Variance Estimation in a Model with Gaussian Sub-Models

Vanja M. Dukić* Edsel A. Peña†

January 16, 2005

Abstract

This paper considers the problem of estimating the dispersion parameter in a Gaussian model which is intermediate between a model where the mean parameter is fully known (fixed) and a model where the mean parameter is completely unknown. One of the goals is to understand the implications of the two-step process of first selecting a model among a finite number of sub-models, and then estimating a parameter of interest after the model selection, but using the same sample data. The estimators are classified into global, two-step, and weighted-type estimators. While the global-type estimators ignore the model space structure, the two-step estimators explore the structure adaptively and can be related to pre-test estimators, and the weighted estimators are motivated by the Bayesian approach. Their performances are compared theoretically and through simulations using their risk functions based on a scale invariant quadratic loss function. It is shown that in the variance estimation problem efficiency gains arise by exploiting the sub-model structure through the use of two-step and weighted estimators, especially when the number of competing sub-models is few; but that this advantage may deteriorate or be lost altogether for some two-step estimators as the number of sub-models increases or as the distance between them decreases. Furthermore, it is demonstrated that weighted estimators, arising from properly chosen priors, outperform two-step estimators when there are many competing sub-models or when the sub-models are close to each other, whereas two-step estimators are preferred when the sub-models are highly distinguishable. The results have implications regarding model averaging and model selection issues.

AMS 2000 Subject Classifications: Primary 62F35, 62F10; Secondary 62C12, 62C25.

Key Words and Phrases: Model selection; inference after model selection, model averaging; Bayes estimators; pre-test estimators; two-step estimators, adaptive estimators.

*V. Dukić is Assistant Professor in the Department of Health Studies, University of Chicago, Chicago, IL 60637. She acknowledges the U.S. EPA STAR Grant R-82940201-0 and NIH 1 R01 DA015223-01A1 for partial support of her research. Her e-mail address is *vanja@uchicago.edu*.

†E. Peña is Professor in the Department of Statistics, University of South Carolina, Columbia, SC 29208. He acknowledges NSF Grant 0102870, NIH Grant GM56182, and NIH COBRE Grant RR17698 for partial support of his research. His e-mail address is *pena@stat.sc.edu*.

The authors thank the Editor, Associate Editor, and Referees for their comments and criticisms which led to a considerably improved paper.

1 Model Selection and Inference

In a variety of settings in statistical practice, it is common to encounter the following situation: we observe data $\boldsymbol{X}$ from a distribution F which is only known to belong to one of p (possibly nested) sub-models $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_p$; and given $\boldsymbol{X}$, we want to estimate a common parameter, or a functional, of F, denoted by $\tau(F)$. For example, we might observe $\boldsymbol{X} \sim F$ where F belongs to either the gamma or Weibull family of distributions, and wish to estimate the mean of F. Or, in a multiple regression setting with p possible predictors, we might want to choose one of the 2^p competing sub-models (Breiman (1992); Zhang (1992b,a)), and then estimate a common parameter such as dispersion or the conditional distribution function of the response variable.

The most frequent strategies for estimating $\tau(F)$ are: (i) utilizing an estimator developed under a larger model $\mathcal{M}$, which contains all sub-models; (ii) using data $\boldsymbol{X}$ to first choose a sub-model, and then applying the estimator developed for the chosen sub-model to the same data X; and (iii) assigning to each sub-model a plausibility measure, possibly using $\boldsymbol{X}$, and then forming a weighted combination of the estimators developed under each of the sub-models. In this paper we are interested in determining whether there is a preferred strategy, and whether that preferred strategy depends on the interplay among the competing sub-models, and possibly the parameter we are estimating.

Issues pertaining to the two-step process of inference after model selection and the consequences of "data double-dipping" in strategy (ii) have been discussed in the econometric literature (Judge, Bock, and Yancey (1974), Leamer (1978), Yancey, Judge, and Mandy (1983), and Wallace (1977)). Further investigations of these issues in other settings are in Potscher (1991), Buhlmann (1999), and Burnham and Anderson (1998). The third strategy has been discussed mostly in the context of model averaging, a notion that naturally arises in the Bayesian paradigm (Madigan and Raftery (1994); Raftery, Madigan, and Hoeting (1997); Hoeting, Madigan, Raftery, and Volinsky (1999); Burnham and Anderson (1998), among many others). The first strategy on the other hand may be viewed as having a nonparametric flavor. Though it is clearly intuitive that the first strategy will entail some loss in efficiency, it is not apparent whether (and when) the second strategy is preferred over the third strategy. Clearly, an examination of this problem in the general framework is important to provide guidance to practitioners regarding which strategy is better in general situations. However, a general treatment of the problem may not yield exact results, and one may need to rely on asymptotics, or local asymptotics such as in the work by Claeskens and Hjort (2003) and Hjort and Claeskens (2003).

In this paper, we focus our attention on a prototype Gaussian model which admits exact results and thereby enables concrete comparison of the three strategies. Though the specific model examined in this paper may be perceived as restrictive, it highlights the difficulties inherent in this problem. In addition, the specific estimation problem examined – the estimation of dispersion parameter – is still the subject of active research (Arnold and Villasenor (1997), Brewster and Zidek (1974), Gelfand and Dey (1977), Maatta and Casella (1990), Ohtani (2001), Pal, Ling, and Lin (1998), Rukhin (1987), Vidaković and DasGupta (1995), and Wallace (1977)).

The paper is outlined as follows. Section 2 will describe the formal setting of the specific problem considered, introduce notation, and present the global-type estimators. Section 3 will present the classical two-step estimators, whereas the Bayes and weighted estimators will be developed in Section 4. Distributional properties and risk comparison will be obtained in Section 5. Concluding remarks are given in Section 6, while Appendix A gathers the technical proofs.

2 Global-Type Estimators

We first describe the specific model examined in this paper. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ be a vector of IID random variables from an unknown distribution function $F(x) = \Pr\{X_1 \leq x\}$ which belongs to the two-parameter normal family of distributions $\mathcal{M} = \{N(\mu, \sigma^2) : (\mu, \sigma^2) \in \Theta = \Re \times \Re_+\}$. If interest is on estimating the variance σ^2, then the uniformly minimum variance unbiased estimator (UMVUE) of σ^2 is

$$\hat{\sigma}^2_{UMVU} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (Lehmann and Casella (1998)). We adopt a decision-theoretic approach for evaluating estimators of σ^2 via the risk function based on the scale invariant quadratic loss function $L : \Re \times \Theta \to \Re$

$$L(a, (\mu, \sigma^2)) = \left(\frac{a - \sigma^2}{\sigma^2}\right)^2. \tag{2}$$

It should be pointed that the appropriateness of this loss function has been questioned, partly because of Stein (1964)'s demonstration that under this loss the UMVUE of σ^2 is inadmissible and dominated by the minimum risk equivariant estimator (MRE)

$$\hat{\sigma}^2_{MRE} = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{3}$$

(which also turns out to be inadmissible). However, quadratic loss functions are still popular when dealing with the estimation of variance (Arnold and Villasenor (1997), Maatta and Casella (1990), Ohtani (2001), Pal et al. (1998), Rukhin (1987), Vidaković and DasGupta (1995), and Wallace (1977)).

If the model is restricted so that $\mu = \mu_0$ where $\mu_0 \in \Re$ is known, so $\mathcal{M}_0 = \{N(\mu, \sigma^2) : (\mu, \sigma^2) \in \Theta_0 = \{\mu_0\} \times \Re_+\}$, the UMVUE and MRE of σ^2 are given, respectively, by

$$\hat{\sigma}^2_{UMVU}(\mu_0) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 \quad \text{and} \quad \hat{\sigma}^2_{MRE}(\mu_0) = \frac{1}{n+2}\sum_{i=1}^{n}(X_i - \mu_0)^2. \tag{4}$$

Incidentally, $\hat{\sigma}^2_{UMVU}(\mu_0)$ is also the minimax estimator of σ^2 since it is a limit of proper Bayes estimators and it has constant risk. Clearly, we are able to improve on the estimators derived under $\mathcal{M}$ by exploiting the knowledge that $\mu = \mu_0$ under $\mathcal{M}_0$: when model $\mathcal{M}_0$ holds, the relative efficiency of the estimator $\hat{\sigma}^2_{UMVU}(\mu_0)$ in (4) with respect to $\hat{\sigma}^2_{UMVU}$ in (1) is $n/(n-1)$. But suppose now that we have a model between $\mathcal{M}$ and $\mathcal{M}_0$. Specifically, let p be a known positive integer, and $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_p\}$ be a set of known real numbers, and consider the estimation of σ^2 under the model

$$\mathcal{M}_p = \mathcal{M}_p(\boldsymbol{\mu}) = \{N(\mu, \sigma^2) : (\mu, \sigma^2) \in \Theta_p \equiv \{\mu_1, \mu_2, \ldots, \mu_p\} \times \Re_+\}.$$

In $\mathcal{M}_p$, in contrast to $\mathcal{M}_0$, there is some information about the possible value of μ, but we are not certain about this value. Model $\mathcal{M}_p$ can be viewed as having p sub-models, with the ith sub-model, $\mathcal{M}_{p,i}$, being the normal class with unknown variance σ^2 and known mean μ_i, that is, $\mathcal{M}_{p,i} = \{N(\mu_i, \sigma^2) : \sigma^2 > 0\}$. This particular model arises in a variety of settings. For example, it includes decision problems with a two-element action space such as in the Neyman-Pearson hypothesis testing setting. If we further allow the possibility that $\mu \in \Re \setminus \{\mu_1, \mu_2, \ldots, \mu_p\}$, we obtain a generalization of the setting utilized by Stein (1964) to derive an estimator dominating $\hat{\sigma}^2_{MRE}$ (see Brewster and Zidek (1974), Wallace (1977), and Maatta and Casella (1990)). The Stein estimator is given by

$$\hat{\sigma}^2_{ST} = \min\left\{\hat{\sigma}^2_{MRE}, \frac{1}{n+2}\sum_{i=1}^{n} X_i^2\right\}. \tag{5}$$

This estimator can be viewed as a preliminary test estimator (Sen and Saleh (1987), Lehmann and Casella (1998), and Sclove, Morris, and Radhakrishnan (1972)). The pre-test estimation approach proceeds by testing a null hypothesis that the parameter equals a certain value, and if it accepts the hypothesis then the estimator based on this parameter value is used; otherwise, an estimator under the general model is used. Since a test is to be performed, a level of significance needs to be specified, and so generally pre-test estimators will depend on such a specified level. Interestingly, the Stein estimator in (5), which could be derived as a pre-test estimator with hypothesis specifying that $\mu = 0$, eliminates this dependence by utilizing an 'optimal' significance level. The approach implemented in our paper differs from that of pre-test estimation, since we altogether avoid the need for testing a hypothesis to derive our estimators.

Note that in problems dealing with the estimation of the normal variance, it is typically assumed that either model $\mathcal{M}$ or model $\mathcal{M}_0$ holds. However, in many settings, the mean can take only a finite number of possible values, such as for example in the Neyman-Pearson lemma, where the variance estimator is typically the sample variance, which does not exploit the fact that there are only two possible means.

3 Classical Two-Step Estimators

Under $\mathcal{M}_p$ the likelihood function for the sample realization $\boldsymbol{X} = \boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ is

$$L(\mu, \sigma^2) = L(\mu, \sigma^2 | \boldsymbol{x}) = \prod_{i=1}^{p} L_i(\mu_i, \sigma^2)^{M_i} \tag{6}$$

where, for $i = 1, 2, \ldots, p$, with $I\{\cdot\}$ denoting the indicator function and $M_i = I\{\mu = \mu_i\}$,

$$L_i(\mu_i, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{n\hat{\sigma}_i^2}{2\sigma^2}\right\} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu_i)^2. \tag{7}$$

$L_i(\mu_i, \sigma^2)$ is maximized with respect to σ^2 at $\hat{\sigma}_i^2$, so $L_i(\mu_i, \hat{\sigma}_i^2) = \sup_{\sigma^2 \in \Re_+} L_i(\mu_i, \sigma^2)$. Define the likelihood-based model selector $\hat{M} = \hat{M}(\boldsymbol{X})$ via

$$\hat{M} = \arg\max_{1 \le i \le p} L_i(\mu_i, \hat{\sigma}_i^2) = \arg\min_{1 \le i \le p} \hat{\sigma}_i^2 = \arg\min_{1 \le i \le p} |\bar{X} - \mu_i|.$$

One could employ model selectors different from $\hat{M}$, such as the highest posterior probability (à la Schwarz' Bayesian criterion (SBC) (Schwarz (1978)) or the Akaike information criterion (AIC) (Akaike (1973), Burnham and Anderson (1998))). In this paper we restrict our attention to the intuitive selector $\hat{M}$, which could actually be viewed also as a highest posterior probability model selector associated with a flat prior distribution. The maximum likelihood estimator (MLE) of σ^2 under $\mathcal{M}_p$ is

$$\hat{\sigma}^2_{p,MLE} = \hat{\sigma}^2_{\hat{M}} = \sum_{i=1}^{p} I\{\hat{M} = i\}\hat{\sigma}_i^2, \tag{8}$$

a two-step estimator, with the first stage selecting the sub-model and the second-stage using the MLE of σ^2 in the chosen sub-model. An alternative to the estimator (8) is to use the sub-model's MRE instead of MLE of σ^2:

$$\hat{\sigma}^2_{p,MRE} = \hat{\sigma}^2_{MRE,\hat{M}} = \sum_{i=1}^{p} I\{\hat{M} = i\}\hat{\sigma}^2_{MRE,i} = \sum_{i=1}^{p} I\{\hat{M} = i\}\frac{n\hat{\sigma}_i^2}{(n+2)}. \tag{9}$$

Note that the label '${}_{p,MRE}$' (and similar labels in the sequel) is a misnomer since this estimator need not be minimum risk equivariant under model $\mathcal{M}_p$. However, we keep the name for clarity.

