Estimation after Model Selection in a Gaussian Model

Vanja M. Dukić* Edsel A. Peña†

November 6, 2002

Abstract

The problem of estimating the variance and the distribution function at a given value for a Gaussian model which is intermediate between a model where the mean parameter is fully known and a model where the mean parameter is completely unknown is considered. This problem is motivated by the desire to understand the theoretical implications of the process of selecting a model among several sub-models, and then estimating a parameter of interest after the model selection, but with these sequential steps using the *same* sample data. This practice is common in many areas such as in regression analysis where a subset of possible predictors is chosen possibly through stepwise regression; in reliability and survival analysis settings where it may only be known that the failure time distribution belongs to either of two parametric classes of distributions functions such as the Weibull and gamma classes; or in goodness-of-fit testing where an embedding approach is utilized to develop the test procedures, such in Neyman's smooth goodness-of-fit approach. Estimators of the variance and the distribution function derived under both the wider model and the intermediate model are obtained. Some of these estimators are related to *pre-test* estimators which improve on classical estimators. The performances of the estimators are compared theoretically and through simulations using their risk functions. It is found that efficiency gains are obtained by exploiting the sub-model structure through the use of adaptive, Bayes, and sub-model weighted estimators, especially when the number of competing sub-models is few, but this advantage may deteriorate and be lost altogether for some adaptive estimators as the number of sub-models increases. In particular, it is demonstrated that weighted estimators motivated via the Bayesian approach perform best in estimating the variance and the distribution function and are preferable over two-step adaptive estimators.

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

Key Words and Phrases: Adaptive estimators; Bayes estimators; estimators of distribution function; model selection; pre-test estimators; risk functions; variance estimation.

*V. Dukić is Assistant Professor in the Department of Health Studies, University of Chicago, Chicago, IL 60637. 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 DMS 0102870 which partially supports his research. His e-mail address is *pena@stat.sc.edu*.

1 Introduction

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ be a vector of independent and identically distributed (IID) random variables from an unknown distribution function $F(x) = \Pr\{X_1 \le x\}$. If it is known that F belongs to the two-parameter normal family of distributions $\mathcal{M} = \{N(\mu, \sigma^2) : (\mu, \sigma^2) \in \Theta = \Re \times \Re_+\}$ with $N(\mu, \sigma^2)$ denoting the normal distribution with mean μ and 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 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (cf., Lehmann and Casella ([11])). More generally, we adopt a decision-theoretic approach for evaluating estimators of σ^2 using the risk function from the quadratic loss function $L_1 : \Re \times \Theta \to \Re$ given by

$$L_1(a, (\mu, \sigma^2)) = \left(\frac{a - \sigma^2}{\sigma^2}\right)^2 . \tag{2}$$

We are of course aware of the controversy in the use of this loss function arising from Stein's demonstration ([21]; cf., Maatta and Casella ([12])) that under this loss function 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 is also inadmissible). At the same time, however, we note that quadratic loss functions are still predominantly the loss functions of choice when dealing with the estimation of the variance (cf., Arnold and Villasenor ([2]), Brewster and Zidek ([3]), Gelfand and Dey ([5]), Maatta and Casella ([12]), Ohtani ([14]), Pal, Ling and Lin ([15]), Rukhin ([19]), Vidaković and DasGupta ([23]), Wallace ([24])).

Shifting gears, suppose also that, for a fixed $t \in \Re$, it is of interest to estimate

$$\tau(t) = \tau(t; \mu, \sigma^2) = F(t; \mu, \sigma^2) = \Pr\{X_1 \le t | \mu, \sigma^2\} = \Phi\left(\frac{t - \mu}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal distribution function. By Lehmann-Scheffe Theorem with assist from Basu's Theorem (cf., Lehmann and Casella ([11])), the UMVUE of $\tau(t; \mu, \sigma^2)$ is

$$\hat{\tau}_{UMVU}(t) = \mathcal{T}\left(\frac{\sqrt{n-2}\,z_1(t)}{\sqrt{1 - z_1(t)^2}}; n-2\right) I\{|z_1(t)| \le 1\} + I\{z_1(t) > 1\} \tag{4}$$

where $z_1(t) = (\sqrt{n}/(n-1))\left((t - \bar{X})/S\right)$, and $\mathcal{T}(\cdot; \nu)$ is the Student's t-distribution function with ν degrees of freedom. From the decision-theoretic framework, estimators of $\tau(t)$ will be evaluated using the risk function from the loss function $L_2 : [0,1] \times \Theta \to \Re$, given by

$$L_2(a, (\mu, \sigma^2)) = (a - \tau(t; \mu, \sigma^2))^2. \tag{5}$$

When we desire a global evaluation of an estimator $\hat{\tau} = \hat{F}$ of the distribution function $\tau = F$, we will use the risk function associated with the integrated loss function $L_3(\hat{F}, F) = \int L_2(\hat{F}(t), F(t))dF(t)$.

