Bayesian variance estimation in the Gaussian sequence model with partial information on the means

Gianluca Finocchio[*] **and Johannes Schmidt-Hieber**[**]

University of Twente
e-mail: [*]g.finocchio@utwente.nl; [**]a.j.schmidt-hieber@utwente.nl

Abstract: Consider the Gaussian sequence model under the additional assumption that a fixed fraction of the means is known. We study the problem of variance estimation from a frequentist Bayesian perspective. The maximum likelihood estimator (MLE) for σ^2 is biased and inconsistent. This raises the question whether the posterior is able to correct the MLE in this case. By developing a new proving strategy that uses refined properties of the posterior distribution, we find that the marginal posterior is inconsistent for any i.i.d. prior on the mean parameters. In particular, no assumption on the decay of the prior needs to be imposed. Surprisingly, we also find that consistency can be retained for a hierarchical prior based on Gaussian mixtures. In this case we also establish a limiting shape result and determine the limit distribution. In contrast to the classical Bernstein-von Mises theorem, the limit is non-Gaussian. We show that the Bayesian analysis leads to new statistical estimators outperforming the correctly calibrated MLE in a numerical simulation study.

Keywords and phrases: Frequentist Bayes, maximum likelihood, semiparametric inference, Gaussian sequence model, Bernstein-von Mises theorems.

1. Introduction

For given $0 \leq \alpha \leq 1$, suppose we observe n independent and normally distributed random variables

$$X_i \sim \mathcal{N}\big(\mu_i^0 \mathbf{1}(i > n\alpha), \sigma_0^2\big), \quad i = 1, \ldots, n. \tag{1.1}$$

The parameters in the model are μ_i^0, $i > n\alpha$ and $\sigma_0 > 0$. The goal is to estimate the variance σ_0^2 while treating the mean vector $\mu_0 := (\mu_{\lceil n\alpha \rceil}^0, \ldots, \mu_n^0)$ as nuisance. For $\alpha = 0$, we recover the Gaussian sequence model. For $\alpha > 0$, this can be viewed as the Gaussian sequence model with additional knowledge that the means of the first $\lfloor n\alpha \rfloor$ observations are known (in which case we can subtract them from the data).

One can think of model (1.1) as a simple prototype of a combined dataset. Using for instance different measurement devices, one often faces merged datasets collected from multiple sources. The different sources might not be of the same

239

quality concerning the underlying parameter, see [24] for an example. An alternative viewpoint is to interpret model (1.1) as a sparse sequence model with known support. Since a $(1 - \alpha)$-fraction of the data is perturbed, we are in the dense regime. Knowledge of the support is then crucial as otherwise there is no consistent estimator for σ_0^2.

If n is even and $\alpha = 1/2$, then (1.1) is equivalent to the Neyman-Scott model [25] up to a reparametrization. Model (1.1) is in this case equivalent to observing $U_i := (X_{n/2+i} + X_i)$ and $V_i := (X_{n/2+i} - X_i)$ for $i = 1, \ldots, n/2$. Since U_i and V_i are independent, this is thus equivalent to observing independent random variables $U_i, V_i \sim \mathcal{N}(\mu_{n/2+i}^0, \widetilde{\sigma}_0^2)$, with $\widetilde{\sigma}_0^2 = 2\sigma_0^2$. Estimation of $\widetilde{\sigma}_0^2$ in the latter model is known as Neyman-Scott problem.

Although σ_0^2 can be estimated with parametric rate based on the first $n\alpha$ observations, a striking feature of the model is that the MLE for σ_0^2 is inconsistent. In fact the MLE $\widehat{\sigma}_{\mathrm{mle}}^2$ converges to $\alpha\sigma_0^2$ therefore underestimating the true variance by the factor α. The reason is that the likelihood of the observations with non-zero mean significantly affects the total likelihood viewed as a function in σ^2.

We study what happens when a Bayesian approach is implemented for the estimation of the variance and whether the posterior distribution can correct for the bias of the MLE. The Bayesian method can be viewed as a weighted likelihood method: instead of taking the parameter with the largest likelihood, the posterior puts mass on parameter sets with large likelihood. Because of this, the posterior can in some cases correct the flaws of the MLE. An example are irregular models, see [15, 11, 26].

In the first part of the paper, we prove that whenever the nuisances are independently generated from a proper distribution, the posterior does not contract around the true variance. This shows that, for a large class of natural priors, the Bayesian method is unable to correct the MLE. In frequentist Bayes, several lower bound techniques have been developed in order to describe when Bayesian methods do not work, [4, 8, 9, 29, 10, 19]. These results can be used for instance to show that a certain decay of the prior is necessary to ensure posterior contraction. Our lower bounds are of a different flavor and do not require a condition on the tail behavior.

Since for the non-zero means no additional structure is assumed, there is no way to say something about one mean from knowledge of all the other means. Therefore, one might be tempted to think that a correlated prior on the means cannot perform better than an i.i.d. prior and consequently must lead to an inconsistent posterior as well. Surprisingly, this is not true and we construct in the second part of the article a Gaussian mixture prior for which the posterior contracts with the parametric rate around the true variance. For this prior we derive the limit distribution in the Bernstein-von Mises sense. In contrast with the classical Bernstein-von Mises theorem, the posterior limit is non-Gaussian in the case of small means. In this case the posterior also incorporates information about the second part of the sample into the estimator and we show in a simulation study that the maximum a posteriori estimate based on the limit distribution outperforms the $\sqrt{n}$-consistent estimator that only uses the observations with zero mean.

Estimation of the variance in model (1.1) can also be interpreted as a semi-parametric problem. The results in this article therefore contribute to the recent efforts to understand frequentist Bayes in semiparametric models. Semiparametric Bernstein-von Mises theorems are derived under various conditions in [27, 5, 3, 7]. For specific priors, it has been observed that there can be a large bias in the posterior limit, see [6, 7, 26]. In all the cases studied so far, it is unclear whether the bias is due to the specific choice of prior or whether this is a fundamental limitation of the Bayesian method. To the best of our knowledge, our results show for the first time, that the posterior can be inconsistent for all natural priors.

Related to model (1.1), [14] studies Bayes for variance estimation of the errors in the nonparametric regression model. It is shown that if the posterior contracts around the true regression function with rate $o(n^{-1/4})$, the marginal posterior for the variance contracts with parametric rate around the true error variance and a Bernstein-von Mises result holds.

The article is organized as follows. In Section 2, we discuss aspects of the problem related to the likelihood and the posterior distribution. A crucial identity for the log-posterior is derived in Section 3. This leads then to the general negative result in Section 4. The Gaussian mixture prior with parametric posterior contraction is constructed in Section 5. This section also contains the limiting shape result and a numerical simulation study. All proofs are deferred to the appendix.

Notation: For a vector $u = (u_1, \dots, u_k)$, we write $\|u\|^2 = \sum_{i=1}^{k} u_i^2$ and $\overline{u^2} = \|u\|^2/k$ for the average of the squares (not to be confused with the squared average). We write $n_1 = \lfloor n\alpha \rfloor$ and $n_2 = n - n_1$. The probability and expectation induced by model (1.1) are denoted by P_0^n and E_0^n.

2. Likelihood and posterior

The MLE For the subsequent analysis, it is convenient to split the data vector $X = (X_1, \dots, X_n)$ in the part with zero means $Y = (X_1, \dots, X_{n_1})$ and the observations with non-zero means $Z = (X_{n_1+1}, \dots, X_n)$ such that $X = (Y, Z)$. The likelihood function of the model is

$$
\begin{aligned}
L(\sigma^2, \mu | Y, Z) &= \underbrace{\frac{1}{(2\pi\sigma^2)^{n_1/2}} e^{-\frac{\|Y\|^2}{2\sigma^2}}}_{L(\sigma^2, \mu | Y)} \underbrace{\frac{1}{(2\pi\sigma^2)^{n_2/2}} e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}}}_{L(\sigma^2, \mu | Z)} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|Y\|^2 + \|Z-\mu\|^2}{2\sigma^2}}.
\end{aligned}
\tag{2.1}
$$

Maximizing over (σ^2, μ) yields the MLE

$$
\left(\widehat{\sigma}_{\mathrm{mle}}^2, \widehat{\mu}_{\mathrm{mle}}\right) = \left(\frac{\|Y\|^2}{n}, Z\right).
\tag{2.2}
$$

If only based on the subsample Y, the MLE for σ_0^2 would be $\|Y\|^2/n_1$ and this converges to σ_0^2 with the parametric rate $n^{-1/2}$. Hence $\|Y\|^2/n$ converges to $\alpha\sigma_0^2$. The MLE for σ_0^2 is therefore inconsistent and misses the true parameter σ_0^2

by a factor α. It is clear that there is very little extractable information about the parameter σ_0^2 in Z. A frequentist estimator can simply discard Z and only use the subsample Y. The MLE also does this but leads to an incorrect scaling of the estimator.

The incorrect scaling factor of the MLE can be explained in different ways. One interpretation is that the MLE can be written as

$$\widehat{\sigma}_{\text{mle}}^2 = \frac{n_1}{n}\widehat{\sigma}_{Y,\text{mle}}^2 + \frac{n_2}{n}\widehat{\sigma}_{Z,\text{mle}}^2,$$

with $\widehat{\sigma}_{Y,\text{mle}}^2 = \|Y\|^2/n_1$ the MLE based on the subsample Y and $\widehat{\sigma}_{Z,\text{mle}}^2 = 0$ the MLE based on the subsample Z. The fact that the overall MLE just forms a linear combination of the MLEs for the subsamples shows again that too much weight is given to Z.

Another explanation for the incorrect scaling of the MLE is to observe that in (2.1) the likelihood based on the second subsample is $L(\sigma^2, \mu|Z) \propto \sigma^{-n_2}$ if $\mu = \widehat{\mu}_{\text{mle}}$. If we would take the likelihood only over the first part of the sample Y we would obtain the optimal estimator $\|Y\|^2/n_1$, but since the likelihhod over the full sample is the product of the likelihood functions for Y and Z, an additional factor σ^{-n_2} occurs in the overall likelihood which leads to the incorrect scaling. More generally, we conjecture that likelihood methods do not perform well for combined datasets where one part of the data is informative about a parameter and the other part is affected by nuisance parameters.

Adjusted profile likelihood For the profile likelihood, we first compute the maximum likelihood estimator of the nuisance parameter for fixed σ^2, denoted by, say $\widehat{\mu}_{\sigma^2}$, and then maximize

$$\sigma^2 \mapsto L\big(\sigma^2, \widehat{\mu}_{\sigma^2}\big|Y, Z\big).$$

Obviously $\widehat{\mu}_{\sigma^2} = Z$ for any $\sigma^2 > 0$ and the profile likelihood estimator coincides with the MLE for σ^2 in the Neyman-Scott problem. If the parameter of interest and the nuisance parameters are orthogonal with respect to the Fisher information, that is,

$$E\Big[\frac{\partial^2}{\partial\sigma^2\partial\mu_j} \log L\big(\sigma^2, \mu\big|Y, Z\big)\Big] = 0, \quad \text{for all } j \tag{2.3}$$

the adjusted profile likelihood estimator [12, 23, 13] is the maximizer of

$$\sigma^2 \mapsto \mathcal{L}(\sigma^2) := \det\big(M(\sigma^2, \widehat{\mu}_{\sigma^2})\big)^{-1/2} L\big(\sigma^2, \widehat{\mu}_{\sigma^2}\big|Y, Z\big) \tag{2.4}$$

for the matrix valued function

$$M(\sigma^2, \mu) := \Big(-\frac{\partial^2}{\partial\mu_j\partial\mu_\ell} \log L\big(\sigma^2, \mu\big|Y, Z\big)\Big)_{j,\ell}$$

and det() the determinant. It is easy to check that (2.3) holds for model (1.1). Since $-\partial^2/(\partial\mu_j\partial\mu_\ell) \, \log L(\sigma^2, \mu|Y, Z) = \sigma^{-2}\mathbf{1}(j = \ell)$, the adjusted profile likelihood estimator for σ^2 coincides with the MLE for the subsample Y,

$$\widehat{\sigma}^2 = \frac{\|Y\|^2}{n_1}.$$

In particular, the adjusted profile likelihood results in an unbiased $\sqrt{n}$-consistent estimator for σ^2.

The posterior distribution From a Bayesian perspective it is quite natural to draw σ^2 and the mean vector μ from independent distributions. Due to the orthogonality with respect to the Fisher information (2.3), we expect no strong interactions of σ^2 and the mean parameters in the likelihood that could be taken care of by a dependent prior. Suppose that $\mu \sim \nu$ and that the prior for σ^2 has Lebesgue density π. The marginal posterior distribution is then given by Bayes formula

$$\pi(\sigma^2|Y, Z) = \frac{L(\sigma^2|Y, Z)\pi(\sigma^2)}{\int_{\mathbb{R}_+} L(\sigma^2|Y, Z)\pi(\sigma^2) \, d\sigma^2}, \tag{2.5}$$

with

$$L(\sigma^2|Y, Z) = \sigma^{-n}e^{-\frac{\|Y\|^2}{2\sigma^2}} \Big(\int_{\mathbb{R}^n} e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}} \, d\nu(\mu) \Big). \tag{2.6}$$

In [28] it has been argued that by using multivariate Laplace approximation,

$$L(\sigma^2|Y, Z) = \mathcal{L}(\sigma^2)\nu(\widehat{\mu}_{\sigma^2})(1 + O_P(n^{-1})) = \mathcal{L}(\sigma^2)\nu(Z)(1 + O_P(n^{-1})), \tag{2.7}$$

with $\mathcal{L}(\sigma^2)$ the adjusted profile likelihood in (2.4). This suggests that the posterior distribution should be centered around the adjusted profile likelihood estimator $\|Y\|^2/n_1$, therefore correcting the MLE.

Associated sequence model with random means For the Gaussian sequence model with partial information (1.1) equipped with the product prior $\pi \otimes \nu$, define the *associated sequence model with random means,* where we observe independent random variables

$$Y_i \sim \mathcal{N}(0, \sigma_0^2), \; i = 1, \ldots, n_1 \; \text{ and } \; Z_i|\mu \sim \mathcal{N}(\mu_i, \sigma_0^2), \; i = n_1 + 1, \ldots, n, \tag{2.8}$$

with $\mu \sim \nu$ and ν known. In this model, the nuisance parameters are replaced by additional randomness. The only parameter in this model is σ_0^2 and the model is therefore parametric.

Remark 2.1. The likelihood function of model (2.8) is $L(\sigma^2|Y, Z)$. Model (1.1) and model (2.8) lead therefore to the same formula for the posterior distribution of σ^2 in terms of Y, Z.

