MLE, DE and NEG priors have an edge over the horseshoe estimator. In similar experiments in (Carvalho, Polson and Scott, 2009; Polson and Scott, 2012a) the horseshoe prior outperformed the DE prior, while behaving similarly to a heavy-tailed discrete mixture. In a wavelet-denoising experiment under several noise levels and loss functions, the horseshoe estimator compared favorably to the discrete wavelet transform and the empirical Bayes model (Polson and Scott, 2010). Bhattacharya et al. (2012) applied several shrinkage priors to data with the underlying mean vector consisting of zeroes and fixed nonzero values and found the posterior median of the horseshoe prior performing better in terms of squared error than the Bayesian Lasso (BL), the Lasso, the posterior median of a point mass mixture prior as in (Castillo and Van der Vaart, 2012) and the empirical Bayes model proposed by Johnstone and Silverman (2004), and comparable to their proposed Dirichlet-Laplace (DL) prior with parameter $\frac{1}{n}$. Results in (Armagan, Dunson and Lee, 2013) are similar. In a second simulation setting, Bhattacharya et al. (2012) generated data of length $n = 1000$, with the first ten means equal to 10, the next 90 equal to a number $A \in \{2, \dots, 7\}$ and the remainder equal to zero. In this simulation, the horseshoe prior beat the BL (except when $A = 2$) and the DL prior with parameter $\frac{1}{n}$ (except when $A = 7$), while performing similarly to the DL prior with parameter $\frac{1}{2}$. It is worthy of note that Koenker (2014) generated data according to the same scheme and applied the empirical Bayes procedures due to Martin and Walker (2014) (EBMW) and Koenker and Mizera (2014) (EBKM) to it. The MSE of EBMW was lower than that of the horseshoe prior for $A \in \{5, 6, 7\}$, while that of EBKM was much lower in all cases.

3. Mean square error and bounds on the posterior variance

In this section, we study the mean square error of the horseshoe estimator, and the spread of the posterior distribution, under the assumption that the number of nonzero parameters p_n is known. Theorem 3.1 provides an upper bound on the mean square error, and shows that for a range of choices of the global parameter τ, the horseshoe estimator attains the minimax ℓ_2 risk, possibly up to a multiplicative constant. Theorem 3.3 states upper bounds on the rate of contraction of the posterior distribution around the underlying mean vector and around the horseshoe estimator, again for a range of values of τ. These upper bounds are equal, up to a multiplicative constant, to the minimax risk. The contraction rate around the truth is sharp, but this may not be the case for the rate of contraction around the horseshoe estimator. Theorems 3.4 and 3.5 provide more insight into the spread of the posterior distribution for various values of τ and indicate that $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$ is a good choice.

Theorem 3.1. *Suppose $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$. Then the estimator $T_\tau(y)$ satisfies*

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \|T_\tau(Y) - \theta_0\|^2 \lesssim p_n \log \frac{1}{\tau} + (n - p_n)\tau \sqrt{\log \frac{1}{\tau}} \tag{2}$$

for $\tau \to 0$, as $n, p_n \to \infty$ and $p_n = o(n)$.

By the minimax risk result in (Donoho et al., 1992), we also have a lower bound:

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \|T_\tau(Y) - \theta_0\|^2 \geq 2\sigma^2 p_n \log \frac{n}{p_n}(1 + o(1)),$$

as $n, p_n \to \infty$ and $p_n = o(n)$. The choice $\tau = (\frac{p_n}{n})^\alpha$, for $\alpha \geq 1$, leads to an upper bound (2) of order $p_n \log(n/p_n)$, with (as can be seen from the proof) a multiplicative constant of at most $4\alpha\sigma^2$. Thus, for this choice of τ, we have:

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \|T_\tau(Y) - \theta_0\|^2 \asymp p_n \log \frac{n}{p_n}.$$

The horseshoe estimator therefore performs well as a point estimator, as it attains the minimax risk (possibly up to a multiplicative constant of at most 2 for $\alpha = 1$). This may seem surprising, as the prior does not include a point mass at zero to account for the assumed sparsity in the underlying mean vector. Theorem 3.1 shows that the pole at zero of the horseshoe prior mimics the point mass well enough, while the heavy tails ensure that large observations are not shrunk too much.

An upper bound on the rate of contraction of the posterior can be obtained through an upper bound on the posterior variance. The posterior variance can be expressed as:

$$\operatorname{Var}(\theta_i \mid y_i) = \frac{\sigma^2}{y_i} T_\tau(y_i) - (T_\tau(y_i) - y_i)^2 + y_i^2 \frac{8\Phi_1\left(\frac{1}{2}, 1, \frac{7}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}{15\Phi_1\left(\frac{1}{2}, 1, \frac{3}{2}; \frac{y_i^2}{2\sigma^2}, 1 - \frac{1}{\tau^2}\right)}.$$

Details on the computation can be found in Lemma A.4. Using a similar approach as when bounding the ℓ_2 risk, we can find an upper bound on the expected value of the posterior variance.

Theorem 3.2. *Suppose $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$. Then the variance of the posterior distribution corresponding to the horseshoe prior satisfies*

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^n \operatorname{Var}(\theta_{0i} \mid Y_i) \lesssim p_n \log \frac{1}{\tau} + (n - p_n)\tau \sqrt{\log \frac{1}{\tau}} \tag{3}$$

for $\tau \to 0$, as $n, p_n \to \infty$ and $p_n = o(n)$.

Again, the choice $\tau = (\frac{p_n}{n})^\alpha$, for $\alpha \geq 1$ leads to an upper bound (3) of the order of the minimax risk. This result indicates that the posterior contracts fast enough to be able to provide a measure of uncertainty of adequate size around the point estimate. Theorems 3.1 and 3.2 combined allow us to find an upper bound on the rate of contraction of the full posterior distribution, both around the underlying mean vector and around the horseshoe estimator.

Theorem 3.3. *Under the assumptions of Theorem 3.1, with $\tau = (\frac{p_n}{n})^\alpha$, $\alpha \geq 1$:*

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \Pi_\tau \left(\theta : \|\theta - \theta_0\|^2 > M_n p_n \log \frac{n}{p_n} \;\middle|\; Y \right) \to 0, \tag{4}$$

and

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \Pi_\tau \left(\theta : \|\theta - T_\tau(Y)\|^2 > M_n p_n \log \frac{n}{p_n} \;\middle|\; Y \right) \to 0, \tag{5}$$

for every $M_n \to \infty$ as $n \to \infty$.

Proof. Combine Markov's inequality with the results of Theorems 3.1 and 3.2 for (4), and only with the result of Theorem 3.2 for (5). □

A remarkable aspect of the preceding Theorems is that many choices of τ, such as $\tau = (\frac{p_n}{n})^\alpha$ for any $\alpha \geq 1$, lead to an upper bound of the order $p_n \log(n/p_n)$ on the worst case ℓ_2 risk and posterior contraction rate. The upper bound on the rate of contraction in (4) is sharp, as the posterior cannot contract faster than the minimax rate around the true mean vector (Ghosal, Ghosh and Van der Vaart, 2000). However, this is not necessarily the case for the upper bound in (5), and for $\tau = (\frac{p_n}{n})^\alpha$ with $\alpha > 1$, the posterior spread may be of smaller order than the rate at which the horseshoe estimator approaches the underlying mean vector. Theorems 3.4 and 3.5 provide more insight into the effect of choosing different values of τ on the posterior spread and mean square error.

Theorem 3.4. *Suppose $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$, $\theta_0 \in \ell_0[p_n]$. Then the variance of the posterior distribution corresponding to the horseshoe prior satisfies*

$$\inf_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^n \mathrm{Var}(\theta_{0i} \mid Y_i) \gtrsim (n - p_n)\tau \sqrt{\log \frac{1}{\tau}} \tag{6}$$

for $\tau \to 0$ and $p_n = o(n)$, as $n \to \infty$. This lower bound is sharp for vectors $\theta_{0,n}$ with p_n entries equal to a_n and the remaining entries equal to zero, if a_n is such that $|a_n| \lesssim 1/\sqrt{\log(1/\tau)}$.

Theorem 3.5. *Suppose $Y \sim \mathcal{N}(\theta_{0,n}, \sigma^2 I_n)$ and $\theta_{0,n} \in \ell_0[p_n]$ is such that p_n entries are equal to $\gamma\sqrt{2\sigma^2 \log(1/\tau)}$, $\gamma \in (0, 1)$, and all remaining entries are equal to zero. Then:*

$$\mathbb{E}_{\theta_{0,n}} \|T_\tau(Y) - \theta_{0,n}\|^2 \asymp p_n \log \frac{1}{\tau} + (n - p_n)\tau \sqrt{\log \frac{1}{\tau}}, \tag{7}$$

and

$$\mathbb{E}_{\theta_{0,n}} \sum_{i=1}^n \mathrm{Var}(\theta_{0,ni} \mid Y_i) \asymp p_n \tau^{(1-\gamma)^2} \left(\log \frac{1}{\tau}\right)^{\gamma - \frac{1}{2}} + (n - p_n)\tau \sqrt{\log \frac{1}{\tau}}, \tag{8}$$

for $\tau \to 0$ and $p_n = o(n)$, as $n \to \infty$.

Consider $\tau = (\frac{p_n}{n})^\alpha$. Three cases can be discerned:

(i) $0 < \alpha < 1$. Lower bound (6) may exceed the minimax rate, implying suboptimal spread of the posterior distribution in the squared ℓ_2 sense.
(ii) $\alpha = 1$. Bounds (3) and (6) differ by a factor $\sqrt{\log(n/p_n)}$, as do (7) and (8). The gap can be closed by choosing $\tau = \frac{p_n}{n}\sqrt{\log \frac{n}{p_n}}$.
(iii) $\alpha > 1$. Bound (6) is not very informative, but Theorem 3.5 exhibits a sequence $\theta_{0,n} \in \ell_0[p_n]$ for which there is a mismatch between the order of the mean square error and the posterior variance. Bounds (7) and (8) are of the

orders $p_n(\log(1/\tau) + \tau^{1-1/\alpha}\sqrt{\log(1/\tau)})$ and $p_n(\tau^{(1-\gamma)^2}(\log(1/\tau))^{\gamma-1/2} + \tau^{1-1/\alpha}\sqrt{\log(1/\tau)})$, respectively. Hence up to logarithmic factors the total posterior variance (8) is a factor $\tau^{(1-1/\alpha)\wedge(1-\gamma)^2}$ smaller than the square distance of the center of the posterior to the truth (7). For $p_n \leq n^c$ for some $c > 0$, this factor behaves as a power of n.

These observations suggest that $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$ is a good choice, because then (2), (3), (6), (7), (8) are all of the order $p_n \log(n/p_n)$, suggesting that the posterior contracts at the minimax rate around both the truth and the horseshoe estimator.

4. Empirical Bayes estimation of τ

A natural follow-up question is how to choose τ in practice, when p_n is unknown. As discussed in Section 2, the full Bayesian approach suggested by Carvalho, Polson and Scott (2010) performs well in simulations. The analysis of such a hierarchical prior would however require different tools than the ones we have used so far. An empirical Bayes estimate of τ would be a natural solution, and allows us in practice to use one of the representations in (1) for computations, instead of an MCMC-type algorithm.

By adapting the approach in Paragraph 6.2 in (Johnstone and Silverman, 2004), we can find conditions under which the horseshoe estimator with an empirical Bayes estimate of τ will still attain the minimax ℓ_2 risk. Based on the consideration of Section 3, we proceed with the choices $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$ and $\tau = \frac{p_n}{n}$. The former is optimal in the sense that the posterior spread is of the order of the minimax risk, but the latter has the simple interpretation of being the proportion of nonzero means, and the difference between the two is only the square root of a log factor.

Theorem 4.1. *Suppose we observe an n-dimensional vector $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$ and we use $T_{\widehat{\tau}}(y)$ as our estimator of θ_0. If $\widehat{\tau} \in (0,1)$ satisfies the following two conditions for $\tau = \frac{p_n}{n}$ or $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$:*

1. *$\mathbb{P}_{\theta_0}(\widehat{\tau} > c\tau) \lesssim \frac{p_n}{n}$ for a constant $c \geq 1$ such that $\tau \leq \frac{1}{c}$;*
2. *There exists a function $g : \mathbb{N} \times \mathbb{N} \to (0,1)$ such that $\widehat{\tau} \geq g(n, p_n)$ with probability one and $-\log(g(n,p_n))\mathbb{P}_{\theta_0}(\widehat{\tau} \leq \tau) \lesssim \log(n/p_n)$,*

then:

$$\sup_{\theta_0 \in \ell_0[p_n]} \mathbb{E}_{\theta_0}\|T_{\widehat{\tau}}(Y) - \theta_0\|^2 \asymp p_n \log\frac{n}{p_n} \tag{9}$$

as $n, p_n \to \infty$ and $p_n = o(n)$. If only the first condition can be verified for an estimator $\widehat{\tau}$, then $\sup\{\frac{1}{n}, \widehat{\tau}\}$ will have an ℓ_2 risk of at most order $p_n \log n$.

The first condition requires that $\widehat{\tau}$ does not overestimate the fraction $\frac{p_n}{n}$ of nonzero means (up to a log factor) too much or by a too large probability. If $p_n \geq 1$, as we have assumed, then it is satisfied already by $\widehat{\tau} = \frac{1}{n}$ (and $c = 1$). According to the last assertion of the theorem, this 'universal threshold' yields the rate $p_n \log n$ (possibly up to a multiplicative constant). This is equal to the rate of the Lasso estimator with the usual choice of $\lambda = 2\sqrt{2\sigma^2 \log n}$ (Bickel, Ritov and Tsybakov, 2009). However, in the framework where $p_n \to \infty$, the estimator $\widehat{\tau} = \frac{1}{n}$ will certainly underestimate the sparsity level. A more natural estimator of $\frac{p_n}{n}$ is:

$$\widehat{\tau} = \frac{\#\{|y_i| \geq \sqrt{c_1 \sigma^2 \log n}, i = 1, \ldots, n\}}{c_2 n}, \tag{10}$$

where c_1 and c_2 are positive constants. By Lemma A.7, this estimator satisfies the first condition for $\tau = \frac{p_n}{n}$ and $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$ if $c_1 > 2, c_2 > 1$ and $p_n \to \infty$ or $c_1 = 2, c_2 > 1$ and $p_n \gtrsim \log n$. Thus $\max\{\widehat{\tau}, \frac{1}{n}\}$ will also lead to a rate of at most order $p_n \log n$ under these conditions. Its behaviour will be explored further in Section 5.