4 Bayes and Weighted Estimators

We focus on the class of prior densities of (μ, σ^2) which consists of the product of a multinomial probability function and an inverse gamma density:

$$\pi(\mu, \sigma^2 | \tilde{\boldsymbol{\theta}}, \kappa, \beta) = \left(\prod_{i=1}^{p} \tilde{\theta}_i^{m_i} \right) \frac{\beta^{\kappa-1}}{\Gamma(\kappa-1)} \left(\frac{1}{\sigma^2} \right)^{\kappa} \exp\left(-\frac{\beta}{\sigma^2} \right), \tag{10}$$

where $\sigma^2 > 0$, $m_i = I\{\mu = \mu_i\}$ so that $\sum_{i=1}^{p} m_i = 1$, and $0 \leq \tilde{\theta}_i \leq 1$ with $\sum_{i=1}^{p} \tilde{\theta}_i = 1$, $\beta > 0$, and $\kappa > 1$. From (10) and (6), we obtain the posterior density of (μ, σ^2) given $\boldsymbol{X} = \boldsymbol{x}$:

$$\pi(\mu, \sigma^2 \mid \boldsymbol{x}) \;=\; C \prod_{i=1}^{p} \left\{ \tilde{\theta}_i \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}+\kappa} \exp\left(-\frac{1}{\sigma^2} \left[\frac{n\hat{\sigma}_i^2}{2} + \beta \right] \right) \right\}^{m_i}. \tag{11}$$

Note that in the "vectorized form," and with a slight abuse of notation, $\pi(\mu, \sigma^2|\boldsymbol{x}) = \pi(\boldsymbol{m}, \sigma^2|\boldsymbol{x})$ because $\{\mu = \mu_i\} = \{\boldsymbol{m} = \mathbf{1}_i\}$, where $\mathbf{1}_i$ is an $n \times 1$ vector with ith component equal to 1 and all others equal 0. It follows that

$$C \;=\; \frac{1}{\Gamma(n/2+\kappa-1)} \left\{ \sum_{i=1}^{p} \frac{\tilde{\theta}_i}{(n\hat{\sigma}_i^2/2+\beta)^{n/2+\kappa-1}} \right\}^{-1}. \tag{12}$$

4.1 Posterior Probabilities

¿From the posterior distribution in (11), the marginal posterior density (with respect to counting measure) of μ, or equivalently of $\boldsymbol{m}$, is

$$\pi(\boldsymbol{m} \mid \mathbf{x}) = C \prod_{i=1}^{p} \left\{ \tilde{\theta}_i \frac{\Gamma(n/2+\kappa-1)}{(n\hat{\sigma}_i^2/2+\beta)^{n/2+\kappa-1}} \right\}^{m_i} = \prod_{i=1}^{p} \{\theta_i(\kappa, \beta, n, \mathbf{x})\}^{m_i},$$

where, for $i = 1, 2, \ldots, p$, the posterior probability that the sub-model $\mathcal{M}_{p,i}$ is true, is

$$\theta_i(\kappa, \beta, n, \boldsymbol{x}) \;=\; \frac{\tilde{\theta}_i (n\hat{\sigma}_i^2/2+\beta)^{-(n/2+\kappa-1)}}{\sum_{j=1}^{p} \tilde{\theta}_j \left(n\hat{\sigma}_j{}^2/2+\beta \right)^{-(n/2+\kappa-1)}}. \tag{13}$$

Note, as expected, that if $\tilde{\theta}_i > 0$ and $\mathcal{M}_{p,i}$ is the true sub-model, $\theta_i(\kappa, \beta, n, \boldsymbol{X})$, when viewed as a function of $\boldsymbol{X}$ with (μ, σ^2) fixed, converges to 1 with probability one (wp1) as $n \to \infty$. This is because if $\mathcal{M}_{p,i}$ is the correct model, $\hat{\sigma}_i^2$ converges wp1 to σ^2 by the strong law of large numbers (SLLN); whereas, for $i' \neq i$, $\hat{\sigma}_{i'}^2$ converges wp1 to $\sigma^2 + (\mu_i - \mu_{i'})^2$.

4.2 Estimators

The marginal posterior density function of σ^2 is directly obtained from (11) to be

$$\pi(\sigma^2 \mid \boldsymbol{x}) \;=\; C \sum_{i=1}^{p} \tilde{\theta}_i \left(\frac{1}{\sigma^2} \right)^{(\kappa+n/2)} \exp\left[-\frac{1}{\sigma^2} \left(\frac{n\hat{\sigma}_i^2}{2} + \beta \right) \right] I\{\sigma^2 > 0\}. \tag{14}$$

The posterior mean, which is the Bayes estimator of σ^2 under the loss function L in (2), is then

$$\hat{\sigma}^2_{p,Bayes}(\kappa,\beta,\boldsymbol{\theta}) = \sum_{i=1}^{p} \theta_i(\kappa,\beta,n,\boldsymbol{x}) \left\{ \left(\frac{n}{n+2(\kappa-2)}\right)\hat{\sigma}_i^2 + \left(\frac{2(\kappa-2)}{n+2(\kappa-2)}\right)\left(\frac{\beta}{\kappa-2}\right)\right\}. \quad (15)$$

Note that $\beta/(\kappa-2)$ is the prior mean of σ^2, provided $\kappa > 2$ (the condition also needed for the prior variance of σ^2 to exist), whereas $\hat{\sigma}_i^2$ is the MLE of σ^2 under the $\mathcal{M}_{p,i}$ model. This estimator mixes in a data-dependent manner, using the posterior probabilities of the p sub-models, the Bayes estimators of σ^2 from each sub-model. Furthermore, the Bayes estimator of σ^2 for the $\mathcal{M}_{p,i}$ sub-model is a convex combination of the $\mathcal{M}_{p,i}$-model MLE and the prior mean of σ^2, a well-known result.

To obtain limiting Bayes estimators for σ^2, we consider improper priors arising by setting $\tilde{\theta}_i = 1/p$, $i = 1,2,\ldots,p$, and $\beta \to 0$. We examine four κ values: $\kappa \to 1$, $\kappa \to 3/2$, $\kappa = 2$, and $\kappa = 3$. The rationale for these choices is as follows: $\kappa \to 1$ amounts to placing Jeffreys' non-informative prior on σ^2 in each of the p sub-models, since Jeffreys' prior for σ^2 (with mean known) is proportional to $1/\sigma^2$ (Robert (2001)); $\kappa \to 3/2$ corresponds to the Jeffreys' prior for σ^2 when the mean is unknown in the normal model, since in this case Jeffreys' prior is proportional to $(1/\sigma^2)^{(3/2)}$; $\kappa = 2$ and $\kappa = 3$ produce (limiting) Bayes estimators that are convex combinations of the sub-models' MLEs and MREs, respectively. Table 1 lists the sub-models' posterior probabilities and the resulting limiting Bayes estimators of σ^2. Each of the sub-models' posterior probabilities

Table 1: Sub-models' posterior probabilities and limiting Bayes estimators of σ^2 for different values of κ when $\tilde{\theta}_i = 1/p$ and $\beta \to 0$.

κ	Sub-model Posterior Probabilities, $\theta_i(\kappa,0,n,\boldsymbol{x}), i = 1,2,\ldots,p$	Limiting Bayes Estimator $\hat{\sigma}^2_{p,LBk}, k = 1,2,3,4$
1	$\theta_{i1} = (\hat{\sigma}_i^2)^{-n/2}/\sum_{j=1}^p (\hat{\sigma}_j^2)^{-n/2}$	$\hat{\sigma}^2_{p,LB1} = \left(\frac{n}{n-2}\right)\sum_{i=1}^p \theta_{i1}\hat{\sigma}_i^2$
3/2	$\theta_{i2} = (\hat{\sigma}_i^2)^{-(n+1)/2}/\sum_{j=1}^p (\hat{\sigma}_j^2)^{-(n+1)/2}$	$\hat{\sigma}^2_{p,LB2} = \left(\frac{n}{n-1}\right)\sum_{i=1}^p \theta_{i2}\hat{\sigma}_i^2$
2	$\theta_{i3} = (\hat{\sigma}_i^2)^{-(n+2)/2}/\sum_{j=1}^p (\hat{\sigma}_j^2)^{-(n+2)/2}$	$\hat{\sigma}^2_{p,LB3} = \sum_{i=1}^p \theta_{i3}\hat{\sigma}_i^2$
3	$\theta_{i4} = (\hat{\sigma}_i^2)^{-(n+4)/2}/\sum_{j=1}^p (\hat{\sigma}_j^2)^{-(n+4)/2}$	$\hat{\sigma}^2_{p,LB4} = \left(\frac{n}{n+2}\right)\sum_{i=1}^p \theta_{i4}\hat{\sigma}_i^2$

associated with $\kappa \in \{1, 3/2, 2, 3\}$ given in Table 1 could also be utilized to form estimators which are convex combinations of the sub-models' MREs. These new estimators need not however be limiting Bayes with respect to our class of priors. These 'weighted' estimators are defined as:

$$\hat{\sigma}^2_{p,PLB1} = \left(\frac{n-2}{n+2}\right)\hat{\sigma}^2_{p,LB1}; \quad \hat{\sigma}^2_{p,PLB2} = \left(\frac{n-1}{n+2}\right)\hat{\sigma}^2_{p,LB2}; \quad \hat{\sigma}^2_{p,PLB3} = \left(\frac{n}{n+2}\right)\hat{\sigma}^2_{p,LB3}. \quad (16)$$

Note also from (15) that the estimators $\tilde{\sigma}^2_{LB,i} = (n/(n-2))\,\hat{\sigma}^2_i$, the ones whose convex combination is being formed in $\hat{\sigma}^2_{p,LB1}$, are the limiting Bayes estimators of σ^2 for each of the p sub-models under Jeffreys' non-informative prior when the sub-model's mean is known (arising from $\kappa \to 1$). The estimators in Table 1 and in (16) have different flavors than the MLE of σ^2 given in (8): in the latter, we choose one among the p estimators of σ^2, while the Bayes and weighted estimators are mixing sub-model estimators according to the sub-models' posterior probabilities.

Finally, we define the two-step estimator based on the sub-models' limiting Bayes estimators:

$$\hat{\sigma}^2_{p,ALB} = \tilde{\sigma}^2_{LB,\hat{M}} = \left(\frac{n}{n-2}\right) \sum_{i=1}^{p} I\{\hat{M} = i\}\hat{\sigma}^2_i. \tag{17}$$

This belongs to the same class of estimators as $\hat{\sigma}^2_{p,MLE}$ and $\hat{\sigma}^2_{p,MRE}$, differing just in the multipliers which are functions of n only. Note that for the purposes of obtaining risk functions, it suffices to derive formulas for the mean and variance functions of $\hat{\sigma}^2_{p,MLE}$.

5 Comparison of Estimators

The goal of this section is to compare the performances of the estimators given in Table 1 and (16), with the estimators developed under $\mathcal{M}$ ($\hat{\sigma}^2_{UMVU}$ and $\hat{\sigma}^2_{MRE}$), with the two-step estimators ($\hat{\sigma}^2_{p,MLE}$, $\hat{\sigma}^2_{p,MRE}$, $\hat{\sigma}^2_{p,MLE}$, and $\hat{\sigma}^2_{p,ALB}$) and, for completeness, with the Stein estimator in (5). Performance will be measured by their risk functions arising from the loss function L in (2). In particular, we address the following questions: (i) How much efficiency is lost by using the estimators developed under the wider model $\mathcal{M}$ when model $\mathcal{M}_p$ holds? (ii) How do the limiting Bayes and weighted estimators $\hat{\sigma}^2_{p,LBk}$ and $\hat{\sigma}^2_{p,PLBk}$ compare with the $\mathcal{M}_p$ MLE-based and MRE-based estimators? (iii) Do the advantages of the $\mathcal{M}_p$-based estimators over $\mathcal{M}$-based estimators decrease as the dimension p increases and/or the spacings among the $\mu_1, ..., \mu_p$ decrease?

5.1 Distributional Representations

It is well-known that, provided $n > 1$, $(n-1)\hat{\sigma}^2_{UMVU}/\sigma^2 \sim \chi^2_{n-1}$, so $\mathbf{E}\{\hat{\sigma}^2_{UMVU}/\sigma^2\} = 1$ and $\mathbf{Var}\{\hat{\sigma}^2_{UMVU}/\sigma^2\} = 2/(n-1)$. Therefore, the risk function of $\hat{\sigma}^2_{UMVU}$ with respect to the loss function L in (2) is $R\left(\hat{\sigma}^2_{UMVU}, (\mu, \sigma^2)\right) = 2/(n-1)$. By exploiting the relationship between $\hat{\sigma}^2_{UMVU}$ and $\hat{\sigma}^2_{MRE}$, the risk function of the latter is easily found to be $R\left(\hat{\sigma}^2_{MRE}, (\mu, \sigma^2)\right) = 2/(n+1)$. This demonstrates the known fact that $\hat{\sigma}^2_{UMVU}$ is inadmissible. To compare estimator performances, we will use $\hat{\sigma}^2_{UMVU}$ as the baseline, so the efficiency of an estimator $\hat{\sigma}^2$ will be given by

$$\mathrm{Eff}(\hat{\sigma}^2 : \hat{\sigma}^2_{UMVU}) = \frac{R(\hat{\sigma}^2_{UMVU}, (\mu, \sigma^2))}{R(\hat{\sigma}^2, (\mu, \sigma^2))}. \tag{18}$$

Thus, in particular, Eff$(\hat{\sigma}^2_{MRE} : \hat{\sigma}^2_{UMVU}) = (n+1)/(n-1) = 1 + 2/(n-1)$.