Bayes with improper uniform prior If the prior on the mean vector in the Bayes formula is chosen as the Lebesgue measure, the formula for the posterior

simplifies to

$$\pi\big(\sigma^2\big|Y,Z\big) \propto \sigma^{-n_1} e^{-\frac{\|Y\|^2}{2\sigma^2}} \pi(\sigma^2). \qquad (2.9)$$

This is the same posterior we would get if we discarded the subsample Z. It follows from the parametric Bernstein-von Mises theorem that if π is positive and continuous in a neighbourhood of σ_0^2, the posterior contracts around the true variance σ_0^2. Notice that in the case of uniform prior, the Laplace approximation in (2.7) is exact and does not involve any remainder terms. Obviously the Lebesgue measure is not a probability measure and the prior is improper. This raises then the question whether there are also proper priors for which the marginal posterior is consistent on the whole parameter space. We will address this problem in the next sections.

3. On the derivative of the log-posterior

We first derive a differential equation for the posterior. Denote by $\mu|(Z,\sigma^2)$ the posterior distribution of μ for the sample Z, that is,

$$d\Pi(\mu|Z,\sigma^2) = \frac{e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}} d\nu(\mu)}{\int_{\mathbb{R}^n} e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}} d\nu(\mu)}. \qquad (3.1)$$

In particular, we set

$$V\big(\mu|(Z,\sigma^2)\big) := \int_{\mathbb{R}^n} \|Z-\mu\|^2 d\Pi(\mu|Z,\sigma^2). \qquad (3.2)$$

The quantity $V(\mu|(Z,\sigma^2))$ measures the spread of $\Pi(\mu|Z,\sigma^2)$ around the vector Z. Recall moreover the definition of $L(\sigma^2|Y,Z)$ in (2.6).

Proposition 3.1. *The marginal posterior satisfies*

$$\partial_{\sigma^2} \log \frac{\pi(\sigma^2|Y,Z)}{\pi(\sigma^2)} = \partial_{\sigma^2} \log L(\sigma^2|Y,Z) = \frac{\|Y\|^2 + V(\mu|(Z,\sigma^2))}{2\sigma^4} - \frac{n}{2\sigma^2}. \quad (3.3)$$

By Remark 2.1, the right hand side is a closed-form expression of the score function for σ^2 in the random means model (2.8). If the MLE in (2.8) does not lie on the boundary, the score function vanishes at the MLE. From the Bernstein-van Mises phenomenon it is conceivable that the posterior will concentrate around this MLE. For the MLE to be close to the truth σ_0^2, the score function evaluated at σ_0^2 must be $o_P(1)$. Since $\|Y\|^2 = n\alpha\sigma_0^2 + O_P(\sqrt{n})$, this leads to the condition

$$\frac{V(\mu|(Z,\sigma_0^2))}{n} = (1-\alpha)\sigma_0^2 + o_P(1).$$

In the next section, we derive a very general negative result. The main part of the argument is to show that the previous equality does not hold in a neighborhood of σ_0^2, see (A.12).

4. Posterior inconsistency for product priors

In this section we study posterior contraction under the following condition.

Prior The prior on μ is independent of the prior on σ^2. Under the prior, each component of the mean vector μ is drawn independently from a distribution ν on $\mathbb{R}$. The prior on σ^2 has a positive and continuously differentiable Lebesgue density on $\mathbb{R}_+$.

So far, ν denoted the prior on the mean vector. By a slight abuse of language we denote the prior on the individual components also by ν. The assumptions on the prior are mild enough to account for proper priors with heavy tails and possibly no moments.

The i.i.d. prior is the natural choice, if we believe that there is no structure in the non-zero means. From (2.8) it follows that the corresponding sequence model with random means is

$$Y_i \sim \mathcal{N}(0, \sigma_0^2), \ i = 1, \ldots, n_1 \text{ and } Z_i | \mu_i \sim \mathcal{N}(\mu_i, \sigma_0^2), \ i = n_1 + 1, \ldots, n, \quad (4.1)$$

with $\mu_i \sim \nu$. For $\alpha = 1/2$ and unknown ν, this model has been studied in [21]. It is shown that the MLE for σ_0^2 and the MLE for the distribution function of the means are consistent. Since the random means model leads to the same posterior distribution as explained in Remark 2.1, this suggests that the posterior might concentrate around the truth.

We now provide a second heuristic that leads to a different conclusion indicating that it makes a huge difference whether the distribution of the means ν is known or unknown. In the framework of (4.1), ν is known. If $\int u^2 d\nu(u) < \infty$, then $\overline{\mu^2} = \int u^2 d\nu(u) + O_P(n^{-1/2})$ and $\overline{Z^2} = \overline{\mu^2} + \sigma_0^2 + O_P(n^{-1/2})$, so we have $\overline{Z^2} - \int u^2 d\nu(u) = \sigma_0^2 + O_P(n^{-1/2})$. This means that model (4.1) carries a lot of information about σ_0^2 in the sense that σ_0^2 can be estimated with parametric rate from the subsample Z only. Since the posterior only sees model (4.1) it is therefore natural to give a lot of weight to the subsample Z as well, which, from a frequentist perspective, is wrong.

This heuristic does not say anything about heavy-tailed priors with $\int u^2 d\nu(u) = \infty$. But even in this case, we will show that the posterior is inconsistent. The first result states that in a neighborhood of σ_0^2 the posterior is increasing extremely fast with high probability.

Proposition 4.1. *Given $\alpha < 1$ and the prior above, then, for all sufficiently large σ_0^2, there exists a mean vector μ_0, such that*

$$\lim_{n \to \infty} P_0^n \left(\left\{ \partial_{\sigma^2} \log \pi(\sigma^2 | Y, Z) \geq \sigma_0^{-2} n, \ \forall \sigma^2 \in \left[\frac{\sigma_0^2}{2}, 2\sigma_0^2 \right] \right\} \right) = 1.$$

The proof of Proposition 4.1 constructs a lower bound on σ_0^2 that is independent of n and moreover guarantees that ν has sufficiently small mass outside $[-\sigma_0^2, \sigma_0^2]$. It therefore depends on the tail behavior of the prior mean distribution ν. The mean vector μ_0 is subsequently chosen with all means being equal to an expression only depending on σ_0^2. Thus the means in μ_0 are uniformly bounded and independent of n as well.

Suppose that almost all posterior mass is close to σ_0^2. By the previous proposition, the posterior is increasing at least up to $2\sigma_0^2$. Hence, there must be even more mass around $2\sigma_0^2$. This is a contradiction and shows that the posterior does not concentrate around σ_0^2. The proof of the next theorem is based on this argument. For this result, the means in the vector μ_0 can again be chosen to be uniformly bounded.

Theorem 4.2. *Given $\alpha < 1$ and the prior above, then, for all sufficiently large σ_0^2, there exists a mean vector μ_0 such that*

$$\lim_{n \to \infty} E_0^n \left[\Pi \left(\left| \frac{\sigma^2}{\sigma_0^2} - 1 \right| \leq \frac{1}{2} \Big| Y, Z \right) \right] = 0.$$

Consequently, the posterior is inconsistent and assigns all its mass outside of a neighbourhood of the true variance.

The posterior is therefore inferior if compared to the frequentist variance estimator $\overline{Y^2}$, which achieves the parametric rate $n^{-1/2}$ in the sense that

$$\sup_{\sigma_0^2 > 0} E_0^n \left[\left| \frac{\overline{Y^2}}{\sigma_0^2} - 1 \right| \right] \lesssim n^{-1/2}.$$

It is remarkable that no conditions on the tail behavior of the prior distribution ν are required for Theorem 4.2. Recall that for the improper uniform prior the posterior contracts around σ_0^2. This shows that for distributions with heavy tailed densities, very sharp bounds are required.

To the best of our knowledge there are no negative results in the nonparametric Bayes literature that hold for such a large class of priors. The proof strategy to establish Proposition 4.1 is based on a highly non-standard shrinkage argument that will be sketched here. By expanding the square term in (3.2) we can lower bound (3.3) by

$$\partial_{\sigma^2} \log \pi(\sigma^2 | Y, Z) \geq \frac{\|Y\|^2}{2\sigma^4} + \frac{\|Z\|^2}{2\sigma^4} - \frac{n}{2\sigma^2} - \frac{1}{\sigma^4} \sum_{i=1}^{n_2} V_i + O_P(1),$$

where $V_i := |Z_i| \int |\mu_i| d\Pi(\mu | Z_i, \sigma^2)$. For σ^2 close to σ_0^2, we have

$$\partial_{\sigma^2} \log \pi(\sigma^2 | Y, Z) \geq \frac{n_2 \overline{\mu_0^2}}{2\sigma_0^4} - \frac{1}{\sigma_0^4} \sum_{i=1}^{n_2} V_i + O_P(\sqrt{n}).$$

For an improper uniform prior, one can check that $V_i \geq Z_i^2$, making the lower bound negative and useless. For a proper prior, there is a shrinkage phenomenon in the sense that for any $c > 0$ there are parameters $(\mu_i^0)^2 \asymp \sigma_0^2$ such that $V_i \leq c Z_i^2$, with high P_0^n-probability. If this is the case then

$$\partial_{\sigma^2} \log \pi(\sigma^2 | Y, Z) \geq \left(\frac{1}{2} - 2c \right) \frac{n_2}{2\sigma_0^2} + O_P(\sqrt{n}),$$

which yields the conclusion by choosing $c > 0$ small enough.

In Proposition 4.1 we showed that the posterior overshoots the true variance σ_0^2 whenever the true means are large enough. By analyzing the Gaussian case in the next section, we see that for small means the posterior will in fact underestimate σ_0^2 and that only for a small range of mean vectors, one can hope that the posterior will be able to concentrate around the true variance.

5. Gaussian mixture priors

5.1. Gaussian priors

To illustrate our approach, we first consider an i.i.d. Gaussian prior on the mean vector

$$\mu_i \sim \mathcal{N}(0, \theta^2), \text{ independently.}$$

From Theorem 4.2 we already know that the posterior will be inconsistent in this case. Nevertheless, the Gaussian assumption yields more explicit formulas and this allows us to build a hierarchical prior resulting in good posterior contraction properties. By Remark 2.1, the marginal likelihood is the same as in the sequence model with random means (4.1). The marginal posterior is therefore

$$\pi\big(\sigma^2 \big| Y, Z\big) \propto \sigma^{-n_1}(\theta^2 + \sigma^2)^{-\frac{n_2}{2}} e^{-\frac{\|Y\|^2}{2\sigma^2}} e^{-\frac{\|Z\|^2}{2(\theta^2+\sigma^2)}} \pi(\sigma^2), \qquad (5.1)$$

which can also be written as the product of two inverse Gamma densities. In view of the Bernstein-von Mises phenomenon, the posterior concentrates around the MLE for parametric problems. Similarly, we can argue here that the posterior will be concentrated around the value $\widehat{\sigma}^2$ maximizing the likelihood part of the posterior (5.1). By differentiation, we find $n_1\widehat{\sigma}^2 + n_2\widehat{\sigma}^4/(\widehat{\sigma}^2 + \theta^2) = \|Y\|^2 + \widehat{\sigma}^4\|Z\|^2/(\theta^2 + \widehat{\sigma}^2)^2$ and rearranging yields

$$\widehat{\sigma}^2 - \overline{Y^2} = \frac{n_2}{n_1}\left(\frac{\widehat{\sigma}^2}{\theta^2 + \widehat{\sigma}^2}\right)^2 \big[\overline{Z^2} - \theta^2 - \widehat{\sigma}^2\big].$$

This can be rewritten as

$$
\begin{aligned}
&\widehat{\sigma}^2 - \sigma_0^2 + O_P(n^{-1/2}) \\
&= \frac{1-\alpha}{\alpha}\big(1 + O(n^{-1})\big)\left(\frac{\widehat{\sigma}^2}{\theta^2 + \widehat{\sigma}^2}\right)^2 \big[\sigma_0^2 - \widehat{\sigma}^2 + \overline{\mu_0^2} + O_P(n^{-1/2}) - \theta^2\big],
\end{aligned}
\qquad (5.2)
$$

where we set

$$\overline{\mu_0^2} = \|\mu_0\|^2/n_2$$

and suppress the dependence of the $O()$ term on σ_0^2 and μ_0. Since θ is fixed, this shows that for $\widehat{\sigma}^2 = \sigma_0^2 + O_P(n^{-1/2})$, we need

$$\overline{\mu_0^2} = \theta^2 + O_P(n^{-1/2}). \qquad (5.3)$$

Differently speaking, to force the maximum $\widehat{\sigma}^2$ to be close to σ_0^2, the variance θ^2 of the prior has to match the empirical variance $\overline{\mu_0^2}$ of the nuisance parameter. We can also deduce from (5.2) that if $|\overline{\mu_0^2} - \theta^2| \gg n^{-1/2}$ and θ is fixed, then also $|\widehat{\sigma}^2 - \sigma_0^2| \gg n^{-1/2}$. More precisely, we even have that $\overline{\mu_0^2} - \theta^2 \gg n^{-1/2}$ implies $\widehat{\sigma}^2 - \sigma_0^2 \gg n^{-1/2}$ and $\overline{\mu_0^2} - \theta^2 \ll -n^{-1/2}$ implies $\widehat{\sigma}^2 - \sigma_0^2 \ll -n^{-1/2}$. This shows that, depending on the size of $\overline{\mu_0^2}$ compared to θ^2, the posterior can either overestimate or underestimate the true variance.

If θ is allowed to vary with n, we can make the right hand side in (5.2) arbitrarily small by letting θ tend to infinity. As θ^2 is the variance of the prior, the behavior resembles then that of the uniform improper prior, which, as we already know, leads to posterior consistency. If we think of a prior as a prior belief on the parameters, then the prior should not change depending on the amount of available data and, in particular, it is unnatural that the prior becomes more vague if the sample size increases. In the next section we show that there are sample size independent mixture priors leading to parametric posterior contraction rates.

5.2. Mixture priors

Section 4 explains the posterior inconsistency for an i.i.d. prior on the nuisance. It seems unintuitive that introducing dependency on the prior of the nuisance parameter can help avoiding posterior inconsistency for σ_0^2. Surprisingly, this is not true. In this section, we first provide some intuition why mixture priors can resolve the issues of i.i.d. priors. Afterwards, we discuss and analyze a specific prior construction.