The rate can be improved to $p_n \log(n/p_n)$ if the second condition is met as well, which ensures that the sparsity level is not underestimated too much or by a too large probability. As we are not aware of any estimators meeting this condition for all θ_0, this condition is currently mostly of theoretical interest. If the true mean vector is very sparse, in the sense that there are relatively few nonzero means or the nonzero means are close to zero, there is not much to be gained in terms of rates by meeting this condition. The extra occurrence of p_n relative to the rate $p_n \log n$ is of interest only if p_n is relatively large. For instance, if $p_n \asymp n^\alpha$ for $\alpha \in (0, 1)$, then $p_n \log(n/p_n) = (1-\alpha)p_n \log n$, which suggests a decrease of the proportionality constant in (9), particularly if α is close to one. Furthermore, when p_n is large, the constant in (9) may be sensitive to the fine properties of $\widehat{\tau}$, as it depends on $g(n, p_n)$ (as can be seen in the proof). If $\widehat{\tau}$ seriously underestimates the sparsity level, the corresponding value of $g(n, p_n)$ from the second condition may be so small that the upper bound on the multiplicative constant before (9) becomes very large. Hence in this case, $\widehat{\tau}$ is required to be close to the proportion $\frac{p_n}{n}$ (up to a log factor) with large probability in order to get an optimal rate.

Datta and Ghosh (2013) warned against the use of an empirical Bayes estimate of τ for the horseshoe prior, because the estimate might collapse to zero. Their references for this statement, Scott and Berger (2010) and Bogdan, Ghosh and Tokdar (2008), indicate that they are thinking of a marginal maximum likelihood estimate of τ. However, an empirical Bayes estimate of τ does not need to be based on this principle. Furthermore, an estimator that satisfies the second condition from Theorem 4.1 or that is truncated from below by $\frac{1}{n}$, would not be susceptible to this potential problem.

5. Simulation study

A simulation study provides more insight into the behaviour of the horseshoe estimator, both when using an empirical Bayes procedure with estimator (10) and when using the fully Bayesian procedure proposed by Carvalho, Polson and Scott (2010) with a half-Cauchy prior on τ. For each data point, 100 replicates

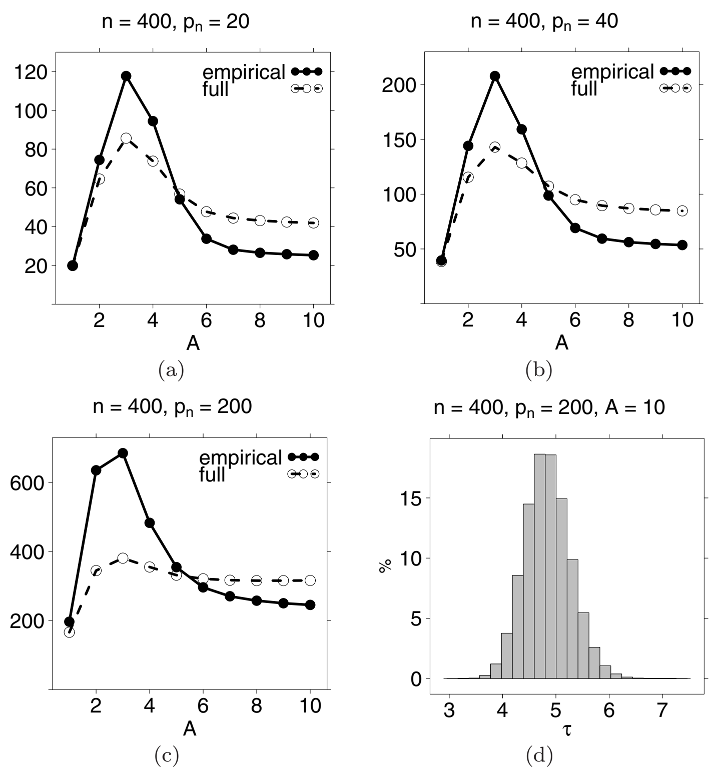

FIG 2. *Average squared error loss over 100 replicates with underlying mean vectors of length $n = 400$ if the nonzero coefficients are taken equal to A, in case 5% (Figure (a)), 10% (Figure (b)) or 50% (Figure (c)) of the means are equal to a nonzero value A. The solid line corresponds to empirical Bayes with (10), $c_1 = 2, c_2 = 1$, the dashed line to full Bayes with a half-Cauchy prior on τ. Figure (d) displays a histogram of all Gibbs samples of τ (after the burn-in) of all replicates in the setting $\tau \sim C^+(0,1)$, $A = 10$, $p_n = 200$.*

of an n-dimensional vector sampled from a $\mathcal{N}(\theta_0, I_n)$ distribution were created, where θ_0 had either 20, 40 or 200 (5%, 10% or 50%) entries equal to an integer A ranging from 1 to 10, and all the other entries equal to zero. The full Bayesian version was implemented using the code provided in (Scott, 2010), and the coordinatewise posterior mean was used as the estimator of θ_0. For the empirical Bayes procedure, the estimator (10) was used with $c_1 = 2$ and $c_2 = 1$. Performance was measured by squared error loss, which was averaged across replicates to create Figure 2.

In all settings, both estimators experience a peak in the ℓ_2 loss for values of A close to the 'universal threshold' of $\sqrt{2 \log 400} \approx 3.5$. This is not unexpected, as in the terminology of Johnstone and Silverman (2004), the horseshoe estimator is a shrinkage rule, and while it is not a thresholding rule in their sense, it does have the bounded shrinkage property which leads to thresholding-like behaviour. The bounded shrinkage property can be derived from Lemma A.3, which yields the following inequality as τ approaches zero:

$$|T_\tau(y) - y| \leq \sqrt{2\sigma^2 \log \frac{1}{\tau}}.$$

With $\tau = \frac{1}{n}$, this leads to the 'universal threshold' of $\sqrt{2\sigma^2 \log n}$, or with $\tau = (\frac{p_n}{n})^\alpha$, a 'threshold' at $\sqrt{2\alpha\sigma^2 \log(n/p_n)}$. Based on this property and the proofs of the main results, we can divide the underlying parameters into three cases:

(i) Those that are exactly or close to zero, where the observations are shrunk close to zero;
(ii) Those that are larger than the threshold, where the horseshoe estimator essentially behaves like the identity;
(iii) Those that are close to the 'threshold', where the horseshoe estimator is most likely to shrink the observations too much.

The horseshoe estimator performs well in cases (i) and (ii) due to its pole at zero and its heavy tails respectively. The hardest parameters to recover from the noise are those that are close to the threshold, and these are the ones that affect the estimation risk the most. This phenomenon explains the peaks in the graphs of Figure 2 around $A = 3.5$.

The full Bayes implementation with a Cauchy prior on τ attains a lower ℓ_2 loss around the universal threshold than the empirical Bayes procedure. This is because estimator (10) counts the number of observations that are above the universal threshold. When all the nonzero means are close to this threshold, $\widehat{\tau}$ may 'miss' some of them, thereby underestimating the sparsity level $\frac{p_n}{n}$ and thus leading to overshrinkage.

For values of A well past the universal threshold, the empirical Bayes estimator does better than the full Bayes version. For such large values of A, the estimator (10) will be equal to the true sparsity level with large probability and hence its good performance is not unexpected. However, an interesting question is why the full Bayes estimator does not do as well as the empirical Bayes estimator, especially because the nonzero means are so far removed from zero that the problem is 'easy'. This could be due to the choice of a half-Cauchy prior for τ: it places no restriction on the possible values of τ and has such heavy tails that values far exceeding the sparsity level $\frac{p_n}{n}$ are possible. This would lead to undershrinkage of the observations corresponding to a zero mean, which would be reflected in the ℓ_2 loss. Figure 2(d) shows a histogram of all Gibbs samples of τ in the setting where 50% of the means are set equal to 10. The range of these values is (3.1, 7.3), which is very far away from $\frac{p_n}{n} = \frac{1}{2}$. This indicates that a full Bayesian version of the horseshoe prior could benefit from a different choice of prior on τ than a half-Cauchy one, for example one that is restricted to [0,1].

6. Concluding remarks

The choice of the global shrinkage parameter τ is critical towards ensuring the right amount of shrinkage of the observations to recover the underlying mean vector. The value of $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$ was found to be optimal. Theorem 4.1

indicates that quite a wide range of estimators for τ will work well, especially in cases where the underlying mean vector is sparse. Of course, it should not come as a surprise that an estimator designed to recover sparse vectors will work especially well if the truth is indeed sparse. An interesting extension to this work would be to investigate whether the posterior concentration properties of the horseshoe prior still remain when a hyperprior is placed on τ. The result that $\tau = \frac{p_n}{n}$ (up to a log factor) yields optimal rates, together with the simulation results, suggests that in a fully Bayesian approach, a prior on τ which is restricted to $[0, 1]$ may perform better than the suggested half-Cauchy prior.

The simulation results also indicate that mean vectors with the nonzero means close to the universal threshold are the hardest to recover. In future simulations involving shrinkage rules, it would therefore be interesting to study the challenging case where all the nonzero parameters are at this threshold. The performance of the empirical Bayes estimator (10) leaves something to be desired around the threshold. In additional numerical experiments (not shown), we tried two other estimators of τ. The first was the 'oracle estimator' $\widehat{\tau} = \frac{p_n}{n}$. For values of the nonzero means well past the 'threshold', the behaviour of this estimator was very similar to that of (10). However, before the threshold, the squared error loss of the empirical procedure with the oracle estimator was between that of the full Bayes estimator and empirical Bayes with estimator (10). The second estimator was the mean of the samples of τ from the full Bayes estimator. The resulting squared error loss was remarkably close to that of the full Bayes estimator, for all values of the nonzero means. Neither of these two estimators is of much practical use. However, their range of behaviours suggests room for improvement over the estimator (10), and it would be worthwhile to study more refined estimators for τ.

An interesting question is what aspects of the horseshoe prior are truly essential towards optimal posterior contraction properties. Our proofs do not elucidate whether the pole at zero of the horseshoe prior is required, or if a prior with heavy tails, and in a sense 'sufficient' mass at zero would work as well. The failure of the Lasso to concentrate around the true mean vector at the minimax rate does indicate that heavy tails in itself may not be sufficient, and adding mass at zero solves this problem (Castillo, Schmidt-Hieber and Van der Vaart, 2014; Castillo and Van der Vaart, 2012). It is possible that the pole at zero is inessential, in particular if the global tuning parameter is chosen carefully, for instance by empirical Bayes. If the tuning parameter is chosen by a full Bayes method, the peak may be more essential, depending on its prior.

The horseshoe estimator has the property that its computational complexity depends on the sparsity level rather than the number of observations. Although there is no point mass at zero to induce sparsity, it still yields good reconstruction in ℓ_2, and a posterior distribution that contracts at an informative rate. None of the estimates will however be exactly zero. Variable selection can be performed by applying some sort of thresholding rule, such as the one suggested in (Carvalho, Polson and Scott, 2010) and analyzed by Datta and Ghosh (2013). The performance of this thresholding rule in simulations in the two works cited has been encouraging.

Acknowledgements

The authors would like to thank two anonymous referees for their helpful suggestions, as well as James Scott for his advice on implementing the full Bayesian version of the horseshoe estimator.

Appendix: Proofs

This section begins with Lemma A.1, providing bounds on some of the degenerate hypergeometric functions appearing in the posterior mean and posterior variance. This is followed by two lemmas that are needed for the proofs of Theorems 3.1 and 3.2: Lemma A.2 provides two upper bounds on the horseshoe estimator and Lemma A.3 gives a bound on the absolute value of the difference between the horseshoe estimator and an observation. We then proceed to the proof of Theorem 3.1, after which Lemma A.4 provides upper bounds on the posterior variance. These upper bounds are then used in the proof of Theorem 3.2. The proof of Theorem 3.4 is given next, followed by Lemmas A.5 and A.6 supporting the proof of Theorem 3.5. This section concludes with the proofs of Theorem 4.1 and Lemma A.7, which both concern the empirical Bayes procedure discussed in Section 4.

Lemma A.1. *Define*

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

Then, for $a > 1$:

$$I_{\frac{3}{2}}(y) \geq \frac{1}{5}\tau^3 + \sigma^2 \frac{\tau}{y^2}\left(e^{\frac{y^2}{2a\sigma^2}} - e^{\tau^2 \frac{y^2}{2\sigma^2}}\right) + \frac{\sigma^2}{\sqrt{a}y^2}\left(e^{\frac{y^2}{2\sigma^2}} - e^{\frac{y^2}{2a\sigma^2}}\right), \tag{11}$$

$$I_{\frac{1}{2}}(y) \geq \frac{1}{3}\tau + \frac{\sigma^2}{y^2}\left(e^{\frac{y^2}{2\sigma^2}} - e^{\tau^2 \frac{y^2}{2\sigma^2}}\right), \tag{12}$$

$$I_{\frac{1}{2}}(y) \leq \frac{2}{3} e^{\tau^2 \frac{y^2}{2\sigma^2}} \tau + 2 e^{\frac{y^2}{2a\sigma^2}}\left(\frac{1}{\sqrt{a}} - \tau\right) + \frac{2\sqrt{a}\sigma^2}{y^2}\left(e^{\frac{y^2}{2\sigma^2}} - e^{\frac{y^2}{2a\sigma^2}}\right), \tag{13}$$

$$\begin{aligned} I_{-\frac{1}{2}}(y) \geq \frac{1}{\tau} + e^{\tau^2 \frac{y^2}{2\sigma^2}}\left(\frac{1}{\tau} - \frac{1}{\sqrt{\tau}}\right) + \frac{a\sqrt{a}\sigma^2}{y^2}\left(e^{\frac{y^2}{2a\sigma^2}} - e^{\tau \frac{y^2}{2\sigma^2}}\right) \\ + \frac{\sigma^2}{y^2}\left(e^{\frac{y^2}{2\sigma^2}} - e^{\frac{y^2}{2a\sigma^2}}\right), \end{aligned} \tag{14}$$

$$\begin{aligned} I_{-\frac{1}{2}}(y) \leq \frac{2e^{\tau^2 \frac{y^2}{2\sigma^2}}}{\tau} + 2e^{\tau \frac{y^2}{2\sigma^2}}\left(\frac{1}{\tau} - \frac{1}{\sqrt{\tau}}\right) + 2e^{\frac{y^2}{2a\sigma^2}}\left(\frac{1}{\sqrt{\tau}} - \sqrt{a}\right) \\ + \frac{2a\sqrt{a}\sigma^2}{y^2}\left(e^{\frac{y^2}{2\sigma^2}} - e^{\frac{y^2}{2a\sigma^2}}\right), \end{aligned} \tag{15}$$

where (11) and (13) hold for $\tau < 1/\sqrt{a}$, (12) holds for $\tau < 1$, and (14) and (15) hold for $\tau < 1/a$.