We present some distributional properties of the estimators which will be used to derive the exact expressions of the risk functions of $\hat{\sigma}^2_{p,MLE}$, and second-order approximations to the risk functions of $\hat{\sigma}^2_{p,LBk}$ and $\hat{\sigma}^2_{p,PLBk}$. Let $Z \sim N(0,1)$ and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)' \sim N_n(\boldsymbol{0}, \boldsymbol{I})$. For the vector of means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ with μ_{i_0} being the true mean ($i_0 \in \{1, 2, \ldots, p\}$), we let

$$\boldsymbol{\Delta} \equiv \boldsymbol{\Delta}(\boldsymbol{\mu}, \sigma) = \frac{\boldsymbol{\mu} - \mu_{i_0}\mathbf{1}}{\sigma} \tag{19}$$

where $\mathbf{1} = (1, 1, \ldots, 1)'$. Note that this will always have a zero component under $\mathcal{M}_p$. In the sequel, the 'equal-in-distribution' relation is denoted by '$\stackrel{d}{=}$'. To achieve a more fluid presentation, formal proofs of lemmas, propositions, theorems, and corollaries are relegated to Appendix A.

Proposition 5.1 *Under $\mathcal{M}_p$ with μ_{i_0} the true mean, $n\hat{\sigma}^2_i/\sigma^2 \stackrel{d}{=} W + V_i^2, i = 1, 2, \ldots, p$, where $W \sim \chi^2_{n-1}$, $\boldsymbol{V} = (V_1, V_2, \ldots, V_p)' \sim N_p(-\sqrt{n}\boldsymbol{\Delta}, \boldsymbol{J} \equiv \mathbf{1}\mathbf{1}')$, and W and $\boldsymbol{V}$ are independent.*

¿From Proposition 5.1, by exploiting the independence between W and $\boldsymbol{V}$ and using the iterated expectation and covariance rules, and by noting that

$$\mathbf{E}\{W^{k/2}\} = (1/2)^{-k/2} \left[\Gamma((n+k-1)/2)/\Gamma((n-1)/2)\right]$$

holds for any $k < n-1$, the following corollary immediately follows.

Corollary 5.1 *Under the conditions of Proposition 5.1, $n\hat{\sigma}^2_i/\sigma^2 \stackrel{d}{=} W(1+T_i^2), i = 1, 2, \ldots, p$, with $\boldsymbol{T} = (T_1, \ldots, T_p)' = \boldsymbol{V}/\sqrt{W}$. The distribution of $\boldsymbol{T}$ depends on $(\boldsymbol{\mu}, \sigma^2)$ only through $\boldsymbol{\Delta}$ and, provided that $n > 3$, the mean vector and covariance matrix of $\boldsymbol{T}$ are given, respectively, by*

$$\mathbf{E}(\boldsymbol{T}) = \boldsymbol{\nu} \equiv -\boldsymbol{\Delta} C_n \quad \textit{and} \quad \mathbf{Cov}(\boldsymbol{T}, \boldsymbol{T}) = \frac{1}{n-3}\boldsymbol{J} + \left(\frac{n}{(n-3)C_n^2} - 1\right)\boldsymbol{\nu}^{\otimes 2}$$

with $C_n = \sqrt{n/2}\left[\Gamma((n-2)/2)/\Gamma((n-1)/2)\right]$.

5.2 Representation and Risk Function of $\hat{\sigma}^2_{p,MLE}$

We now give a representation of $\hat{\sigma}^2_{p,MLE}$ and obtain the exact expressions for its mean, variance, and risk. For a given $\boldsymbol{\Delta}$, let $\Delta_{(1)} < \Delta_{(2)} < \ldots < \Delta_{(p)}$ denote the associated ordered values.

Theorem 5.1 *Let μ_{i_0} be the true mean. Then under $\mathcal{M}_p$,*

$$n\hat{\sigma}^2_{p,MLE}/\sigma^2 \stackrel{d}{=} W + \sum_{i=1}^{p} I\{L(\Delta_{(i)}, \boldsymbol{\Delta}) < Z < U(\Delta_{(i)}, \boldsymbol{\Delta})\}(Z - \sqrt{n}\Delta_{(i)})^2$$

where, under the convention that $\Delta_{(0)} = -\infty$ *and* $\Delta_{(p+1)} = +\infty$,

$$L(\Delta_{(i)}, \boldsymbol{\Delta}) = (\sqrt{n}/2)\left[\Delta_{(i)} + \Delta_{(i-1)}\right] \quad \textit{and} \quad U(\Delta_{(i)}, \boldsymbol{\Delta}) = (\sqrt{n}/2)\left[\Delta_{(i)} + \Delta_{(i+1)}\right],$$

$W \sim \chi^2_{n-1}$, $Z \sim N(0,1)$, *and* W *and* Z *are independent.*

Define the events $\Omega_{(i)} = \left\{L(\Delta_{(i)}, \boldsymbol{\Delta}) < Z < U(\Delta_{(i)}, \boldsymbol{\Delta})\right\}, i = 1, 2, \ldots, p$. The collection of sets $\{\{\hat{M} = i\}, i = 1, 2, \ldots, p\}$ is in one-to-one correspondence with the collection $\{\Omega_{(1)}, \Omega_{(2)}, \ldots, \Omega_{(p)}\}$, as can be seen from Theorem 5.1, and thus $\Omega_{(i)}, i = 1, 2, \ldots, p$, are disjoint. Using Theorem 5.1, we can now obtain expressions of the mean and variance of $\hat{\sigma}^2_{p,MLE}$. For $i = 1, 2, \ldots, p$, we let

$$P_{(i)}(\boldsymbol{\Delta}) \equiv \Pr\{\Omega_{(i)}\} = \Phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i+1)})\right) - \Phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i-1)})\right), \tag{20}$$

where $\Phi(\cdot)$ is the standard normal distribution function. In the sequel, we let $\phi(\cdot)$ denote the density function of a standard normal random variable.

Theorem 5.2 *Under the conditions of Theorem 5.1,*

$$\begin{aligned} EpMLE(\boldsymbol{\Delta}) \equiv \mathbf{E}\left\{\frac{\hat{\sigma}^2_{p,MLE}}{\sigma^2}\right\} &= 1 - \frac{2}{\sqrt{n}}\sum_{i=1}^{p} \Delta_{(i)}[\phi(L(\Delta_{(i)}, \boldsymbol{\Delta})) - \phi(U(\Delta_{(i)}, \boldsymbol{\Delta}))] + \\ &\sum_{i=1}^{p} \Delta^2_{(i)}[\Phi(U(\Delta_{(i)}, \boldsymbol{\Delta})) - \Phi(L(\Delta_{(i)}, \boldsymbol{\Delta}))]. \end{aligned}$$

Next, we present an expression for the variance function of the estimator $\hat{\sigma}^2_{p,MLE}$. Towards this end, we introduce some notation to simplify the presentation. For $k \in \mathcal{Z}_+ = \{0, 1, 2, \ldots\}$, define

$$\xi(k; \Omega_{(i)}) \equiv \mathbf{E}\left\{Z^k I(\Omega_{(i)})\right\} = \int_{L(\Delta_{(i)}, \boldsymbol{\Delta})}^{U(\Delta_{(i)}, \boldsymbol{\Delta})} z^k \phi(z) dz.$$

Using this, observe that by the binomial expansion, for $m \in \mathcal{Z}_+$,

$$\zeta_{(i)}(m) \equiv \mathbf{E}\left\{I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^m\right\} = \sum_{k=0}^{m} (-1)^{(m-k)} \binom{m}{k} \left(\sqrt{n}\Delta_{(i)}\right)^{(m-k)} \xi(k; \Omega_{(i)}). \tag{21}$$

To compute the quantity $\xi(k; \Omega_{(i)})$, observe that for $k \in \mathcal{Z}_+$ and $t \in \Re$,

$$\int_{-\infty}^{t} z^k \phi(z) dz = (-1)^k \frac{2^{(k-1)/2}}{\sqrt{2\pi}} \Gamma((k+1)/2) \times \begin{cases} \Pr\{\chi^2_{k+1} > t^2\} & \text{if } t < 0 \\ \left[1 + (-1)^k \Pr\{\chi^2_{k+1} < t^2\}\right] & \text{if } t \geq 0 \end{cases}.$$

Using the above formulas, we obtain $\xi(k; \Omega_{(i)})$ according to

$$\xi(k; \Omega_{(i)}) = \int_{-\infty}^{U(\Delta_{(i)}, \boldsymbol{\Delta})} z^k \phi(z) dz - \int_{-\infty}^{L(\Delta_{(i)}, \boldsymbol{\Delta})} z^k \phi(z) dz. \tag{22}$$

Theorem 5.3 *Under the conditions of Theorem 5.1,*

$$VpMLE(\mathbf{\Delta}) \equiv \mathbf{Var}\left\{\frac{\hat{\sigma}^2_{p,MLE}}{\sigma^2}\right\} = \frac{1}{n}\left\{2\left(1-\frac{1}{n}\right)+\frac{1}{n}\left[\sum_{i=1}^{p}\zeta_{(i)}(4)-\left(\sum_{i=1}^{p}\zeta_{(i)}(2)\right)^2\right]\right\}.$$

In the situation where there are only two sub-models so $p = 2$, the expressions for the mean and variance functions of $\hat{\sigma}^2_{p,MLE}/\sigma^2$ simplify. These simplified forms are provided in the following corollary. The proofs of these results are straightforward, hence to conserve space, we omit them but instead refer the reader to the more detailed technical report by Dukić and Peña (2003).

Corollary 5.2 *If* $p = 2$ *so* $\mathbf{\Delta} = (0, \Delta)$, *then under the conditions of Theorem 5.1,*

$$\begin{aligned} EpMLE(\Delta) &= 1-\left(\frac{2}{\sqrt{n}}|\Delta|\right)\left\{\phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)-\left(\frac{\sqrt{n}}{2}|\Delta|\right)\left[1-\Phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)\right]\right\}; \\ VpMLE(\Delta) &= \frac{2}{n}+|\Delta|^4\Phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)\left[1-\Phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)\right]-\frac{4}{\sqrt{n}}|\Delta|^3\Phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)\times \\ &\quad \int_{\frac{\sqrt{n}}{2}|\Delta|}^{\infty} z\phi(z)dz-\frac{4}{n^{3/2}}|\Delta|\left\{\int_{\frac{\sqrt{n}}{2}|\Delta|}^{\infty} z^3\phi(z)dz-\int_{\frac{\sqrt{n}}{2}|\Delta|}^{\infty} z\phi(z)dz\right\}+ \\ &\quad \frac{1}{n}|\Delta|^2\left\{6\int_{\frac{\sqrt{n}}{2}|\Delta|}^{\infty} z^2\phi(z)dz-4\left(\int_{\frac{\sqrt{n}}{2}|\Delta|}^{\infty} z\phi(z)dz\right)^2-2\left[1-\Phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)\right]\right\}. \end{aligned}$$

We note from the expression in Corollary 5.2 that, for a fixed n, $\lim_{|\Delta|\to 0}$ EpMLE$(\Delta) \to 1$ and $\lim_{|\Delta|\to 0}$ VpMLE$(\Delta) \to 2/n = \mathbf{Var}\{\hat{\sigma}^2_{MLE}/\sigma^2\}$, where $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_{i_0})^2$, the ML (also UMVU, minimax) estimator of σ^2 under the true model. Also, for a fixed Δ, we see that $\lim_{n\to\infty}$ EpMLE$(\Delta) \to 1$ and $\lim_{n\to\infty}\{n(\text{VpMLE}(\Delta))\} \to 2$.

The next result in Corollary 5.3 shows that even though the sub-models' MLEs are each unbiased for σ^2, the two-step estimator $\hat{\sigma}^2_{p,MLE}$, which employs the MLE of the sub-model selected by the model selector $\hat{M}$, is a negatively biased estimator of σ^2. The result is an immediate consequence of Corollary 5.2 by noting that the continuous function $g(u) = \phi(u) - u[1-\Phi(u)]$ is positive by virtue of the facts that $\lim_{u\downarrow 0} h(u) > 0$, $\lim_{u\to\infty} h(u) = 0$, and $g'(u) = \phi'(u) + u\phi(u) - [1-\Phi(u)] = -[1-\Phi(u)] < 0$ since $\phi'(u) = -u\phi(u)$.

Corollary 5.3 *Under the conditions of Corollary 5.2 with* $\Delta \neq 0$, $\mathbf{E}\{\hat{\sigma}^2_{p,MLE}\} < \sigma^2$, *that is,* $\hat{\sigma}^2_{p,MLE}$ *is negatively biased for* σ^2.

Now that we have the exact expressions for the mean and variance of $\hat{\sigma}^2_{p,MLE}/\sigma^2$, we could obtain the risk function of $\hat{\sigma}^2_{p,MLE}$ under $\mathcal{M}_p$ and loss L in (2) as

$$R\left(\hat{\sigma}^2_{p,MLE}, (\mu_{i_0}, \sigma^2)\right) = \text{VpMLE}(\mathbf{\Delta}) + [\text{EpMLE}(\mathbf{\Delta}) - 1]^2. \tag{23}$$

Finally, for $\hat{\sigma}^2_{p,MLE}$, we address the question of what happens when p increases and the spacings in $\mathbf{\Delta}$ decrease. This will indicate whether we will lose the advantage of $\mathcal{M}_p$-based estimators over $\mathcal{M}$-based estimators. The proof of Theorem 5.4 is rather lengthy and hence omitted; instead we refer the reader to the technical report by Dukić and Peña (2003).