If the model is restricted according to $\mu = \mu_0$ 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

$$\hat{\sigma}^2_{UMVU}(\mu_0) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 \quad \text{and} \quad \hat{\sigma}^2_{MRE}(\mu_0) = \frac{1}{n+2}\sum_{i=1}^{n}(X_i - \mu_0)^2. \tag{6}$$

And with $z_2(t) = z_2(t; \mu_0) = (1/\sqrt{n})(t - \mu_0)/\hat{\sigma}_{UMVU}(\mu_0)$, the UMVUE of $\tau(t)$ under $\mathcal{M}_0$ is

$$\hat{\tau}_{UMVU}(t; \mu_0) = \mathcal{T}\left(\frac{\sqrt{n-1}z_2(t)}{\sqrt{1 - z_2(t)^2}}; n - 1\right) I\{|z_2(t)| \le 1\} + I\{z_2(t) > 1\}. \tag{7}$$

Since $\mathcal{M}_0$ is contained in $\mathcal{M}$, and has no uncertainty regarding the value of the mean parameter, we are able to improve on the estimators derived under $\mathcal{M}$ by exploiting the knowledge about μ under $\mathcal{M}_0$. For instance, when model $\mathcal{M}_0$ holds, the relative efficiency of the estimator $\hat{\sigma}^2_{UMVU}(\mu_0)$ in (6) with respect to $\hat{\sigma}^2_{UMVU}$ in (1) is $n/(n-1)$.

Suppose however 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. Consider the estimation of σ^2 and $\tau(t)$ under the model

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

In this $\mathcal{M}_p$ model, in contrast to $\mathcal{M}_0$, there is some information about the possible value of μ, but we are not certain about this value. This situation arises in a variety of practical settings. For example, it includes decision problems with only two possible actions such as 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 model Stein utilized ([21]; cf., Zacks ([26]), Brewster and Zidek ([3]), Wallace ([24]), and Maatta and Casella ([12])) to derive a pre-test estimator dominating $\hat{\sigma}^2_{MRE}$. Stein considered the case where $p = 1$ and $\mu_1 = 0$.

The situation is of theoretical interest in the context of model selection since $\mathcal{M}_p$ can be viewed as consisting of p sub-models, only one of which is the correct model. In the process of estimating parameters, such as σ^2 and $\tau(t)$, three possible strategies are as follows:

1. one may simply utilize estimators developed for a larger model such as $\mathcal{M}$;

2. one may use the sample data to first select a sub-model, and then utilize an estimator developed for the chosen sub-model, but using the same data;

3. one may assign to each sub-model a plausibility measure (either using the data or not), and then form a weighted combination of sub-model estimators.

The first strategy is clearly inefficient. It is thus of interest to compare strategies 2 and 3, and to decide when to recommend one over the other. As it is intuitive to expect that the advantages of strategies 2 and 3 with respect to strategy 1 will degrade as the number of sub-models increases, one may also ask whether this advantage could disappear entirely at some point. For instance, there will be no gain in knowing that μ belongs to the set of rational numbers, since $(\bar{X}, S^2)$ is complete sufficient and $\mathcal{M}$-estimators suffice. Some of the issues pertaining to the two-step process of inference after model selection have been discussed in the econometric literature, cf., Judge, Bock and Yancey ([7]), Wallace ([24]), Leamer ([9]), and Yancey, Judge and Mandy ([25]).

A related and more complicated situation arises in the context of multiple regression, when we have several possible predictors to choose from (using the stepwise method for example), and having chosen the predictors, perform inference such as estimation of regression coefficients, prediction of future observations, etc., using the same data utilized in the predictor selection step. For an interesting discussion of the consequences of such "data double-dipping" in an econometric setting see Wallace ([24]) and references therein.

Another situation, which initially motivated this paper, pertains to goodness-of-fit testing. One approach to developing goodness-of-fit tests (cf., Neyman ([13]), Rayner and Best ([17]), Peña ([16])) is to embed the null class of distributions into a wider class. The formation of this wider class involves a smoothing order which determines the number of basis functions. This creates a family of possible models of different dimensions. The goodness-of-fit tests are usually score tests derived under the wider class. However, the power properties of the tests depend on the chosen smoothing order. To alleviate this problem the smoothing order could be adaptively chosen via, for example, the Schwarz information criterion (Schwartz ([20])), or some other information-based selection methods (cf., Ledwina ([10]), Kallenberg and Ledwina ([8])). Having chosen the smoothing order, one proceeds by using the test for the chosen order, but again using the *same* data. The properties of the test for a fixed order are different from those where the order is chosen adaptively.

It is our assessment that a variety of issues pertaining to the re-use of the sample data in the process of choosing a model (selection) and doing inference (estimation) have not been fully examined. Variance estimation is certainly of independent research interest and has had a wide literature (see Arnold and Villasenor ([2]), Gelfand and Dey ([5]), Pal, Ling and Lin ([15]), Vidaković and DasGupta ([23]), and Yancey, Judge and Mandy ([25])). Most often in the variance estimation problems however, models $\mathcal{M}$ or $\mathcal{M}_0$ are postulated. Even when there are only two possible populations, as in the setting of the Neyman-Pearson Lemma, the typical variance estimator does not exploit the fact that there are only two possible means! The present paper, which considers the model $\mathcal{M}_p$ as the initial specimen, will shed some light onto the variance estimation problem, as well as the estimation of a distribution function, when there is partial information about the population mean.