Analyzing Gaussian priors above, (5.3) suggests that for any nuisance parameter vector μ_0, there exists an i.i.d. prior which seems to work. This i.i.d. prior does, however, depend on the unknown μ_0 and can therefore not be chosen without knowledge of the data. Intuitively, if the posterior had the chance to see all possible i.i.d. priors on μ, instead of just one, it is conceivable that it would automatically select one that is adapted to the unknown nuisance parameter and consequently leads to posterior consistency for the parameter of interest. De Finetti's theorem [18] states that an exchangeable prior ν over the infinite sequence $\mu = (\mu^1, \mu^2, \dots)$ can be written as a mixture over i.i.d. priors in the sense that

$$\nu(A^1 \times \cdots \times A^k) := \int_{\mathcal{P}(\mathbb{R})} Q(A^1) \cdots Q(A^k) \lambda(dQ),$$

with λ a probability measure on the set of probability densities $\mathcal{P}(\mathbb{R})$ on $\mathbb{R}$. Assuming interchangeability of the integrals, the posterior (2.5) then becomes

$$\pi(\sigma^2 | Y, Z) \propto \pi(\sigma^2) \int_{\mathbb{R}^n} \frac{L(\sigma^2, \mu | Y, Z)}{L(\sigma_0^2, \mu_0 | Y, Z)} \nu(\mu) d\mu,$$

$$= \pi(\sigma^2) \int_{\mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}^n} \frac{L(\sigma^2, \mu | Y, Z)}{L(\sigma_0^2, \mu_0 | Y, Z)} \prod_{i=1}^{n} q(\mu^i) d\mu^i \right) \lambda(dq),$$

where q denotes the probability density function of Q. Let q_0 be the i.i.d. prior maximizing the interior integral. Suppose that this is a unique maximum and that the outer integral is determined by the behavior of the integrand in a suitable neighborhood $\mathcal{S}$ of q_0. This means that

$$
\begin{aligned}
\pi\big(\sigma^2\big|Y,Z\big) &\propto \pi(\sigma^2) \int_{\mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}^n} \frac{L(\sigma^2,\mu|Y,Z)}{L(\sigma_0^2,\mu_0|Y,Z)} \prod_{i=1}^n q(\mu^i) d\mu^i \right) \lambda(dq) \\
&\approx \pi(\sigma^2) \int_{\mathcal{S}} \left(\int_{\mathbb{R}^n} \frac{L(\sigma^2,\mu|Y,Z)}{L(\sigma_0^2,\mu_0|Y,Z)} \prod_{i=1}^n q(\mu^i) d\mu^i \right) \lambda(dq) \\
&\approx \pi(\sigma^2) \left(\int_{\mathbb{R}^n} \frac{L(\sigma^2,\mu|Y,Z)}{L(\sigma_0^2,\mu_0|Y,Z)} \prod_{i=1}^n q_0(\mu^i) d\mu^i \right) \int_{\mathcal{S}} \lambda(dq).
\end{aligned}
$$

The right hand side is the posterior density of σ^2 for i.i.d. prior $\prod_{i=1}^n q_0(\mu^i)$ on the components.

Although this argument is only a sketch, it suggests that something might be gained by mixing over i.i.d. priors instead of just choosing one. Maximizing the marginalized likelihood in (5.1) over θ^2 yields

$$
\theta^2 = \overline{Z^2} - \sigma^2, \tag{5.4}
$$

if the r.h.s. is non-negative. For this choice of θ^2, (5.1) becomes $\pi(\sigma^2|Y,Z) \propto \sigma^{-n_1} \exp(-\|Y\|^2/(2\sigma^2))\pi(\sigma^2)$. The posterior therefore coincides with the posterior density based on the first part of the sample only, which we know has good posterior contraction properties.

Prior In a first step generate $\theta^2 \sim \gamma$, with γ a positive Lebesgue density on $\mathbb{R}_+$. Given θ^2, each non-zero mean is drawn independently from a centered normal distribution with variance θ^2, that is, $\mu_i|\theta^2 \sim \mathcal{N}(0,\theta^2)$, $i > n_1$.

Another heuristic about the posterior properties for this prior can again be derived by making the link to the associated sequence model with random means (2.8). For the prior considered here, the random means model has the form

$$
Y_i \sim \mathcal{N}(0,\sigma_0^2), \ i = 1,\dots,n_1 \text{ and } Z_i|\theta^2 \sim \mathcal{N}(0,\theta^2+\sigma_0^2), \ i = n_1+1,\dots,n, \tag{5.5}
$$

with $\theta^2 \sim \gamma$. If θ^2 were a second parameter and not generated from γ, the variance σ_0^2 would not be identifiable if only the Z_i's are observed. In model (5.5) we know the density γ, but this is not enough to consistently reconstruct σ_0^2 from the subsample Z. By Remark 2.1, this model leads to the same posterior for σ^2. The posterior should therefore realize that there is little extractable information about σ_0^2 in Z and discard these observations. We will see in the limiting shape result below that this is roughly what happens.

We denote by $\ell(\sigma^2|Y)$ and $\ell(\sigma^2+\theta^2|Z)$ the log-likelihoods of the sub-samples Y and Z coming from model (5.5) with σ^2 replacing σ_0^2, that is

$$
\begin{aligned}
\ell(\sigma^2|Y) &= -\frac{n_1}{2}\log(2\pi\sigma^2) - \frac{n_1\overline{Y^2}}{2\sigma^2}, \\
\ell(\sigma^2+\theta^2|Z) &= -\frac{n_2}{2}\log(2\pi(\sigma^2+\theta^2)) - \frac{n_2\overline{Z^2}}{2(\sigma^2+\theta^2)}.
\end{aligned}
\tag{5.6}
$$

The log-likelihoods appearing in (5.6) can be written in terms of inverse-gamma distributions. We denote by $\mathrm{IG}(\gamma,\beta)$ the inverse-gamma distribution with shape $\gamma > 0$ and scale $\beta > 0$. The corresponding p.d.f. is

$$
f_{\mathrm{IG}(\gamma,\beta)}(x) = \frac{\beta^\gamma}{\Gamma(\gamma)}x^{-\gamma-1}e^{-\frac{\beta}{x}},
\tag{5.7}
$$

where $\Gamma(\cdot)$ is the Gamma function. Rewriting the posterior, we have that

Lemma 5.1. *Under the Gaussian mixture prior, the marginal posterior density has the form*

$$
\pi(\sigma^2|Y,Z) \propto f_{\mathrm{IG}(\gamma_1,\beta_1)}(\sigma^2)\left(\int_0^{+\infty} f_{\mathrm{IG}(\gamma_2,\beta_2)}(\sigma^2+\theta^2)\gamma(\theta^2)d\theta^2\right)\pi(\sigma^2),
\tag{5.8}
$$

with $\gamma_1 = n_1/2 - 1$, $\beta_1 = n_1\overline{Y^2}/2$ and $\gamma_2 = n_2/2 - 1$, $\beta_2 = n_2\overline{Z^2}/2$. The $\mathrm{IG}(\gamma_1,\beta_1)$-distribution has mode $\beta_1/(\gamma_1+1) = \overline{Y^2}$ and variance $\beta_1^2/(\gamma_1-1)^2(\gamma_1 - 2) = O(n^{-1})$, whereas the $\mathrm{IG}(\gamma_2,\beta_2)$-distribution has mode $\beta_2/(\gamma_2+1) = \overline{Z^2}$ and variance $\beta_2^2/(\gamma_2-1)^2(\gamma_2-2) = O(n^{-1})$.

Starting from Lemma 5.1, we can develop a heuristic argument on how to recover the shape of the limit posterior distribution. We interpret the posterior $\Pi(\cdot|Y,Z)$ with density (5.8) as the marginalized version, over the set $\theta^2 \in (0,+\infty)$, of the distribution $\widetilde{\Pi}(\cdot|Y,Z)$ whose density is given by

$$
\widetilde{\pi}(\sigma^2,\theta^2|Y,Z) \propto f_{\mathrm{IG}(\gamma_1,\beta_1)}(\sigma^2)f_{\mathrm{IG}(\gamma_2,\beta_2)}(\sigma^2+\theta^2)\gamma(\theta^2)\pi(\sigma^2),
\tag{5.9}
$$

and refer to $\widetilde{\Pi}(\cdot|Y,Z)$ as the joint posterior on $(\sigma^2,\theta^2) \in (0,+\infty)^2$. The first step is double localization. Thanks to the exponential tails of the inverse Gamma distribution, the joint posterior $\widetilde{\Pi}(\cdot|Y,Z)$ asymptotically concentrates on the set $\{\sigma^2 \in B_1\} \cap \{\theta^2 \in B_2\}$, with B_1 a $O(\zeta_n)$-ball centered at $\overline{Y^2}$ and B_2 a $O(\zeta_n)$-ball around $0 \vee (\overline{Z^2} - \overline{Y^2})$ for a sequence $\zeta_n \asymp \sqrt{\log n/n}$. This also implies that the joint posterior (5.9) is arbitrarily close, in total variation distance, to the truncated posterior distribution with density $\widetilde{\pi}(\sigma^2,\theta^2|Y,Z)\mathbf{1}(\{\sigma^2 \in B_1\} \cap \{\theta^2 \in B_2\})$. In particular, this means that the hyperparameter θ^2 concentrates on a neighborhood of the maximal value derived in (5.4).

Arguing as in the classical proof of the Bernstein-von Mises theorem, we can then show that the truncated posterior distribution will asymptotically not depend on the prior and prove that the posterior given by (5.8) behaves asymptotically like

$$
\pi_1(\sigma^2|Y,Z) = \mathbf{1}(\sigma^2 \in B_1)f_{\mathrm{IG}(\gamma_1,\beta_1)}(\sigma^2)\int_{B_2} f_{\mathrm{IG}(\gamma_2,\beta_2)}(\sigma^2+\theta^2)d\theta^2.
\tag{5.10}
$$

Using essentially Laplace approximation, we show that the log-likelihoods $\ell(\sigma^2|Y)$ and $\ell(\sigma^2 + \theta^2|Z)$ in (5.6) can be always uniformly approximated by a second-order Taylor expansion around their maxima $\overline{Y^2}$ and $\overline{Z^2} - \sigma^2$, and thus the localized posterior converges in total variation distance to a distribution with density

$$\pi_2(\sigma^2|Y,Z) \propto \mathbf{1}(\sigma^2 \in B_1)e^{-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2} \int_{B_2} e^{-\frac{n_2}{4(\sigma_0^2 + \mu_0^2)^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2} d\theta^2, \quad (5.11)$$

whose factors are a truncated Gaussian density with mode $\overline{Y^2}$ and variance $2\sigma_0^4/n_1 = O(n^{-1})$ and the integral of a truncated Gaussian density with mode $\overline{Z^2} - \sigma^2$ and variance $2(\sigma_0^2 + \overline{\mu_0^2})^2/n_2 = O(n^{-1})$. By undoing the localization argument, we can show that the restriction to the sets B_1 and B_2 can be removed from (5.11) and the posterior given by (5.8) converges in total variation distance to the posterior limit distribution

$$\pi_\infty(\sigma^2|Y,Z) \propto \mathbf{1}(\sigma^2 \geq 0)e^{-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2}\left[1 - \Phi\left(\frac{\sqrt{n_2}(\sigma^2 - \overline{Z^2})}{\sqrt{2}(\sigma_0^2 + \overline{\mu_0^2})}\right)\right], \quad (5.12)$$

with Φ the c.d.f. of the standard normal distribution. Recall that $\overline{Z^2} \approx \sigma_0^2 + \overline{\mu_0^2}$. This suggests that the term involving Φ in the posterior limit distribution should asymptotically disappear if $\overline{\mu_0^2} \gg n^{-1/2}$. The limit of the posterior should then be the truncated Gaussian

$$\widetilde{\pi}_\infty(\sigma^2|Y) \propto \mathbf{1}(\sigma^2 \geq 0)\exp\left(-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2\right), \quad (5.13)$$

with mode $\overline{Y^2}$ and variance $2\sigma_0^4/n_1 = O(n^{-1})$.

The next result is a formal statement of the arguments mentioned above. To pass to (5.13) involves an additional $\log n$-factor in the signal strength of $\overline{\mu_0^2}$. Denote by $\|\cdot\|_{\mathrm{TV}}$ the total variation distance and recall that the expectation E_0^n is taken with respect to model (1.1).

Theorem 5.2. *Let $\Pi_\infty(\cdot|Y,Z)$ and $\widetilde{\Pi}_\infty(\cdot|Y)$ be the distributions corresponding to the densities (5.12) and (5.13), respectively. If the prior densities $\gamma, \pi : [0,\infty) \to (0,\infty)$ are positive and uniformly continuous, then, for any compact sets $K \subset (0,\infty), K' \subset (-\infty,\infty)$, and $n \to \infty$,*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} E_0^n\left[\left\|\Pi(\cdot|Y,Z) - \Pi_\infty(\cdot|Y,Z)\right\|_{\mathrm{TV}}\right] \to 0.$$

Moreover, if $\inf_{\mu_i^0 \in K', \forall i} |\mu_i^0| \gg (\log n/n)^{1/4}$, then

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} E_0^n\left[\left\|\Pi(\cdot|Y,Z) - \widetilde{\Pi}_\infty(\cdot|Y)\right\|_{\mathrm{TV}}\right] \to 0.$$

As a corollary of the proof, posterior contraction around the true variance σ_0^2 with contraction rate $O(\sqrt{\log n/n})$ can be established. In the case of large

means this is an immediate consequence of the posterior limit $\widetilde{\Pi}_\infty(\cdot|Y)$ and the parametric Bernstein-von Mises theorem. For small means it is less obvious because of the non-standard limit of the posterior.

Corollary 5.3. *There exists a constant $M = M(\alpha)$, such that*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} E_0^n \left[\Pi \left(\left| \frac{\sigma^2}{\sigma_0^2} - 1 \right| \geq M \sqrt{\frac{\log n}{n}} \middle| Y, Z \right) \right] \to 0.$$

The posterior limit distribution is closely related to the class of skew normal distributions, see [1, 2]. We now derive an alternative characterization of the limit distribution. From the argumentation above, the p.d.f.

$$\propto \mathbf{1}(\sigma^2, \theta^2 \geq 0) e^{-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2} e^{-\frac{n_2}{4(\sigma_0^2 + \overline{\mu_0^2})^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2} \tag{5.14}$$

can be viewed as the joint posterior limit of (σ^2, θ^2). In particular, the posterior limit distribution is the marginal distribution with respect to σ^2. As this is clear from the context, we do not write explicitly that the following distributions are conditional on Y, Z, that is, Y, Z are assumed to be fixed.

Lemma 5.4. *Let*

$$\xi \sim \mathcal{N}\left(\overline{Y^2}, \frac{2\sigma_0^4}{n_1}\right), \quad \eta \sim \mathcal{N}\left(\overline{Z^2}, \frac{2(\sigma_0^2 + \overline{\mu_0^2})^2}{n_2}\right).$$

be independent. The distribution with p.d.f. (5.14) coincides with the distribution of

$$(\xi, \eta - \xi) \big| (0 \leq \xi \leq \eta).$$

In particular, the posterior limit distribution $\Pi_\infty(\cdot|Y, Z)$ coincides with the distribution of

$$\xi \big| (0 \leq \xi \leq \eta).$$

If the standard deviations of η, ξ are small compared to the means, the posterior limit distribution essentially compares the means $\overline{Y^2}$ and $\overline{Z^2}$. This behavior is very reasonable because if $\overline{\mu_0^2}$ is small, $\overline{Y^2} \approx \overline{Z^2}$ and the subsample Z becomes informative about σ^2.