Proof. Write $\xi = y^2/(2\sigma^2)$. We first note that for $z \geq \tau^2$, we have $z \leq \tau^2 + (1-\tau^2)z \leq 2z$, while for $z \leq \tau^2$, we have $\tau^2 \leq \tau^2 + (1-\tau^2)z \leq 2\tau^2$. Hence, we can bound I_k from above by:

$$I_k(y) \leq \frac{1}{\tau^2}\int_0^{\tau^2} z^k e^{\xi z} dz + \int_{\tau^2}^1 z^{k-1} e^{\xi z} dz,$$

and from below by half of that quantity. We bound the integral over $[0, \tau^2]$ in all cases by bounding the factor $e^{\xi z}$ by 1 or $e^{\tau^2 \xi}$. For the integral over $[\tau^2, 1]$, we first substitute $u = \xi z$, yielding: $\int_{\tau^2}^1 z^{k-1} e^{\xi z} dz = \xi^{-k} \int_{\tau^2 \xi}^{\xi} u^{k-1} e^u du$. For (11) and (13), we split the domain of integration into $[\tau^2\xi, \frac{\xi}{a}]$ and $[\frac{\xi}{a}, \xi]$. For $I_{\frac{3}{2}}$, we bound by:

$$I_{\frac{3}{2}}(y) \geq \frac{1}{2}\left(\frac{1}{\tau^2}\int_0^{\tau^2} z^{\frac{3}{2}} dz + \xi^{-\frac{3}{2}}(\tau^2\xi)^{\frac{1}{2}} \int_{\tau^2\xi}^{\frac{\xi}{a}} e^u du + \xi^{-\frac{3}{2}}\left(\frac{\xi}{a}\right)^{\frac{1}{2}} \int_{\frac{\xi}{a}}^{\xi} e^u du\right),$$

yielding (11). Similarly, for $I_{\frac{1}{2}}$:

$$I_{\frac{1}{2}}(y) \leq \frac{1}{\tau^2} e^{\tau^2\xi} \int_0^{\tau^2} z^{\frac{1}{2}} dz + \xi^{-\frac{1}{2}} e^{\frac{\xi}{a}} \int_{\tau^2\xi}^{\frac{\xi}{a}} u^{-\frac{1}{2}} du + \xi^{-\frac{1}{2}}\left(\frac{\xi}{a}\right)^{-\frac{1}{2}} \int_{\frac{\xi}{a}}^{\xi} e^u du,$$

resulting in (13). The bound (12) is obtained similarly, but without splitting up $[\tau^2\xi, \xi]$ further, by the inequality

$$I_{\frac{1}{2}}(y) \geq \frac{1}{2\tau^2}\int_0^{\tau^2} z^{\frac{1}{2}} dz + \frac{1}{2}\xi^{-1}\int_{\tau^2\xi}^{\xi} e^u du.$$

For the bounds on $I_{-\frac{1}{2}}$, we split up the domain of integration $[\tau^2\xi, \xi]$ into $[\tau^2\xi, \tau\xi], [\tau\xi, \frac{\xi}{a}]$ and $[\frac{\xi}{a}, \xi]$, and then bound by:

$$\begin{aligned} I_{-\frac{1}{2}}(y) \geq \frac{1}{2}\Bigg(&\frac{1}{\tau^2}\int_0^{\tau^2} z^{-\frac{1}{2}} dz + \xi^{\frac{1}{2}} e^{\tau^2\xi} \int_{\tau^2\xi}^{\tau\xi} u^{-\frac{3}{2}} du + \xi^{\frac{1}{2}}\left(\frac{\xi}{a}\right)^{-\frac{3}{2}} \int_{\tau\xi}^{\frac{\xi}{a}} e^u du \\ &+ \xi^{\frac{1}{2}}\xi^{-\frac{3}{2}} \int_{\frac{\xi}{a}}^{\xi} e^u du\Bigg), \end{aligned}$$

yielding (14), and by:

$$\begin{aligned} I_{-\frac{1}{2}}(y) \leq \frac{1}{\tau^2} e^{\tau^2\xi} &\int_0^{\tau^2} z^{-\frac{1}{2}} dz + \xi^{\frac{1}{2}} e^{\tau\xi} \int_{\tau^2\xi}^{\tau\xi} u^{-\frac{3}{2}} du + \xi^{\frac{1}{2}} e^{\frac{\xi}{a}} \int_{\tau\xi}^{\frac{\xi}{a}} u^{-\frac{3}{2}} du \\ &+ \xi^{\frac{1}{2}}\left(\frac{\xi}{a}\right)^{-\frac{3}{2}} \int_{\frac{\xi}{a}}^{\xi} e^u du, \end{aligned}$$

to find (15). ∎

Lemma A.2. *If $\tau^2 < 1$, the posterior mean of the horseshoe prior can be bounded above by:*

1. $T_\tau(y) \leq y e^{\frac{y^2}{2\sigma^2}} f(\tau)$, *where* f *is such that* $f(\tau) \leq \frac{2}{3}\tau$;
2. $$T_\tau(y) \leq y \frac{\frac{2}{3} e^{\tau^2 \frac{y^2}{2\sigma^2}} \tau + 2 e^{\frac{y^2}{2a\sigma^2}} (\frac{1}{\sqrt{a}} - \tau) + \frac{2\sqrt{a}\sigma^2}{y^2} (e^{\frac{y^2}{2\sigma^2}} - e^{\frac{y^2}{2a\sigma^2}})}{\frac{1}{\tau} + e^{\tau^2 \frac{y^2}{2\sigma^2}} (\frac{1}{\tau} - \frac{1}{\sqrt{\tau}}) + \frac{a\sigma^2\sqrt{a}}{y^2} (e^{\frac{y^2}{2a\sigma^2}} - e^{\tau \frac{y^2}{2\sigma^2}}) + \frac{\sigma^2}{y^2} (e^{\frac{y^2}{2\sigma^2}} - e^{\frac{y^2}{2a\sigma^2}})},$$
for any $a > 1$ *and* $\tau < \frac{1}{a}$.

Proof. We bound the integrals in the numerator and denominator of expression (1). For the first upper bound, we will use the fact that for $0 \leq z \leq 1$, $e^{\frac{y^2}{2\sigma^2}z}$ is bounded below by 1 and above by $e^{\frac{y^2}{2\sigma^2}}$. The posterior mean can therefore be bounded by:

$$T_\tau(y) \leq y e^{\frac{y^2}{2\sigma^2}} \frac{\int_0^1 z^{\frac{1}{2}} \frac{1}{\tau^2 + (1-\tau^2)z} dz}{\int_0^1 z^{-\frac{1}{2}} \frac{1}{\tau^2 + (1-\tau^2)z} dz} = y e^{\frac{y^2}{2\sigma^2}} f(\tau),$$

where

$$f(\tau) = \frac{\tau}{1-\tau^2} \left(\frac{\sqrt{1-\tau^2}}{\arctan\left(\frac{\sqrt{1-\tau^2}}{\tau}\right)} - \tau \right).$$

By Shafer's inequality for the arctangent (Shafer, 1966):

$$\frac{f(\tau)}{\tau} = \frac{1}{1-\tau^2} \left(\frac{\sqrt{1-\tau^2}}{\arctan\left(\frac{\sqrt{1-\tau^2}}{\tau}\right)} - \tau \right) < \frac{2}{3} \frac{1}{1+\tau} \leq \frac{2}{3},$$

which completes the proof for the first upper bound.

For the second inequality, we note that, in the notation of Lemma A.1, $T_\tau(y) = y \frac{I_{\frac{1}{2}}(y)}{I_{-\frac{1}{2}}(y)}$. The bounds in Lemma A.1 yield the stated inequality. □

Lemma A.3. *For* $\tau^2 < 1$, *the absolute value of the difference between the horseshoe estimator and an observation* y *can be bounded by a function* $h(y, \tau)$ *such that for any* $c > 2$:

$$\lim_{\tau \downarrow 0} \sup_{|y| > \sqrt{c\sigma^2 \log \frac{1}{\tau}}} h(y, \tau) = 0.$$

Proof. We assume $y > 0$ without loss of generality. By a change of variables of $x = 1 - z$:

$$|T_\tau(y) - y| = y \frac{\int_0^1 e^{-\frac{y^2}{2\sigma^2}x} x(1-x)^{-\frac{1}{2}} \frac{1}{1-(1-\tau^2)x} dx}{\int_0^1 e^{-\frac{y^2}{2\sigma^2}x} (1-x)^{-\frac{1}{2}} \frac{1}{1-(1-\tau^2)x} dx}.$$

By following the proof of Watson's lemma provided in Miller (2006), we can find bounds on the numerator and denominator of the above expression. First define $g(x) = (1-x)^{-\frac{1}{2}} \frac{1}{1-(1-\tau^2)x}$ and note that by Taylor's theorem, $g(x) = g(0) + x g'(\xi_x)$, where ξ_x is between 0 and x. Let s be any number between 0 and 1. Because $g''(x)$ is not negative for $x \in [0,1)$, we have that for $x \in [0,s]$, $s \in (0,1)$: $g'(0) \leq g'(x) \leq g'(s)$. The numerator can then be bounded by:

$$\begin{aligned}\int_0^1 e^{-\frac{y^2}{2\sigma^2}x} x g(x)dx &= \int_0^s e^{-\frac{y^2}{2\sigma^2}x} x g(0)dx + \int_0^s e^{-\frac{y^2}{2\sigma^2}x} x^2 g'(\xi_x)dx \\ &\quad + \int_s^1 e^{-\frac{y^2}{2\sigma^2}x} x g(x)dx \\ &\leq \frac{1}{y^4} h_1(y,\sigma,s) + \frac{g'(s)}{y^6} h_2(y,\sigma,s) + 2e^{-\frac{sy^2}{2\sigma^2}} h_3(\tau),\end{aligned}$$

where $h_1(y,\sigma,s) = 4\sigma^4 - 2\sigma^2(sy^2+2\sigma^2)e^{-\frac{sy^2}{2\sigma^2}}$, $h_2(y,\sigma,s) = 16\sigma^6 - 2\sigma^2(s^2y^4 + 4s\sigma^2y^2 + 8\sigma^4)e^{-\frac{sy^2}{2\sigma^2}}$ and $h_3(\tau) = \arctan(\frac{\sqrt{1-\tau^2}}{\tau})\tau^{-1}(1-\tau^2)^{-\frac{3}{2}} - (1-\tau^2)^{-1}$. The denominator can similarly be bounded by:

$$\begin{aligned}\int_0^1 e^{-\frac{y^2}{2\sigma^2}x} g(x)dx &= \int_0^s e^{-\frac{y^2}{2\sigma^2}x} g(0)dx + \int_0^s e^{-\frac{y^2}{2\sigma^2}x} x g'(\xi_x)dx \\ &\quad + \int_s^1 e^{-\frac{y^2}{2\sigma^2}x} g(x)dx \\ &\geq \frac{1}{y^2} h_4(y,\sigma,s) + \frac{g'(0)}{y^4} h_5(y,\sigma,s) + 0,\end{aligned}$$

where $h_4(y,\sigma,s) = 2\sigma^2 - 2\sigma^2 e^{-\frac{sy^2}{2\sigma^2}}$ and $h_5(y,\sigma,s) = 4\sigma^4 - 2\sigma^2 e^{-\frac{sy^2}{2\sigma^2}}(sy^2+2\sigma^2)$. Hence:

$$|T_\tau(y) - y| \leq \frac{\frac{1}{y}h_1(y,\sigma,s) + \frac{g'(s)}{y^3}h_2(y,\sigma,s) + 2y^3 e^{-\frac{sy^2}{2\sigma^2}} h_3(\tau)}{h_4(y,\sigma,s) + \frac{g'(0)}{y^2}h_5(y,\sigma,s)}.$$

For any fixed τ, this bound tends to zero as y tends to infinity. If $\tau \to 0$, the term containing $h_3(\tau)$ could potentially diverge. For $s = \frac{2}{3}$ and $y = \sqrt{c\sigma^2 \log(1/\tau)}$, where c is a positive constant, this term displays the following limiting behaviour as $\tau \to 0$:

$$\begin{aligned}\lim_{\tau \downarrow 0} y^3 e^{-\frac{1}{3\sigma^2}y^2} h_3(\tau) &= \lim_{\tau \downarrow 0} \left(c\sigma^2 \log\frac{1}{\tau}\right)^{\frac{3}{2}} \tau^{\frac{c}{3}-1} \left(\frac{\arctan\left(\frac{\sqrt{1-\tau^2}}{\tau}\right)}{(1-\tau^2)^{\frac{3}{2}}} - \frac{\tau}{1-\tau^2}\right) \\ &= \begin{cases} 0 & c > 3 \\ \infty & \text{otherwise,}\end{cases}\end{aligned}$$

because $\lim_{\tau\downarrow 0} \arctan(\frac{\sqrt{1-\tau^2}}{\tau})(1-\tau^2)^{-\frac{3}{2}} = \frac{\pi}{2}$, $\lim_{\tau \downarrow 0} \frac{\tau}{1-\tau^2} = 0$ and the factor $(c\sigma^2\log(1/\tau))^{\frac{3}{2}}\tau^{\frac{c}{3}-1}$ tends to zero as $\tau \downarrow 0$ if $\frac{c}{3} - 1 > 0$ and infinity otherwise. The condition $c > 3$ is related to the choice of $s = \frac{2}{3}$ and can be improved to any constant strictly greater than 2 by choosing s appropriately close to one. Hence, we find that the absolute value of the difference between the posterior mean and an observation can be bounded by a function $h(y,\tau)$ with the desired property. $\square$

Proof of Theorem 3.1

Proof. Suppose that $Y \sim \mathcal{N}(\theta, \sigma^2 I_n)$, $\theta \in \ell_0[p_n]$ and $\tilde{p}_n = \#\{i : \theta_i \neq 0\}$. Note that $\tilde{p}_n \leq p_n$. Assume without loss of generality that for $i = 1, \ldots, \tilde{p}_n$, $\theta_i \neq 0$, while for $i = \tilde{p}_n + 1, \ldots, n$, $\theta_i = 0$. We split up the expectation $\mathbb{E}_\theta \|T_\tau(Y) - \theta\|^2$ into the two corresponding parts:

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_i}(T_\tau(Y_i) - \theta_i)^2 = \sum_{i=1}^{\tilde{p}_n} \mathbb{E}_{\theta_i}(T_\tau(Y_i) - \theta_i)^2 + \sum_{i=\tilde{p}_n+1}^{n} \mathbb{E}_0 T_\tau(Y_i)^2.$$

We will now show that these two terms can be bounded by $\tilde{p}_n(1 + \log \frac{1}{\tau})$ and $(n - \tilde{p}_n)\sqrt{\log(1/\tau)}\tau$ respectively, up to multiplicative constants only depending on σ, for any choice of τ such that $\tau \in (0, 1)$.