2 Likelihood-Based Estimators

2.1 Estimators of σ^2

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

$$L(\mu, \sigma^2) = L(\mu, \sigma^2 | \boldsymbol{x}) = \prod_{i=1}^{p} L_i(\mu_i, \sigma^2)^{M_i} \tag{8}$$

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

$$L_i(\mu_i, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{n\hat{\sigma}_i^2}{2\sigma^2}\right\} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu_i)^2. \tag{9}$$

$L_i(\mu_i, \sigma^2)$ 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 Schwartz' Bayesian information criterion (BIC) ([20])) or the Akaike information criterion (AIC) ([1]). In this paper we restrict our attention to the selector $\hat{M}$. This selector could also be viewed as a highest posterior probability model selector associated with a flat prior distribution. The maximum likelihood estimator (MLE) of (μ, σ^2) under $\mathcal{M}_p$ is $(\hat{\mu}_{p,MLE}, \hat{\sigma}^2_{p,MLE}) = (\mu_{\hat{M}}, \hat{\sigma}^2_{\hat{M}}) = \sum_{i=1}^{p} I\{\hat{M} = i\}(\mu_i, \hat{\sigma}_i^2)$. It follows that the MLE of σ^2 is

$$\hat{\sigma}^2_{p,MLE} = \hat{\sigma}^2_{\hat{M}} = \sum_{i=1}^{p} I\{\hat{M} = i\}\hat{\sigma}_i^2, \tag{10}$$

a two-stage 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 (10) is to use the sub-model's MRE instead of MLE of σ^2:

$$\hat{\sigma}^2_{p,MRE} = \hat{\sigma}^2_{MRE,\hat{M}} = \sum_{i=1}^{p} I\{\hat{M} = i\}\hat{\sigma}^2_{MRE,i} = \sum_{i=1}^{p} I\{\hat{M} = i\}\frac{n\hat{\sigma}_i^2}{(n+2)}. \tag{11}$$

At this point, we remark 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, with a little sacrifice in literal precision we are able to indicate the type or origin of the estimator, an acceptable price to pay.

2.2 Estimators of $\tau(t)$

By the invariance principle, the MLE of $\tau(t)$ under $\mathcal{M}_p$ is

$$\hat{\tau}_{p,MLE}(t) = \Phi\left(\frac{t-\mu_{\hat{M}}}{\hat{\sigma}_{\hat{M}}}\right) = \sum_{i=1}^{p} I\{\hat{M}=i\}\Phi\left(\frac{t-\mu_i}{\hat{\sigma}_i}\right). \tag{12}$$

This estimator is also a two-stage estimator formed by first selecting a sub-model, then using an estimator of $\tau(t)$ in the chosen sub-model. If we define $z_{3i}(t) = (1/\sqrt{n})\left((t-\mu_i)/\hat{\sigma}_i\right), i = 1, 2, \ldots, p$, the UMVUE of $\tau(t)$ under $\mathcal{M}_{p,i} = \{N(\mu_i, \sigma^2) : \sigma^2 \in \Re_+\}$ is

$$\hat{\tau}_{UMVU,i}(t) = \mathcal{T}\left(\frac{\sqrt{n-1}z_{3i}(t)}{\sqrt{1-z_{3i}^2(t)}}; n-1\right) I\{|z_{3i}(t)| \le 1\} + I\{z_{3i}(t) > 1\}. \tag{13}$$

By using these sub-models UMVUEs, an estimator of $\tau(t)$ under $\mathcal{M}_p$ can be formed via

$$\hat{\tau}_{p,UMVU}(t) = \hat{\tau}_{UMVU,\hat{M}}(t) = \sum_{i=1}^{p} I\{\hat{M}=i\}\hat{\tau}_{UMVU,i}(t). \tag{14}$$

3 Bayes and Weighted Estimators

The class of prior densities of (μ, σ^2) we choose consists of the product of a multinomial probability function and an inverse gamma density:

$$\pi(\mu, \sigma^2|\tilde{\boldsymbol{\theta}}, \kappa, \beta) = \left(\prod_{i=1}^{p} \tilde{\theta}_i^{m_i}\right) \frac{\beta^{\kappa-1}}{\Gamma(\kappa-1)} \left(\frac{1}{\sigma^2}\right)^{\kappa} \exp\left(-\frac{\beta}{\sigma^2}\right), \tag{15}$$

where $\sigma^2 > 0$, $m_i = I\{\mu = \mu_i\}$ so that $\sum_1^p m_i = 1$, and $0 \le \tilde{\theta}_i \le 1$ with $\sum_1^p \tilde{\theta}_i = 1$, $\beta > 0$, and $\kappa > 1$. The density in (15) is with respect to the product measure induced by the counting measure on $\Re^p$ and Lebesgue measure on $\Re^1$.