Theorem 5.4 *Given n fixed, if $p \to \infty$, $\max_{2\le i\le p} |\Delta_{(i)} - \Delta_{(i-1)}| \to 0$, with $\Delta_{(1)} \to \Delta_{min} \in (-\infty, 0]$, and $\Delta_{(p)} \to \Delta_{max} \in [0,\infty)$, then*

$$\begin{aligned}
&EpMLE(\mathbf{\Delta}) \to 1 - \frac{1}{n}\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} w^2\phi(w)dw + \frac{2}{\sqrt{n}}\left\{\Delta_{min}\phi(\sqrt{n}\Delta_{min}) - \Delta_{max}\phi(\sqrt{n}\Delta_{max})\right\} + \\
&\quad \left\{(\Delta_{min})^2\Phi(\sqrt{n}\Delta_{min}) + (\Delta_{max})^2[1-\Phi(\sqrt{n}\Delta_{max})]\right\};\\
&VpMLE(\mathbf{\Delta}) \to \frac{2}{n}\left(1-\frac{1}{n}\right) + \\
&\quad \frac{1}{n^2}\Big[\Big\{\mathbf{E}[(Z-\sqrt{n}\Delta_{min})^4 I(Z<\sqrt{n}\Delta_{min})] + \mathbf{E}[(Z-\sqrt{n}\Delta_{max})^4 I(Z>\sqrt{n}\Delta_{max})]\Big\} - \\
&\quad \Big\{\mathbf{E}[(Z-\sqrt{n}\Delta_{min})^2 I(Z<\sqrt{n}\Delta_{min})] + \mathbf{E}[(Z-\sqrt{n}\Delta_{max})^2 I(Z>\sqrt{n}\Delta_{max})]\Big\}^2\Big].
\end{aligned}$$

Letting $\Delta_{min} \to -\infty$ and $\Delta_{max} \to \infty$, $EpMLE(\mathbf{\Delta}) \to 1 - 1/n$ and $VpMLE(\mathbf{\Delta}) \to (2/n)(1-1/n)$.

Using Theorem 5.4 we could now address the issue of whether we lose the advantage by utilizing the two-step estimator which was developed under model $\mathcal{M}_p$ over the estimator developed under the more general model $\mathcal{M}$ when p increases. For this purpose we have the following corollary.

Corollary 5.4 *With $n > 1$ fixed, if as $p \to \infty$, $\max_{2\le i\le p} |\Delta_{(i)} - \Delta_{(i-1)}| \to 0$, $\Delta_{(1)} \to -\infty$, and $\Delta_{(p)} \to \infty$, then (i) Eff$\left(\hat{\sigma}^2_{p,MLE} : \hat{\sigma}^2_{UMVU}\right) \to 2n^2/[(n-1)(2n-1)] > 1$; (ii) Eff$\left(\hat{\sigma}^2_{p,MRE} : \hat{\sigma}^2_{UMVU}\right) \to 2(n+2)^2/[(n-1)(2n+7)] > 1$; (iii) Eff$\left(\hat{\sigma}^2_{p,MRE} : \hat{\sigma}^2_{p,MLE}\right) \to (2n-1)(n+2)^2/[(2n+7)n^2] > 1$; and (iv) Eff$\left(\hat{\sigma}^2_{p,MRE} : \hat{\sigma}^2_{MRE}\right) \to 2(n+2)^2/[(n+1)(2n+7)] < 1$. In addition, $\hat{\sigma}^2_{p,ALB}$ is dominated by $\hat{\sigma}^2_{UMVU}$.*

The fourth result in Corollary 5.4 indicates that when the number of sub-models increases indefinitely the estimator $\hat{\sigma}^2_{MRE}$ (which is the minimum risk estimator under the general model $\mathcal{M}$) dominates the two-step estimator $\hat{\sigma}^2_{p,MRE}$ (which was developed by exploiting the sub-model structure of $\mathcal{M}_p$). Using the limiting results for $p = 2$ and as $|\Delta| \to 0$, stated after Corollary 5.2, we find that the limiting risk function of $\hat{\sigma}^2_{p,MRE}$ is $2/(n+2)$, which is smaller than $2/(n+1)$, the risk function of $\hat{\sigma}^2_{MRE}$. This shows that when the number of sub-models is small, we can gain efficiency by using the two-step estimator developed under model $\mathcal{M}_p$. These results agree with our intuition: when the number of sub-models increases it is better to utilize the best estimator developed under the more general model. However, as it will be seen in the simulation studies reported later in the paper, the weighted and Bayes-type estimators' performance seems not degraded by an increase in the number of sub-models.

5.3 Representation of Limiting Bayes and Weighted Estimators

We now provide distributional representations useful for the limiting Bayes estimators $\hat{\sigma}^2_{p,LBk}$s and the weighted estimators $\hat{\sigma}^2_{p,PLBk}$s under $\mathcal{M}_p$, in order to find an approximation to the risk functions of these estimators. For $\alpha > 0$, define the "umbrella" estimator as

$$\hat{\sigma}^2_{p,LB} \equiv \hat{\sigma}^2_{p,LB}(\alpha) = \sum_{i=1}^{p} \left\{ \frac{(\hat{\sigma}^2_i)^{-\alpha}}{\sum_{j=1}^{p} (\hat{\sigma}^2_j)^{-\alpha}} \right\} \hat{\sigma}^2_i. \tag{24}$$

Individual estimators are easily derived from this umbrella estimator by choosing an appropriate α. For example:

$$\hat{\sigma}^2_{p,LB1} = \left(\frac{n}{n-2}\right) \hat{\sigma}^2_{p,LB}(n/2) \quad \text{and} \quad \hat{\sigma}^2_{p,PLB1} = \left(\frac{n}{n+2}\right) \hat{\sigma}^2_{p,LB}(n/2).$$

Theorem 5.5 *Under $\mathcal{M}_p$ where μ_{i_0} is the true mean, for a fixed $\alpha > 0$, $n\hat{\sigma}^2_{p,LB}/\sigma^2 \stackrel{d}{=} W(1+H(\boldsymbol{T}))$, where*

$$H(\boldsymbol{T}) \equiv H(\boldsymbol{T};\alpha) = \sum_{i=1}^{p} \theta_i(\boldsymbol{T}) T_i^2 \quad \textit{with} \quad \theta_i(\boldsymbol{T}) \equiv \theta_i(\boldsymbol{T};\alpha) = \frac{(1+T_i^2)^{-\alpha}}{\sum_{j=1}^{p}(1+T_j^2)^{-\alpha}}, \quad i = 1, 2, \ldots, p.$$

Consequently, the distribution of $\hat{\sigma}^2_{p,LB}/\sigma^2$ depends on $(\boldsymbol{\mu}, \sigma^2)$ only through $\boldsymbol{\Delta}$.

¿From the distributional representation in Theorem 5.5, a closed-form expression for the risk function of $\hat{\sigma}^2_{p,LB}$ will be difficult to obtain because of the adaptive, i.e., data-dependent, nature of the mixing probabilities $\theta_i(\boldsymbol{T})$ and the fact that these are rational functions of $\boldsymbol{T}$. To obtain an approximation to the risk function of $\hat{\sigma}^2_{p,LB}$ we used a second-order Taylor expansion of the function $H(\boldsymbol{T})$ about $\boldsymbol{T} = \boldsymbol{\nu}$, the mean vector of $\boldsymbol{T}$. For notation, let

$$H \equiv H(\boldsymbol{\nu}); \quad \boldsymbol{H}^{(1)} \equiv \nabla_{\boldsymbol{T}} H(\boldsymbol{T})|_{\boldsymbol{T}=\boldsymbol{\nu}}; \quad \text{and} \quad \boldsymbol{H}^{(2)} \equiv \frac{\partial^2}{\partial \boldsymbol{T} \partial \boldsymbol{T}'} H(\boldsymbol{T})|_{\boldsymbol{T}=\boldsymbol{\nu}}.$$

A second-order Taylor approximation for $\hat{\sigma}^2_{p,LB}/\sigma^2$ is provided by

$$\frac{\hat{\sigma}^2_{p,LB}}{\sigma^2} \stackrel{d}{\approx} \frac{W}{n} \left\{ 1 + H + \boldsymbol{H}^{(1)\prime}(\boldsymbol{T}-\boldsymbol{\nu}) + \frac{1}{2}(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu}) \right\}. \tag{25}$$

¿From this approximate representation, we are able to obtain approximate expressions for the mean and variance of the Bayes estimator. These mean and variance expressions, which involve the constant C_n defined in Corollary 5.1, are given in the next two theorems. The proofs require several intermediate results (contained in lemmas), and these are presented in Appendix A.

Theorem 5.6 *Under $\mathcal{M}_p$, a second-order approximation to the mean of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is*

$$E_2(\boldsymbol{\Delta}) \equiv \left(1-\frac{1}{n}\right)(1+H)+\frac{1}{2}\left(\frac{1}{C_n^2}-1+\frac{3}{n}\right)(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})-\frac{1}{n}\left\{(\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu})-\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})\right\}.$$

Theorem 5.7 *Under $\mathcal{M}_p$, a second-order approximation to the variance of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is $V_2(\boldsymbol{\Delta}) \equiv \frac{1}{n}\{\mathrm{VE}(\boldsymbol{\Delta})+\mathrm{EV}(\boldsymbol{\Delta})\}$, where*

$$\begin{aligned}
\mathrm{VE}(\boldsymbol{\Delta}) &\equiv 2\left(1-\frac{1}{n}\right)\left(1+H-\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}+\frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2+\left(\frac{n-1}{C_n^2}-\frac{(n-2)^2}{n}\right)\times \\
&\quad \left(\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}-\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2+2\left(1-\frac{2}{n}\right)\left(1+H-\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}+\frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)\left(\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}-\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right); \\
\mathrm{EV}(\boldsymbol{\Delta}) &\equiv \frac{1}{2n}\left(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1}\right)^2+\left(1-\frac{1}{n}\right)\left(\boldsymbol{H}^{(1)\prime}\mathbf{1}-\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2+ \\
&\quad 2\left(1-\frac{2}{n}\right)\left(\boldsymbol{H}^{(1)\prime}\mathbf{1}-\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)\left(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)+\frac{1}{C_n^2}\left(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2.
\end{aligned}$$

¿From these expressions, we can compute second-order approximations to the risk functions of $\hat{\sigma}^2_{p,LB1}$ according to the formula

$$R\left(\hat{\sigma}^2_{p,LB1},(\mu_{i_0},\sigma^2)\right) \approx \left(\frac{n}{n-2}\right)^2 V_2(\boldsymbol{\Delta};\alpha=n/2)+\left[\left(\frac{n}{n-2}\right)E_2(\boldsymbol{\Delta};\alpha=n/2)-1\right]^2, \qquad (26)$$

where $E_2(\boldsymbol{\Delta};\alpha) \equiv E_2(\boldsymbol{\Delta})$ and $V_2(\boldsymbol{\Delta};\alpha) \equiv V_2(\boldsymbol{\Delta})$ are given in Theorem 5.6 and Theorem 5.7, respectively. For other limiting Bayes and weighted estimators in Table 1 and (16), analogous approximate risk expressions can be obtained similarly as for $\hat{\sigma}^2_{p,LB1}$.

Lastly, still for a given $\alpha > 0$, we present a few expressions for the components $H_k^{(1)}(\boldsymbol{T}), k \in \{1,2,\ldots,p\}$ of the $p \times 1$ vector $\boldsymbol{H}^{(1)}(\boldsymbol{T})$ and the components $H_{kl}^{(2)}(\boldsymbol{T}), k,l \in \{1,2,\ldots,p\}$ of the $p \times p$ matrix $\boldsymbol{H}^{(2)}(\boldsymbol{T})$, which when evaluated at $\boldsymbol{T}=\boldsymbol{\nu}$ yield $\boldsymbol{H}^{(1)}$ and $\boldsymbol{H}^{(2)}$, respectively. From the expressions for $H(\boldsymbol{T})$ and $\theta_i(\boldsymbol{T})$ in Corollary 5.1, we find that for $j,k \in \{1,2,\ldots,p\}$,

$$\begin{gathered}
H_k^{(1)}(\boldsymbol{T}) \equiv \tfrac{\partial}{\partial T_k}H(\boldsymbol{T}) = 2\theta_k(\boldsymbol{T})T_k+\textstyle\sum_{i=1}^p \theta_{ik}^{(1)}(\boldsymbol{T})T_i^2; \\
H_{kl}^{(2)}(\boldsymbol{T}) \equiv \tfrac{\partial^2}{\partial T_k \partial T_l}H(\boldsymbol{T}) = 2\theta_k(\boldsymbol{T})I\{k=l\}+2[\theta_{kl}^{(1)}(\boldsymbol{T})T_k+\theta_{lk}^{(1)}(\boldsymbol{T})T_l]+\textstyle\sum_{i=1}^p \theta_{ikl}^{(2)}(\boldsymbol{T})T_i^2;
\end{gathered}$$

where, for $i,k \in \{1,2,\ldots,p\}$, $\theta_{ik}^{(1)}(\boldsymbol{T}) = (2\alpha)\left(T_k/(1+T_k^2)\right)\theta_k(\boldsymbol{T})[\theta_i(\boldsymbol{T})-I\{k=i\}]$; and, for $i,k,l \in \{1,2,\ldots,p\}$,

$$\begin{aligned}
\theta_{ikl}^{(2)}(\boldsymbol{T}) = (2\alpha)\Bigg\{&I\{k=l\}\left(\frac{1-T_k^2}{(1+T_k^2)^2}\right)\theta_k(\boldsymbol{T})[\theta_i(\boldsymbol{T})-I\{k=i\}]+ \\
&\left(\frac{T_k}{(1+T_k^2)^2}\right)\left[\theta_{kl}^{(1)}(\boldsymbol{T})[\theta_i(\boldsymbol{T})-I\{k=i\}]+\theta_k(\boldsymbol{T})\theta_{il}^{(1)}(\boldsymbol{T})\right]\Bigg\}.
\end{aligned}$$

5.4 Assessing the Second-Order Approximations via Simulation

To assess the goodness of the second-order approximations, we compared the values of the means, variances, and risks of $\hat{\sigma}^2_{p,LB}(\alpha = n/2)/\sigma^2$ based on 10,000 simulated datasets to their second-order approximations. The results revealed the same pattern across all choices of $\boldsymbol{\Delta}$. For $n = 3$ the approximation performs rather poorly, gradually improving with increasing n, to finally become almost identical to the simulation-based values when n is 30. In one of the worst-case scenarios, when $n = 3$ and $\boldsymbol{\Delta}$ is symmetric with a medium-size spread (such as $\boldsymbol{\Delta} = (-0.25, 0, 0.25)$), the approximate mean values lie generally within 20% of the simulated ones. Similar behavior is shown by variances and risks also. Furthermore, as the model dimension p increases or as the separations among the sub-models' means become smaller, the differences between simulated values and approximations also seem to diminish. With increasing n the accuracy of the approximations improves. Therefore, the second-order approximation appears to work well overall, but when n is small (less than 15) it seems better to use simulations. In the remainder of this paper, all analyses involving risks of the limiting Bayes ($\hat{\sigma}^2_{p,LBk}$s) and weighted ($\hat{\sigma}^2_{p,PLBk}$s) estimators are based on simulations.