Nonzero parameters

Denote $\zeta_\tau = \sqrt{2\sigma^2 \log(1/\tau)}$. We will show

$$\mathbb{E}_{\theta_i}(T_\tau(Y_i) - \theta_i)^2 \lesssim \sigma^2 + \zeta_\tau^2. \tag{16}$$

for all nonzero θ_i, which can be done by bounding $\sup_y |T_\tau(y) - y|$:

$$\begin{aligned}
\mathbb{E}_{\theta_i}(T_\tau(Y_i) - \theta_i)^2 &= \mathbb{E}_{\theta_i}((T_\tau(Y_i) - Y_i) + (Y_i - \theta_i))^2 \\
&\leq 2\mathbb{E}_{\theta_i}(Y_i - \theta_i)^2 + 2\mathbb{E}_{\theta_i}(T_\tau(Y_i) - Y_i)^2 \\
&\leq 2\sigma^2 + 2\left(\sup_y |T_\tau(y) - y|\right)^2,
\end{aligned}$$

Lemma A.3 yields the following bound on the difference between the observation and the horseshoe estimator: $|T_\tau(y) - y| \leq h(y, \tau)$, where $h(y, \tau)$ is such that $\lim_{\tau \downarrow 0} \sup_{|y| > c\zeta_\tau} h(y, \tau) = 0$ for any $c > 1$. Combining this with the inequality $|T_\tau(y) - y| \leq |y|$, we have as $\tau \to 0$:

$$\arg\max_y |T_\tau(y) - y| \lesssim \zeta_\tau, \tag{17}$$

which implies (16), as $|T_\tau(y)| \leq |y|$:

$$\left(\sup_y |T_\tau(y) - y|\right)^2 \lesssim \zeta_\tau^2.$$

Parameters equal to zero

We split up the term for the zero means into two parts:

$$\mathbb{E}_0 T_\tau(Y)^2 = \mathbb{E}_0 T_\tau(Y)^2 \mathbf{1}_{|Y| \leq \zeta_\tau} + \mathbb{E}_0 T_\tau(Y)^2 \mathbf{1}_{|Y| > \zeta_\tau},$$

where $\zeta_\tau = \sqrt{2\sigma^2 \log(1/\tau)}$. For the first term, we have, by the first bound in Lemma A.2:

$$\mathbb{E}_0 T_\tau(Y)^2 \mathbf{1}_{\{|Y| \leq \zeta_\tau\}} = \int_{-\zeta_\tau}^{\zeta_\tau} T_\tau(y)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$\leq \int_{-\zeta_\tau}^{\zeta_\tau} y^2 e^{\frac{y^2}{\sigma^2}} f(\tau)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy = \frac{f(\tau)^2}{\sqrt{2\pi\sigma^2}} \int_{-\zeta_\tau}^{\zeta_\tau} y^2 e^{\frac{y^2}{2\sigma^2}} dy$$
$$\leq \sqrt{\frac{2}{\pi}} \sigma f(\tau)^2 \zeta_\tau \frac{1}{\tau} \leq \sqrt{\frac{2}{\pi}} \sigma \frac{4}{9} \zeta_\tau \tau \lesssim \zeta_\tau \tau,$$

where the identity $\frac{d}{dy} y e^{\frac{y^2}{2\sigma^2}} = \frac{y^2}{\sigma^2} e^{\frac{y^2}{2\sigma^2}} + e^{\frac{y^2}{2\sigma^2}}$ was used to bound $\int_{-\zeta_\tau}^{\zeta_\tau} y^2 e^{\frac{y^2}{2\sigma^2}} dy$. For the second term, because $|T_\tau(y)| \leq |y|$ for all y, we have by the identity $y^2\phi(y) = \phi(y) - \frac{d}{dy}[y\phi(y)]$, and by Mills' ratio:

$$\mathbb{E}_0 T_\tau(Y)^2 \mathbf{1}_{\{|Y|>\zeta_\tau\}} \leq \mathbb{E}_0 Y^2 \mathbf{1}_{\{|Y|>\zeta_\tau\}} = 2\int_{\frac{\zeta_\tau}{\sigma}}^{\infty} \sigma^2 y^2 \phi(y) dy$$
$$\leq 2\sigma\zeta_\tau \phi\left(\frac{\zeta_\tau}{\sigma}\right) + 2\sigma^3 \frac{\phi\left(\frac{\zeta_\tau}{\sigma}\right)}{\zeta_\tau} \leq 4\sigma\zeta_\tau\phi\left(\frac{\zeta_\tau}{\sigma}\right) = 4\sigma\zeta_\tau \frac{1}{\sqrt{2\pi}}\tau,$$

where the last inequality holds for $\zeta_\tau > \sigma^2$. If we apply this inequality and combine this upper bound with the upper bound on the first term, we find, for $\zeta_\tau > \sigma^2$ (corresponding to $\tau < e^{-\frac{\sigma^2}{2}}$):

$$\mathbb{E}_0 T_\tau(Y)^2 = \mathbb{E}_0 T_\tau(Y)^2 \mathbf{1}_{\{|Y|\leq\zeta_\tau\}} + \mathbb{E}_0 T_\tau(Y)^2 \mathbf{1}_{\{|Y|>\zeta_\tau\}} \lesssim \zeta_\tau \tau. \tag{18}$$

Hence, for $\tau < e^{-\frac{\sigma^2}{2}}$:

$$\sum_{i=p_n+1}^{n} \mathbb{E}_0 T_\tau(Y_i)^2 \lesssim (n - p_n)\zeta_\tau \tau. \tag{19}$$

Conclusion

By (16) and (19), we find for $\tau < e^{-\frac{\sigma^2}{2}}$:

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_i}(T_\tau(Y_i) - \theta_i)^2 \lesssim \tilde{p}_n(1 + \zeta_\tau^2) + (n - \tilde{p}_n)\tau\zeta_\tau. \qquad \square$$

Lemma A.4. *The posterior variance when using the horseshoe prior can be expressed as:*

$$\mathrm{Var}(\theta \mid y) = \frac{\sigma^2}{y} T_\tau(y) - (T_\tau(y) - y)^2 + y^2 \frac{\int_0^1 (1-z)^2 z^{-\frac{1}{2}} \frac{1}{\tau^2 + (1-\tau^2)z} e^{\frac{y^2}{2\sigma^2}z} dz}{\int_0^1 z^{-\frac{1}{2}} \frac{1}{\tau^2 + (1-\tau^2)z} e^{\frac{y^2}{2\sigma^2}z} dz}, \tag{20}$$

and bounded from above by:

1. $\mathrm{Var}(\theta \mid y) \leq \sigma^2 + y^2$;
2. $\mathrm{Var}(\theta \mid y) \leq (\frac{\sigma^2}{y} + y) T_\tau(y) - T_\tau(y)^2$.

Proof. As proven in Pericchi and Smith (1992):

$$\mathrm{Var}(\theta \mid y) = \sigma^2 + \sigma^4 \frac{d^2}{dy^2} \log m(y) = \sigma^2 - \left(\sigma^2 \frac{m'(y)}{m(y)}\right)^2 + \sigma^4 \frac{m''(y)}{m(y)},$$

where $m(y)$ is the density of the marginal distribution of y. Equality (20) can be found by combining the expressions

$$m(y) = \frac{1}{\sqrt{2\pi^3}\sigma\tau} e^{-\frac{y^2}{2\sigma^2}} \int_0^1 z^{-\frac{1}{2}} \frac{1}{1-\left(1-\frac{1^2}{\tau^2}\right)z} e^{\frac{y^2}{2\sigma^2}z} dz$$
$$m''(y) = \frac{1}{y} m'(y) + \frac{1}{\sqrt{2\pi^3}\sigma\tau} \frac{y^2}{\sigma^4} e^{-\frac{y^2}{2\sigma^2}} \int_0^1 z^{-\frac{1}{2}} (1-z)^2 \frac{1}{1-\left(1-\frac{1}{\tau^2}\right)z} e^{\frac{y^2}{2\sigma^2}z} dz$$

with the equality $T_\tau(y) = y + \sigma^2 \frac{m'(y)}{m(y)}$. The first upper bound is implied by the property $|T_\tau(y)| < |y|$ and the fact that $(1-z)^2 \leq 1$ for $z \in [0,1]$. The second upper bound can be demonstrated by noting that $(1-z)^2 \leq 1-z$ for $z \in [0,1]$ and hence:

$$\mathrm{Var}(\theta \mid y) \leq \frac{\sigma^2}{y} T_\tau(y) - (y - T_\tau(y))^2 + y^2 \left(1 - \frac{1}{y} T_\tau(y)\right). \qquad \square$$

Proof of Theorem 3.2

Proof. As in the proof of Theorem 3.1 we assume that $\theta_i \neq 0$ for $i = 1, \ldots, \tilde{p}_n$ and $\theta_i = 0$ for $i = \tilde{p}_n + 1, \ldots, n$, where $\tilde{p}_n \leq p_n$ by assumption. We consider the posterior variances for the zero and nonzero means separately. Denote $\zeta_\tau = \sqrt{2\sigma^2 \log(1/\tau)}$.

Nonzero means

By applying the same reasoning as in Lemma A.3 to the final term of $\mathrm{Var}(\theta|y)$ in (20), we can find a function $\tilde{h}(y,t)$ such that $\mathrm{Var}(\theta|y) \leq \tilde{h}(y,\tau)$, where $\tilde{h}(y,\tau) \to \sigma^2$ as $y \to \infty$ for any fixed τ. If $\tau \to 0$, the function $\tilde{h}(y,\tau)$ displays the following limiting behaviour for any $c > 1$:

$$\lim_{\tau \downarrow 0} \sup_{|y| > c\zeta_\tau} \tilde{h}(y,\tau) = \sigma^2.$$

Hence, as $\tau \to 0$: $\mathrm{Var}(\theta|y) \lesssim \sigma^2$, for any $|y|$ that increases as least as fast as ζ_τ when τ decreases. Now suppose $|y| \leq \zeta_\tau$. Then, by the bound $\mathrm{Var}(\theta \mid y) \leq \sigma^2 + y^2$ from Lemma A.4, we find:

$$\mathrm{Var}(\theta \mid y) \leq \sigma^2 + \zeta_\tau^2.$$

Therefore:

$$\sum_{i=1}^{\tilde{p}_n} \mathbb{E}_{\theta_i} \mathrm{Var}(\theta_i \mid Y_i) \lesssim \tilde{p}_n (1 + \zeta_\tau^2). \tag{21}$$

Zero means

By the bound $\mathrm{Var}(\theta \mid y) \leq \sigma^2 + y^2$, we find for $c \geq 1$:

$$\mathbb{E}_0 \mathrm{Var}(\theta \mid Y) \mathbf{1}_{\{|Y| > c\zeta_\tau\}} \leq 2 \int_{c\zeta_\tau}^\infty (\sigma^2 + y^2) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$= 2\sigma^2 \Phi^c\left(\frac{c\zeta_\tau}{\sigma}\right) + 2\int_{\frac{c\zeta_\tau}{\sigma}}^{\infty} \sigma^2 x^2 \phi(x)dx$$
$$\leq 4\sigma^3 \frac{\phi\left(\frac{c\zeta_\tau}{\sigma}\right)}{c\zeta_\tau} + 2\sigma c\zeta_\tau \phi\left(\frac{c\zeta_\tau}{\sigma}\right) \lesssim \frac{\tau}{\zeta_\tau} + \zeta_\tau \tau.$$

For $|y| < c\zeta_\tau$, we consider the upper bound $\mathrm{Var}(\theta \mid y) \leq (\frac{\sigma^2}{y} + y)T_\tau(y) - T_\tau(y)^2$ from Lemma A.4. From this bound, we get $\mathrm{Var}(\theta \mid y) \leq \frac{\sigma^2}{y}T_\tau(y) + yT_\tau(y)$. Hence:

$$\mathbb{E}_0 \mathrm{Var}(\theta \mid Y)\mathbf{1}_{\{|Y| \leq c\zeta_\tau\}} \leq \sigma^2 \int_{-c\zeta_\tau}^{c\zeta_\tau} \frac{1}{y} T_\tau(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy$$
$$+ \int_{-c\zeta_\tau}^{c\zeta_\tau} yT_\tau(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy. \tag{22}$$