Multiplying the prior density function in (15) and the likelihood function in (8), and subsuming constants into a factor C, the posterior density of (μ, σ^2), given $\boldsymbol{X} = \boldsymbol{x}$, is

$$\pi(\mu, \sigma^2 \mid \boldsymbol{x}) \quad = \quad C\prod_{i=1}^{p}\left\{\tilde{\theta}_i\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\kappa}\exp\left(-\frac{1}{\sigma^2}\left[\frac{n\hat{\sigma}_i^2}{2}+\beta\right]\right)\right\}^{m_i}. \tag{16}$$

Note that $\pi(\mu, \sigma^2|\boldsymbol{x}) = \pi(\boldsymbol{m}, \sigma^2|\boldsymbol{x})$ 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's. Since $\sum_{\mathbf{m} \in \mathbf{M}} \int_0^\infty \pi(\mathbf{m}, \sigma^2 \mid \boldsymbol{x})\, d\sigma^2 = 1$, where $\mathbf{M} = \{\mathbf{m} = (m_1, ..., m_p) : \ m_i \in \{0, 1\}, \ \sum_{i=1}^p m_i = 1\}$, it follows that

$$C \;=\; \frac{1}{\Gamma(n/2+\kappa-1)} \left\{ \sum_{i=1}^p \frac{\tilde{\theta}_i}{(n\hat{\sigma}_i^2/2+\beta)^{n/2+\kappa-1}} \right\}^{-1}. \tag{17}$$

3.1 Posterior Probabilities of Sub-Models

From the posterior distribution in (16), the marginal posterior density (with respect to counting measure) of μ, or equivalently of $\boldsymbol{M} = (M_1, ..., M_p)$, is

$$\pi(M_1 = m_1, ..., M_p = m_p \mid \mathbf{x}) = C \prod_{i=1}^p \left\{ \tilde{\theta}_i \frac{\Gamma(n/2+\kappa-1)}{(n\hat{\sigma}_i^2/2+\beta)^{n/2+\kappa-1}} \right\}^{m_i} = \prod_{i=1}^p \{\theta_i(\kappa, \beta, n, \mathbf{x})\}^{m_i},$$

where, for $i = 1, 2, \ldots, p$,

$$\theta_i(\kappa, \beta, n, \boldsymbol{x}) \;=\; \frac{\tilde{\theta}_i (n\hat{\sigma}_i^2/2+\beta)^{-(n/2+\kappa-1)}}{\sum_{j=1}^p \tilde{\theta}_j \left(n\hat{\sigma}_j{}^2/2+\beta\right)^{-(n/2+\kappa-1)}}. \tag{18}$$

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

3.2 Posterior Distribution of σ^2

The marginal posterior density function of σ^2 is directly obtained from (16) to be

$$\pi(\sigma^2 \mid \boldsymbol{x}) \;=\; C \sum_{i=1}^p \tilde{\theta}_i \left(\frac{1}{\sigma^2}\right)^{(\kappa+n/2)} \exp\left[-\frac{1}{\sigma^2} \left(\frac{n\hat{\sigma}_i^2}{2} + \beta \right) \right] I\{\sigma^2 > 0\}. \tag{19}$$

A straightforward derivation reveals that the posterior mean of σ^2 is

$$E(\sigma^2 \mid \boldsymbol{x}) \;=\; \sum_{i=1}^p \theta_i(\kappa, \beta, n, \boldsymbol{x}) \left\{ \left(\frac{n/2}{n/2+\kappa-2} \right) \hat{\sigma}_i^2 + \left(\frac{\kappa-2}{n/2+\kappa-2} \right) \left(\frac{\beta}{\kappa-2} \right) \right\}.$$

Note that $\beta/(\kappa-2)$ is the prior mean of σ^2, provided $\kappa > 2$ (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. Therefore, the Bayes estimator of σ^2 under the loss function L_1 in (2) is

$$\hat{\sigma}^2_{p,Bayes}(\kappa, \beta, \boldsymbol{\theta}) \;=\; \sum_{i=1}^p \theta_i(\kappa, \beta, n, \boldsymbol{x}) \left\{ \left(\frac{n}{n+2(\kappa-2)} \right) \hat{\sigma}_i^2 + \left(\frac{2(\kappa-2)}{n+2(\kappa-2)} \right) \left(\frac{\beta}{\kappa-2} \right) \right\}. \tag{20}$$

This estimator mixes adaptively, 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 set κ to four values: $\kappa \to 1$, $\kappa \to 3/2$, $\kappa = 2$, and $\kappa = 3$. The rationale for these choices of κ is as follows. Letting $\kappa \to 1$ 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 $1/\sigma^2$ (cf., Roberts ([18]). The choice $\kappa \to 3/2$ on the other hand leads to Jeffreys' prior for σ^2 when the mean is unknown in the normal model, because in this case Jeffreys' prior is proportional to $(1/\sigma^2)^{(3/2)}$. Setting $\kappa = 2$ and $\kappa = 3$ produce (limiting) Bayes estimators that are convex combinations of the sub-models' MLEs and MREs, respectively. The table below lists the sub-models' posterior probabilities and the resulting limiting Bayes estimators of σ^2:

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	$\theta_{i4} = (\hat{\sigma}_i^2)^{-(n+4)/2} / \sum_{j=1}^p (\hat{\sigma}_j^2)^{-(n+4)/2}$	$\hat{\sigma}^2_{p,LB4} = \left(\frac{n}{n+2}\right) \sum_{i=1}^p \theta_{i4} \hat{\sigma}_i^2$

Each of the set of sub-models' posterior probabilities 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 new 'weighted' estimators are defined according to:

$$\hat{\sigma}^2_{p,PLB1} = \left(\frac{n-2}{n+2}\right) \hat{\sigma}^2_{p,LB1}; \quad \hat{\sigma}^2_{p,PLB2} = \left(\frac{n-1}{n+2}\right) \hat{\sigma}^2_{p,LB2}; \quad \hat{\sigma}^2_{p,PLB3} = \left(\frac{n}{n+2}\right) \hat{\sigma}^2_{p,LB3}. \tag{21}$$

The limiting Bayes estimators in Table 1 and the weighted estimators in (21) will be compared later with the $\mathcal{M}$-based estimators and the two-step estimators $\hat{\sigma}^2_{p,MLE}$ and $\hat{\sigma}^2_{p,MRE}$.