The posterior limit depends on unknown quantities. A frequentist estimator mimicking the posterior would be to estimate σ^2 from the MLE for zero means $\overline{X^2}$ in the case that the means are small. To detect whether small means are present, we can check whether $\overline{Y^2} \geq \overline{Z^2}$, which leads then to the estimator

$$\widetilde{\sigma}^2 = \begin{cases} \overline{Y^2}, & \text{if } \overline{Y^2} < \overline{Z^2}, \\ \overline{X^2}, & \text{otherwise}. \end{cases}$$

5.3. Finite sample analysis

We compare the estimators $\widehat{\sigma}_Y^2 = \overline{Y^2}$ and $\widetilde{\sigma}^2$ to the maximum $\widehat{\sigma}_{\mathrm{map},\infty}^2$ and the mean $\widehat{\sigma}_{\mathrm{mean},\infty}^2$ of the limit density $\sigma^2 \mapsto \pi_\infty(\sigma^2|Y,Z)$ for sample sizes $n \in \{10, 100, 1000\}$. As discussed above, we expect to see some differences for small means. We study the performances for $\sigma_0^2 = 1$ and μ the vector with all entries equal to $t/n^{1/4}$ for the values $t \in \{0,1,2,5\}$. Since $\widehat{\sigma}_Y^2$ does not depend on the means, the estimator performs equally well in all setups. Table 1 reports the average of the squared errors and the corresponding standard errors based on 10.000 repetitions. The rescaled MLE $\widehat{\sigma}_Y^2$ performs worse than any of the other estimators for small signals. Among the other estimators there is no clear 'winner'. For $t = 5$, the risk of all estimators is nearly the same. For larger values of t, our simulation experiments did not show any changes compared to $t = 5$ and the results are therefore omitted from the table.

TABLE 1

Comparison of the estimators for $(\sigma_0^2, \mu_0) = (1, (t/n^{1/4}, \ldots, t/n^{1/4}))$ and $t \in \{0,1,2,5\}$.

Estim.	n	0	1	2	5
$\widehat{\sigma}_Y^2$	10	0.414 ($\pm$ 8.7e-03)	0.411 ($\pm$ 8.6e-03)	0.386 ($\pm$ 8.2e-03)	0.399 ($\pm$ 8.4e-03)
	100	0.040 ($\pm$ 5.9e-04)	0.040 ($\pm$ 5.9e-04)	0.390 ($\pm$ 5.7e-04)	**0.041** ($\pm$ 6.4e-04)
	1000	0.004 ($\pm$ 5.7e-05)	0.004 ($\pm$ 5.6e-05)	0.004 ($\pm$ 5.8e-05)	**0.004** ($\pm$ 5.8e-05)
$\widetilde{\sigma}^2$	10	0.235 ($\pm$ 3.1e-03)	0.268 ($\pm$ 4.2e-03)	0.336 ($\pm$ 6.2e-03)	0.399 ($\pm$ 8.4e-03)
	100	**0.028** ($\pm$ 3.8e-04)	**0.031** ($\pm$ 4.2e-04)	0.037 ($\pm$ 5.2e-04)	**0.041** ($\pm$ 6.4e-05)
	1000	**0.003** ($\pm$ 4.3e-05)	**0.003** ($\pm$ 4.4e-05)	0.004 ($\pm$ 5.4e-05)	**0.004** ($\pm$ 5.8e-05)
$\widehat{\sigma}_{\mathrm{map},\infty}^2$	10	0.337 ($\pm$ 3.3e-03)	0.330 ($\pm$ 4.6e-03)	0.359 ($\pm$ 6.9e-03)	0.398 ($\pm$ 8.3e-03)
	100	0.036 ($\pm$ 4.3e-04)	0.032 ($\pm$ 4.2e-04)	**0.034** ($\pm$ 4.7e-04)	**0.041** ($\pm$ 6.3e-04)
	1000	**0.003** ($\pm$ 4.9e-05)	**0.003** ($\pm$ 4.5e-05)	**0.003** ($\pm$ 4.9e-05)	**0.004** ($\pm$ 5.8e-05)
$\widehat{\sigma}_{\mathrm{mean},\infty}^2$	10	**0.167** ($\pm$ 2.1e-03)	**0.182** ($\pm$ 3.8e-03)	**0.232** ($\pm$ 5.9e-03)	**0.283** ($\pm$ 7.0e-03)
	100	0.040 ($\pm$ 4.5e-04)	0.034 ($\pm$ 4.3e-04)	**0.034** ($\pm$ 4.7e-04)	**0.041** ($\pm$ 6.2e-04)
	1000	0.004 ($\pm$ 5.1e-05)	**0.003** ($\pm$ 4.6e-05)	**0.003** ($\pm$ 4.9e-05)	**0.004** ($\pm$ 5.8e-05)

There has been a long-standing debate whether Bayesian methods perform well if interpreted as frequentist methods. Results like the complete class theorem and the Bernstein-von Mises theorem have been foundational in this regard, see [22, 16]. Our theory highlights another instance where Bayes leads to new estimators with good finite sample properties. The analysis moreover shows that the construction of a prior resulting in a posterior with good frequentist properties can be highly non-intuitive.

Appendix A: Proofs

A.1. Proofs for Section 3

Proof of Proposition 3.1. By direct computation,

$$\partial_{\sigma^2} \log L(\sigma^2|Y,Z) = -\frac{n}{2\sigma^2} + \frac{\|Y\|^2}{2\sigma^4} + \frac{\partial_{\sigma^2}\left(\int e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}} d\nu(\mu)\right)}{\int e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}} d\nu(\mu)}.$$

Since

$$\partial_{\sigma^2}\Big(\int e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}}d\nu(\mu)\Big) = \int \frac{\|Z-\mu\|^2}{2\sigma^4}e^{-\frac{\|Z-\mu\|^2}{2\sigma^2}}d\nu(\mu),$$

we recover (3.3). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

A.2. Proofs for Section 4

Proof of Proposition 4.1. It is enough to show that the following statements hold for sufficiently large sample size n. Let $Q(u) = \nu([-u, u]^c)/\nu([-u, u])$. Since ν is a distribution function $Q(u) \to 0$ for $u \to \infty$. We work on $I = [\sigma_0^2/2, 2\sigma_0^2]$, where σ_0^2 is chosen such that

$$Q(\sigma_0) \le \exp\Big(-48\Big(17 + 2e^2 + \frac{24}{1-\alpha}\Big)\Big), \tag{A.1}$$

and α denotes the fraction of known zero means in the model. Notice that

$$\frac{\sigma^2}{2} \le \sigma_0^2 \le 2\sigma^2 \quad \text{for all} \ \ \sigma^2 \in I. \tag{A.2}$$

Let

$$R := \frac{\sigma_0}{\sqrt{6}}\sqrt{\log\Big(\frac{1}{Q(\sigma_0)}\Big)}. \tag{A.3}$$

We choose the non-zero means to be

$$\mu_i^0 := \frac{R}{2}. \tag{A.4}$$

The interval I is compact and the prior π is continuous and positive on $\mathbb{R}_+$, $\inf_{\sigma^2 \in I} \pi(\sigma^2) > 0$. Since we also assumed that π' is continuous, we find that

$$\sup_{\sigma^2 \in I} \frac{\sigma_0^2|\pi'(\sigma^2)|}{n\pi(\sigma^2)} \le 1$$

for all sufficiently large n. With (3.3) and (A.2),

$$\begin{aligned}
\inf_{\sigma^2 \in I} \partial_{\sigma^2} \log \pi(\sigma^2|Y, Z) &\ge \frac{n}{\sigma_0^2} \inf_{\sigma^2 \in I}\Big(\frac{\sigma_0^2 V(\mu|(Z, \sigma^2))}{2n\sigma^4} - \frac{\sigma_0^2}{2\sigma^2} - 1\Big) \\
&\ge \frac{n}{\sigma_0^2}\Big(\frac{\inf_{\sigma^2 \in I} V(\mu|(Z, \sigma^2))}{8\sigma_0^2 n} - 2\Big).
\end{aligned} \tag{A.5}$$

Using (3.1) and (3.2), we expand $V(\mu|(Z, \sigma^2))$,

$$\begin{aligned}
\frac{V(\mu|(Z, \sigma^2))}{n} &= \frac{\|Z\|^2}{n} + \frac{1}{n}\int_{\mathbb{R}^n}(\|\mu\|^2 - 2Z^\top\mu)\pi(\mu|Z, \sigma^2)d\mu \\
&= \frac{\|Z\|^2}{n} + \frac{1}{n}\sum_{i=1}^n\int_{\mathbb{R}}(\mu_i^2 - 2Z_i\mu_i)\pi(\mu_i|Z_i, \sigma^2)d\mu_i.
\end{aligned}$$

Since the integrands in the latter display are positive for $|\mu_i| \geq 2|Z_i|$, we can set $V_i := |Z_i| \int_{|\mu| \leq 2|Z_i|} |\mu| \pi(\mu|Z_i, \sigma^2) d\mu$ and bound

$$\frac{V(\mu|(Z, \sigma^2))}{n} \geq \frac{\|Z\|^2}{n} - \frac{2}{n} \sum_{i=1}^{n_2} Z_i \int_{|\mu_i| \leq 2|Z_i|} \mu_i \pi(\mu_i|Z_i, \sigma^2) d\mu_i$$

$$\geq \frac{\|Z\|^2}{n} - \frac{2}{n} \sum_{i=1}^{n_2} V_i.$$

As a next step in the proof, we show

$$\inf_{\sigma^2 \in I} \frac{V(\mu|(Z, \sigma^2))}{n} \geq \frac{\|Z\|^2}{2n} - \frac{16}{n} \left\| Z - \frac{R}{2} \right\|^2 - \frac{2n_2}{n} \sigma_0^2 e^2. \qquad (A.6)$$

To prove this inequality, we distinguish the cases $|Z_i| > R$ and $|Z_i| \leq R$, decomposing

$$V_i =: |Z_i|(A_i + B_i) \qquad (A.7)$$

with

$$A_i := \mathbf{1}(|Z_i| > R) \int_{|\mu| \leq 2|Z_i|} |\mu| \pi(\mu|Z_i, \sigma^2) d\mu$$
$$B_i := \mathbf{1}(|Z_i| \leq R) \int_{|\mu| \leq 2|Z_i|} |\mu| \pi(\mu|Z_i, \sigma^2) d\mu. \qquad (A.8)$$

For the term A_i of (A.8), observe that $A_i \leq 2|Z_i| \mathbf{1}(|Z_i| > R)$. If $|Z_i| > R$, $|Z_i| \leq 2|Z_i| - R \leq 2|Z_i - R/2|$ and therefore,

$$|Z_i| A_i \leq 8 \left(Z_i - \frac{R}{2} \right)^2. \qquad (A.9)$$

Next, we bound the term B_i in (A.8). In the sequel, we frequently make use of the fact that $\sigma^2 \in I$. The idea is to split the domain of integration $0 \leq |\mu| \leq 2|Z_i|$ into sets $|\mu| \leq \sigma_0$ and $\sigma_0 < |\mu| \leq 2|Z_i|$. The contribution of the first part can be bounded by σ_0. More work is needed for the second part. By expanding the square $(\mu - Z_i)^2$ in the exponent, the Z_i^2-terms in the numerator and denominator cancel against each other, as they do not depend on μ, and we have

$$B_i = \mathbf{1}(|Z_i| \leq R) \frac{\int_{|\mu| \leq 2|Z_i|} |\mu| e^{-\frac{(\mu - Z_i)^2}{2\sigma^2}} d\nu(\mu)}{\int e^{-\frac{(\mu - Z_i)^2}{2\sigma^2}} d\nu(\mu)}$$

$$\leq \sigma_0 + \mathbf{1}(|Z_i| \leq R) \frac{\int_{\sigma_0 < |\mu| \leq 2R} |\mu| e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu Z_i}{\sigma^2}} d\nu(\mu)}{\int e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu Z_i}{\sigma^2}} d\nu(\mu)}.$$

We now treat numerator and denominator separately. For the numerator, the function $y \mapsto y e^{-y^2/2}$ attains its maximum at $y = 1$ and is bounded by $e^{-1/2}$.

This means that $|\mu|e^{-\frac{\mu^2}{2\sigma^2}} \le \sigma e^{-1/2} \le \sigma_0$, where the last step follows from (A.2). Together with (A.2), we obtain

$$\mathbf{1}(|Z_i| \le R) \int_{\sigma_0 < |\mu| \le 2R} |\mu| e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu Z_i}{\sigma^2}} \nu(\mu) d\mu \le \sigma_0 e^{\frac{4R^2}{\sigma_0^2}} \nu\big([-\sigma_0, \sigma_0]^c\big),$$

using $\mu Z_i/\sigma^2 \le 4R^2/\sigma_0^2$ to bound the exponent in the integral. To derive a lower bound of the denominator, we replace the integral over $\mathbb{R}$ by an integral over $[-\sigma_0, \sigma_0]$. On this interval, $e^{-\mu^2/(2\sigma^2)} \ge e^{-1}$ and $\mathbf{1}(|Z_i| \le R)e^{\frac{\mu Z_i}{\sigma^2}} \ge e^{-R^2/\sigma^2} \ge e^{-2R^2/\sigma_0^2}$, since $\sigma_0 \le R$. We obtain

$$\mathbf{1}(|Z_i| \le R) \int_{\mathbb{R}} e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu Z_i}{\sigma^2}} d\nu(\mu) \ge e^{-1} e^{-\frac{2R^2}{\sigma_0^2}} \nu\big([-\sigma_0, \sigma_0]\big).$$

Combining this with the upper bound for the numerator yields, with (A.1), (A.3) and the definition of the function $Q(u)$,

$$B_i \le e^{1+\frac{6R^2}{\sigma_0^2}} Q(\sigma_0)\sigma_0 = e^{1-\log Q(\sigma_0)} Q(\sigma_0)\sigma_0 = e\sigma_0 \quad \text{for all } \sigma^2 \in I. \qquad \text{(A.10)}$$

Together with (A.9) and (A.7),

$$V_i \le 8\Big(Z_i - \frac{R}{2}\Big)^2 + |Z_i|\sigma_0 e, \quad \text{for all } \sigma^2 \in I.$$

With $|Z_i|\sigma_0 e \le Z_i^2/4 + \sigma_0^2 e^2$, we finally obtain (A.6).