We bound the first integral from (22) by applying the first bound on $T_\tau(y)$ from Lemma A.2:

$$\sigma^2 \int_{-c\zeta_\tau}^{c\zeta_\tau} \frac{1}{y} T_\tau(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy \leq \sigma^2 \int_{-c\zeta_\tau}^{c\zeta_\tau} f(\tau) \frac{1}{\sqrt{2\pi\sigma^2}} dy$$
$$= \sqrt{\frac{2\sigma}{\pi}} c\zeta_\tau f(\tau) \lesssim \zeta_\tau \tau,$$

because $f(\tau) \leq \frac{2}{3}\tau$. For the second term in (22), we first note that the second bound from Lemma A.2 can be relaxed to:

$$T_\tau(y) \leq \tau y \left(\frac{2}{3}\tau e^{\tau^2 \frac{y^2}{2\sigma^2}} + \frac{2}{\sqrt{a}} e^{\frac{y^2}{2a\sigma^2}} + 2\sqrt{a}\sigma^2 \frac{1}{y^2} e^{\frac{y^2}{2\sigma^2}}\right) \tag{23}$$

for any $a > 1$ and $\tau < \frac{1}{a}$. By plugging this bound into the second integral of (22), we get three terms, which we will name I_1, I_2 and I_3 respectively. We then find, bounding above by the integral over $\mathbb{R}$ instead of $[-c\zeta_\tau, c\zeta_\tau]$ for I_1 and I_2:

$$I_1 = \frac{2}{3}\tau^2 \int_{-c\zeta_\tau}^{c\zeta_\tau} y^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1-\tau^2)\frac{y^2}{2\sigma^2}} dy \leq \frac{2}{3}\tau^2 \frac{\sigma^2}{(1-\tau^2)^{\frac{3}{2}}} \lesssim \tau^2.$$
$$I_2 = \frac{2}{\sqrt{a}}\tau \int_{-c\zeta_\tau}^{c\zeta_\tau} y^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{a-1}{a}\frac{y^2}{2\sigma^2}} dy \leq \frac{2a\sigma^2}{(a-1)^{\frac{3}{2}}}\tau \lesssim \tau.$$
$$I_3 = 2\sqrt{a}\sigma^2\tau \int_{-c\zeta_\tau}^{c\zeta_\tau} \frac{1}{\sqrt{2\pi\sigma^2}} dy = \frac{2\sqrt{2a}c\sigma}{\sqrt{\pi}} \zeta_\tau \tau \lesssim \zeta_\tau \tau.$$

And thus:

$$\sum_{i=\tilde{p}_n+1}^{n} \mathbb{E}_0 \mathrm{Var}(\theta_i \mid Y_i) \lesssim (n - \tilde{p}_n)(\zeta_\tau + \tau + 1)\tau. \tag{24}$$

Conclusion

By (21) and (24):

$$\mathbb{E}_\theta \sum_{i=1}^{n} \mathrm{Var}(\theta_i \mid Y_i) \lesssim \tilde{p}_n(1 + \zeta_\tau^2) + (n - \tilde{p}_n)(\zeta_\tau + \tau + 1)\tau. \qquad \square$$

Proof of Theorem 3.4

Proof. By expanding $(1-z)^2 z^{-\frac{1}{2}} = z^{-\frac{1}{2}} - 2z^{\frac{1}{2}} + z^{\frac{3}{2}}$, we see that the final term in (20) is equal to:

$$y^2 - 2yT_\tau(y) + y^2 \frac{\int_0^1 z^{\frac{3}{2}} \frac{1}{\tau^2+(1-\tau^2)z} e^{\frac{y^2}{2\sigma^2}z} dz}{\int_0^1 z^{-\frac{1}{2}} \frac{1}{\tau^2+(1-\tau^2)z} e^{\frac{y^2}{2\sigma^2}z} dz}.$$

As $\frac{T_\tau(y)}{y}$ is non-negative, we can bound the posterior variance from below by the final two terms in (20). By the above equality, this yields the following lower bound:

$$\operatorname{Var}(\theta \mid y) \geq y^2 \frac{I_{\frac{3}{2}}(y)}{I_{-\frac{1}{2}}(y)} - T_\tau(y)^2 = y^2 \left(\frac{I_{\frac{3}{2}}(y)}{I_{-\frac{1}{2}}(y)} - \left(\frac{I_{\frac{1}{2}}(y)}{I_{-\frac{1}{2}}(y)} \right)^2 \right),$$

where I_k is as in Lemma A.1. We now use the bounds from Lemma A.1 with $a = 2$ and take ξ equal to $c\log(1/\tau)$ for some nonnegative constant c. Then $e^\xi = \frac{1}{\tau^c}$ and $e^{\frac{\xi}{2}} = \frac{1}{\tau^{\frac{c}{2}}}$. Taking for each bound on I_k, $k \in \{\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}\}$, the term that diverges fastest as τ approaches zero, we find that the lower bound is asymptotically of the order:

$$2\sigma^2\xi \left(\frac{\frac{1}{2\sqrt{2}\xi}\frac{1}{\tau^c}}{\max\left\{\frac{2e^{\tau\xi}}{\tau}, \frac{2\sqrt{2}}{\xi}\frac{1}{\tau^c}\right\}} - \left(\frac{\frac{\sqrt{2}}{\xi}\frac{1}{\tau^c}}{\max\{\frac{e^{\tau^2\xi}}{\tau}, \frac{1}{2\xi}\frac{1}{\tau^c}\}} \right)^2 \right).$$

For $c \leq 1$, this reduces to:

$$\frac{\sigma^2}{2\sqrt{2}} e^{-\tau\xi}\tau^{1-c} - \frac{4\sigma^2}{\xi} e^{-2\tau^2\xi}\tau^{2-2c}.$$

The second term is negligible compared to the first. Hence, we will use the term $\frac{\sigma^2}{2\sqrt{2}} e^{-\tau\xi}\tau^{1-c}$ as our lower bound on $\operatorname{Var}(\theta \mid y)$ for $y = \pm\sqrt{2c\sigma^2\log(1/\tau)} = \sqrt{c}\zeta_\tau$, where $\zeta_\tau = \sqrt{2\sigma^2\log(1/\tau)}$. To find the lower bound on $\sum_{i=1}^n \mathbb{E}_{\theta_i}\operatorname{Var}(\theta_i \mid Y_i)$, we only need to consider the parameters equal to zero:

$$\sum_{i=1}^n \mathbb{E}_{\theta_i}\operatorname{Var}(\theta_i \mid Y_i) \geq (n-p_n)\mathbb{E}_0\operatorname{Var}(\theta_i \mid Y_i)\mathbf{1}_{\{|Y_i|\leq\zeta_\tau\}}. \tag{25}$$

By the substitution $x = y^2/\zeta_\tau^2, dy = \frac{\sigma\sqrt{\log(1/\tau)}}{\sqrt{2x}}dx$, we find:

$$\begin{aligned} E_0\operatorname{Var}(\theta_i \mid Y_i)\mathbf{1}_{\{|Y_i|\leq\zeta_\tau\}} &\geq 2\int_0^{\zeta_\tau} \frac{\sigma^2}{2\sqrt{2}} e^{-\tau\frac{y^2}{2\sigma^2}} \tau^{1-\frac{y^2}{\zeta_\tau^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy \\ &= \frac{\sigma}{4\sqrt{\pi}}\tau\zeta_\tau \int_0^1 \frac{\tau^{\tau x}}{\sqrt{x}}dx \geq \frac{\sigma}{2\sqrt{\pi}} e^{-\frac{1}{e}}\tau\zeta_\tau, \end{aligned} \tag{26}$$

where in the last step, we used $\tau^{\tau x} \geq \tau^{\tau} \geq e^{-\frac{1}{e}}$ for $x \in [0,1], \tau \in (0,1]$. By plugging this into (25), we find that as $\tau \to 0$:

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_i} \mathrm{Var}(\theta_i \mid Y_i) \gtrsim (n - p_n)\tau\zeta_\tau, \tag{27}$$

finishing the proof for the first statement of the theorem.

We now consider θ such that $\theta_i = a_n$ for $i = 1, \ldots, p_n$, and $\theta_i = 0$ for $i = p_n + 1, \ldots, n$, and assume without loss of generality that $a_n > 0$. We wish to find conditions on a_n such that the lower bound (27) is sharp (up to a constant factor). Denoting $\zeta_\tau = \sqrt{2\sigma^2 \log(1/\tau)}$, as before, it is sufficient if we can find a_n such that $\mathbb{E}_{\theta_i = a_n} \mathrm{Var}(\theta_i \mid Y_i) \lesssim \tau\zeta_\tau$, because in combination with the bound (24), this will yield $\sum_{i=1}^{n} \mathbb{E}_{\theta_i} \mathrm{Var}(\theta_i \mid Y_i) \lesssim n\tau\zeta_\tau$, which is of the same order as (27), as $p_n = o(n)$. Sufficient conditions on a_n can be found by adapting the proof for the 'zero means' case of Theorem 3.2.

We first consider $|y_i| > \zeta_\tau$. By the first bound of Lemma A.4:

$$\begin{aligned}\mathbb{E}_{\theta_i} \mathrm{Var}(\theta_i \mid Y_i) \mathbf{1}_{\{|Y_i| > \zeta_\tau\}} \leq \int_{\zeta_\tau}^{\infty} (\sigma^2 + y^2) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy \\ + \int_{-\infty}^{-\zeta_\tau} (\sigma^2 + y^2) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy.\end{aligned} \tag{28}$$

The first integral from (28) can be split into two parts by splitting up the factor $\sigma^2 + y^2$, the first of which can be bounded, by substituting $x = (y - a_n)/\sigma$ and applying Mills' ratio:

$$\sigma^2 \int_{(\zeta_\tau - a_n)/\sigma}^{\infty} \phi(x) dx = \sigma^2 \Phi^c \left(\frac{\zeta_\tau - a_n}{\sigma} \right) \leq \frac{\sigma^3}{\zeta_\tau - a_n} \phi \left(\frac{\zeta_\tau - a_n}{\sigma} \right). \tag{29}$$

The second of these integrals is, by $y^2 = (y - a_n)^2 - a_n^2 + 2a_n y$, equal to:

$$\begin{aligned}\int_{\zeta_\tau}^{\infty} (y - a_n)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy - a_n^2 \int_{\zeta_\tau}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy \\ + a_n \int_{\zeta_\tau}^{\infty} y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy.\end{aligned} \tag{30}$$

The second integral of (30) can be bounded from below by zero, and the third from above by $a_n \mathbb{E}_{\theta_i} Y_i = a_n^2$. Again substituting $x = (y - a_n)/\sigma$ yields the following upper bound on (30): $\sigma^2 \int_{(\zeta_\tau - a_n)/\sigma}^{\infty} x^2 \phi(x) dx + a_n^2$. Now using the equality $x^2\phi(x) = \phi(x) - \frac{d}{dx}[x\phi(x)]$ and again Mills' ratio, and combining with (29), we find the following upper bound on the first integral from (28):

$$\frac{2\sigma^3}{\zeta_\tau - a_n} \phi \left(\frac{\zeta_\tau - a_n}{\sigma} \right) + \sigma(\zeta_\tau - a_n) \phi \left(\frac{\zeta_\tau - a_n}{\sigma} \right) + a_n^2. \tag{31}$$

By substituting $x = -y$ in the second integral from (28) and then applying the same inequalities to it as to the first integral, the following bound is obtained:

$$\frac{2\sigma^3}{\zeta_\tau + a_n} \phi \left(\frac{\zeta_\tau + a_n}{\sigma} \right) + \sigma(\zeta_\tau + a_n) \phi \left(\frac{\zeta_\tau + a_n}{\sigma} \right). \tag{32}$$

This bound does not include a term a_n^2, because in the step equivalent to (30), the identity $y^2 = (y+a_n)^2 - a_n^2 - 2ya_n$ is used, and thus only the integral $\int_{\zeta_\tau}^{\infty}(y+a_n)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+a_n)^2}{2\sigma^2}} dy$ needs to be bounded in that step. $\mathbb{E}_{\theta_i}\mathrm{Var}(\theta \mid Y)\mathbf{1}_{\{|Y|>\zeta_\tau\}}$ can thus be bounded by the sum of (31) and (32). The factor $\phi((\zeta_\tau + a_n)/\sigma)$ can be bounded from above by $\phi(\zeta_\tau/\sigma) = \tau/\sqrt{2\pi}$. The factor $\phi((\zeta_\tau - a_n)/\sigma)$ is equal to $\frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta_\tau^2}{2\sigma^2}} e^{-\frac{a_n^2}{2\sigma^2}} e^{\frac{\zeta_\tau a_n}{\sigma}} = \frac{\tau}{\sqrt{2\pi}} e^{-\frac{a_n^2}{2\sigma^2}} e^{\frac{\zeta_\tau a_n}{\sigma}}$. Hence we arrive at the following upper bound:

$$\frac{\sigma}{\sqrt{2\pi}}\left[\left(\frac{2\sigma^2}{\zeta_\tau - a_n} + \zeta_\tau - a_n\right) e^{-\frac{a_n^2}{2\sigma^2}} e^{\frac{\zeta_\tau a_n}{\sigma}} + \frac{2\sigma^2}{\zeta_\tau + a_n} + \zeta_\tau + a_n\right]\tau + a_n^2. \tag{33}$$

If $a_n \lesssim 1/\zeta_\tau$, then $e^{-\frac{a_n^2}{2\sigma^2}} e^{\frac{\zeta_\tau a_n}{\sigma}} = \mathcal{O}(1)$ and $\zeta_\tau \pm a_n = \mathcal{O}(\zeta_\tau)$, yielding an upper bound on (33) of order $\tau\zeta_\tau$.