Notice also from (20) that the estimators $\tilde{\sigma}^2_{LB,i} = (n/(n-2))\, \hat{\sigma}_i^2$, 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 (21) have different flavors than the MLE of σ^2 given in (10) since, in the latter, we are picking out one particular estimator 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, an adaptive estimator utilizing the sub-models' limiting Bayes estimators $\tilde{\sigma}^2_{LB,i}$'s is

$$\hat{\sigma}^2_{p,ALB} = \tilde{\sigma}^2_{LB,\hat{M}} = \left(\frac{n}{n-2}\right) \sum_{i=1}^{p} I\{\hat{M} = i\} \hat{\sigma}^2_i. \tag{22}$$

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 $\hat{\sigma}^2_{p,MLE}$, from which the mean and variance functions of the other two estimators are obtained. The risk functions are then obtained in an obvious manner since the loss functions are quadratic.

3.3 Bayes Estimators of $\tau(t)$

First, note that the parametric function $\tau(t) = \Pr\{X_i \leq t\}$ is expressible in the form

$$\tau(t) = \prod_{i=1}^{p} \left[\Phi\left(\frac{t-\mu_i}{\sigma}\right)\right]^{M_i}, \tag{23}$$

where $(M_1, ..., M_p) \in \boldsymbol{M}$ with $M_i = I\{\mu = \mu_i\}$. For the prior distribution in (15), we obtain the prior mean of $\tau(t)$.

Proposition 3.1 *For the prior distribution of* (μ, σ^2) *in (15), if* $\kappa > 1$*, the prior mean of* $\tau(t) = \Phi((t-\mu)/\sigma)$ *is* $E\{\tau(t)\} = \sum_{i=1}^{p} \tilde{\theta}_i \ \mathcal{T}\left(\sqrt{\kappa-1}(t-\mu_i)/\sqrt{\beta}; 2(\kappa-1)\right)$.

Proof: Using the representation in (23) of $\tau(t)$, we have

$$\begin{aligned} E\{\tau(t)\} &= \sum_{\mathbf{m}\in\mathbf{M}} \int_0^\infty \left\{\prod_{i=1}^{p}\left[\Phi\left(\frac{t-\mu_i}{\sigma}\right)^{m_i}\right]\left[\prod_{i=1}^{p}\tilde{\theta}_i^{m_i}\right]\frac{\beta^{\kappa-1}}{\Gamma(\kappa-1)}\left(\frac{1}{\sigma^2}\right)^{\kappa} e^{-\beta/\sigma^2}\right\} d\sigma^2 \\ &= \sum_{i=1}^{p}\tilde{\theta}_i \int_0^\infty \Phi\left(\frac{t-\mu_i}{\sigma}\right)\frac{\beta^{\kappa-1}}{\Gamma(\kappa-1)}\left(\frac{1}{\sigma^2}\right)^{\kappa} e^{-\beta/\sigma^2} d\sigma^2. \end{aligned}$$

The result follows immediately upon applying Lemma 3.1 below, whose proof is straightforward and hence omitted. ||

Lemma 3.1 *If* $\kappa > 1$*,* $\beta > 0$*, and* $\mu \in \Re$*, then*

$$\int_0^\infty \Phi\left(\frac{t-\mu}{\sigma}\right)\frac{\beta^{\kappa-1}}{\Gamma(\kappa-1)}\left(\frac{1}{\sigma^2}\right)^{\kappa} e^{-\beta/\sigma^2} d\sigma^2 = \mathcal{T}\left(\frac{\sqrt{\kappa-1}(t-\mu)}{\sqrt{\beta}}; 2(\kappa-1)\right).$$

The Bayes estimator of $\tau(t)$ with respect to the loss function in (5) is the posterior mean of $\tau(t)$, with respect to the posterior distribution of (μ, σ^2) given in (16):

$$
\begin{aligned}
E\{\tau(t) \mid \boldsymbol{x}\} &= E\left\{\prod_{i=1}^{p}\left[\Phi\left(\frac{t-\mu_i}{\sigma}\right)\right]^{M_i} \mid \boldsymbol{x}\right\} \\
&= C\sum_{m\in\mathbf{M}}\int_0^{\infty}\left\{\prod_{i=1}^{p}\left[\Phi\left(\frac{t-\mu_i}{\sigma}\right)\right]^{m_i}\right\}\left\{\prod_{i=1}^{p}\left[\tilde{\theta}_i\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\kappa}\exp\left(-\frac{1}{\sigma^2}\left[\frac{n}{2}\hat{\sigma}_i^2+\beta\right]\right)\right]^{m_i}\right\}d\sigma^2 \\
&= \sum_{i=1}^{p}\theta_i(\kappa,\beta,n,\mathbf{x})\int_0^{\infty}\Phi\left(\frac{t-\mu_i}{\sigma}\right)\frac{\beta_i^{\kappa^*-1}}{\Gamma(\kappa^*-1)}\left(\frac{1}{\sigma^2}\right)^{\kappa^*}\exp\left(-\frac{1}{\sigma^2}\beta_i\right)\,d\sigma^2
\end{aligned}
$$