5.5 Comparison of Relative Efficiencies

We now carry out the comparison of relative efficiencies of the variance estimators with respect to $\hat{\sigma}^2_{UMVU}$, using simulated datasets with a variety of $\boldsymbol{\Delta}$ values for $n \in \{3, 10, 30\}$. The results are summarized in Tables 2 and 3 and Figures 1 and 2.

Table 2 focuses on the differences in relative efficiencies between symmetric and asymmetric $\boldsymbol{\Delta}$ cases. As can be seen, there does not seem to be a strong effect of the asymmetry of $\boldsymbol{\Delta}$ on the estimators. In all $\boldsymbol{\Delta}$ cases that we have chosen, $\hat{\sigma}^2_{p,PLB1}$, $\hat{\sigma}^2_{p,PLB2}$, $\hat{\sigma}^2_{p,PLB3}$, $\hat{\sigma}^2_{p,LB4}$ perform best, with the two-step estimator $\hat{\sigma}^2_{p,MRE}$ following. Clearly, the best among the $\mathcal{M}_p$-based estimators dominate the $\mathcal{M}$-based estimators $\hat{\sigma}^2_{UMVU}$ and $\hat{\sigma}^2_{MRE}$, with the gain in efficiency being quite impressive for small sample sizes.

Table 3 is designed to examine the impact of increasing number of sub-models. We see that when p is large the two-step estimator $\hat{\sigma}^2_{p,MRE}$ becomes less efficient than the global-type estimator $\hat{\sigma}^2_{MRE}$. This result is consistent with the theoretical result of Corollary 5.4. Note also that even for p as low as 33, the ratios of the relative efficiency values from Table 3 start to agree (to the third decimal) with the limiting relative efficiencies predicted by Corollary 5.4. The weighted estimators do not seem to be affected much by the increasing p, faring much better than the two-step estimators.

Figure 1 presents two contour plots of the relative efficiencies of $\hat{\sigma}^2_{p,MRE}$ with respect to $\hat{\sigma}^2_{MRE}$ as a function of p (where $p > 3$) and the range of the values in the $\boldsymbol{\Delta}$ vector. In the top and bottom contour plots symmetric and asymmetric $\boldsymbol{\Delta}$ cases are considered separately. The top contour plot reveals a structure that is consistent with Corollary 5.4, especially in the regions of the contour plot where $p \to \infty$ and $\Delta_{max} \to \infty$ (hence $\Delta_{min} \to -\infty$) which fall in the top and bottom right corners, where $\hat{\sigma}^2_{MRE}$ starts to dominate $\hat{\sigma}^2_{p,MRE}$. Note that the 98% relative efficiency in this region is very close to the limiting 97% from Corollary 5.4. The bottom contour plot is constructed using $\boldsymbol{\Delta}$ that are quite asymmetric (all $\Delta_{min} = 0$), and therefore could not be compared to the predictions of Corollary 5.4. From this plot we can see however that $\hat{\sigma}^2_{p,MRE}$ seems to dominate $\hat{\sigma}^2_{MRE}$ everywhere.

Figure 2 explores the case when $p = 2$ only, so $\boldsymbol{\Delta} = (0, \Delta)$, and for sample sizes $n = 3$ and $n = 10$. The plots depict the efficiency of the leading estimators, together with that of the Stein estimator in (5), as a function of the magnitude of the Δ parameter. As can be seen, $\hat{\sigma}^2_{p,MRE}$ performs best when $|\Delta|$ is large, with $\hat{\sigma}^2_{p,LB4}$ giving a very comparable performance. The estimator $\hat{\sigma}^2_{p,PLB1}$ performs better than $\hat{\sigma}^2_{p,MRE}$ when $|\Delta|$ is closer to zero, but degrades in performance when $|\Delta|$ becomes large. Thus, we see that the estimators' performances and regions where they perform well will depend to a large extent on the magnitude of Δ. In particular, it appears that the best among the weighted estimators perform very well when the $|\Delta|$ is neither too large nor too small, while the two-step estimator performs very well when $|\Delta|$ is large. This points to the following intuitive explanation: when $|\Delta|$ is large, the two models are well-separated and the model selection is easier; however, when $|\Delta|$ is neither too small nor too large, it is not so clear which model to choose, and it seems better to average over the sub-models' estimators. When $|\Delta|$ is quite close to zero, i.e. when there is not much difference among the sub-models, either approach to estimation works well. With regards to the Stein estimator, observe that though it dominates $\hat{\sigma}^2_{MRE}$, as is expected from theory, it is at the same time dominated by the estimators $\hat{\sigma}^2_{p,MRE}$, $\hat{\sigma}^2_{p,PLB1}$, and $\hat{\sigma}^2_{p,LB4}$.

Overall, based on the results of the risk comparison, the estimators performing best are the Bayes-type or weighted estimators $\hat{\sigma}^2_{p,PLB1}$ and $\hat{\sigma}^2_{p,LB4}$, and the two-step estimator $\hat{\sigma}^2_{p,MRE}$. We give a slight preference to the weighted estimators because their performance does not degrade much even when the number of sub-models increases, in contrast to the two-step estimator which becomes dominated by the $\mathcal{M}$-estimator $\hat{\sigma}^2_{MRE}$ when p, the number of sub-models, increases.

Finally, a cautionary note arising from these efficiency studies is that one ought to be very careful in the choice of prior parameters. At least in the situation when one is concerned with variance estimation, the limiting Bayes estimators $\hat{\sigma}^2_{p,LB1}$ and $\hat{\sigma}^2_{p,LB2}$, corresponding to the limiting cases of $\kappa \to 1$ and $\kappa \to 3/2$ respectively, perform quite poorly, especially for small sample sizes. These two estimators are dominated by the estimator $\hat{\sigma}^2_{UMVU}$ in terms of risk function. However, these improper priors associated with the limiting values of κ are most likely the worst-case scenarios, and other, more carefully chosen and meaningful priors should result in improved performance.

6 Concluding Remarks

We have examined some of the issues arising when considering a model with a finite number of sub-models, where the goal is to make inference about a common parameter among these sub-models, based on a single realization of a sample. It is of interest to determine which of the three possible strategies is preferable: (i) to utilize a wider model that encompasses all competing sub-models; (ii) to adopt a two-step approach: select the sub-model, and then do inference within this chosen sub-model, but with both steps utilizing the same sample data; (iii) to do a sub-model averaging scheme where the inference procedure is formed by weighting the sub-models' procedures, with the weights being also data-dependent. The second strategy may be labeled the classical approach, while the third strategy coincides or is motivated by the Bayesian approach.

Through a simple model prototype with a finite number of Gaussian sub-models with common variance but different means, we have studied each of the strategies, with regards to the estimation of variance. Based on the theoretical and simulated comparison of the different types of estimators, and with the estimator performance evaluated through risk functions based on quadratic loss, we have reached the following conclusions: (i) There could be considerable improvement in using estimators developed by exploiting the structure of the sub-models, over the strategy of simply using estimators from a wider model. (ii) However, the properties of these resulting estimators may be quite difficult to obtain. Furthermore, some desirable properties of the sub-model estimators, such as unbiasedness and minimum variance, may not carry-over when they are combined to form the estimator for the full model of interest. (iii) Based on the theoretical and simulated results for the variance parameter σ^2 considered in this paper, the weighted estimators, which were motivated and/or derived via the Bayesian approach, seem preferable over the two-step estimators even though these estimators were derived using improper priors. (iv) When the number of sub-models increases and two-step estimators are employed, it appears that their performance could degrade relative to estimators developed under a wider model, but that the weighted estimators' performances are not necessarily affected. (v) And, finally, when developing weighted estimators through the Bayesian framework, caution must be observed in assigning prior parameters as a particular prior specification may also lead to poor estimators.

Approaches similar to the one presented here could be a useful first step in many contexts where model selection is often done and where there exist a natural notion of "distance" among models: regression, survival analysis, or goodness-of-fit testing. For example, techniques and results presented here could be extended to settings of regression with p possible predictor variables, where the goal is estimation of dispersion parameters associated with error components for the 2^p competing sub-models. There is clearly a need for studies of more complicated situations in varied settings, where multiple parameters are being estimated simultaneously and/or where sub-models are of different dimensions (Claeskens and Hjort (2003)). As was pointed out earlier, for these more general settings, exact risk expressions may not be possible and asymptotic analysis may be needed, in contrast to the situation considered in this paper where the Gaussian distributional assumption enabled us to obtain concrete results. Also, many other interesting alternative options need to be examined: for example, in the two-step approach, would it have been better to subdivide the sample data into two parts and use the first part for model selection and the second part for making inference in the chosen sub-model, an issue alluded for instance in Hastie, Tibshirani, and Friedman (2001) and investigated by Yang (2003)? Finally, we have observed in Dukić and Peña (2003) that in the case of estimation of the distribution function in this same specific model, a different conclusion holds with respect to which of the three strategies is preferable. This is consistent with recent work by Claeskens and Hjort (2003)) where they advocate the use of a focussed information criterion in model selection problems that is tailored to the specific parameter of interest.

A Appendix: Proofs of Selected Results

In this appendix we gather the technical proofs of the results presented in earlier sections.

Proof of Proposition 5.1: With $\bar{Z} = (\boldsymbol{Z}\boldsymbol{1}')/n,$ we have $n\hat{\sigma}_i^2/\sigma^2 = \|\boldsymbol{X} - \mu_i\boldsymbol{1}/\sigma\|^2 = \|[(\boldsymbol{X} - \mu_{i_0}\boldsymbol{1}) - (\mu_i - \mu_{i_0})\boldsymbol{1}]/\sigma\|^2 \stackrel{d}{=} \|\boldsymbol{Z} - \Delta_i\boldsymbol{1}\|^2 = \|\boldsymbol{Z} - \bar{Z}\boldsymbol{1}\|^2 + n(\bar{Z} - \Delta_i)^2$. Letting $W = \|\boldsymbol{Z} - \bar{Z}\boldsymbol{1}\|^2$ and $V_i = \sqrt{n}(\bar{Z} - \Delta_i), i = 1, 2, \ldots, p$, it follows that $W \sim \chi^2_{n-1}$ with W and $\boldsymbol{V} = (V_1, V_2, \ldots, V_p)'$ independent. Furthermore, since $\bar{Z} \sim N(0, n^{-1})$, $\boldsymbol{V}$ has representation

$$\boldsymbol{V} = Z\boldsymbol{1} - \sqrt{n}\boldsymbol{\Delta}, \tag{27}$$

so $\boldsymbol{V} \sim N_p(\boldsymbol{0}, \boldsymbol{J})$ since $\mathbf{Cov}(\boldsymbol{V}, \boldsymbol{V}) = \mathbf{Cov}\{Z\boldsymbol{1} - \sqrt{n}\boldsymbol{\Delta}, Z\boldsymbol{1} - \sqrt{n}\boldsymbol{\Delta}\} = \boldsymbol{1}\mathbf{Var}(Z)\boldsymbol{1}' = \boldsymbol{J}$.

Proof of Theorem 5.1: By Prop. 5.1, $\hat{\sigma}^2_{p,MLE}/\sigma^2 \stackrel{d}{=} \left[W + \sum_{i=1}^p I\{\hat{M} = i\}V_i^2\right]/n$. Now,

$$\begin{aligned}
\{\hat{M} = i\} &= \left\{\frac{\hat{\sigma}_i^2}{\sigma^2} < \frac{\hat{\sigma}_j^2}{\sigma^2},\ j = 1, 2, \ldots, p; j \neq i\right\} \stackrel{d}{=} \left\{V_i^2 < V_j^2,\ j = 1, 2, \ldots, p; j \neq i\right\} \\
&= \left\{(Z - \sqrt{n}\Delta_i)^2 < (Z - \sqrt{n}\Delta_j)^2,\ j = 1, 2, \ldots, p; j \neq i\right\} \quad \text{by (27)} \\
&= \left\{(\Delta_j - \Delta_i)Z < \frac{\sqrt{n}}{2}(\Delta_j + \Delta_i)(\Delta_j - \Delta_i),\ j = 1, 2, \ldots, p; j \neq i\right\}
\end{aligned}$$

$$= \left\{ Z < \frac{\sqrt{n}}{2}(\Delta_j + \Delta_i),\ \forall (j \neq i, \Delta_j > \Delta_i) \right\} \cap \left\{ Z > \frac{\sqrt{n}}{2}(\Delta_j + \Delta_i),\ \forall (j \neq i, \Delta_j < \Delta_i) \right\}$$
$$= \left\{ \frac{\sqrt{n}}{2} \left(\Delta_i + \max_{\{j:\ \Delta_j < \Delta_i\}} \Delta_j \right) < Z < \frac{\sqrt{n}}{2} \left(\Delta_i + \min_{\{j:\ \Delta_j > \Delta_i\}} \Delta_j \right) \right\}.$$