In a final step of the proof, we derive, on an event with large probability, a deterministic lower bound for the right hand side in (A.6). Let $U_1, \ldots, U_{n_2}$ be independent random variables. Rewriting Chebyshev's inequality yields $P(n^{-1} \sum_{i=1}^{n_2} U_i > n^{-1} \sum_{i=1}^{n_2} (E[U_i] - \sigma_0^2)) \ge 1 - \sum_{i=1}^{n_2} \text{Var}(U_i)/(n_2\sigma_0^2)^2$. We aply this with $U_i = Z_i^2/2 - 16(Z_i - R/2)^2$. Recall that $Z_i \sim \mathcal{N}(R/2, \sigma_0^2)$. Therefore, $E_0[Z_i^2] = R^2/4 + \sigma_0^2$ and $E[(Z_i - R/2)^2] = \sigma_0^2$. For the variance, $\text{Var}_0(Z_i^2) = R^2\sigma_0^2 + \sigma_0^4$ and $\text{Var}((Z_i - R/2)^2) = \sigma_0^4$. Since by assumption $\alpha < 1$, Chebyshev's inequality yields then $P_0^n(\mathcal{A}_n) \to 1$ when $n \to \infty$ for the set

$$\mathcal{A}_n := \Big\{ \frac{\|Z\|^2}{2n} - \frac{16}{n}\Big\|Z - \frac{R}{2}\Big\|^2 \ge \frac{n_2}{n}\Big(\frac{R^2 + 4\sigma_0^2}{8} - 17\sigma_0^2\Big) \Big\}. \qquad \text{(A.11)}$$

On $\mathcal{A}_n$, we have using (A.3), (A.6) and $Q(\sigma_0) \le \exp(-48(17 + 2e^2 + 24/(1-\alpha)))$,

$$\inf_{\sigma^2 \in I} \frac{V(\mu|(Z, \sigma^2))}{8\sigma_0^2 n} \ge \frac{n_2}{8\sigma_0^2 n}\Big(\frac{R^2}{8} - \sigma_0^2(17 + 2e^2)\Big) \ge 3. \qquad \text{(A.12)}$$

The assertion follows with (A.5). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 4.2. Proposition 4.1 shows that

$$\inf_{\sigma^2 \in [\sigma_0^2/2, 2\sigma_0^2]} \partial_{\sigma^2} \log \pi(\sigma^2|Y, Z) \ge \frac{n}{\sigma_0^2}$$

has P_0^n-probability tending to one. This means that for $\sigma^2, \widetilde{\sigma}^2 \in [\sigma_0^2/2, 2\sigma_0^2]$, with $\sigma^2 \leq \widetilde{\sigma}^2$, we must have $\log \pi(\sigma^2|Y, Z) \leq \log \pi(\widetilde{\sigma}^2|Y, Z) - n(\widetilde{\sigma}^2 - \sigma^2)/\sigma_0^2$. Exponentiating this inequality for $\widetilde{\sigma}^2 = \sigma^2 + \sigma_0^2/2$, yields

$$
\Pi\Big(\sigma^2 \in \Big[\frac{\sigma_0^2}{2}, 3\frac{\sigma_0^2}{2}\Big] \Big| Y, Z\Big) = \int_{\sigma_0^2/2}^{3\sigma_0^2/2} \pi_n(\sigma^2|Y, Z)d\sigma^2
$$

$$
\leq e^{-n/2} \int_{\sigma_0^2}^{2\sigma_0^2} \pi_n(\sigma^2|Y, Z)d\sigma^2 \leq e^{-n/2}
$$

and this completes the proof since $|\sigma^2/\sigma_0^2 - 1| \leq 1/2$ is equivalent to $\sigma^2 \in [\sigma_0^2/2, 3\sigma_0^2/2]$. □

A.3. Proofs for Section 5

Proof of Lemma 5.1. We can write the posterior as

$$
\pi(\sigma^2|Y, Z) \propto \mathbf{1}(\sigma^2 \geq 0)e^{\ell(\sigma^2|Y)} \int_0^\infty e^{\ell(\sigma^2+\theta^2|Z)}\gamma(\theta^2)d\theta^2 \pi(\sigma^2). \tag{A.13}
$$

By using (5.6) and (5.7) we obtain (5.8). □

We now prepare for the proof of the limiting shape result. From (5.8), the density (5.9) of the joint posterior is

$$
\widetilde{\pi}(\sigma^2, \theta^2|Y, Z) \propto \mathbf{1}(\sigma^2 \geq 0, \theta^2 \geq 0)e^{\ell(\sigma^2|Y)}e^{\ell(\sigma^2+\theta^2|Z)}\gamma(\theta^2)\pi(\sigma^2).
$$

With

$$
\zeta_n := 4\sqrt{\Big(1 + \Big(\frac{\alpha}{1-\alpha} \vee \frac{1-\alpha}{\alpha}\Big)\Big)\frac{\log n}{n_1 \wedge n_2}} \wedge 1, \tag{A.14}
$$

define

$$
B_1 := \Big[\frac{\overline{Y^2}}{1+\zeta_n}, \frac{\overline{Y^2}}{1-\zeta_n}\Big],
$$
$$
B_2 := \Big[0 \vee \Big(\frac{\overline{Z^2}}{1+\zeta_n} - \frac{\overline{Y^2}}{1-\zeta_n}\Big), \frac{\overline{Z^2}}{1-\zeta_n} - \frac{\overline{Y^2}}{1+\zeta_n}\Big]. \tag{A.15}
$$

It is shown below that the posterior concentrates on $\{\sigma^2 \in B_1\}$ and $\{\theta^2 \in B_2\}$. The posterior can consequently be approximated by the distribution $\Pi_1(\cdot|Y, Z)$ defined through its density (5.10). On the localized set $(\sigma^2, \theta^2) \in B_1 \times B_2$, we are able to replace the log-likelihoods by a quadratic expansion. This then allows us to approximate the posterior by $\Pi_2(\cdot|Y, Z)$ which is defined as the distribution with density (5.11). We now state the single steps formally and provide the proofs.

Proposition A.1. *If the prior densities $\gamma, \pi : [0, \infty) \to (0, \infty)$ are positive and uniformly continuous, then there exists a sequence of sets $(A_n)_n$ such that for any compact sets $K \subset (0, \infty), K' \subset (-\infty, \infty)$,*

(i) $\lim_{n \to \infty} \sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} P_0^n(A_n^c) = 0.$

(ii) *With B_1, B_2 as defined in (A.15), we have for $n \to \infty$,*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \widetilde{\Pi}\big(\{\sigma^2 \notin B_1\} \cup \{\theta^2 \notin B_2\} \,|Y, Z\big) \mathbf{1}\big((Y, Z) \in A_n\big) \to 0.$$

(iii) *For $n \to \infty$,*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \widetilde{\Pi}\big(\sigma^2 \in \cdot |Y, Z\big) - \Pi_1(\cdot |Y, Z) \right\|_{\mathrm{TV}} \mathbf{1}\big((Y, Z) \in A_n\big) \to 0.$$

(iv) *For $n \to \infty$,*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \Pi_1(\cdot |Y, Z) - \Pi_2(\cdot |Y, Z) \right\|_{\mathrm{TV}} \mathbf{1}\big((Y, Z) \in A_n\big) \to 0.$$

(v) *For $n \to \infty$,*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \Pi_2(\cdot |Y, Z) - \Pi_\infty(\cdot |Y, Z) \right\|_{\mathrm{TV}} \mathbf{1}\big((Y, Z) \in A_n\big) \to 0.$$

(vi) *For $n \to \infty$, and $\inf_{\mu_i^0 \in K'} |\mu_i^0| \gg (\log n / n)^{1/4}$,*

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \Pi_\infty(\cdot |Y, Z) - \widetilde{\Pi}_\infty(\cdot |Y) \right\|_{\mathrm{TV}} \mathbf{1}\big((Y, Z) \in A_n\big) \to 0.$$

Proof of Proposition A.1. Recall the definition of ζ_n in (A.14) and set

$$\delta_n := C^{-1}\zeta_n = \sqrt{2\frac{\log n}{n_1 \wedge n_2}} \wedge C^{-1}, \text{ with } C^2 := 16 + 16\Big(\frac{\alpha}{1-\alpha} \vee \frac{1-\alpha}{\alpha}\Big). \tag{A.16}$$

Let $\underline{\sigma}_0^2 = \inf\{\sigma_0^2 \in K\} > 0$. Define the event

$$A_n := \Big\{\overline{Z^2} > \frac{\overline{Y^2}}{1 + \delta_n/2}\Big\} \cap \Big\{\Big|\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2} - 1\Big| + \Big|\frac{\overline{Y^2}}{\sigma_0^2} - 1\Big| \le \delta_n\Big\}. \tag{A.17}$$

Since $\delta_n \le 1/2$, this implies in particular that on A_n, $\overline{Y^2} \wedge \overline{Z^2} \ge \underline{\sigma}_0^2/2$.

Proof of (i). We simplify the notation by introducing the events

$$B_n := \Big\{\overline{Z^2} > \frac{\overline{Y^2}}{1 + \delta_n/2}\Big\}, \quad D_n := \Big\{\Big|\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2} - 1\Big| + \Big|\frac{\overline{Y^2}}{\sigma_0^2} - 1\Big| \le \delta_n\Big\},$$

so that $A_n = B_n \cap D_n$. Thus $P_0^n(A_n^c) \leq P_0^n(B_n^c) + P_0^n(D_n^c)$. We show that both $P_0^n(B_n^c)$ and $P_0^n(D_n^c)$ tend to zero uniformly over compact sets of parameters. By Chebyshev's inequality,

$$P_0^n(D_n^c) \leq P_0^n\left(\left|\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2} - 1\right| > \frac{\delta_n}{2}\right) + P_0^n\left(\left|\frac{\overline{Y^2}}{\sigma_0^2} - 1\right| > \frac{\delta_n}{2}\right)$$

$$\leq 4\frac{\text{Var}_0\left(\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2}\right) + \text{Var}_0\left(\frac{\overline{Y^2}}{\sigma_0^2}\right)}{\delta_n^2}.$$

Since

$$\text{Var}_0\left(\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2}\right) = \frac{2}{n_2} + \frac{4\overline{\mu_0^2}}{n_2\sigma_0^2}, \quad \text{Var}_0\left(\frac{\overline{Y^2}}{\sigma_0^2}\right) = \frac{2}{n_1},$$

we find

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} P_0^n(D_n^c) \leq \frac{8}{n_1\delta_n^2} + \frac{8}{n_2\delta_n^2} + \frac{16H}{n_2\delta_n^2}$$

with $H := \sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i}(\mu_i^0)^2/\sigma_0^2$. Notice that H is a finite constant since $K \subset (0, \infty)$ and K' are compact sets. Because $\delta_n = O(\sqrt{\log n/n})$, the previous probability tends to zero as n increases. We now bound $P_0^n(B_n^c)$. Rewriting B_n^c, we obtain

$$B_n^c = \left\{\left(1 + \frac{\delta_n}{2}\right)\left(\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2} - 1\right) + 1 - \frac{\overline{Y^2}}{\sigma_0^2} \leq -\frac{\delta_n}{2} - \left(1 + \frac{\delta_n}{2}\right)\frac{\overline{\mu_0^2}}{\sigma_0^2}\right\},$$

and again by Chebyshev's inequality

$$P_0^n(B_n^c) \leq \frac{\left(1 + \frac{\delta_n}{2}\right)^2 \text{Var}_0\left(\frac{\overline{Z^2} - \overline{\mu_0^2}}{\sigma_0^2} - 1\right) + \text{Var}_0\left(1 - \frac{\overline{Y^2}}{\sigma_0^2}\right)}{\left(\frac{\delta_n}{2} + \left(1 + \frac{\delta_n}{2}\right)\frac{\overline{\mu_0^2}}{\sigma_0^2}\right)^2}$$

$$\leq \left(1 + \frac{\delta_n}{2}\right)^2\left(\frac{8}{n_2\delta_n^2} + \frac{16H}{n_2\delta_n^2}\right) + \frac{8}{n_1\delta_n^2},$$

which again tends to zero for $n \to \infty$ uniformly over $\sigma_0^2 \in K, \mu_i^0 \in K', \forall i$.

Proof of (ii). We work on the event A_n defined in (A.17) deriving deterministic lower and upper bounds for the denominator and numerator in the Bayes formula. We start with

$$\widetilde{\Pi}(B_1^c \times \mathbb{R}_+ | Y, Z) = \frac{\int_{B_1^c} e^{\ell(\sigma^2|Y)} \int_0^\infty e^{\ell(\sigma^2 + \theta^2|Z)}\gamma(\theta^2)d\theta^2 \pi(\sigma^2)d\sigma^2}{\int_0^\infty e^{\ell(\sigma^2|Y)} \int_0^\infty e^{\ell(\sigma^2 + \theta^2|Z)}\gamma(\theta^2)d\theta^2 \pi(\sigma^2)d\sigma^2}, \quad (A.18)$$

and show that on the event A_n this quantity tends to 0 when n tends to infinity. The first part of the proof provides a lower bound for the denominator. For

that, we restrict $\sigma^2 \in \Sigma := [\overline{Y^2}/(1+\delta_n), \overline{Y^2}/(1+\delta_n/2)]$ and $\theta^2 \in \Theta(\sigma^2) := [\overline{Z^2} - \sigma^2, \overline{Z^2}(1+\delta_n) - \sigma^2] \subset (0,\infty)$, where the last inclusion follows since by definition of the event A_n in (A.17), $Z^2 - \sigma^2 \geq Z^2 - \overline{Y^2}/(1+\delta_n/2) \geq 0$. The inner integral in the denominator of (A.18) can be lower bounded by

$$\int_0^\infty e^{\ell(\sigma^2+\theta^2|Z)}\gamma(\theta^2)d\theta^2 \geq \int_{\Theta(\sigma^2)} e^{\ell(\sigma^2+\theta^2|Z)}d\theta^2 \inf_{\theta^2 \leq \overline{Z^2}(1+\delta_n)} \gamma(\theta^2).$$

Thanks to the definition of A_n in (A.17) and $\delta_n \leq 1$, we have $\overline{Z^2} \leq \overline{\mu_0^2} + \sigma_0^2(1+\delta_n)$, so that $\overline{Z^2}(1+\delta_n) \leq 2\overline{\mu_0^2} + 4\sigma_0^2$. We then set

$$\underline{\gamma} := \inf_{\theta^2 \leq \sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} 2\overline{\mu_0^2}+4\sigma_0^2} \gamma(\theta^2) \leq \inf_{\theta^2 \leq \overline{Z^2}(1+\delta_n)} \gamma(\theta^2).$$

Since K, K' are compact sets and γ is continuous and positive, we must have $\underline{\gamma} > 0$. Differentiating (5.6) gives $\partial_{\theta^2}\ell(\sigma^2+\theta^2|Y) = \frac{1}{2}n_2(\overline{Z^2}-\sigma^2-\theta^2)/(\sigma^2+\theta^2)^2$, so the function $\theta^2 \mapsto \ell(\sigma^2+\theta^2|Y)$ is decreasing on $\Theta(\sigma^2)$ for any σ^2. As a direct consequence of (5.6), we obtain

$$\ell(\overline{Z^2}(1+\delta_n)|Z) = \ell(\overline{Z^2}|Z) + \frac{n_2}{2}(\delta_n/(1+\delta_n) - \log(1+\delta_n)). \qquad (A.19)$$