where $\kappa^* = \kappa + n/2$ and $\beta_i = (n/2)\hat{\sigma}_i^2 + \beta, i = 1, 2, ..., p$. Applying Lemma 3.1, this becomes

$$
E\{\tau(t) \mid \boldsymbol{x}\} = \sum_{i=1}^{p}\theta_i(\kappa,\beta,n,\mathbf{x})\mathcal{T}\left(\frac{\sqrt{\kappa^*-1}(t-\mu_i)}{\sqrt{\beta_i}};2(\kappa^*-1)\right).
$$

The Bayes estimator of $\tau(t) = \Phi((t-\mu)/\sigma)$ under $\mathcal{M}_p$ for the prior distribution in (15) is therefore

$$
\hat{\tau}_{p,Bayes}(t;\kappa,\beta,\boldsymbol{\theta}) = \sum_{i=1}^{p}\theta_i(\kappa,\beta,n,\boldsymbol{x})\mathcal{T}\left(\frac{\sqrt{\kappa-1+n/2}(t-\mu_i)}{\sqrt{(n/2)\hat{\sigma}_i^2+\beta}};\ 2\,(\kappa-1+n/2)\right). \tag{24}
$$

Specializing to $\tilde{\theta}_i = 1/p, i = 1, 2, ..., p$, $\beta \to 0$, and $\kappa \in \{1, 3/2, 2, 3\}$, the four choices considered in Table 1, we obtain the limiting Bayes estimators of $\tau(t)$ given in the following table.

Table 2: Limiting Bayes estimators of $\tau(t)$ for different values of κ when $\tilde{\theta}_i = 1/p$ and $\beta \to 0$.

κ	Limiting Bayes Estimator of $\tau(t)$ $\hat{\tau}_{p,LBk}(t), k = 1, 2, 3, 4$
1	$\hat{\tau}_{p,LB1}(t) = \sum_{i=1}^{p}\theta_{i1}\mathcal{T}\left(\frac{t-\mu_i}{\hat{\sigma}_i};n\right)$
3/2	$\hat{\tau}_{p,LB2}(t) = \sum_{i=1}^{p}\theta_{i2}\mathcal{T}\left(\sqrt{\frac{n+1}{n}}\left(\frac{t-\mu_i}{\hat{\sigma}_i}\right);n+1\right)$
2	$\hat{\tau}_{p,LB3}(t) = \sum_{i=1}^{p}\theta_{i3}\mathcal{T}\left(\sqrt{\frac{n+2}{n}}\left(\frac{t-\mu_i}{\hat{\sigma}_i}\right);n+2\right)$
3	$\hat{\tau}_{p,LB4}(t) = \sum_{i=1}^{p}\theta_{i4}\mathcal{T}\left(\sqrt{\frac{n+4}{n}}\left(\frac{t-\mu_i}{\hat{\sigma}_i}\right);n+4\right)$

4 Comparisons of σ^2 Estimators

From Sections 2 and 3, we obtained competing estimators of σ^2. For clarity, we re-enumerate some of these estimators. The first two estimators, developed under $\mathcal{M}$, are

$$
\hat{\sigma}^2_{UMVU} = \frac{1}{n-1}\sum_{j=1}^{n}(X_j-\bar{X})^2 \quad \text{and} \quad \hat{\sigma}^2_{MRE} = \frac{n-1}{n+1}\hat{\sigma}^2_{UMVU},
$$

while the other estimators were developed under the model $\mathcal{M}_p$. With $\hat{M} = \arg\min_{1\le i\le p} |\bar{X} - \mu_i|$ and $\hat{\sigma}_i^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \mu_i)^2$, we have the estimators:

$$\hat{\sigma}^2_{p,MLE} = \sum_{i=1}^p I\{\hat{M} = i\}\hat{\sigma}_i^2; \qquad \hat{\sigma}^2_{p,MRE} = \tfrac{n}{n+2}\hat{\sigma}^2_{p,MLE}; \qquad \hat{\sigma}^2_{p,ALB} = \tfrac{n}{n-2}\hat{\sigma}^2_{p,MLE}.$$

The limiting Bayes ($\hat{\sigma}^2_{p,LBk}$'s) and the weighted estimators ($\hat{\sigma}^2_{p,PLBk}$'s) are given in Table 1 and (21), respectively. The goal of this section is compare the performances of these estimators via their risk functions arising from the loss function L_1 in (2). In particular, we address the following questions:

- How much loss in efficiency is incurred by using the estimators developed under the wider model $\mathcal{M}$ when model $\mathcal{M}_p$ holds?
- How do the limiting Bayes and weighted estimators $\hat{\sigma}^2_{p,LBk}$'s and $\hat{\sigma}^2_{p,PLBk}$'s compare with the $\mathcal{M}_p$ MLE-based and MRE-based estimators?
- 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?