But,

$$\sum_{i=1}^{p} I \left\{ \frac{\sqrt{n}}{2} \left(\Delta_i + \max_{\{j:\ \Delta_j < \Delta_i\}} \Delta_j \right) < Z < \frac{\sqrt{n}}{2} \left(\Delta_i + \min_{\{j:\ \Delta_j > \Delta_i\}} \Delta_j \right) \right\} V_i^2$$
$$= \sum_{i=1}^{p} I \left\{ \frac{\sqrt{n}}{2} \left(\Delta_{(i)} + \max_{\{j:\ \Delta_j < \Delta_{(i)}\}} \Delta_j \right) < Z < \frac{\sqrt{n}}{2} \left(\Delta_{(i)} + \min_{\{j:\ \Delta_j > \Delta_{(i)}\}} \Delta_j \right) \right\} (Z - \sqrt{n}\Delta_{(i)})^2$$
$$= \sum_{i=1}^{p} I \left\{ \frac{\sqrt{n}}{2} \left(\Delta_{(i)} + \Delta_{(i-1)} \right) < Z < \frac{\sqrt{n}}{2} \left(\Delta_{(i)} + \Delta_{(i+1)} \right) \right\} (Z - \sqrt{n}\Delta_{(i)})^2$$
$$= \sum_{i=1}^{p} I \left\{ L(\Delta_{(i)}, \mathbf{\Delta}) < Z < U(\Delta_{(i)}, \mathbf{\Delta}) \right\} (Z - \sqrt{n}\Delta_{(i)})^2.$$

Proof of Theorem 5.2: From Theorem 5.1, using $\int_a^b z\phi(z)dz = \phi(a) - \phi(b)$ for $a < b$,

$$\text{EpMLE}(\mathbf{\Delta}) = \frac{1}{n} \left\{ \mathbf{E}(W) + \mathbf{E} \left\{ \sum_{i=1}^{p} I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2 \right\} \right\}$$
$$= \frac{1}{n} \left\{ (n-1) + \mathbf{E}(Z^2) - 2\sqrt{n} \sum_{i=1}^{p} \Delta_{(i)} \mathbf{E}\{ZI(\Omega_{(i)})\} + n \sum_{i=1}^{p} \Delta_{(i)}^2 P_{(i)}(\mathbf{\Delta}) \right\}$$
$$= \frac{1}{n} \left\{ n - 2\sqrt{n} \sum_{i=1}^{p} \Delta_{(i)} \int_{L(\Delta_{(i)}, \mathbf{\Delta})}^{U(\Delta_{(i)}, \mathbf{\Delta})} z\phi(z)dz + n \sum_{i=1}^{p} \Delta_{(i)}^2 P_{(i)}(\mathbf{\Delta}) \right\}$$
$$= 1 - \frac{2}{\sqrt{n}} \sum_{i=1}^{p} \Delta_{(i)} \left[\phi(L(\Delta_{(i)}, \mathbf{\Delta})) - \phi(U(\Delta_{(i)}, \mathbf{\Delta})) \right] + \sum_{i=1}^{p} \Delta_{(i)}^2 \left[\Phi(U(\Delta_{(i)}, \mathbf{\Delta})) - \Phi(L(\Delta_{(i)}, \mathbf{\Delta})) \right].$$

Proof of Theorem 5.3: ¿From Theorem 5.1, by independence of W and Z, and because $\Omega_{(i)}$s depend only on Z, we have $\text{VpMLE}(\mathbf{\Delta}) = n^{-2} \left\{ \mathbf{Var}(W) + \mathbf{Var} \left[\sum_{i=1}^{p} I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2 \right] \right\}$. We already know that $\mathbf{Var}(W) = 2(n-1)$. On the other hand, $\mathbf{Var} \left\{ \sum_{i=1}^{p} I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2 \right\} = \sum_{i=1}^{p} \mathbf{Var} \left\{ I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2 \right\} + \sum_{i \neq j} \mathbf{Cov} \left\{ I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2, I(\Omega_{(j)})(Z - \sqrt{n}\Delta_{(j)})^2 \right\}$. By the variance formula and definition of $\zeta_{(i)}(k)$, $\mathbf{Var} \left\{ I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2 \right\} = \zeta_{(i)}(4) - [\zeta_{(i)}(2)]^2$; and $\mathbf{Cov} \left\{ I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2, I(\Omega_{(j)})(Z - \sqrt{n}\Delta_{(j)})^2 \right\} = -\zeta_{(i)}(2)\zeta_{(j)}(2)$, as $I(\Omega_{(i)})I(\Omega_{(j)}) = 0$ whenever $i \neq j$. Consequently, $\mathbf{Var} \left\{ \sum_{i=1}^{p} I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^2 \right\} = \sum_{i=1}^{p} \left[\zeta_{(i)}(4) - [\zeta_{(i)}(2)]^2 \right] + \sum_{i \neq j} \left[-\zeta_{(i)}(2)\zeta_{(j)}(2) \right] = \sum_{i=1}^{p} \zeta_{(i)}(4) - \left(\sum_{i=1}^{p} \zeta_{(i)}(2) \right)^2$.

Proof of Corollary 5.4: From Theorem 5.4, $R\left(\hat{\sigma}^2_{p,MLE}, (\mu_{i_0}, \sigma^2)\right) = \text{VpMLE}(\mathbf{\Delta}) + [\text{EpMLE}(\mathbf{\Delta}) - 1]^2 \rightarrow (2/n)(1 - 1/n) + [(1 - 1/n) - 1]^2 = (2n-1)/n^2$. Also, since $\hat{\sigma}^2_{p,MRE} = (n/(n+2))\,\hat{\sigma}^2_{p,MLE}$, we have $R\left(\hat{\sigma}^2_{p,MRE}, (\mu_{i_0}, \sigma^2)\right) = (n/(n+2))^2\,\mathrm{VpMLE}(\mathbf{\Delta}) + [(n/(n+2))\,\mathrm{EpMLE}(\mathbf{\Delta}) - 1]^2 \rightarrow (n/(n+2))^2\,(2/n)\,(1-1/n) + [(n/(n+2))\,(1-1/n) - 1]^2 = (2n+7)[(n+2)^2]$. The efficiency expressions relative to $\hat{\sigma}^2_{UMVU}$ of $\hat{\sigma}^2_{p,MLE}$ and $\hat{\sigma}^2_{p,MRE}$ are obtained by dividing $R\left(\hat{\sigma}^2_{UMVU}, (\mu_{i_0}, \sigma^2)\right) = 2/(n-1)$ by the preceding risk expressions. Both of the resulting efficiency expressions are easily shown to exceed 1. Taking the ratio of $R\left(\hat{\sigma}^2_{p,MLE}, (\mu_{i_0}, \sigma^2)\right)$ and $R\left(\hat{\sigma}^2_{p,MRE}, (\mu_{i_0}, \sigma^2)\right)$ yields the third expression, an expression easily shown to exceed 1. In comparing $\hat{\sigma}^2_{MRE}$ and $\hat{\sigma}^2_{p,MRE}$, we observe that $\mathrm{Eff}\left(\hat{\sigma}^2_{p,MRE} : \hat{\sigma}^2_{MRE}\right) = 2(n+2)^2/[(n+1)(2n+7)]$. Since $2(n+2)^2/[(n+1)(2n+7)] - 1 = -(n-1)/[(n+1)(2n+7)] < 0$ for $n > 1$, then we have established that, as $p \rightarrow \infty$, $\hat{\sigma}^2_{MRE}$ is more efficient than $\hat{\sigma}^2_{p,MRE}$. To show that $\hat{\sigma}^2_{p,ALB}$ is dominated by $\hat{\sigma}^2_{UMVU}$, note that $R(\hat{\sigma}^2_{p,ALB}, (\mu_{i_0}, \sigma^2)) = (2n-1)/[(n-2)^2]$, which is easily shown to exceed $2/(n-1)$ whenever $n > 2$.

Proof of Theorem 5.5: Starting from (24) and using Corollary 5.1, we obtain

$$\frac{\hat{\sigma}^2_{p,LB}}{\sigma^2} \stackrel{d}{=} \sum_{i=1}^{p} \left\{ \frac{\left[\frac{W}{n}(1+T_i^2)\right]^{-\alpha}}{\sum_{j=1}^{p}\left[\frac{W}{n}(1+T_j^2)\right]^{-\alpha}} \right\} \left[\frac{W}{n}(1+T_i^2)\right] = \frac{W}{n}\left\{1 + \sum_{i=1}^{p} \theta_i(\boldsymbol{T}) T_i^2\right\}.$$

The following lemma will be needed for proving Theorems 5.6 and 5.7.

Lemma A.1 *Under the conditions of Proposition 5.1, (i)* $\mathbf{E}\{\boldsymbol{T}-\boldsymbol{\nu}|W\} = \left(\sqrt{n}/(\sqrt{W}C_n) - 1\right)\boldsymbol{\nu}$; *(ii)* $\mathbf{E}\{(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})'|W\} = \boldsymbol{J}/W + \left(\sqrt{n}/(\sqrt{W}C_n) - 1\right)^2 \boldsymbol{\nu}^{\otimes 2}$; *(iii)* $\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})\} = -\boldsymbol{\nu}$; *and (iv)* $\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})'\} = \boldsymbol{J} + n\left(1/C_n^2 - 1 + 3/n\right)\boldsymbol{\nu}^{\otimes 2}$, *with constant* C_n *given in Corollary 5.1.*

Proof of Lemma A.1: Since $\mathbf{E}(\boldsymbol{T}|W) = \mathbf{E}(\boldsymbol{V}/\sqrt{W}|W) = \mathbf{E}(\boldsymbol{V})/\sqrt{W} = \sqrt{n}\boldsymbol{\nu}/(\sqrt{W}C_n)$, then the first result immediately follows. Using that $\mathbf{E}(W) = n-1$ and $\mathbf{E}(\sqrt{W}) = (n-2)C_n/\sqrt{n}$, the third result follows trivially from the first identity. To prove the second result, observe that

$$\mathbf{E}(\boldsymbol{T}\boldsymbol{T}'|W) = \frac{1}{W}\mathbf{E}(\boldsymbol{V}\boldsymbol{V}') = \frac{1}{W}\mathbf{E}\{(Z\mathbf{1} - \sqrt{n}\mathbf{\Delta})^{\otimes 2}\} = \frac{1}{W}\left(\boldsymbol{J} + \frac{n}{C_n^2}\boldsymbol{\nu}^{\otimes 2}\right);$$

and $\mathbf{E}\{\boldsymbol{T}\boldsymbol{\nu}'|W\} = -\sqrt{n}\mathbf{\Delta}\boldsymbol{\nu}'/\sqrt{W} = \sqrt{n}\boldsymbol{\nu}^{\otimes 2}/(\sqrt{W}C_n)$. Consequently,

$$\mathbf{E}\{(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})'|W\} = \mathbf{E}(\boldsymbol{T}\boldsymbol{T}'|W) - 2\mathbf{E}(\boldsymbol{T}\boldsymbol{\nu}'|W) + \boldsymbol{\nu}^{\otimes 2} = \frac{\boldsymbol{J}}{W} + \left(\frac{\sqrt{n}}{\sqrt{W}C_n} - 1\right)^2 \boldsymbol{\nu}^{\otimes 2}.$$

Finally, by the iterated expectation rule, and using the expressions for $\mathbf{E}(W)$ and $\mathbf{E}(\sqrt{W})$,

$$\begin{aligned}\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})'\} &= \mathbf{E}\{W\mathbf{E}[(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})'|W]\} \\ &= \mathbf{E}\left\{\boldsymbol{J} + \left(\frac{n}{C_n^2} - \frac{2\sqrt{n}}{C_n}\sqrt{W} + W\right)\boldsymbol{\nu}^{\otimes 2}\right\} = \boldsymbol{J} + \left(\frac{n}{C_n^2} - 2(n-2) + (n-1)\right)\boldsymbol{\nu}^{\otimes 2}.\end{aligned}$$

Simplifying leads to the expression given in the statement of the lemma.

Proof of Theorem 5.6: First, note that by using the fourth result in Lemma A.1, we have

$$\begin{aligned}&\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})\} = \mathbf{E}\{\mathrm{tr}[\boldsymbol{H}^{(2)}W(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})']\}\\ &= \ \mathrm{tr}\left(\boldsymbol{H}^{(2)}\left[\boldsymbol{J}+n\left(\frac{1}{C_n^2}-1+\frac{3}{n}\right)\boldsymbol{\nu}^{\otimes 2}\right]\right) = (\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})+n\left(\frac{1}{C_n^2}-1+\frac{3}{n}\right)(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}).\end{aligned}$$

¿From (25), and using Lemma A.1 and the preceding result, the approximation to the mean of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is obtained to be

$$E_2(\boldsymbol{\Delta}) \ = \ \frac{1}{n}\left\{(1+H)(n-1)-\boldsymbol{H}^{(1)'}\boldsymbol{\nu}+\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})+\frac{n}{2}\left(\frac{1}{C_n^2}-1+\frac{3}{n}\right)(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\},$$

which simplifies to the expression in the statement of the theorem.

To obtain second-order approximation to the variance of $\hat{\sigma}^2_{p,LB}/\sigma^2$, we first establish two intermediate lemmas concerning the conditional mean and variance, given W, of the variable

$$Q \equiv 1+H+\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu})+\frac{1}{2}(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu}), \tag{28}$$

which is the second-order Taylor approximation of the function $1+H(\boldsymbol{T})$ (see equation (25)).

Lemma A.2 *Under conditions of Proposition 5.1,*

$$\begin{aligned}\mathbf{E}(Q|W) = &\left\{1+H-(\boldsymbol{H}^{(1)'}\boldsymbol{\nu})+\frac{1}{2}(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\}+\\ &\frac{\sqrt{n}}{C_n}\left\{(\boldsymbol{H}^{(1)'}\boldsymbol{\nu})-(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\}\frac{1}{\sqrt{W}}+\frac{1}{2}\left\{(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})+\frac{n}{C_n^2}(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\}\frac{1}{W}.\end{aligned}$$

Proof of Lemma A.2: From the proof of and results in Lemma A.1, it follows that

$$\mathbf{E}\left\{(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\right\} = \tfrac{1}{W}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})+\left(\tfrac{n}{WC_n^2}-\tfrac{2\sqrt{n}}{\sqrt{W}C_n}+1\right)(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}).$$

¿From first result in Lemma A.1 and the preceding result, $\mathbf{E}(Q|W) = 1+H+\left(\frac{\sqrt{n}}{\sqrt{W}C_n}-1\right)\boldsymbol{H}^{(1)'}\boldsymbol{\nu}+\frac{1}{2}\left\{\frac{1}{W}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})+\left(\frac{n}{WC_n^2}-\frac{2\sqrt{n}}{\sqrt{W}C_n}+1\right)(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\}$, and the result follows.