Consequently, for any $\sigma^2 \in \Sigma$,

$$\int_0^\infty e^{\ell(\sigma^2+\theta^2|Z)}\gamma(\theta^2)d\theta^2 \geq \underline{\gamma}\,\overline{Z^2}\delta_n e^{\ell(\overline{Z^2}|Z)+\frac{n_2}{2}(\delta_n/(1+\delta_n)-\log(1+\delta_n))}$$
$$\geq \frac{1}{2}\underline{\gamma}\sigma_0^2\delta_n e^{\ell(\overline{Z^2}|Z)-\frac{n_2}{4}\delta_n^2}, \qquad (A.20)$$

where the last inequality follows since $\overline{Z^2} \geq \sigma_0^2/2$ on A_n, $\delta_n \leq 1$, and $-\log(1+\delta_n) \geq -\delta_n$ for $\delta_n \leq 1$. The right hand side does not depend on σ^2 anymore. To lower bound the first integral in the denominator of (A.18) we apply a similar argument. By (5.6), $\partial_{\sigma^2}\ell(\sigma^2|Y) = n_1(\overline{Y^2} - \sigma^2)/(2\sigma^4)$. This means that the function $\sigma^2 \mapsto \ell(\sigma^2|Y)$ is increasing on Σ and (5.6) yields

$$\ell(\overline{Y^2}/(1+\delta_n)) = \ell(\overline{Y^2}|Y) + \frac{n_1}{2}(\log(1+\delta_n) - \delta_n).$$

On A_n, $\overline{Y^2} \leq \sigma_0^2(1+\delta_n)$ and therefore $\overline{Y^2}/(1+\delta_n/2) \leq 2\sigma_0^2$. Set

$$\underline{\pi} := \inf_{\sigma^2 \leq \sup_{\sigma_0^2 \in K} 2\sigma_0^2} \pi(\sigma^2) \leq \inf_{\sigma^2 \leq \overline{Y^2}/(1+\delta_n/2)} \pi(\sigma^2),$$

so that $\underline{\pi} > 0$ because K is a compact set and π is continuous and positive. We bound

$$\int_0^\infty e^{\ell(\sigma^2|Y)}\pi(\sigma^2)d\sigma^2 \geq \inf_{\sigma^2 \in \Sigma}\pi(\sigma^2)\frac{\delta_n}{2}\overline{Y^2}e^{\ell(\overline{Y^2}/(1+\delta_n)|Y)}$$
$$\geq \underline{\pi}\frac{\delta_n}{2}\overline{Y^2}e^{\ell(\overline{Y^2}|Y)+\frac{n_1}{2}(\log(1+\delta_n)-\delta_n)} \qquad (A.21)$$
$$\geq \frac{1}{4}\underline{\pi}\delta_n\sigma_0^2 e^{\ell(\overline{Y^2}|Y)-\frac{n_1}{16}\delta_n^2},$$

using that on A_n, $\overline{Y^2} \geq \sigma_0^2/2$ and $\log(1 + \delta_n) \geq \delta_n - \delta_n^2/8$ for $0 \leq \delta_n \leq 1$. The product of the lower bounds obtained in (A.20) and (A.21) is then a lower bound for the denominator of (A.18).

In the next step we upper bound the numerator of (A.18). Firstly, observe that $\ell(\sigma^2 + \theta^2|Z) \leq \ell(\overline{Z^2}|Z)$ and

$$\int_0^\infty e^{\ell(\sigma^2 + \theta^2|Z)} \gamma(\theta^2) d\theta^2 \leq e^{\ell(\overline{Z^2}|Z)}. \tag{A.22}$$

Secondly, since $\sigma^2 \mapsto \ell(\sigma^2|Y)$ is increasing on $(0, \overline{Y^2}]$ and decreasing on $[\overline{Y^2}, \infty)$,

$$\int_0^{\overline{Y^2}/(1+\zeta_n)} e^{\ell(\sigma^2|Y)} \pi(\sigma^2) d\sigma^2 \leq e^{\ell(\overline{Y^2}/(1+\zeta_n)|Y)}$$

$$= e^{\ell(\overline{Y^2}|Y) + \frac{n_1}{2}(\log(1+\zeta_n) - \zeta_n)}$$

$$\leq e^{\ell(\overline{Y^2}|Y) - \frac{n_1}{16}\zeta_n^2}, \tag{A.23}$$

$$\int_{\overline{Y^2}/(1-\zeta_n)}^\infty e^{\ell(\sigma^2|Y)} \pi(\sigma^2) d\sigma^2 \leq e^{\ell(\overline{Y^2}/(1-\zeta_n)|Y)} = e^{\ell(\overline{Y^2}|Y) + \frac{n_1}{2}(\log(1-\zeta_n) + \zeta_n)}$$

$$\leq e^{\ell(\overline{Y^2}|Y) - \frac{n_1}{16}\zeta_n^2}.$$

The numerator of (A.18) is upper bounded by the product of the bounds obtained in (A.22) and (A.23). Together with the bounds on the denominator in (A.20) and (A.21), and $\zeta_n = C\delta_n$, we derive, on the event A_n, the following bound for (A.18):

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \widetilde{\Pi}\left(\sigma^2 \notin B_1 \big| Y, Z\right) \leq \frac{16}{\underline{\pi}\,\underline{\gamma}\,\underline{\sigma_0^4}\delta_n^2} e^{-(C^2 n_1 - 4n_2 - n_1)\delta_n^2/16} \to 0. \tag{A.24}$$

The convergence to zero follows since by definition of the constant C in (A.16), $n_1 C^2 - 4n_2 - n_1 > 4n_1$ and because of $\delta_n = O(\sqrt{\log n/n})$.

Along similar lines, we show now that, on the event A_n, $\widetilde{\Pi}(\theta^2 \notin B_2|Y, Z) \to 0$ as n tends to infinity. Since $\{\theta^2 \notin B_2\} \subset \{\sigma^2 \notin B_1\} \cup (\{\sigma^2 \in B_1\} \cap \{\theta^2 \notin B_2\})$, and $\widetilde{\Pi}(\sigma^2 \notin B_1|Y, Z)$ tends to zero by (A.24), it is sufficient to establish convergence of

$$\widetilde{\Pi}(B_1 \times B_2^c|Y, Z) = \frac{\int_{B_1} e^{\ell(\sigma^2|Y)} \int_{B_2^c} e^{\ell(\sigma^2 + \theta^2|Z)} \gamma(\theta^2) d\theta^2 \pi(\sigma^2) d\sigma^2}{\int_0^\infty e^{\ell(\sigma^2|Y)} \int_0^\infty e^{\ell(\sigma^2 + \theta^2|Z)} \gamma(\theta^2) d\theta^2 \pi(\sigma^2) d\sigma^2} \tag{A.25}$$

to zero. We can argue similarly as for the upper bound above using that $\ell(\sigma^2|Y) \leq \ell(\overline{Y^2}|Y)$. By following the same steps as for (A.22) and (A.23) and using that $a \mapsto \ell(a|Z)$ is increasing on $(0, \overline{Z^2}]$ and decreasing on $[\overline{Z^2}, \infty)$, the numerator in (5.9) integrated over the set $\{\sigma^2 \in B_1\} \cap \{\theta^2 \notin B_2\}$ is upper bounded by

$$\leq e^{\ell(\overline{Y^2}|Y)} \sup_{\sigma^2 \in B_1} \int_{B_2^c} e^{\ell(\sigma^2 + \theta^2|Z)} \gamma(\theta^2) d\theta^2 \leq 2e^{\ell(\overline{Y^2}|Y) + \ell(\overline{Z^2}|Z) - \frac{n_2}{16}\zeta_n^2}.$$

Together with the lower bounds for the denominator in (A.20) and (A.21), we upper bound (A.25), on the event A_n, by

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \widetilde{\Pi}(B_1 \times B_2^c | Y, Z) \leq \frac{32}{\underline{\pi \gamma} \underline{\sigma}_0^4 \delta_n^2} e^{-(C^2 n_2 - 4n_2 - n_1)\delta_n^2/16}. \qquad (A.26)$$

By definition (see (A.16)), the constant $C^2 > 0$ satisfies $n_2 C^2 - 4n_2 - n_1 > 4n_2$. Since $\delta_n = O(\sqrt{\log n/n})$, this implies that the right hand side of (A.26) is bounded above by $\lesssim n \exp(-n_2 \delta_n^2/4) \to 0$, as $n \to \infty$. Together with (A.24), this completes the proof for part (ii).

Proof of (iii). It is well-known that for probability measures P, Q defined on the same measurable space $\mathcal{X}$,

$$\|P - P(\cdot|A)\|_{\text{TV}} \leq 2P(A^c), \qquad (A.27)$$

see Lemma E.1 in [26]. With $A = B_1 \cap B_2$, $P = \widetilde{\Pi}(\cdot|Y, Z)$ and $\Pi_0(\cdot|Y, Z)$ the distribution with density

$$\pi_0(\sigma^2, \theta^2 | Y, Z) = \frac{e^{\ell(\sigma^2|Y)} e^{\ell(\sigma^2 + \theta^2|Z)} \mathbf{1}(\sigma^2 \in B_1, \theta^2 \in B_2)}{\int_{B_1} e^{\ell(\sigma^2|Y)} (\int_{B_2} e^{\ell(\sigma^2 + \theta^2|Z)} d\theta^2) d\sigma^2},$$

we have that

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \widetilde{\Pi}(\sigma^2 \in \cdot \,|Y, Z) - \Pi_0(\sigma^2 \in \cdot \,|Y, Z) \right\|_{\text{TV}}$$

$$\leq \sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \widetilde{\Pi}(\sigma^2 \in \cdot, \theta^2 \in \cdot \,|Y, Z) - \Pi_0(\sigma^2 \in \cdot, \theta^2 \in \cdot \,|Y, Z) \right\|_{\text{TV}} \to 0.$$

By bounding the L^1-distance between the densities, we now show that $\Pi_0(\sigma^2 \in \cdot|Y, Z)$ and $\Pi_1(\sigma^2 \in \cdot|Y, Z)$ are close in total variation using the following lemma.

Lemma A.2 (Lemma E.3 in [26]). *If $h(\sigma^2) \propto d\Pi_0(\sigma^2 \in \cdot|Y, Z)/d\Pi_1(\sigma^2 \in \cdot|Y, Z)$ exists and $\int |h(\sigma^2) - 1| d\Pi_1(\sigma^2|Y, Z) \leq \delta$ for some $\delta \in (0, 1)$, then also*

$$\left\| \Pi_0(\sigma^2 \in \cdot \,|Y, Z) - \Pi_1(\sigma^2 \in \cdot \,|Y, Z) \right\|_{\text{TV}} \leq \frac{\delta}{1 - \delta}.$$

As h is the Radon-Nikodym derivative up to a multiplicative factor, we can choose

$$h(\sigma^2) = \frac{\pi(\sigma^2) \int_{B_2} e^{\ell(\sigma^2 + \theta^2|Z)} \gamma(\theta^2) d\theta^2}{\inf_{\widetilde{\sigma}^2 \in B_1, \widetilde{\theta}^2 \in B_2} \pi(\widetilde{\sigma}^2) \gamma(\widetilde{\theta}^2) \int_{B_2} e^{\ell(\sigma^2 + \theta^2|Z)} d\theta^2} \mathbf{1}(\sigma^2 \in B_1).$$

Then,

$$1 \leq h(\sigma^2) \leq \frac{\sup_{\sigma^2 \in B_1, \theta^2 \in B_2} \pi(\sigma^2) \gamma(\theta^2)}{\inf_{\widetilde{\sigma}^2 \in B_1, \widetilde{\theta}^2 \in B_2} \pi(\widetilde{\sigma}^2) \gamma(\widetilde{\theta}^2)}. \qquad (A.28)$$

Using the argument above, it remains to prove that $\sup_{\sigma^2 \in B_1} |h(\sigma^2) - 1| \to 0$ for $n \to \infty$. By the definition of A_n and due to $\delta_n \leq \zeta_n$,

$$B_1 \subseteq B_1' := [\kappa_n \sigma_0^2, \kappa_n^{-1} \sigma_0^2] \quad \text{with } \kappa_n := \frac{1 - \zeta_n}{1 + \zeta_n} = 1 - 2\zeta_n + O(\zeta_n^2). \quad \text{(A.29)}$$

Recall that K is a compact set. Since π is positive and uniformly continuous,

$$\sup_{\sigma_0^2 \in K} \sup_{\sigma^2, \widetilde{\sigma}^2 \in [\kappa_n \sigma_0^2, \kappa_n^{-1} \sigma_0^2]} \left| \frac{\pi(\sigma^2)}{\pi(\widetilde{\sigma}^2)} - 1 \right| \to 0. \quad \text{(A.30)}$$

Similarly, we have on the event A_n,

$$B_2 \subseteq B_2' := \left[\frac{\overline{\mu_0^2}}{1 + \zeta_n} + \left(\kappa_n - \frac{1}{\kappa_n} \right) \sigma_0^2, \frac{\overline{\mu_0^2}}{1 - \zeta_n} + \left(\frac{1}{\kappa_n} - \kappa_n \right) \sigma_0^2 \right]. \quad \text{(A.31)}$$

Since $\mu_i^0 \in K'$ for all i, the average of the squares $\overline{\mu_0^2}$ lies in the convex hull of K' and

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \sup_{\theta^2, \widetilde{\theta}^2 \in B_2'} \left| \frac{\gamma(\theta^2)}{\gamma(\widetilde{\theta}^2)} - 1 \right| \to 0.$$

For real numbers u, v, $uv = (u-1)(v-1) + (u-1) + (v-1) + 1$. We therefore obtain with (A.28) and (A.30), $\sup_{\sigma^2 \in B_1} |h(\sigma^2) - 1| \to 0$ for $n \to \infty$. This completes the proof of (iii).