We now consider $|y_i| \leq \zeta_\tau$. We use the second bound of Lemma A.4:

$$\begin{aligned}\mathbb{E}_{\theta_i}\mathrm{Var}(\theta_i \mid Y_i)\mathbf{1}_{\{|Y_i|\leq\zeta_\tau\}} &\leq \sigma^2 \int_{-\zeta_\tau}^{\zeta_\tau} \frac{1}{y} T_\tau(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy \\ &\quad + \sigma^2 \int_{-\zeta_\tau}^{\zeta_\tau} y T_\tau(y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy. \end{aligned} \tag{34}$$

Applying inequality $\frac{1}{y}T_\tau(y) \leq \frac{2}{3}\tau e^{\frac{y^2}{2\sigma^2}}$ from Lemma A.2 to the first integral yields the bound:

$$\frac{\sqrt{2}\sigma}{3\sqrt{\pi}}\tau \int_{-\zeta_\tau}^{\zeta_\tau} e^{\frac{y^2}{2\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy = \frac{\sqrt{2}\sigma}{3\sqrt{\pi}}\tau e^{-\frac{a_n^2}{2\sigma^2}} \int_{-\zeta_\tau}^{\zeta_\tau} e^{\frac{a_n y}{\sigma^2}} dy \leq \frac{\sqrt{2}\sigma}{3\sqrt{\pi}}\tau e^{-\frac{a_n^2}{2\sigma^2}} 2\zeta_\tau e^{\frac{a_n\zeta_\tau}{\sigma^2}}.$$

If $a_n \lesssim 1/\zeta_\tau$, we have $a_n\zeta_\tau = \mathcal{O}(1)$ and thus this term will be of order $\tau\zeta_\tau$. For the second integral from (34), we use bound (23). This leads to three integrals to be bounded, I_1, I_2 en I_3.

$$\begin{aligned} I_1 &= \frac{\sigma}{\sqrt{2\pi}} \frac{2}{3}\tau^2 e^{\frac{\tau^2}{1-\tau^2}\frac{a_n^2}{2\sigma^2}} \int_{-\zeta_\tau}^{\zeta_\tau} y^2 e^{-\frac{1}{2\sigma^2/(1-\tau^2)}\left(y - \frac{a_n}{1-\tau^2}\right)^2} dy \\ &\leq \frac{2}{3} e^{\frac{\tau^2}{1-\tau^2}\frac{a_n^2}{2\sigma^2}} \frac{\sigma^2}{(1-\tau^2)^{3/2}}\left(\sigma^2 + \frac{a_n^2}{1-\tau^2}\right)\tau^2. \\ I_2 &= \frac{2\sigma}{\sqrt{a}\sqrt{2\pi}}\tau e^{\frac{a_n^2}{(b-1)2\sigma^2}} \int_{-\zeta_\tau}^{\zeta_\tau} y^2 e^{-\frac{1}{2\sigma^2 \frac{a}{a-1}}\left(y - \frac{a}{a-1}a_n\right)^2} \\ &\leq \frac{2}{\sqrt{b}} e^{\frac{a_n^2}{(b-1)2\sigma^2}} \sigma^2 \left(\frac{a}{a-1}\right)^{3/2}\left(\sigma^2 + \frac{a}{a-1}a_n^2\right)\tau. \\ I_3 &= \frac{2\sqrt{a}\sigma^3}{\sqrt{2\pi}}\tau \int_{-\zeta_\tau}^{\zeta_\tau} e^{\frac{y^2}{2\sigma^2}} e^{-\frac{(y-a_n)^2}{2\sigma^2}} dy \leq \frac{2\sqrt{2a}\sigma^3}{\sqrt{\pi}} e^{-\frac{a_n^2}{2\sigma^2}} e^{\frac{a_n\zeta_\tau}{\sigma^2}} \tau\zeta_\tau. \end{aligned}$$

I_1, I_2 and I_3 will all be of no larger order than $\tau\zeta_\tau$ if $a_n \lesssim 1/\zeta_\tau$. ∎

Lemma A.5. *For all $k \in \mathbb{R}$, $\int_1^y u^k e^u du = y^k e^y(1+\mathcal{O}(1/y))$, as $y \to \infty$.*

Proof. For $k = 0$, the statement is immediate. By integration by parts the integral is seen to be equal to $y^k e^y - e - \int_1^y k u^{k-1} e^u du$. For $k \neq 0$, the latter integral is bounded above by

$$|k| \int_1^{y/2} (1 \vee y/2)^{k-1} e^u du + |k| \int_{y/2}^{y} (y/2 \vee y)^{k-1} e^u du.$$

This is further bounded above by a multiple of $(1 \vee y^{k-1}) e^{y/2} + y^{k-1} e^y$. □

Lemma A.6. *Let I_k be as in Lemma A.1. There exist functions R_k with $\sup_{\zeta_\tau/4 \le y \le 4\zeta_\tau} |R_k(y)| \to 0$ for $k > 0$ and $k = -\frac{1}{2}$, such that,*

$$I_k(y) = \left(\tau^{2k} \int_0^1 \frac{z^k}{1+z} dz + \frac{2\sigma^2}{y^2} e^{\frac{y^2}{2\sigma^2}}\right)(1 + R_k(y)), \quad \textit{for } k > 0,$$

$$I_{-\frac{1}{2}}(y) = \left(\tau^{-1} \int_0^\infty \frac{1}{\sqrt{z}(1+z)} dz + \frac{2\sigma^2}{y^2} e^{\frac{y^2}{2\sigma^2}}\right)\left(1 + R_{-\frac{1}{2}}(y)\right).$$

Proof. We split the integral in the definition of I_k over the intervals $[0,\tau^2]$ and $[\tau^2, 1]$. The first interval contributes, uniformly in $y\tau \to 0$,

$$\begin{aligned}\int_0^{\tau^2} \frac{z^k e^{\frac{y^2}{2\sigma^2} z}}{\tau^2 + (1-\tau^2)z} dz &= \int_0^{\tau^2} \frac{z^k}{\tau^2 + (1-\tau^2)z} dz\,(1+o(1)) \\ &= \tau^{2k} \int_0^1 \frac{u^k}{1+(1-\tau^2)u} du\,(1+o(1)),\end{aligned} \tag{35}$$

by the substitution $u = z/\tau^2$. The integral tends to $\int_0^1 \frac{u^k}{1+u}\, du$, by the dominated convergence theorem, for any $k > -1$. The second interval contributes, with the substitution $u = (y^2/2\sigma^2)z$:

$$\int_{\tau^2}^1 \frac{z^k e^{\frac{y^2}{2\sigma^2} z}}{\tau^2 + (1-\tau^2)z} dz = \left(\frac{2\sigma^2}{y^2}\right)^k \left(\int_{\frac{y^2}{2\sigma^2}\tau^2}^1 + \int_1^{\frac{y^2}{2\sigma^2}}\right) \frac{u^k e^u}{\frac{y^2}{2\sigma^2}\tau^2 + (1-\tau^2)u} du.$$

In the second integral the argument satisfies $u \ge 1$, and hence $u/((y^2\tau^2/(2\sigma^2) + (1-\tau^2)) \to 1$, uniformly in u and $y\tau \to 0$. Hence

$$\begin{aligned}\left(\frac{2\sigma^2}{y^2}\right)^k \int_1^{\frac{y^2}{2\sigma^2}} \frac{u^k e^u}{\frac{y^2}{2\sigma^2}\tau^2 + (1-\tau^2)u} du &\asymp \left(\frac{2\sigma^2}{y^2}\right)^k \int_1^{\frac{y^2}{2\sigma^2}} u^{k-1} e^u du \\ &\asymp \frac{2\sigma^2}{y^2} e^{\frac{y^2}{2\sigma^2}} (1+o(1))\end{aligned}$$

as $y \to \infty$, by Lemma A.5. For the first integral we separately consider the cases $k > 0$ and $k = -1/2$. If $k > 0$, then $\int_0^1 u^{k-1} e^u du$ converges, and hence, by the dominated convergence theorem, uniformly in $y\tau \to 0$,

$$\left(\frac{2\sigma^2}{y^2}\right)^k \int_{\tau^2 \frac{y^2}{2\sigma^2}}^1 \frac{u^k e^u}{\frac{y^2}{2\sigma^2}\tau^2 + (1-\tau^2)u} du \to \left(\frac{2\sigma^2}{y^2}\right)^k \int_0^1 u^{k-1} e^u du.$$

If $k = -1/2$, then we substitute $v = 2\sigma^2 u/(\tau^2 y^2)$ and rewrite the integral as

$$\left(\frac{2\sigma^2}{y^2}\right)^{-\frac{1}{2}} \int_1^{\frac{2\sigma^2}{\tau^2 y^2}} \frac{v^{-\frac{1}{2}} e^{\frac{\tau^2 y^2}{2\sigma^2} v}}{1+(1-\tau^2)v} \left(\frac{\tau^2 y^2}{2\sigma^2}\right)^{-\frac{1}{2}} dv = \frac{1}{\tau} \int_1^{\infty} \frac{v^{-1/2}}{1+v} dv(1+o(1)).$$

This combines with the integral (35). □

Proof of Theorem 3.5

Proof. Denote $\zeta_\tau = \sqrt{2\sigma^2 \log(1/\tau)}$ and assume that $\theta_i = \gamma\zeta_\tau$ for $i = 1, \ldots, p_n$ and $\theta_i = 0$ for $i = p_n + 1, \ldots, n$. We prove (7) by proving that there exists a positive constant $c_1(\gamma)$ such that

$$\mathbb{E}_{\theta=\gamma\zeta_\tau} T_\tau(Y) = \tau^{(1-\gamma)^2} \zeta_\tau^{2\gamma-2} c_1(\gamma)(1+o(1)). \tag{36}$$

If (36) holds, we have, by Jensen's inequality:

$$\sum_{i=1}^{p_n} \mathbb{E}_{\theta_i}(T_\tau(Y_i) - \theta_i)^2 \geq p_n(\tau^{(1-\gamma)^2} \zeta_\tau^{2\gamma-2} c_1(\gamma) - \gamma\zeta_\tau)^2 \gtrsim p_n \zeta_\tau^2, \tag{37}$$

as $\tau \to 0$. In addition, we have $T_\tau(y) = y I_{\frac{1}{2}}(y)/I_{-\frac{1}{2}}(y)$. For $|y| = \sqrt{2\sigma^2 c \log(1/\tau)}$, with $c > 1$, the lower bound (12) on $I_{\frac{1}{2}}(y)$ behaves as $(\sigma^2/y^2) e^{\frac{y^2}{2\sigma^2}}$, while the upper bound (15) on $I_{-\frac{1}{2}}(y)$ behaves as $(2a\sqrt{a}\sigma^2/y^2) e^{\frac{y^2}{2\sigma^2}}$, as $\tau \to 0$. Therefore, for $|y| > \zeta_\tau$, we have $T_\tau(y) \gtrsim y$. Thus, we can bound by:

$$\begin{aligned}
\sum_{i=p_n+1}^{n} \mathbb{E}_{\theta_i} T_\tau(Y_i)^2 &\geq (n-p_n) \mathbb{E}_{\theta=0} T_\tau(Y)^2 \mathbf{1}_{\{|Y|>\zeta_\tau\}} \gtrsim (n-p_n) \int_{\frac{\zeta_\tau}{\sigma}}^{\infty} y^2 \phi(y) dy \\
&= (n-p_n) \left(\int_{\frac{\zeta_\tau}{\sigma}}^{\infty} \phi(y) dy + \frac{\zeta_\tau}{\sigma} \phi\left(\frac{\zeta_\tau}{\sigma}\right) \right) \gtrsim (n-p_n) \zeta_\tau \phi\left(\frac{\zeta_\tau}{\sigma}\right) \\
&= (n-p_n) \frac{1}{\sqrt{2\pi}} \tau \zeta_\tau.
\end{aligned} \tag{38}$$

By combining the lower bounds (37) and (38) with the upper bound (2), we arrive at (7). For the posterior variance, we already have $\sum_{i=p_n+1}^{n} \mathrm{Var}(\theta_i \mid Y_i) \asymp (n-p_n)\tau\zeta_\tau$ by (24) and (26). Expression (8) can therefore be proven by showing that there exists a positive constant $c_2(\gamma)$ such that:

$$\mathbb{E}_{\theta=\gamma\zeta_\tau} \mathrm{Var}(\theta \mid Y) = \tau^{(1-\gamma)^2} \zeta_\tau^{2\gamma-1} c_2(\gamma)(1+o(1)). \tag{39}$$

Proof of (36)

The expected value $\mathbb{E}_{\theta=\gamma\zeta_\tau} T_\tau(Y)$ is equal to

$$\frac{1}{\sigma} \left(\int_{-\infty}^{-\frac{\zeta_\tau}{2}} + \int_{-\frac{\zeta_\tau}{2}}^{3\zeta_\tau} + \int_{3\zeta_\tau}^{\infty} \right) (\zeta_\tau + y) \frac{I_{\frac{1}{2}}(\zeta_\tau + y)}{I_{-\frac{1}{2}}(\zeta_\tau + y)} \phi\left(\frac{y + (1-\gamma)\zeta_\tau}{\sigma}\right) dy. \tag{40}$$

We shall show that the first and third integrals are negligible, while the second gives the approximation in (36). On the domain of the second integral, we have $\zeta_\tau/4 \le \zeta_\tau + y \le 4\zeta_\tau$, so we can apply Lemma A.6 to see that this integral is asymptotic to

$$\frac{1}{\sigma}\int_{-\frac{\zeta_\tau}{2}}^{3\zeta_\tau}(\zeta_\tau+y)\frac{c_2\tau^2(\zeta_\tau+y)^2+2\sigma^2 e^{\frac{y^2+2y\zeta_\tau}{2\sigma^2}}}{c_1(y+\zeta_\tau)^2+2\sigma^2 e^{\frac{y^2+2y\zeta_\tau}{2\sigma^2}}}\phi\left(\frac{y+(1-\gamma)\zeta_\tau}{\sigma}\right)dy, \tag{41}$$

where $c_1 = \int_0^\infty z^{-1/2}(1-z)^{-1}dz$ and $c_2 = \int_0^1 z^{1/2}(1-z)^{-1}dz$. On $[-\zeta_\tau/2, 3\zeta_\tau]$:

$$\begin{aligned} c_2\tau^2(\zeta_\tau+y)^2\phi\left(\frac{y+(1-\gamma)\zeta_\tau}{\sigma}\right) &\le \frac{c_2}{\sqrt{2\pi}}\tau^2(4\zeta_\tau)^3 e^{\frac{(1/2-\gamma)^2\zeta_\tau^2}{2\sigma^2}} \\ &= \frac{64c_2}{\sqrt{2\pi}}\zeta_\tau^3\tau^{2-(1/2-\gamma)^2}, \end{aligned}$$

so (41) is asymptotic to:

$$\mathcal{O}(\tau) + \frac{2\sigma}{\sqrt{2\pi}}e^{-\frac{(1-\gamma)^2\zeta_\tau^2}{2\sigma^2}}\int_{-\frac{\zeta_\tau}{2}}^{3\zeta_\tau}\frac{(\zeta_\tau+y)e^{\frac{\gamma\zeta_\tau y}{\sigma^2}}}{c_1(y+\zeta_\tau)^2+2\sigma^2 e^{\frac{y^2+2y\zeta_\tau}{2\sigma^2}}}dy.$$