Lemma A.3 *Under conditions of Proposition 5.1,*

$$\begin{aligned}\mathbf{Var}(Q|W) = &\left\{(\boldsymbol{H}^{(1)'}\mathbf{1})-(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\}^2\frac{1}{W}+\\ &\frac{2\sqrt{n}}{C_n}\left\{(\boldsymbol{H}^{(1)'}\mathbf{1})-(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\frac{1}{W^{3/2}}+\left\{\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2+\frac{n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\}\frac{1}{W^2}.\end{aligned}$$

Proof of Lemma A.3: First, observe that we have

$$\begin{aligned}
\mathbf{Var}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu})|W\} &= \frac{1}{W}\boldsymbol{H}^{(1)'}\mathbf{Cov}(\boldsymbol{V})\boldsymbol{H}^{(1)} = \frac{1}{W}\boldsymbol{H}^{(1)'}\boldsymbol{J}\boldsymbol{H}^{(1)} = \frac{1}{W}(\boldsymbol{H}^{(1)'}\mathbf{1})^2;\\
\mathbf{Var}\{(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} &= \mathbf{Var}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} + 4\mathbf{Var}\{\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\}\\
&\quad -4\mathbf{Cov}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\};\\
\mathbf{Var}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} &= \mathbf{Var}\left\{\frac{\boldsymbol{V}'}{\sqrt{W}}\boldsymbol{H}^{(2)}\frac{\boldsymbol{V}}{\sqrt{W}}|W\right\} = \frac{1}{W^2}\mathbf{Var}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}\}.
\end{aligned}$$

¿From the representation of $\boldsymbol{V}$ in (27), we obtain

$$\mathbf{Var}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}\} = \mathbf{Var}\{(Z\mathbf{1}-\sqrt{n}\boldsymbol{\Delta})'\boldsymbol{H}^{(2)}(Z\mathbf{1}-\sqrt{n}\boldsymbol{\Delta})\}2(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2 + \tfrac{4n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2$$

since $\mathbf{Var}(Z^2) = 2$, $\mathbf{Var}(Z) = 1$, and $\mathbf{Cov}(Z^2, Z) = 0$. Thus,

$$\mathbf{Var}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} = \frac{4}{W^2}\left\{\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2 + \frac{n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\}.$$

We also have

$$\begin{aligned}
\mathbf{Var}\{\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} = \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\mathbf{Cov}(\boldsymbol{T}|W)\tfrac{1}{W}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{J}\boldsymbol{H}^{(2)}\boldsymbol{\nu} = \tfrac{1}{W}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2;\\
\mathbf{Cov}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} = \tfrac{1}{W^{3/2}}\mathbf{Cov}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{V}\}.
\end{aligned}$$

Again, by utilizing the representation for $\boldsymbol{V}$ in (27), we obtain

$$\mathbf{Cov}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{V}\} = \frac{2\sqrt{n}}{C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2,$$

so $\mathbf{Cov}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} = \frac{2\sqrt{n}}{W^{3/2}C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2$. Combining these results, we obtain

$$\begin{aligned}
&\mathbf{Var}\{(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} =\\
&\quad \frac{4}{W^2}\left\{\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2 + \frac{n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\} + 4\left\{\frac{1}{W}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\} - 4\left\{\frac{2\sqrt{n}}{W^{3/2}C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\}.
\end{aligned}$$

Also,

$$\mathbf{Cov}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu}), (\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} = \boldsymbol{H}^{(1)'}\left[\mathbf{Cov}(\boldsymbol{T}, \boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W) - 2\mathbf{Cov}(\boldsymbol{T}|W)\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right].$$

But, once again,

$$\mathbf{Cov}(\boldsymbol{T}, \boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W) = \tfrac{1}{W^{3/2}}\mathbf{Cov}(\boldsymbol{V}, \boldsymbol{V}\boldsymbol{H}^{(2)}\boldsymbol{V}) = -\tfrac{2\sqrt{n}}{W^{3/2}}\mathbf{1}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\Delta}) = \tfrac{2\sqrt{n}}{W^{3/2}C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\mathbf{1},$$

while $\boldsymbol{H}^{(1)'}\mathbf{Cov}(\boldsymbol{T}|W)\boldsymbol{H}^{(2)}\boldsymbol{\nu} = \frac{1}{W}(\boldsymbol{H}^{(1)'}\mathbf{1})(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})$. Thus, we have

$$\mathbf{Cov}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu}), (\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} = \frac{2\sqrt{n}}{W^{3/2}C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\mathbf{1} - \frac{2}{W}(\boldsymbol{H}^{(1)'}\mathbf{1})(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}).$$

Now, by using these intermediate results, we find that

$$\begin{aligned}\mathbf{Var}(Q|W) &= \mathbf{Var}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu})\} + \frac{1}{4}\mathbf{Var}\{(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} + \\ &\quad (2)\left(\frac{1}{2}\right)\mathbf{Cov}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu}), (\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} \\ &= \frac{1}{W}(\boldsymbol{H}^{(1)'}\mathbf{1})^2 + \frac{1}{4}\left\{\frac{4}{W^2}\left[\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2 + \frac{n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right] + \right. \\ &\quad \left. \frac{4}{W}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2 - \frac{(4)(2)}{W^{3/2}}\frac{\sqrt{n}}{C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\} + \\ &\quad \frac{2}{W^{3/2}}\frac{\sqrt{n}}{C_n}(\boldsymbol{H}^{(1)'}\mathbf{1})(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}) - \frac{2}{W}(\boldsymbol{H}^{(1)'}\mathbf{1})(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}).\end{aligned}$$

and the result follows after simplification.

Proof of Theorem 5.7: From (25) and (28), and by the iterated variance rule, an approximate variance of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is $V_2(\boldsymbol{\Delta}) \equiv \mathbf{Var}\left\{\frac{1}{n}WQ\right\} = \frac{1}{n}\left[\mathbf{Var}\left\{\frac{W}{\sqrt{n}}\mathbf{E}(Q|W)\right\} + \mathbf{E}\left\{\frac{W^2}{n}\mathbf{Var}(Q|W)\right\}\right] \equiv \frac{1}{n}\left\{\mathrm{VE}(\boldsymbol{\Delta}) + \mathrm{EV}(\boldsymbol{\Delta})\right\}$. The final expression of $\mathrm{VE}(\boldsymbol{\Delta})$ follows from the Lemma A.2, the identities $\mathbf{Var}(W) = 2(n-1)$, $\mathbf{Var}(\sqrt{W}) = (n-1) - (n-2)^2C_n^2/n$, and $\mathbf{Cov}(W, \sqrt{W}) = (n-2)C_n/\sqrt{n}$, and

$$\begin{aligned}&\mathrm{VE}(\boldsymbol{\Delta}) = \frac{1}{n}\left[\mathbf{Var}\left\{W\left[1 + H - \boldsymbol{H}^{(1)'}\boldsymbol{\nu} + \frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right] + \sqrt{W}\frac{\sqrt{n}}{C_n}\left(\boldsymbol{H}^{(1)'}\boldsymbol{\nu} - \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)\right\}\right] \\ &= \frac{1}{n}\left\{\left[1 + H - \boldsymbol{H}^{(1)'}\boldsymbol{\nu} + \frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right]^2\mathbf{Var}(W) + \frac{n}{C_n^2}\left(\boldsymbol{H}^{(1)'}\boldsymbol{\nu} - \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2\mathbf{Var}(\sqrt{W})\right. \\ &\left. + \frac{2\sqrt{n}}{C_n}\left[1 + H - \boldsymbol{H}^{(1)'}\boldsymbol{\nu} + \frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right]\left(\boldsymbol{H}^{(1)'}\boldsymbol{\nu} - \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)\mathbf{Cov}(W, \sqrt{W})\right\}.\end{aligned}$$

Also, from Lemma A.3,

$$\begin{aligned}&\mathrm{EV}(\boldsymbol{\Delta}) = \frac{1}{n}\left\{\mathbf{E}(W)\left(\boldsymbol{H}^{(1)'}\mathbf{1} - \mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2 + \right. \\ &\left. \frac{2\sqrt{n}}{C_n}\left(\boldsymbol{H}^{(1)'}\mathbf{1} - \mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\mathbf{E}(\sqrt{W}) + \frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2 + \frac{n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\},\end{aligned}$$

which simplifies to the expression of $\mathrm{EV}(\boldsymbol{\Delta})$ in the statement of the theorem upon substituting the expressions $\mathbf{E}(W) = n-1$ and $\mathbf{E}(\sqrt{W}) = (n-2)C_n/\sqrt{n}$. This completes the proof.

References

Akaike, H. *Information theory and the maximum likelihood principle*, 610–624. Budapest: Akademiai Kiado (1973).

Arnold, B. and Villasenor, J. "Variance estimation with diffuse prior information." *Stat. Prob. Letters*, 33:35–39 (1997).

Breiman, L. “The Little Bootstrap and Other Methods for Dimensionality Selection in Regression: X-Fixed Prediction Error.” *J. Amer. Statist. Assoc.*, 87:738–754 (1992).

Brewster, J. and Zidek, J. “Improving on equivariant estimators.” *Ann. Statist.*, 2:21–38 (1974).

Buhlmann, P. “Efficient and adaptive post-model-selection estimators.” *Journal of Statistical Plannning and Inference*, 79:1–9 (1999).

Burnham, K. and Anderson, D. *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer (1998).

Claeskens, G. and Hjort, N. “The Focussed Information Criterion (with discussion).” *Journal of the American Statistical Assoiation*, 98:900–945 (2003).

Dukić, V. and Peña, E. “Estimation after Model Selection in a Gaussian Model.” Technical report, Department of Health Studies, University of Chicago, Department of Health Studies (2003).

Gelfand, A. and Dey, D. “Improved estimation of the disturbance variance in a linear regression model.” *J. Econometrics*, 39:387–395 (1977).

Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer (2001).

Hjort, N. and Claeskens, G. “Frequentist Model Average Estimators.” *Journal of the American Statistical Association*, 98:879–899 (2003).

Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. “Bayesian Model Averaging.” *Statistical Science*, 14:382–401 (1999).

Judge, G., Bock, M., and Yancey, T. “Post Data Model Evaluation.” *The Review of Economics and Statistics*, 56:245–253 (1974).

Leamer, E. *Specification Searches*. New York: Wiley (1978).

Lehmann, E. and Casella, G. *Theory of Point Estimation*. New York: Springer (1998).

Maatta, J. and Casella, G. “Developments in decision-theoretic variance estimation (with discussions).” *Statist. Sci.*, 5:90–120 (1990).

Madigan, D. and Raftery, A. “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam’s Window.” *J. Amer. Statist. Assoc.*, 89:1535–1546 (1994).

Ohtani, K. “MSE dominance of the pre-test iterative variance estimator over the iterative variance estimator in regression.” *Stat. Prob. Letters*, 54:331–340 (2001).

Pal, N., Ling, C., and Lin, J. “Estimation of a normal variance – a critical review.” *Statist. Papers*, 39:389–404 (1998).

Potscher, B. “Effects of Model Selection on Inference.” *Econometric Theory*, 7:163–185 (1991).

Raftery, A., Madigan, D., and Hoeting, J. “Bayesian Model Averaging for Linear Regression Models.” *J. Amer. Statist. Assoc.*, 92:179–191 (1997).

Robert, C. *The Bayesian Choice*. New York: Springer, 2nd edition edition (2001).

Rukhin, A. “How much better are better estimators of a normal variance.” *J. Amer. Statist. Assoc.*, 82:925–928 (1987).

Schwarz, G. “Estimating the dimension of a model.” *Ann. Statist.*, 6:461–464 (1978).

Sclove, S., Morris, C., and Radhakrishnan, R. “Nonoptimality of preliminary-test estimators for the mean of a multivariate normal distribution.” *Ann. Math. Statist.*, 43:1481–1490 (1972).

Sen, P. . and Saleh, A. “On preliminary test and shrinkage M-estimation in linear models.” *Ann. Statist.*, 15:1580–1592 (1987).

Stein, C. “Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean.” *Ann. Inst. Statist. Math.*, 16:155–160 (1964).

Vidaković, B. and DasGupta, A. “Lower bounds on Bayes risks for estimating a normal variance: with applications.” *Canad. J. Statist.*, 23:269–282 (1995).

Wallace, T. “Pretest estimation in regression: A survey.” *Amer. J. Agr. Econ.*, 59:431–443 (1977).

Yancey, T., Judge, G., and Mandy, D. “The sampling performance of pre-test estimators of the scale parameter under squared error loss.” *Economics Letters*, 12:181–186 (1983).

Yang, Y. “Regression with multiple candidate models: selecting or mixing?” *Statistica Sinica*, 13:783–809 (2003).

Zhang, P. “Inference after variable selection in linear regression models.” *Biometrika*, 79:741–746 (1992a).

—. “On the Distributional Properties of Model Selection Criteria.” *J. Amer. Statist. Assoc.*, 87:732–737 (1992b).

Table 2: Efficiencies (relative to of the UMVU estimator $\hat{\sigma}^2_{UMVU}$) of the different variance estimators for different combinations of p, $\mathbf{\Delta}$, and n. For the limiting Bayes and weighted estimators, the values are based on simulation studies with 10000 replications for each combination.