4.1 Distributional Representations

It is well-known from distribution theory 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_2 in (2) is $R(\hat{\sigma}^2_{UMVU}, (\mu, \sigma^2)) = 2/(n-1)$. 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 $R(\hat{\sigma}^2_{MRE}, (\mu, \sigma^2)) = 2/(n+1)$. This demonstrates that $\hat{\sigma}^2_{UMVU}$ is inadmissible, a well-known result. For purposes of comparing the performance of the estimators, we will use $\hat{\sigma}^2_{UMVU}$ as the baseline, so the efficiency of an estimator $\hat{\sigma}^2$ will be given by

$$\text{Eff}(\hat{\sigma}^2 : \hat{\sigma}^2_{UMVU}) = \frac{R(\hat{\sigma}^2_{UMVU}, (\mu, \sigma^2))}{R(\hat{\sigma}^2, (\mu, \sigma^2))}. \tag{25}$$

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

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 $\hat{\sigma}^2_{p,LBk}$'s and $\hat{\sigma}^2_{p,PLBk}$'s. For notation, let $Z \sim N(0,1)$ and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)' \sim N_n(\mathbf{0}, \boldsymbol{I})$. For the vector of means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ with μ_{i_0} being the true mean with $i_0 \in \{1, 2, \ldots, p\}$, we let

$$\boldsymbol{\Delta} = \frac{\boldsymbol{\mu} - \mu_{i_0}\mathbf{1}}{\sigma} \tag{26}$$

where $\mathbf{1} = (1, 1, \dots, 1)'$. Note that this will always have a zero component under $\mathcal{M}_p$. In the sequel, the 'equal-in-distribution' relation is denoted by '$\stackrel{d}{=}$'.

Proposition 4.1 *Under $\mathcal{M}_p$ with μ_{i_0} the true mean, $n\hat{\sigma}_i^2/\sigma^2 \stackrel{d}{=} W + V_i^2, i = 1, 2, \dots, p$, where $W \sim \chi^2_{n-1}$, $\boldsymbol{V} = (V_1, V_2, \dots, V_p)' \sim N_p(-\sqrt{n}\boldsymbol{\Delta}, \boldsymbol{J} \equiv \mathbf{1}\mathbf{1}')$, and W and $\boldsymbol{V}$ are stochastically independent.*

Proof: We have

$$\begin{aligned} n\hat{\sigma}_i^2/\sigma^2 &= \|\boldsymbol{X} - \mu_i \mathbf{1}/\sigma\|^2 = \|[(\boldsymbol{X} - \mu_{i_0}\mathbf{1}) - (\mu_i - \mu_{i_0})\mathbf{1}]/\sigma\|^2 \\ &\stackrel{d}{=} \|\boldsymbol{Z} - \Delta_i \mathbf{1}\|^2 = \|(\boldsymbol{Z} - \bar{Z}\mathbf{1}) + (\bar{Z} - \Delta_i)\mathbf{1}\|^2 \quad \text{with} \quad \bar{Z} = (\boldsymbol{Z}\mathbf{1}')/n \\ &= \|\boldsymbol{Z} - \bar{Z}\mathbf{1}\|^2 + (\bar{Z} - \Delta_i)^2\|\mathbf{1}\|^2 = \|\boldsymbol{Z} - \bar{Z}\mathbf{1}\|^2 + n(\bar{Z} - \Delta_i)^2 \quad \text{since } \|\mathbf{1}\|^2 = n. \end{aligned}$$

Letting $W = \|\boldsymbol{Z} - \bar{Z}\mathbf{1}\|^2$ and $V_i = \sqrt{n}(\bar{Z} - \Delta_i), i = 1, 2, \dots, p$, from normal distribution theory, it follows that $W \sim \chi^2_{n-1}$ with W and $\boldsymbol{V} = (V_1, V_2, \dots, V_p)'$ stochastically independent. Furthermore, since $\bar{Z} \sim N(0, n^{-1})$, $\boldsymbol{V}$ has representation

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

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

Corollary 4.1 *Under the conditions of Proposition 4.1, $n\hat{\sigma}_i^2/\sigma^2 \stackrel{d}{=} W(1 + T_i^2), i = 1, 2, \dots, p$, with $\boldsymbol{T} = (T_1, \dots, T_p)' = \boldsymbol{V}/\sqrt{W}$. 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 $C_n = \sqrt{n/2}\,[\Gamma((n-2)/2)/\Gamma((n-1)/2)]$.

Proof: The first part follows from Proposition 4.1. The mean vector and the covariance matrix of $\boldsymbol{T}$ are derived using the independence between W and $\boldsymbol{V}$, the iterated covariance rule, and the formula $\mathbf{E}\{W^{k/2}\} = (1/2)^{-k/2}\,[\Gamma((n+k-1)/2)/\Gamma((n-1)/2)]$ which holds for any k such that $n + k - 1 > 0$. ‖

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

A distributional representation of the estimator $\hat{\sigma}^2_{p,MLE}$ will now be provided which will enable us to obtain the exact expressions for the mean and variance, and hence the risk. For a given $\boldsymbol{\Delta}$, let $\Delta_{(1)} < \Delta_{(2)} < \dots < \Delta_{(p)}$ denote the associated ordered values.