Proof of (iv). We use the same strategy as in the proof of part (iii), applying Lemma A.2 to

$$h(\sigma^2) = \mathbf{1}(\sigma^2 \in B_1) e^{\ell(\sigma^2|Y) - \ell(\overline{Y^2}|Y) + \frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2} \frac{\int_{B_2} e^{\ell(\sigma^2 + \theta^2|Z) - \ell(\overline{Z^2}|Z)} d\theta^2}{\int_{B_2} e^{-\frac{n_2}{4(\sigma_0^2 + \overline{\mu_0^2})^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2} d\theta^2},$$

which is a constant multiple of the likelihood ratio of $\Pi_1(\sigma^2 \in \cdot|Y, Z)$ and $\Pi_2(\sigma^2 \in \cdot|Y, Z)$. To verify the assumptions of Lemma A.2, we have to show that $\sup_{\sigma_0^2 \in K} |h(\sigma^2) - 1| \to 0$ for $n \to \infty$. Using again the identity $uv = (u-1)(v-1) + (u-1) + (v-1) + 1$ and the fact that $|\int f / \int g - 1| \leq \sup |f/g - 1|$, we find that it is enough to prove that on the event A_n,

$$\sup_{\sigma_0^2 \in K} \sup_{\sigma^2 \in B_1} \left| \ell(\sigma^2|Y) - \ell(\overline{Y^2}|Y) + \frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2 \right| \to 0. \quad \text{(A.32)}$$

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \sup_{\sigma^2 \in B_1, \theta^2 \in B_2} \left| \ell(\sigma^2 + \theta^2|Z) - \ell(\overline{Z^2}|Z) + \frac{n_2(\theta^2 + \sigma^2 - \overline{Z^2})^2}{4(\sigma_0^2 + \overline{\mu_0^2})^2} \right| \to 0. \quad \text{(A.33)}$$

To verify (A.32), differentiating (5.6) gives

$$\partial_{\sigma^2} \ell(\sigma^2|Y) = \frac{n_1}{2\sigma^4}(\overline{Y^2} - \sigma^2), \quad \partial_{\sigma^2} \ell(\overline{Y^2}|Y) = 0,$$

$$\partial_{\sigma^2}^2 \ell(\sigma^2|Y) = \frac{n_1}{2\sigma^6}(\sigma^2 - 2\overline{Y^2}), \quad \partial_{\sigma^2}^2 \ell(\overline{Y^2}|Y) = -\frac{n_1}{2\overline{Y^2}^2} < 0,$$

$$\partial_{\sigma^2}^3 \ell(\sigma^2|Y) = \frac{n_1}{\sigma^8}(3\overline{Y^2} - \sigma^2),$$

and by a third-order Taylor expansion around the maximum $\overline{Y^2}$,

$$\ell(\sigma^2|Y) - \ell(\overline{Y^2}|Y)$$
$$= \frac{1}{2}\partial_{\sigma^2}^2 \ell(\overline{Y^2}|Y)(\sigma^2 - \overline{Y^2})^2 + \frac{1}{6}\partial_{\sigma^2}^3 \ell(s^2|Y)(\sigma^2 - \overline{Y^2})^3$$
$$= -\frac{n_1}{4\overline{Y^2}^2}(\sigma^2 - \overline{Y^2})^2 + \frac{n_1}{6s^8}(3\overline{Y^2} - s^2)(\sigma^2 - \overline{Y^2})^3$$
$$= -\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2 + \frac{n_1(\overline{Y^2} + \sigma_0^2)}{4\sigma_0^4\overline{Y^2}^2}(\overline{Y^2} - \sigma_0^2)(\sigma^2 - \overline{Y^2})^2$$
$$\quad + \frac{n_1}{6s^8}(3\overline{Y^2} - s^2)(\sigma^2 - \overline{Y^2})^3,$$

for some s^2 between σ^2 and $\overline{Y^2}$. We now control the smaller order terms uniformly over $\sigma^2 \in B_1$. Observe that also $\overline{Y^2}, s^2 \in B_1$. With (A.29), $\sup_{\sigma^2,\widetilde{\sigma}^2 \in B_1} |\sigma^2 - \widetilde{\sigma}^2| = O(\zeta_n)$ and $\sigma_0^2/2 \le \sigma^2 \le 2\sigma_0^2$ for all $\sigma^2 \in B_1$. Moreover, since $K \subset (0,\infty)$ is compact, $\inf \sigma_0^2 \in K > 0$. Together this shows that

$$\sup_{\sigma_0^2 \in K} \sup_{\sigma^2 \in B_1} \left| \ell(\sigma^2|Y) - \ell(\overline{Y^2}|Y) + \frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2 \right| = O(n_1\zeta_n^3) \to 0,$$

establishing (A.32). To prove (A.33) we argue similarly. Differentiating (5.6) gives

$$\partial_{\theta^2} \ell(\sigma^2 + \theta^2|Z) = \frac{n_2}{2(\sigma^2 + \theta^2)^2}(\overline{Z^2} - \sigma^2 - \theta^2), \quad \partial_{\theta^2} \ell(\overline{Z^2}|Z) = 0,$$
$$\partial_{\theta^2}^2 \ell(\sigma^2 + \theta^2|Z) = \frac{n_2}{2(\sigma^2 + \theta^2)^3}(\theta^2 + \sigma^2 - 2\overline{Z^2}), \quad \partial_{\theta^2}^2 \ell(\overline{Z^2}|Z) = -\frac{n_2}{2\overline{Z^2}^2} < 0,$$
$$\partial_{\theta^2}^3 \ell(\sigma^2 + \theta^2|Z) = \frac{n_2}{(\sigma^2 + \theta^2)^4}(3\overline{Z^2} - \sigma^2 - \theta^2),$$

and by a third-order Taylor expansion around the maximum $\theta_*^2 = \overline{Z^2} - \sigma^2$,

$$\ell(\sigma^2 + \theta^2|Z) - \ell(\overline{Z^2}|Z)$$
$$= \frac{1}{2}\partial_{\theta^2}^2 \ell(\overline{Z^2}|Z)(\theta^2 + \sigma^2 - \overline{Z^2})^2 + \frac{1}{6}\partial_{\theta^2}^3 \ell(\sigma^2 + s^2|Z)(\theta^2 + \sigma^2 - \overline{Z^2})^3$$
$$= -\frac{n_2}{4\overline{Z^2}^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2 + \frac{n_2}{6(\sigma^2 + s^2)^4}(3\overline{Z^2} - \sigma^2 - s^2)(\theta^2 + \sigma^2 - \overline{Z^2})^3$$
$$= -\frac{n_2}{4(\sigma_0^2 + \overline{\mu_0^2})^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2$$
$$\quad + \frac{n_2(\overline{Z^2} + \sigma_0^2 + \overline{\mu_0^2})}{4(\sigma_0^2 + \overline{\mu_0^2})^2\overline{Z^2}^2}(\overline{Z^2} - \sigma_0^2 - \overline{\mu_0^2})(\theta^2 + \sigma^2 - \overline{Z^2})^2$$
$$\quad + \frac{n_2}{6(\sigma^2 + s^2)^4}(3\overline{Z^2} - \sigma^2 - s^2)(\theta^2 + \sigma^2 - \overline{Z^2})^3,$$

for some s^2 between θ^2 and $\overline{Z^2} - \sigma^2$. If $(\sigma^2, \theta^2) \in B_1 \times B_2$, then, on A_n, both $\overline{Z^2} - \sigma^2$ and s^2 are in B_2'. With (A.29) and (A.31), we have $\sup_{u,v \in B_2'} |u - v| = O(\zeta_n)$ and $(\sigma_0^2 + \overline{\mu_0^2})/2 \le \sigma^2 + s^2 \le 2(\sigma_0^2 + \overline{\mu_0^2})$ for sufficiently large n. Together with the reasoning for (A.32), this leads to

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \sup_{\sigma^2 \in B_1, \theta^2 \in B_2} \left| \ell(\sigma^2 + \theta^2 | Z) - \ell(\overline{Z^2} | Z) \right.$$
$$\left. + \frac{n_2}{4(\sigma_0^2 + \overline{\mu_0^2})^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2 \right|$$

being bounded by $\lesssim n\zeta_n^3$ and thus converging to zero.

Proof of (v). Define $\Pi_3(\cdot | Y, Z)$ as the distribution on $(0, \infty)^2$, with density (5.14), that is,

$$\pi_3(\sigma^2, \theta^2 | Y, Z) \propto \mathbf{1}(\sigma^2 \ge 0, \theta^2 \ge 0) e^{-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2} e^{-\frac{n_2}{4(\sigma_0^2 + \overline{\mu_0^2})^2}(\theta^2 + \sigma^2 - \overline{Z^2})^2}.$$

and $\widetilde{\Pi}_3(\cdot | Y, Z)$ as the localization of $\Pi_3(\cdot | Y, Z)$ on $B_1 \times B_2$, that is, the distribution with density

$$\widetilde{\pi}_3(\sigma^2, \theta^2 | Y, Z) \propto \mathbf{1}(\sigma^2 \in B_1, \theta^2 \in B_2) e^{-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2} e^{-\frac{n_2}{4}(\sigma_0^2 + \overline{\mu_0^2})^{-2}(\theta^2 + \sigma^2 - \overline{Z^2})^2}.$$

Here B_1, B_2 are as defined in (A.15). The marginal distributions of $\widetilde{\Pi}_3(\cdot | Y, Z)$ and $\Pi_3(\cdot | Y, Z)$ with respect to σ^2 are $\Pi_2(\cdot | Y, Z)$ and $\Pi_\infty(\cdot | Y, Z)$, respectively. Applying (A.27) yields

$$\begin{aligned}
\left\| \Pi_2(\cdot | Y, Z) - \Pi_\infty(\cdot | Y, Z) \right\|_{\mathrm{TV}} &\le \left\| \widetilde{\Pi}_3(\cdot | Y, Z) - \Pi_3(\cdot | Y, Z) \right\|_{\mathrm{TV}} \\
&\le 2\Pi_3\big(\{\sigma^2 \notin B_1\} \cup \{\theta^2 \notin B_2\} \big| Y, Z\big).
\end{aligned} \tag{A.34}$$

To prove (v), it remains to show that for $n \to \infty$,

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \Pi_3\big(\{\sigma^2 \notin B_1\} \cup \{\theta^2 \notin B_2\} \big| Y, Z\big) \mathbf{1}\big((Y, Z) \in A_n\big) \to 0. \tag{A.35}$$

By Lemma 5.4, it is enough to prove that on A_n,

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} P\big(\xi \notin B_1 \big| (0 \le \xi \le \eta)\big) + P\big(\eta - \xi \notin B_2 \big| (0 \le \xi \le \eta)\big) \to 0, \tag{A.36}$$

for independent $\xi \sim \mathcal{N}(\overline{Y^2}, 2\sigma_0^4/n_1), \eta \sim \mathcal{N}(\overline{Z^2}, 2(\sigma_0^2 + \overline{\mu_0^2})^2/n_2)$. Recall that this and all the following statements in (v) should be understood conditionally on Y, Z.

To bound the terms, we heavily rely on the exponential bounds for tail probabilities of Gaussian variables given by Mill's ratio [17]

$$\left(\frac{x^2}{1 + x^2} \right) \frac{e^{-x^2/2}}{\sqrt{2\pi}x} \le P\big(\mathcal{N}(0, 1) > x\big) \le \frac{e^{-x^2/2}}{\sqrt{2\pi}x}, \quad \forall x > 0. \tag{A.37}$$

In a first step we derive a lower bound on $P(0 \leq \xi \leq \eta)$. Using that on A_n, $\overline{Y^2}/(1 + \delta_n/2) \leq \overline{Z^2} = E[\eta]$, the definition of ξ, the symmetry properties of the $\mathcal{N}(0,1)$ distribution, $\sigma_0^2/2 \leq \overline{Y^2} \leq 2\sigma_0^2$ on A_n, and Mill's ratio, we find

$$
\begin{aligned}
P\big(0 \leq \xi \leq \eta\big) &\geq P\Big(0 \leq \xi \leq \frac{\overline{Y^2}}{1 + \delta_n/2}\Big) P\big(\overline{Z^2} \leq \eta\big) \\
&= \frac{1}{2} P\Big(\mathcal{N}(0,1) \in \Big[-\frac{\sqrt{n_1}\overline{Y^2}}{\sqrt{2}\sigma_0^2}, -\frac{\sqrt{n_1}\delta_n\overline{Y^2}}{2\sqrt{2}(1 + \delta_n/2)\sigma_0^2} \Big]\Big) \\
&= \frac{1}{2} P\Big(\mathcal{N}(0,1) \in \Big[\frac{\sqrt{n_1}\delta_n\overline{Y^2}}{2\sqrt{2}(1 + \delta_n/2)\sigma_0^2}, \frac{\sqrt{n_1}\overline{Y^2}}{\sqrt{2}\sigma_0^2} \Big]\Big) \\
&\geq \frac{1}{2} P\Big(\mathcal{N}(0,1) \in \Big[\frac{\sqrt{n_1}\delta_n}{\sqrt{2}}, \frac{\sqrt{n_1}}{2\sqrt{2}} \Big]\Big) \\
&= P\Big(\mathcal{N}(0,1) \geq \frac{\sqrt{n_1}\delta_n}{\sqrt{2}}\Big) - P\Big(\mathcal{N}(0,1) \geq \frac{\sqrt{n_1}}{2\sqrt{2}}\Big) \\
&\geq \frac{1}{2\sqrt{\pi n_1}\delta_n} e^{-\frac{n_1\delta_n^2}{4}} - \frac{2}{\sqrt{\pi n_1}} e^{-\frac{n_1}{16}}.
\end{aligned}
\tag{A.38}
$$

where in the last inequality we used that $x^2/(1 + x^2) > \frac{1}{2}$ for $x > 1$.

We now derive an upper bound for $P(\xi \notin B_1)$. Using the definition of ξ, $\zeta_n \leq 1$, $\overline{Y^2} \geq \sigma_0^2/2$, and Mill's ratio (A.37),

$$
\begin{aligned}
P(\xi \notin B_1) &= P\Big(\mathcal{N}(0,1) \notin \Big[-\frac{\sqrt{n_1}\zeta_n\overline{Y^2}}{\sqrt{2}(1 + \zeta_n)\sigma_0^2}, \frac{\sqrt{n_1}\zeta_n\overline{Y^2}}{\sqrt{2}(1 - \zeta_n)\sigma_0^2} \Big]\Big) \\
&\leq 2P\Big(\mathcal{N}(0,1) > \frac{\sqrt{n_1}\zeta_n\overline{Y^2}}{\sqrt{2}(1 + \zeta_n)\sigma_0^2}\Big) \\
&\leq 2P\Big(\mathcal{N}(0,1) > \frac{\sqrt{n_1}\zeta_n}{4\sqrt{2}}\Big) \\
&\leq \frac{8}{\sqrt{\pi n_1}\zeta_n} e^{-\frac{n_1\zeta_n^2}{64}}.
\end{aligned}
\tag{A.39}
$$

Next, we obtain a similar bound for $P(\eta - \xi \notin B_2, \xi \leq \eta, \xi \in B_1)$. If we define the difference of two sets U, V as $U - V := \{u - v : u \in U, v \in V\}$, then, $B_2 = ([\overline{Z^2}/(1 + \zeta_n), \overline{Z^2}/(1 - \zeta_n)] - B_1) \cap \mathbb{R}_+$. On the event $\xi \leq \eta, \xi \in B_1$, we have that $\eta \in [\overline{Z^2}/(1 + \zeta_n), \overline{Z^2}/(1 - \zeta_n)]$ implies that $\eta - \xi \in B_2$, which is equivalent to saying that $\eta - \xi \notin B_2$ implies $\eta \notin [\overline{Z^2}/(1 + \zeta_n), \overline{Z^2}/(1 - \zeta_n)]$. On A_n, $|\overline{Z^2} - \overline{\mu_0^2} - \sigma_0^2| \leq \sigma_0^2\delta_n$ by definition. Because of $\delta_n \leq 1/2$, we obtain $\overline{Z^2} \geq (\overline{\mu_0^2} + \sigma_0^2)/2$. Together with the symmetry properties of the normal distribution, $\zeta_n \leq 1$, and Mill's ratio (A.37), this yields

$$
\begin{aligned}
&P\big(\eta - \xi \notin B_2, \xi \leq \eta, \xi \in B_1\big) \\
&\leq P\Big(\eta \notin \Big[\frac{\overline{Z^2}}{1 + \zeta_n}, \frac{\overline{Z^2}}{1 - \zeta_n} \Big]\Big)
\end{aligned}
$$

$$= P\left(\mathcal{N}(0,1) \notin \left[-\frac{\sqrt{n_2}\zeta_n \overline{Z^2}}{\sqrt{2}(\mu_0^2 + \sigma_0^2)(1 + \zeta_n)}, \frac{\sqrt{n_2}\zeta_n \overline{Z^2}}{\sqrt{2}(\mu_0^2 + \sigma_0^2)(1 - \zeta_n)} \right]\right) \quad \text{(A.40)}$$

$$\leq 2P\left(\mathcal{N}(0,1) > \frac{\sqrt{n_2}\zeta_n}{4\sqrt{2}}\right)$$

$$\leq \frac{8}{\sqrt{\pi n_2}\zeta_n} e^{-\frac{n_2 \zeta_n^2}{64}}.$$

To prove (A.36), we bound

$$P\big(\xi \notin B_1 \big| (0 \leq \xi \leq \eta)\big) \leq \frac{P(\xi \notin B_1)}{P(0 \leq \xi \leq \eta)}$$

and

$$P\big(\eta - \xi \notin B_2 \big| (0 \leq \xi \leq \eta)\big) \leq \frac{P(\eta - \xi \notin B_2, \xi \in B_1, 0 \leq \xi \leq \eta) + P(\xi \notin B_1)}{P(0 \leq \xi \leq \eta)}.$$

Now (A.36) (and therefore (A.35)) follow from the inequalities (A.38), (A.39), (A.40) and the definition of δ_n. This completes the proof of (v).