Combinations of p and $\mathbf{\Delta}$	n	Efficiency %										
		MRE	pMLE	pMRE	ALB	LB1	LB2	LB3	LB4	PLB1	PLB2	PLB3
$\mathbf{\Delta}$=(-0.25,0,0.25) p=3	3	200	174	222	14	10	60	152	230	240	236	233
	10	122	117	124	71	62	89	113	128	129	129	128
	30	107	106	107	91	89	99	107	105	112	112	112
$\mathbf{\Delta}$=(-0.5,0,0.5) p=3	3	200	183	209	17	11	66	160	217	235	226	221
	10	122	119	124	73	59	86	110	127	129	128	127
	30	107	106	110	89	84	95	104	109	111	111	111
$\mathbf{\Delta}$=(0,0.25, 0.50) p=3	3	200	164	226	13	9	51	135	227	238	234	232
	10	122	114	127	66	53	78	103	131	129	128	128
	30	107	105	109	88	81	92	100	107	108	108	108
$\mathbf{\Delta}$=(0,0.5,1) p=3	3	200	166	222	13	8	47	128	222	233	228	224
	10	122	115	128	65	51	76	102	126	126	126	126
	30	107	105	110	87	83	94	103	109	110	110	110
$\mathbf{\Delta}$=(-0.25:2^{-4}:0.25) p=9	3	200	174	222	14	10	58	149	234	241	238	235
	10	122	117	123	71	61	88	112	127	130	130	129
	30	107	105	106	91	88	98	106	109	111	111	111
$\mathbf{\Delta}$=(-0.25:2^{-5}:0.25) p=17	3	200	174	222	14	10	57	145	234	239	237	234
	10	122	117	123	71	60	86	109	130	127	127	126
	30	107	105	106	91	87	97	105	108	110	110	110
$\mathbf{\Delta}$=(0:2^{-4}:0.5) p=9	3	200	164	225	13	9	52	137	230	243	240	238
	10	122	114	126	66	53	77	102	129	128	127	127
	30	107	104	108	88	76	87	97	109	107	107	107
$\mathbf{\Delta}$=(0:2^{-5}:0.5) p=17	3	200	164	225	13	10	55	145	235	249	246	243
	10	122	114	126	66	54	79	106	130	134	133	133
	30	107	104	108	88	76	87	97	111	107	107	108

Table 3: Efficiencies (relative to the UMVU estimator $\hat{\sigma}^2_{UMVU}$) of the different variance estimators for different combinations of p, $\mathbf{\Delta}$, and n. For the limiting Bayes and weighted estimators, 10000 simulation replications were performed for each combination.

Combinations of p and $\mathbf{\Delta}$	n	Efficiency %										
		MRE	pMLE	pMRE	ALB	LB1	LB2	LB3	LB4	PLB1	PLB2	PLB3
$\mathbf{\Delta}$=(0, 1) p=2	3	200	170	232	13	8	52	143	229	243	237	234
	10	122	115	134	63	55	81	107	130	129	130	130
	30	107	104	111	85	85	96	105	109	112	112	112
$\mathbf{\Delta}$=(-1, 0, 1) p=3	3	200	195	216	18	10	63	162	221	255	237	228
	10	122	120	134	67	51	77	103	129	126	127	128
	30	107	104	111	85	84	95	103	110	111	111	111
$\mathbf{\Delta}$=(-1: 2^{-1} :1) p=5	3	200	185	199	19	11	66	162	208	246	230	221
	10	122	119	124	73	52	78	103	125	126	126	125
	30	107	106	110	89	81	92	101	107	108	108	108
$\mathbf{\Delta}$=(-1: 2^{-2} :1) p=9	3	200	182	195	20	11	64	153	210	232	219	211
	10	122	118	120	74	56	82	106	126	126	125	125
	30	107	106	107	91	82	92	100	105	106	106	106
$\mathbf{\Delta}$=(-1: 2^{-3} :1) p=17	3	200	181	194	20	11	65	155	207	234	220	213
	10	122	117	119	75	55	81	104	126	125	124	123
	30	107	105	106	92	80	90	99	107	107	107	107
$\mathbf{\Delta}$=(-1: 2^{-4} :1) p=33	3	200	181	194	20	12	72	168	206	240	226	217
	10	122	117	119	75	57	82	106	125	126	125	124
	30	107	105	105	92	82	93	101	105	108	108	108
$\mathbf{\Delta}$=(-1: 2^{-5} :1) p=65	3	200	181	194	20	11	65	156	208	234	221	214
	10	122	117	119	75	58	84	109	123	128	127	127
	30	107	105	105	92	82	93	102	107	109	109	109
$\mathbf{\Delta}$=(-1: 2^{-6} :1) p=129	3	200	181	194	20	11	69	163	207	237	224	216
	10	122	117	119	75	56	82	106	126	126	125	125
	30	107	105	105	92	81	91	99	108	106	106	106
$\mathbf{\Delta}$=(-1: 2^{-7} :1) p=257	3	200	181	194	20	11	67	159	205	236	223	215
	10	122	117	119	75	57	83	107	124	128	127	126
	30	107	105	105	92	82	92	101	108	108	108	108
$\mathbf{\Delta}$=(-1: 2^{-8} :1) p=513	3	200	181	194	20	11	66	156	206	233	221	213
	10	122	117	119	75	58	84	108	127	128	127	126
	30	107	105	105	92	80	91	99	106	107	107	107

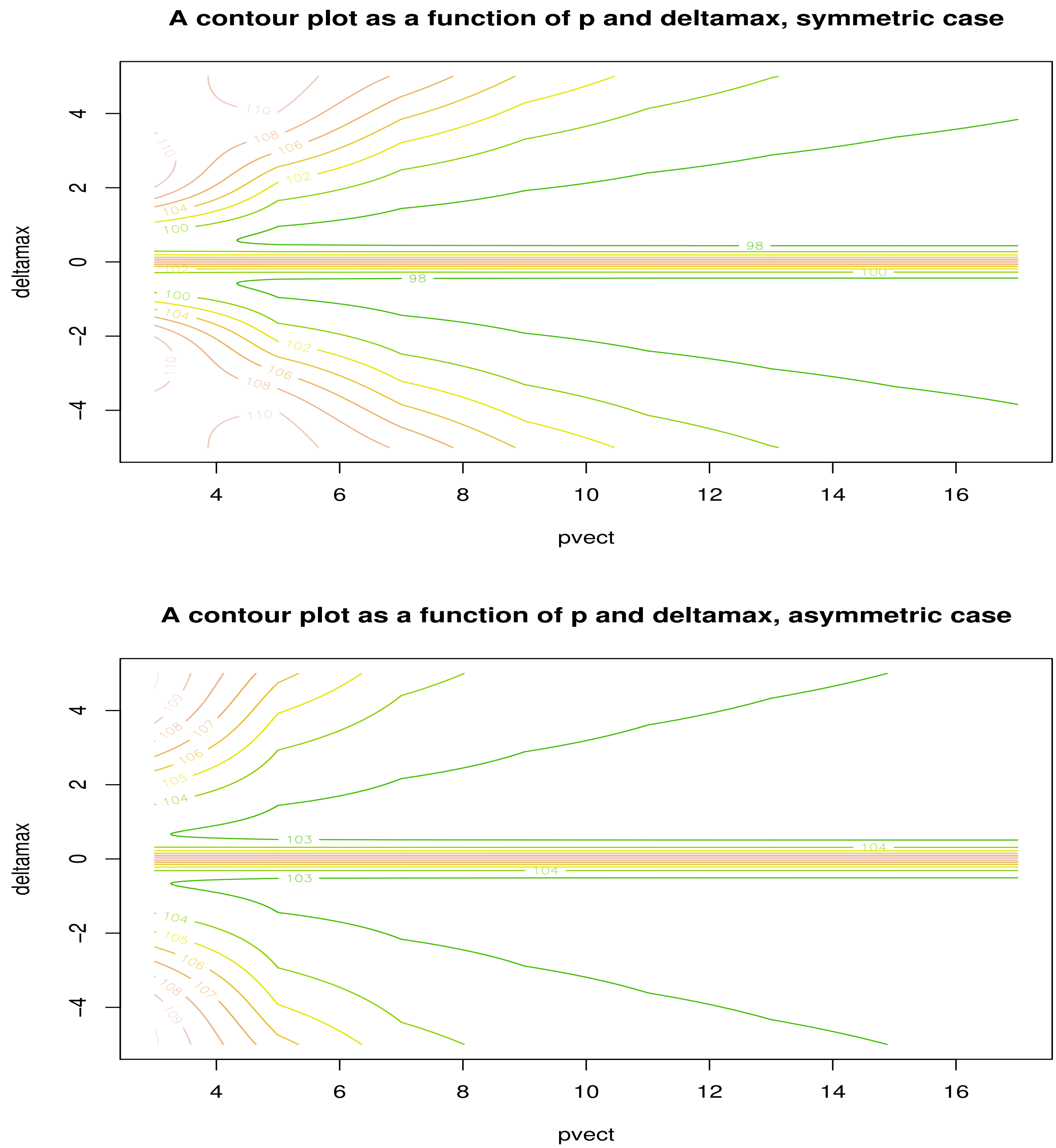

Figure 1: Relative efficiencies of pMRE with respect to MRE in a symmetric and asymmetric Δ cases, as a function of $\Delta_{\max}$ and number of sub-models p for sample size of $n = 10$. The symmetric case is of form $\boldsymbol{\Delta} = [-\Delta_{max} : \Delta_{max}/(p-1) : \Delta_{max}]$, while the asymmetric case is of form $\boldsymbol{\Delta} = [0 : \Delta_{max}/(2(p-1)) : \Delta_{max}]$.

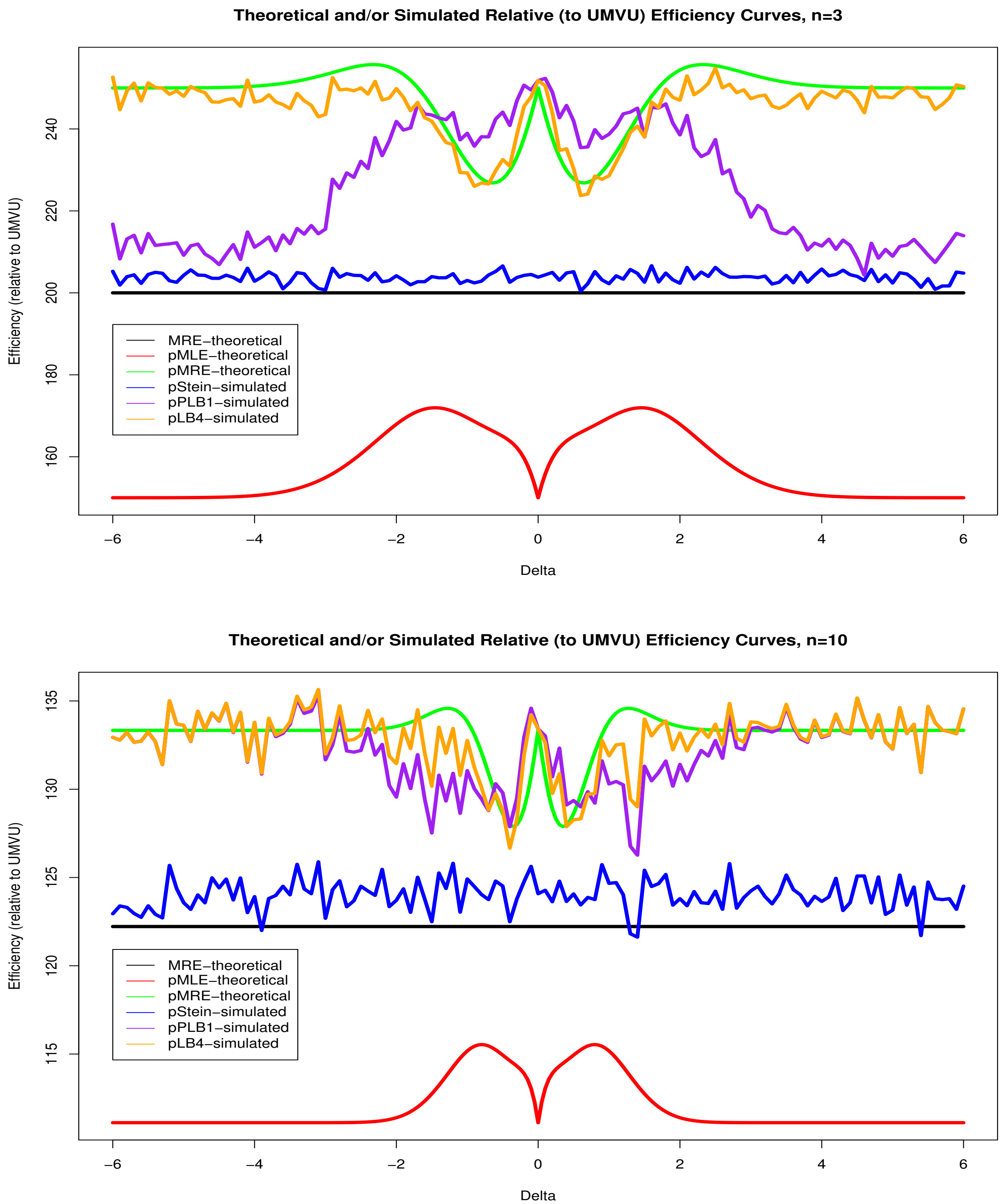

Figure 2: Efficiencies of the leading estimators and Stein's estimator of σ^2 relative to the UMVUE $\hat{\sigma}^2_{UMVUE}$ for $p = 2$ and $\mathbf{\Delta} = (0, \Delta)$, with Δ varying, for $n = 3, 10$. The connected scatterplots represent the simulated relative efficiencies based on 20000 replications for each Δ for the limiting Bayes (pLB4, lighter of the top two), weighted (pPLB1, darker of the top two), and Stein (bottom) estimators. The smooth curves correspond to theoretical efficiencies of $\hat{\sigma}^2_{p,MRE}$ (top), $\hat{\sigma}^2_{MRE}$ (middle), and $\hat{\sigma}^2_{p,MLE}$ (bottom).