By the substitution $u = \zeta_\tau y - 2\sigma^2\log\zeta_\tau$, the remaining integral is equal to, with $a_\tau = -\frac{\zeta_\tau^2}{2} - 2\sigma^2\log\zeta_\tau$ and $b_\tau = 3\zeta_\tau^2 - 2\sigma^2\log\zeta_\tau$:

$$\begin{aligned} &\frac{2\sigma}{\sqrt{2\pi}}\tau^{(1-\gamma)^2}\frac{1}{\zeta_\tau}\int_{a_\tau}^{b_\tau}\frac{(\zeta_\tau+\zeta_\tau^{-1}(u+2\sigma^2\log\zeta_\tau))e^{\frac{\gamma u}{\sigma^2}}\zeta_\tau^{2\gamma}}{c_1(\zeta_\tau+\zeta_\tau^{-1}(u+2\sigma^2\log\zeta_\tau))^2+2\sigma^2 e^{\frac{u}{\sigma^2}}\zeta_\tau^2 e^{\frac{(u+2\sigma^2\log\zeta_\tau)^2}{2\sigma^2\zeta_\tau^2}}}du \\ &\qquad \sim \frac{2\sigma}{\sqrt{2\pi}}\tau^{(1-\gamma)^2}\frac{1}{\zeta_\tau}\int_{-\infty}^{\infty}\frac{\zeta_\tau e^{\frac{\gamma u}{\sigma^2}}\zeta_\tau^{2\gamma}}{(c_1+2\sigma^2 e^{\frac{u}{\sigma^2}})\zeta_\tau^2}du, \end{aligned}$$

by the dominated convergence theorem. This yields the approximation in (36), with $c_1(\gamma) = (2\sigma/\sqrt{2\pi})\int_{-\infty}^{\infty} e^{\frac{\gamma u}{\sigma^2}}/(c_1+2\sigma^2 e^{\frac{u}{\sigma^2}})du$.

For the first integral in (40), we use bound 1 from Lemma A.2, and obtain a bound on its absolute value equal to

$$\begin{aligned} &\frac{1}{\sigma}\int_{-\infty}^{-\frac{\zeta_\tau}{2}}|\zeta_\tau+y|\tau e^{\frac{(\zeta_\tau+y)^2}{2\sigma^2}}\phi\left(\frac{y+(1-\gamma)\zeta_\tau}{\sigma}\right)dy \\ &\qquad = \frac{2}{3\sqrt{2\pi}\sigma}\tau^{(1-\gamma)^2}\int_{-\infty}^{-\frac{\zeta_\tau}{2}}|\zeta_\tau+y|e^{\frac{\gamma\zeta_\tau y}{\sigma^2}}dy \lesssim \tau^{(1-\gamma)^2}e^{-\frac{\gamma\zeta_\tau^2}{2\sigma^2}} = \tau^{(1-\gamma)^2+\gamma}, \end{aligned} \tag{42}$$

where the last inequality follows by integration by parts. This is of much smaller order than the second integral from (40). In the third integral of (40), we bound $I_{\frac{1}{2}}(\zeta_\tau+y)/I_{-\frac{1}{2}}(\zeta_\tau+y)$ by 1, giving the upper bound

$$\frac{1}{\sigma}\int_{3\zeta_\tau}^{\infty}(\zeta_\tau+y)\phi\left(\frac{y+(1-\gamma)\zeta_\tau}{\sigma}\right)dy \lesssim \phi\left(\frac{3\zeta_\tau+(1-\gamma)\zeta_\tau}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\tau^{4-\gamma},$$

by Mills' ratio. This is also of much smaller order than the second integral from (40), thus concluding the proof of (36).

Proof of (39)

By expanding the term $(1-z)^2$ in the numerator of the final term of (20), the posterior variance can be seen to be equal to:

$$\operatorname{Var}(\theta \mid y) = \sigma^2 \frac{I_{\frac{1}{2}}(y)}{I_{-\frac{1}{2}}(y)} + y^2 \left[\frac{I_{\frac{3}{2}}(y)}{I_{-\frac{1}{2}}(y)} - \left(\frac{I_{\frac{1}{2}}(y)}{I_{-\frac{1}{2}}(y)} \right)^2 \right]. \tag{43}$$

Because $I_{\frac{1}{2}}(y)/I_{-\frac{1}{2}}(y)$ can be interpreted as the mean of the density proportional to $z \to z^{-1/2} e^{y^2 z/(2\sigma^2)}/(\tau^2+(1-\tau^2)z)$, and $I_{\frac{3}{2}}(y)/I_{-\frac{1}{2}}(y)$ as the second moment, it follows that the term in square brackets in (43) is nonnegative. By (43), we write:

$$\begin{aligned}
\mathbb{E}_{\theta=\gamma\zeta_\tau} \operatorname{Var}(\theta \mid Y) = \sigma \int & \frac{I_{\frac{1}{2}}(\zeta_\tau + y)}{I_{-\frac{1}{2}}(\zeta_\tau + y)} \phi\left(\frac{y + (1-\gamma)\zeta_\tau}{\sigma}\right) dy \\
&+ \frac{1}{\sigma} \left(\int_{-\infty}^{-\frac{\zeta_\tau}{2}} + \int_{-\frac{\zeta_\tau}{2}}^{3\zeta_\tau} + \int_{3\zeta_\tau}^{\infty} \right) (\zeta_\tau + y)^2 \\
&\cdot \left[\frac{I_{\frac{3}{2}}(\zeta_\tau + y)}{I_{-\frac{1}{2}}(\zeta_\tau + y)} - \left(\frac{I_{\frac{1}{2}}(\zeta_\tau + y)}{I_{-\frac{1}{2}}(\zeta_\tau + y)} \right)^2 \right] \phi\left(\frac{y + (1-\gamma)\zeta_\tau}{\sigma}\right) dy.
\end{aligned} \tag{44}$$

The first term in (44) is as (40), except without the factor $(\zeta_\tau + y)$. Following the same steps as the proof of (36), we see that it is smaller than a multiple of ζ_τ^{-1} times the bound on (40), so it is of the order $\zeta_\tau^{2\gamma-3}\tau^{(1-\gamma)^2}$. The first and third integrals of the second term of (44) are also negligible. For the first, we use that the expression in square brackets is nonnegative and bounded above by $I_{\frac{3}{2}}(y)/I_{-\frac{1}{2}}(y)$, which in turn is bounded above by $I_{\frac{1}{2}}(y)/I_{-\frac{1}{2}}(y)$. We bound as in (42), with the difference that the leading factor is $(\zeta_\tau+y)^2$ instead of $(\zeta_\tau+y)$. This leads to the order $\zeta_\tau \tau^{(1-\gamma)^2+\gamma}$, much smaller than the claimed rate. For the third integral, we can bound the term in square brackets by 1 and use Mills' ratio to see that it is of the order $\zeta_\tau \tau^{(4-\gamma)^2}$.

We are left with the middle integral of the second term of (44). On the domain of this integral, by Lemma A.6:

$$\frac{I_{\frac{3}{2}}(\zeta_\tau + y)}{I_{-\frac{1}{2}}(\zeta_\tau + y)} = \frac{c_3 \tau^4 (\zeta_\tau + y)^2 + 2\sigma^2 e^{\frac{y^2 + 2y\zeta_\tau}{2\sigma^2}}}{c_1(\zeta_\tau + y)^2 + 2\sigma^2 e^{\frac{y^2+2y\zeta_\tau}{2\sigma^2}}}(1 + o(1)),$$

where $c_3 = \int_0^1 z^{3/2}(1+z)^{-1} dz$, and c_1 is as in (41). We see that $I_{\frac{3}{2}}(y)/I_{-\frac{1}{2}}(y)$ and $I_{\frac{1}{2}}(y)/I_{-\frac{1}{2}}(y)$ are asymptotic to the same function on this domain. Since $A/(A+B) - A^2/(A+B)^2 = AB/(A+B)^2$, it follows that up to $\mathcal{O}(\tau)$, the middle integral is asymptotic to

$$\frac{1}{\sigma}\int_{-\frac{\zeta_\tau}{2}}^{3\zeta_\tau}(\zeta_\tau+y)^2\frac{c_1(\zeta_\tau+y)^2 2\sigma^2 e^{\frac{y^2+2y\zeta_\tau}{2\sigma^2}}}{\left(c_1(\zeta_\tau+y)^2+2\sigma^2 e^{\frac{y^2+2y\zeta_\tau}{2\sigma^2}}\right)^2}\phi\left(\frac{y+(1-\gamma)\zeta_\tau}{\sigma}\right)dy$$

$$=\frac{2\sigma c_1}{\sqrt{2\pi}}\tau^{(1-\gamma)^2}\int_{-\frac{\zeta_\tau}{2}}^{3\zeta_\tau}\frac{(\zeta_\tau+y)^4 e^{\frac{\gamma\zeta_\tau y}{\sigma^2}}}{\left(c_1(\zeta_\tau+y)^2+2\sigma^2 e^{\frac{y^2+2\zeta_\tau y}{2\sigma^2}}\right)^2}dy.$$

We substitute $u=\zeta_\tau y-2\sigma^2\log\zeta_\tau$ to reduce this to

$$\frac{2\sigma c_1}{\sqrt{2\pi}}\tau^{(1-\gamma)^2}\frac{1}{\zeta_\tau}\int_{-\frac{\zeta_\tau^2}{2}-2\sigma^2\log\zeta_\tau}^{3\zeta_\tau^2-2\sigma^2\log\zeta_\tau}\frac{(\zeta_\tau+\zeta_\tau^{-1}(u+2\sigma^2\log\zeta_\tau))^4 e^{\frac{\gamma u}{\sigma^2}}\zeta_\tau^{2\gamma}}{\left(c_1(\zeta_\tau+\zeta_\tau^{-1}(u+2\sigma^2\log\zeta_\tau))^2+2\sigma^2 e^{\frac{u}{\sigma^2}}\zeta_\tau^2\right)^2}du$$

$$\sim\frac{2\sigma c_1}{\sqrt{2\pi}}\tau^{(1-\gamma)^2}\frac{1}{\zeta_\tau}\int_{-\infty}^{\infty}\frac{\zeta_\tau^4 e^{\frac{\gamma u}{\sigma^2}}\zeta_\tau^{2\gamma}}{\left(c_1\zeta_\tau^2+2\sigma^2\zeta_\tau^2 e^{\frac{u}{\sigma^2}}\right)^2}du.$$

This is asymptotic to expression (39), with $c_2(\gamma)=(2\sigma c_1/\sqrt{2\pi})\int_{-\infty}^{\infty}e^{\frac{\gamma u}{\sigma^2}}/(c_1+2\sigma^2 e^{\frac{u}{\sigma^2}})^2du$. □

Proof of Theorem 4.1

Proof. Suppose that $Y\sim\mathcal{N}(\theta,\sigma^2 I_n)$, $\theta\in\ell_0[p_n]$. We adapt the approach in Paragraph 6.2 in (Johnstone and Silverman, 2004). We first derive the following inequality for events A such that $\widehat{\tau}>\tau$ holds with probability one on A:

$$\begin{aligned}\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_A&\le 2\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-Y_i)^2\mathbf{1}_A+2\mathbb{E}_\theta(Y_i-\theta_i)^2\mathbf{1}_A\\&\lesssim 2\mathbb{E}_\theta\zeta_{\widehat{\tau}}^2\mathbf{1}_A+2\sigma^2\mathbb{E}_\theta Z^2\mathbf{1}_A\end{aligned}\tag{45}$$

where (17) was used in the second line, and Z follows a standard normal distribution. If A is such that $\widehat{\tau}>\tau$ holds with probability one on A, we can use the inequality $\zeta_{\widehat{\tau}}<\zeta_\tau$ if $\widehat{\tau}>\tau$ to find:

$$\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_A\lesssim 2\zeta_\tau^2\mathbb{P}_\theta(A)+2\sigma^2\mathbb{E}_\theta Z^2\mathbf{1}_A,\tag{46}$$

We now consider the nonzero and zero parameters separately. For both cases, we split up the expected ℓ_2 loss as follows:

$$\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2=\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_{\{\widehat{\tau}>c\tau\}}+\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_{\{\widehat{\tau}\le c\tau\}},$$

and then bound each of terms on the right hand side. For the nonzero means, we take $c=1$, while for the zero means, we consider $c\ge 1$. Note that for $\zeta_{\widehat{\tau}}$ to be well-defined, we need $\widehat{\tau}\le 1$ and consequently, when we consider $\widehat{\tau}>c\tau$, we must have $c\tau<1$.