Proof of (vi). Recall the definitions of the densities

$$\pi_\infty(\sigma^2|Y, Z) \propto \mathbf{1}(\sigma^2 \geq 0) \exp\left(-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2 \right) \left(1 - \Phi\left(\frac{\sqrt{n_2}(\sigma^2 - \overline{Z^2})}{\sqrt{2}(\sigma_0^2 + \mu_0^2)} \right) \right),$$

$$\widetilde{\pi}_\infty(\sigma^2|Y) \propto \mathbf{1}(\sigma^2 \geq 0) \exp\left(-\frac{n_1}{4\sigma_0^4}(\sigma^2 - \overline{Y^2})^2 \right),$$

and let

$$\pi_{\infty, B_1}(\sigma^2|Y, Z) \propto \pi_\infty(\sigma^2|Y, Z)\mathbf{1}(\sigma^2 \in B_1),$$

$$\widetilde{\pi}_{\infty, B_1}(\sigma^2|Y) \propto \widetilde{\pi}_\infty(\sigma^2|Y)\mathbf{1}(\sigma^2 \in B_1),$$

be their localised versions on B_1. It is enough to show that, on A_n,

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \Pi_\infty(\cdot|Y, Z) - \Pi_{\infty, B_1}(\cdot|Y, Z) \right\|_{\text{TV}} \xrightarrow{n \to \infty} 0, \quad \text{(A.41)}$$

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \widetilde{\Pi}_\infty(\cdot|Y) - \widetilde{\Pi}_{\infty, B_1}(\cdot|Y) \right\|_{\text{TV}} \xrightarrow{n \to \infty} 0, \quad \text{(A.42)}$$

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \left\| \Pi_{\infty, B_1}(\cdot|Y, Z) - \widetilde{\Pi}_{\infty, B_1}(\cdot|Y) \right\|_{\text{TV}} \xrightarrow{n \to \infty} 0. \quad \text{(A.43)}$$

For (A.41), we apply (A.27) and the fact that $\Pi_\infty(\cdot|Y, Z)$ is the marginal distribution of $\Pi_3(\cdot|Y, Z)$, finding

$$\left\| \Pi_\infty(\cdot|Y, Z) - \Pi_{\infty, B_1}(\cdot|Y, Z) \right\|_{\text{TV}} \leq 2\Pi_\infty(B_1^c|Y, Z) = 2\Pi_3(B_1^c|Y, Z).$$

In (v) we proved that the right hand side converges to zero uniformly over $\sigma_0^2 \in K, \mu_i^0 \in K', \forall i$. For (A.42), we argue similarly, using that

$$\left\| \widetilde{\Pi}_\infty(\cdot|Y) - \widetilde{\Pi}_{\infty, B_1}(\cdot|Y) \right\|_{\text{TV}} \leq 2\widetilde{\Pi}_\infty(B_1^c|Y) = 2P(\xi \notin B_1),$$

with $\xi \sim \mathcal{N}(\overline{Y^2}, 2\sigma_0^4/n_1)$. Using (A.39), we see that the right hand side converges to zero, uniformly over $\sigma_0^2 \in K, \mu_i^0 \in K', \forall i$.

For (A.43), we apply Lemma A.2. On A_n, the likelihood ratio of $\Pi_{\infty, B_1}(\cdot|Y, Z)$ and $\widetilde{\Pi}_{\infty, B_1}(\cdot|Y)$ is given by

$$h(\sigma^2|Y, Z) := \left(1 - \Phi\left(\frac{\sqrt{n_2}(\sigma^2 - \overline{Z^2})}{\sqrt{2}(\mu_0^2 + \sigma_0^2)} \right) \right) \mathbf{1}(\sigma^2 \in B_1).$$

On A_n,

$$\sup_{\sigma^2 \in B_1} \sigma^2 - \overline{Z^2} = \frac{\overline{Y^2}}{1 - \zeta_n} - \overline{Z^2} \leq \frac{\sigma_0^2(1 + \delta_n)}{1 - \zeta_n} - \overline{\mu_0^2} - \sigma_0^2(1 - \delta_n).$$

Uniformly over $\sigma_0^2 \in K$ and $\inf_{\mu_i^0 \in K'} |\mu_i^0|^2 \gg \zeta_n$, the right hand side can be further upper bounded by $-\overline{\mu_0^2}/2$ for sufficiently large n. Thus,

$$|h(\sigma^2|Y, Z) - 1| = P\left(\mathcal{N}(0, 1) \leq \frac{\sqrt{n_2}(\sigma^2 - \overline{Z^2})}{\sqrt{2}(\overline{\mu_0^2} + \sigma_0^2)} \right)$$

$$\leq P\left(\mathcal{N}(0, 1) \geq \frac{\sqrt{n_2}\overline{\mu_0^2}}{2\sqrt{2}(\overline{\mu_0^2} + \sigma_0^2)} \right).$$

Since $n\overline{\mu_0^2} \gg n\zeta_n \to \infty$ for $n \to \infty$,

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} P\left(\mathcal{N}(0, 1) \geq \frac{\sqrt{n_2}\overline{\mu_0^2}}{2\sqrt{2}(\overline{\mu_0^2} + \sigma_0^2)} \right) \xrightarrow{n \to \infty} 0.$$

This concludes the proof of *(vi)*. $\qquad\square$

Proof of Theorem 5.2. We insert $1 = \mathbf{1}((Y, Z) \in A_n) + \mathbf{1}((Y, Z) \notin A_n)$ in the expectation. Since the total variation distance of probability measures is bounded, the result follows from Proposition A.1. $\qquad\square$

Proof of Corollary 5.3. Recall that the posterior is the marginal distribution of $\widetilde{\Pi}(\cdot|Y, Z)$ with respect to σ^2. By Proposition A.1 (ii), we have that

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \Pi\left(\sigma^2 \notin \left[\frac{\overline{Y^2}}{1 + \zeta_n}, \frac{\overline{Y^2}}{1 - \zeta_n} \right] \middle| Y, Z \right) \mathbf{1}((Y, Z) \in A_n) \to 0.$$

Using that on A_n, $\sigma_0^2(1 - \delta_n) \leq \overline{Y^2} \leq \sigma_0^2(1 + \delta_n)$, and $\delta_n = C^{-1}\zeta_n = O(\sqrt{\log n/n})$, we obtain

$$\sup_{\sigma_0^2 \in K, \mu_i^0 \in K', \forall i} \Pi\left(\left| \frac{\sigma^2}{\sigma_0^2} - 1 \right| \geq M\sqrt{\frac{\log n}{n}} \middle| Y, Z \right) \mathbf{1}((Y, Z) \in A_n) \to 0$$

for a constant $M = M(\alpha)$ that is chosen to be sufficiently large. The claim follows by splitting the expected posterior, inserting $1 = \mathbf{1}((Y, Z) \in A_n) + \mathbf{1}((Y, Z) \notin A_n)$ in the expectation and using Proposition A.1 (i). $\qquad\square$

Proof of Lemma 5.4. To prove the result, we derive an expression for the joint density of $(\xi, \eta - \xi) \big| (0 \leq \xi \leq \eta)$. Observe that

$$P\big(\xi \leq s, \eta - \xi \leq t \big| (0 \leq \xi \leq \eta)\big) = \frac{P(\xi \leq s, \eta - \xi \leq t, 0 \leq \xi \leq \eta)}{P(0 \leq \xi \leq \eta)}$$
$$\propto P\big((\eta - t) \vee 0 \leq \xi \leq \eta \wedge s\big).$$

The right hand side is zero if $s \leq 0$. Suppose now that $0 \leq s \leq t$. Conditioning on η, the right hand side can be rewritten as

$$= \int_0^s P\big(0 \leq \xi \leq u\big) f_\eta(u) du + \int_s^t P\big(0 \leq \xi \leq s\big) f_\eta(u) du$$
$$+ \int_t^{t+s} P\big(u - t \leq \xi \leq s\big) f_\eta(u) du.$$

Taking derivatives $\partial_s \partial_t$, the density of $(\xi, \eta - \xi) \big| (0 \leq \xi \leq \eta)$ at point (s, t) equals up to a multiplicative constant $f_\xi(s) f_\eta(t + s)$. Which completes the proof for the case $0 \leq s \leq t$.

The case $0 \leq t \leq s$ is similar and the proof for this case therefore omitted.

Since the posterior limit distribution is the marginal over the first component of the joint distribution in (5.14), it must coincide with the distribution of $\xi \big| (0 \leq \xi \leq \eta)$. □

Acknowledgment

References

[1] Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12 (2)**, 171–178. MR0808153

[2] Azzalini, A. and Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83 (4)**, 715–726. MR1440039

[3] Bickel, P. J. and Kleijn J. K. (2012) The semiparametric Bernstein-von Mises theorem. *Ann. Statist.*, **40 (1)**, 206–237. MR3013185

[4] Castillo, I. (2008) Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.*, **2**, 1281–1299. MR2471287

[5] Castillo, I. (2012) A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields*, **152**, 53–99. MR2875753

[6] Castillo, I. (2012) Semiparametric Bernstein–von Mises theorem and bias, illustrated with Gaussian process priors. *Sankhya A*, **74 (2)**, 194–221. MR3021557

[7] Castillo, I. and Rousseau. J. (2015) A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.*, **43 (6)**, 2353–2383. MR3405597

[8] Castillo, I. and van der Vaart, A. (2012) Lower bounds for posterior rates with Gaussian process priors. *Ann. Statist.*, **40 (4)**, 2069–2101. MR3021557

[9] Castillo, I., Schmidt-Hieber J. and van der Vaart, A. (2012) Bayesian linear regression with sparse priors. *Ann. Statist.*, **43 (5)**, 1986–2018. MR3375874

[10] Castillo, I. and Mismer, R. (2018) Empirical Bayes analysis of spike and slab posterior distributions. *ArXiv e-prints*, **40 (1)**, 206–237. https://arxiv.org/abs/1801.01696. MR3885271

[11] Chernozhukov, V. and Hong, H. (2004) Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica*, **72 (5)**, 1445–1480. MR2077489

[12] Cox, D. R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B*, **49 (1)**, 1–39. MR0893334

[13] Cox, D. R. and Reid, N. (1993) A note on the calculation of adjusted profile likelihood. *J. Roy. Statist. Soc. Ser. B*, **55 (2)**, 467–471. MR1224410

[14] de Jonge, R. and van Zanten, H. (2013) Semiparametric Bernstein-von Mises for the error standard deviation. *Electron. J. Stat.*, **7**, 217–243. MR3020419

[15] Ghosal, S. and Samanta, T. (1995) Asymptotic behaviour of Bayes estimates and posterior distributions in multiparameter nonregular cases. *Mathematical Methods of Statistics*, **4 (4)**, 361–388. MR1372011

[16] Ghosal, S. and van der Vaart, A. (2017) Fundamentals of Nonparametric Bayesian Inference. *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge. MR3587782

[17] Gordon, R. D. (1941) Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Statist.*, **12 (3)**, 364–366. MR0005558

[18] Hewitt, E., Savage, L. J. (1955) Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, **80 (1)**, 470–501. MR0076206

[19] Hoffmann, M., Rousseau, J. and Schmidt-Hieber, J. (2015) On adaptive posterior concentration rates. *Ann. Statist.*, **43 (5)**, 2259–2295. MR3396985

[20] Johnstone, I. M. and Onatski, A. (2015) Testing in high-dimensional spiked models. *ArXiv e-prints*, https://arxiv.org/abs/1509.07269.

[21] Kiefer, J. and Wolfowitz, J. (1956) Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, **27 (4)**, 887–906. MR0086464

[22] Le Cam, L. (1986) Asymptotic methods in statistical decision theory. *Springer Series in Statistics*, Springer-Verlag, New York. MR0856411

[23] McCullagh, P. and Tibshirani, R. (1990) A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. Ser. B*, **52 (2)**, 325–344. MR1064420

[24] Meijer, E., Rohwedderm S. and Wansbeek T. (2012) Measurement error in earnings data: using a mixture model approach to combine survey and

register data. *Journal of Business & Economic Statistics*, **30 (2)**, 191–201. MR2934935

[25] Neyman, J. and Scott, E. L. (1948) Consistent estimates based on partially consistent observations. *Econometrica*, **16 (1)**, 1–32. MR0025113

[26] Reiss, M. and Schmidt-Hieber, J. (2017) Nonparametric Bayesian analysis of the compound Poisson prior for support boundary recovery. To appear in Annals of Statistics. *ArXiv e-prints*, https://arxiv.org/abs/1809.04140.

[27] Shen, X. (2002) Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.*, **97**, 222–235. MR1947282

[28] Sweeting, T.J. (1987) Discussion of "Parameter orthogonality and approximate conditional inference". *J. Roy. Statist. Soc. Ser. B*, **49 (1)**, 20–21. MR0893334

[29] van der Pas, S. L., Salomond, J. B. and Schmidt-Hieber, J. (2016) Conditions for posterior contraction in the sparse normal means problem. *Electron. J. Stat.*, **10 (1)**, 976–1000. MR3486423