Nonzero means

By (46), we find:

$$\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_{\{\widehat{\tau}>\tau\}}\lesssim 2\zeta_\tau^2+2\sigma^2.\tag{47}$$

If $\widehat{\tau} \leq \tau$, the inequality $\zeta_{\widehat{\tau}}^2 \leq \zeta_\tau^2$ needed for (46) does not hold. For this case, we assume that $\widehat{\tau} \geq g(n, p_n)$ with probability one, for some function $g(n, p_n)$, corresponding to $\zeta_{\widehat{\tau}} \leq \sqrt{-2\sigma^2 \log g(n, p_n)}$. Then we find by (45):

$$\begin{aligned}\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i) - \theta_i)^2 \mathbf{1}_{\{\widehat{\tau} \leq \tau\}} &\lesssim 2\mathbb{E}_\theta \zeta_{\widehat{\tau}}^2 \mathbf{1}_{\{\widehat{\tau} \leq \tau\}} + 2\sigma^2 \\ &\leq -4\sigma^2 \log(g(n, p_n))\mathbb{P}_\theta(\widehat{\tau} \leq \tau) + 2\sigma^2. \qquad (48)\end{aligned}$$

By (47) and (48), we have for $\theta_i \neq 0$:

$$\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i) - \theta_i)^2 \lesssim 1 + \zeta_\tau^2 - \log(g(n, p_n))\mathbb{P}_\theta(\widehat{\tau} \leq \tau). \qquad (49)$$

Zero means

We first establish an inequality for $\mathbb{E}_\theta[Z^2 \mathbf{1}_A]$, where A is an event and Z a standard normal random variable. By Young's inequality, we have for any positive x and y:

$$xy \leq \int_0^x (e^s - 1)ds + \int_0^y \log(s+1)ds = e^x - x - 1 + (y+1)\log(y+1) - y.$$

By this inequality combined with the inequality $\log(y+1) < y$, we have:

$$\mathbb{E}_\theta Z^2 \mathbf{1}_A \leq cd\mathbb{E}_\theta \left[e^{\frac{Z^2}{c}} - \frac{Z^2}{c} - 1 \right] + cd\mathbb{P}_\theta(A) \left(\frac{1}{d} \log\left(\frac{1}{d} + 1\right) - \frac{1}{d} \right).$$

With $c = 3$ and $d = \mathbb{P}_\theta(A)$, we find:

$$\begin{aligned}\mathbb{E}_\theta Z^2 \mathbf{1}_A &\leq (3\sqrt{3} - 4)\mathbb{P}_\theta(A) + 3\mathbb{P}_\theta(A) \log\left(1 + \frac{1}{\mathbb{P}_\theta(A)}\right) \\ &< 5\mathbb{P}_\theta(A) \log\left(1 + \frac{1}{\mathbb{P}_\theta(A)}\right). \qquad (50)\end{aligned}$$

By (46) and (50), we get for any $c \geq 1$ such that $c\tau < 1$:

$$\begin{aligned}\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i) - \theta_i)^2 \mathbf{1}_{\{\widehat{\tau} > c\tau\}} &\lesssim 2\zeta_\tau^2 \mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right) \\ &\quad + 10\sigma^2 \mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right) \log\left(1 + \frac{1}{\mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right)}\right). \qquad (51)\end{aligned}$$

Now suppose $\widehat{\tau} \leq c\tau$ for some $c \geq 1$ such that $c\tau < 1$. First note that $|T_\tau(y)|$ increases monotonically in τ, as is clear from

$$T_\tau(y_i) = \mathbb{E}[(1 - \kappa_i)y_i \mid y_i, \tau] = \mathbb{E}\left[\frac{\tau^2\lambda_i^2}{1 + \tau^2\lambda_i^2} y_i \;\middle|\; y_i, \tau \right].$$

Because $\mathrm{sign}(T_{\widehat{\tau}}(y_i)) = \mathrm{sign}(T_{c\tau}(y_i))$ and $0 \leq |T_{\widehat{\tau}}(y_i)| \leq |T_{c\tau}(y_i)|$, we have:

$$(T_{\widehat{\tau}}(y_i) - \theta_i)^2 \leq \max\{\theta_i^2, (T_{c\tau}(y_i) - \theta_i)^2\} \leq \theta_i^2 + (T_{c\tau}(y_i) - \theta_i)^2.$$

Hence:

$$\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_{\{\widehat{\tau}\le c\tau\}} \le \theta_i^2 + \mathbb{E}_\theta(T_{c\tau}(Y_i)-\theta_i)^2.$$

And thus, by (18), we have for $\theta_i = 0$:

$$\mathbb{E}_\theta(T_{\widehat{\tau}}(Y_i)-\theta_i)^2\mathbf{1}_{\{\widehat{\tau}\le c\tau\}} \lesssim \zeta_{c\tau}c\tau \lesssim \zeta_\tau\tau. \tag{52}$$

Combining (51) and (52), we find:

$$\mathbb{E}_\theta T_{\widehat{\tau}}(Y_i)^2 \lesssim \zeta_\tau\tau + \zeta_\tau^2\mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right) + \mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right)\log\left(1+\frac{1}{\mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right)}\right). \tag{53}$$

Conclusion

We can now bound the expected ℓ_2 loss. We assume that $\theta_i \neq 0$ for $i = 1,\dots,\tilde{p}_n$ and $\theta_i = 0$ for $i = \tilde{p}_n+1,\dots,n$, where $\tilde{p}_n \le p_n$. By combining (49) and (53), we find:

$$\begin{aligned}\mathbb{E}_\theta\|T_{\widehat{\tau}}(Y)-\theta\|^2 &\lesssim \tilde{p}_n\left(1+\zeta_\tau^2 - \log(g(n,p_n))\mathbb{P}_\theta(\widehat{\tau}\le\tau)\right) + (n-\tilde{p}_n)\zeta_\tau\tau \\ &\quad + (n-\tilde{p}_n)\mathbb{P}_\theta\left(\widehat{\tau} > c\tau\right)\left(\zeta_\tau^2 + \log\left(1+\frac{1}{\mathbb{P}_\theta\left(\widehat{\tau}>c\tau\right)}\right)\right).\end{aligned} \tag{54}$$

The function $x\log(1+\frac{1}{x})$ is monotonically increasing in x for $x \in [0,1]$. Hence, with the choice $\tau = \frac{p_n}{n}$ or $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$, the conditions stated in the theorem are sufficient for (54) to be bounded by the minimax squared error rate in the worst case.

If an estimator $\widehat{\tau}$ satisfies only the first condition, then $\sup\{\frac{1}{n},\widehat{\tau}\}$ satisfies the second condition with $-\log g(n,p_n) = \log n$. By the assumption $p_n \to \infty$, we have $\mathbb{P}_\theta(\sup\{\frac{1}{n},\widehat{\tau}\} > c\frac{p_n}{n}) \le \mathbb{P}_\theta(\widehat{\tau} > c\frac{p_n}{n})$. Plugging this into inequality (54) yields an ℓ_2 risk of at most order $p_n \log n$. ☐

Lemma A.7. *Suppose $Y_i \sim \mathcal{N}(\theta_i,\sigma^2), i = 1,\dots,p_n$ and $Y_i \sim \mathcal{N}(0,\sigma^2), i = p_n+1,\dots,n$ and define*

$$\widehat{\tau} = \frac{\#\{|y_i| \ge \sqrt{c_1\sigma^2\log n}, i = 1,\dots,n\}}{c_2 n}$$

for some $c_2 > 1$. Then $\mathbb{P}_\theta(\widehat{\tau} > \tau) \lesssim \frac{p_n}{n}$ as $p_n, n \to \infty$, $p_n = o(n)$ if $c_1 > 2$, or $c_1 = 2$ and $p_n \lesssim \log n$ for $\tau = \frac{p_n}{n}$ or $\tau = \frac{p_n}{n}\sqrt{\log(n/p_n)}$.

Proof. We only need to consider $\mathbb{P}_\theta(\widehat{\tau} > \frac{p_n}{n})$, as we assume $p_n = o(n)$ and thus, for large n, $\mathbb{P}_\theta(\widehat{\tau} > \frac{p_n}{n}\sqrt{\log(n/p_n)}) \le \mathbb{P}_\theta(\widehat{\tau} > \frac{p_n}{n})$. Define $A_i = \{|y_i| \ge \sqrt{c_1\sigma^2\log n}\}, i = 1,\dots,n$. For $i = p_n+1,\dots,n$, $\mathbf{1}_{A_i}$ follows a Bernoulli distribution with parameter $q_n = 2\Phi^c(\sqrt{c_1\log n})$, which by Mills' ratio can be bounded from above by $\sqrt{\frac{2}{c_1\pi}}(\log n)^{-\frac{1}{2}}n^{-\frac{c_1}{2}}$. For $X \sim \text{Bin}(n,p)$, we have the bound $\mathbb{P}(X \geq k) \leq (\frac{enp}{k})^k$ as a consequence of Theorem 1 in (Chernoff, 1952). Hence:

$$\mathbb{P}_\theta\left(\widehat{\tau} > \frac{p_n}{n}\right) \leq \mathbb{P}_\theta\left(\sum_{i=p_n+1}^{n} \mathbf{1}_{A_i} > (c_2-1)p_n\right) \leq \left(\frac{e(n-p_n)q_n}{(c_2-1)p_n+1}\right)^{(c_2-1)p_n+1}$$
$$\leq \left(\sqrt{\frac{2e^2}{c_1\pi}}\frac{1}{(c_2-1)p_n+1}\frac{1}{\sqrt{\log n}}n^{1-\frac{c_1}{2}}\right)^{(c_2-1)p_n+1}. \tag{55}$$

The inequality $\mathbb{P}_\theta(\widehat{\tau} > \frac{p_n}{n}) \lesssim \frac{p_n}{n}$ holds if $-\log \mathbb{P}_\theta(\widehat{\tau} > \frac{p_n}{n}) \geq \log \frac{n}{p_n} + c$ holds for some positive constant c. The negative logarithm of bound (55) is:

$$((c_2-1)p_n+1)\left(\frac{1}{2}\log\frac{c_1\pi}{2e^2} + \log((c_2-1)p_n+1) + \frac{1}{2}\log\log n + \left(\frac{c_1}{2}-1\right)\log n\right).$$

For $c_1 = 2$, this quantity will exceed $\log \frac{n}{p_n}$ if $p_n \gtrsim \log n$. If $c_1 > 2$, we require $((c_2-1)p_n+1)(\frac{c_1}{2}-1) \geq 1$, which is certainly satisfied if $p_n \to \infty$. □

References

Armagan, A., Dunson, D. B. and Lee, J. (2013). Generalized Double Pareto Shrinkage. *Statistica Sinica* **23** 119–143. MR3076161

Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B. (2012). Bayesian Shrinkage. arXiv:1212.6088.

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *The Annals of Statistics* **37** 1705–1732. MR2533469

Bogdan, M., Ghosh, J. K. and Tokdar, S. T. (2008). A Comparison of the Benjamini-Hochberg Procedure with Some Bayesian Rules for Multiple Testing. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen.* The Institute of Mathematical Statistics.

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

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The Horseshoe Estimator for Sparse Signals. *Biometrika* **97** 465–480. MR2650751

Castillo, I., Schmidt-Hieber, J. and Van der Vaart, A. W. (2014). Bayesian Linear Regression with Sparse Priors. arXiv:1403.0735.

Castillo, I. and Van der Vaart, A. W. (2012). Needles and Straw in a Haystack: Posterior Concentration for Possibly Sparse Sequences. *The Annals of Statistics* **40** 2069–2101. MR3059077

Chernoff, H. (1952). A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations. *The Annals of Mathematical Statistics* **23** 493–507. MR0057518

Datta, J. and Ghosh, J. K. (2013). Asymptotic Properties of Bayes Risk for the Horseshoe Prior. *Bayesian Analysis* **8** 111–132. MR3036256

Donoho, D. L., Johnstone, I. M., Hoch, J. C. and Stern, A. S. (1992). Maximum Entropy and the Nearly Black Object (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)* **54** 41–81. MR1157714

Efron, B. (2008). Microarrays, Empirical Bayes and the Two-Groups Model. *Statistical Science* **23** 1–22. MR2431866

Ghosal, S., Ghosh, J. K. and Van der Vaart, A. W. (2000). Convergence Rates of Posterior Distributions. *The Annals of Statistics* **28** 500–531. MR1790007

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of Integrals, Series and Products*. Academic Press.

Griffin, J. E. and Brown, P. J. (2010). Inference with Normal-Gamma Prior Distributions in Regression Problems. *Bayesian Analysis* **5** 171–188. MR2596440

Jiang, W. and Zhang, C.-H. (2009). General Maximum Likelihood Empirical Bayes Estimation of Normal Means. *The Annals of Statistics* **37** 1647–1684. MR2533467

Johnstone, I. M. and Silverman, B. W. (2004). Needles and Straw in Haystacks: Empirical Bayes Estimates of Possibly Sparse Sequences. *The Annals of Statistics* **32** 1594–1649. MR2089135

Koenker, R. (2014). A Gaussian Compound Decision Bakeoff. *Stat.* **3** 12–16.

Koenker, R. and Mizera, I. (2014). Convex Optimization, Shape Constraints, Compound Decisions and Empirical Bayes Rules. *Journal of the American Statistical Association* **109** 674–685. MR3223742

Martin, R. and Walker, S. G. (2014). Asymptotically Minimax Empirical Bayes Estimation of a Sparse Normal Mean Vector. *Electronic Journal of Statistics* **8** 2188–2206. MR3273623

Miller, P. D. (2006). *Applied Asymptotic Analysis*. *Graduate Studies in Mathematics* **75**. The American Mathematical Society. MR2238098

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association* **83** 1023–1032. MR0997578

Pericchi, L. R. and Smith, A. F. M. (1992). Exact and Approximate Posterior Moments for a Normal Location Parameter. *Journal of the Royal Statistical Society. Series B (Methodological)* **54** 793–804. MR1185223

Polson, N. G. and Scott, J. G. (2010). Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction. In *Bayesian Statistics 9* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) Oxford University Press. MR3204017

Polson, N. G. and Scott, J. G. (2012a). Good, Great or Lucky? Screening for Firms with Sustained Superior Performance Using Heavy-Tailed Priors. *The Annals of Applied Statistics* **6** 161–185. MR2951533

Polson, N. G. and Scott, J. G. (2012b). On the Half-Cauchy Prior for a Global Scale Parameter. *Bayesian Analysis* **7** 887–902. MR3000018

Scott, J. G. (2010). Parameter Expansion in Local-Shrinkage Models. arXiv:1010.5265.

Scott, J. G. (2011). Bayesian Estimation of Intensity Surfaces on the Sphere via Needlet Shrinkage and Selection. *Bayesian Analysis* **6** 307–328. MR2806246

Scott, J. G. and Berger, J. O. (2010). Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem. *The Annals of Statistics* **38** 2587–2619. MR2722450

Shafer, R. E. (1966). Elementary Problems. Problem E 1867. *The American Mathematical Monthly* **73** 309.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58** 267–288. MR1379242

Yuan, M. and Lin, Y. (2005). Efficient Empirical Bayes Variable Selection and Estimation in Linear Models. *Journal of the American Statistical Association* **100** 1215–1225. MR2236436