Theorem 4.1 *Let μ_{i_0} be the true mean. Then under $\mathcal{M}_p$,*

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

where, under the convention that $\Delta_{(0)} = -\infty$ and $\Delta_{(p+1)} = +\infty$, $L(\Delta_{(i)}, \mathbf{\Delta}) = (\sqrt{n}/2)\left[\Delta_{(i)} + \Delta_{(i-1)}\right]$ and $U(\Delta_{(i)}, \mathbf{\Delta}) = (\sqrt{n}/2)\left[\Delta_{(i)} + \Delta_{(i+1)}\right]$, $W \sim \chi^2_{n-1}$, $Z \sim N(0,1)$, and W and Z are stochastically independent.

Proof: We have

$$\frac{\hat{\sigma}^2_{p,MLE}}{\sigma^2} = \sum_{i=1}^{p} I\{\hat{M} = i\}\frac{\hat{\sigma}^2_i}{\sigma^2} \stackrel{d}{=} \frac{1}{n}\left[W + \sum_{i=1}^{p} I\{\hat{M} = i\}V_i^2\right] \quad \text{by Prop. 4.1.}$$

Now,

$$\begin{aligned}
\{\hat{M} = i\} &= \left\{\frac{\hat{\sigma}^2_i}{\sigma^2} < \frac{\hat{\sigma}^2_j}{\sigma^2},\ j = 1,2,\ldots,p; j \neq i\right\} \\
&\stackrel{d}{=} \left\{V_i^2 < V_j^2,\ j = 1,2,\ldots,p; j \neq i\right\} \\
&= \left\{(Z - \sqrt{n}\Delta_i)^2 < (Z - \sqrt{n}\Delta_j)^2,\ j = 1,2,\ldots,p; j \neq i\right\} \quad \text{by (26)} \\
&= \left\{(\Delta_j - \Delta_i)Z < \frac{\sqrt{n}}{2}(\Delta_j + \Delta_i)(\Delta_j - \Delta_i),\ j = 1,2,\ldots,p; j \neq i\right\} \\
&= \left\{Z < \frac{\sqrt{n}}{2}(\Delta_j + \Delta_i), \forall(j \neq i)(\Delta_j > \Delta_i)\right\} \cap \left\{Z > \frac{\sqrt{n}}{2}(\Delta_j + \Delta_i), \forall(j \neq i)(\Delta_j < \Delta_i)\right\} \\
&= \left\{\frac{\sqrt{n}}{2}\left(\Delta_i + \max_{\{j:\ \Delta_j < \Delta_i\}} \Delta_j\right) < Z < \frac{\sqrt{n}}{2}\left(\Delta_i + \min_{\{j:\ \Delta_j > \Delta_i\}} \Delta_j\right)\right\}.
\end{aligned}$$

But,

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

This completes the proof of the theorem. ||

Define the events $\Omega_{(i)} = \left\{L(\Delta_{(i)}, \mathbf{\Delta}) < Z < U(\Delta_{(i)}, \mathbf{\Delta})\right\}, i = 1,2,\ldots,p$. 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 4.1. Since the former is a partition, so is the latter, and therefore the sets $\Omega_{(i)}, i = 1, 2, \ldots, p$, are disjoint. From the representation in Theorem 4.1, we can now obtain expressions of the mean and variance of $\hat{\sigma}^2_{p,MLE}$. For $i = 1, 2, \ldots, p$, we let

$$P_{(i)}(\mathbf{\Delta}) \equiv \Pr\{\Omega_{(i)}\} = \Phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i+1)})\right) - \Phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i-1)})\right), \tag{28}$$

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 4.2 *Under the conditions of Theorem 4.1,*

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

Proof: From the representation in Theorem 4.1, we obtain

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

where we used the fact that for $a < b$, $\int_a^b z\phi(z)dz = \phi(a) - \phi(b)$. ‖

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

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

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

$$\zeta_{(i)}(m) \equiv \mathbf{E}\left\{I(\Omega_{(i)})(Z - \sqrt{n}\Delta_{(i)})^m\right\} = \sum_{k=0}^{m}(-1)^{(m-k)}\binom{m}{k}\left(\sqrt{n}\Delta_{(i)}\right)^{(m-k)}\xi(k; \Omega_{(i)}). \tag{29}$$

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

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

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

$$\xi(k;\Omega_{(i)}) = \int_{-\infty}^{U(\Delta_{(i)},\mathbf{\Delta})} z^k\phi(z)dz - \int_{-\infty}^{L(\Delta_{(i)},\mathbf{\Delta})} z^k\phi(z)dz. \tag{30}$$

Theorem 4.3 *Under the conditions of Theorem 4.1,*

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

Proof: From the representation in Theorem 4.1, by virtue of the independence between W and Z, and the fact that the events $\Omega_{(i)}$'s depend only on Z, we have

$$\text{VpMLE}(\mathbf{\Delta}) = \frac{1}{n^2}\left\{\mathbf{Var}(W) + \mathbf{Var}\left[\sum_{i=1}^p I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2\right]\right\}.$$

We already know that $\mathbf{Var}(W) = 2(n-1)$. On the other hand, we have

$$\mathbf{Var}\left\{\sum_{i=1}^p I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2\right\} = \sum_{i=1}^p \mathbf{Var}\left\{I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2\right\} + \sum_{i\neq j}\mathbf{Cov}\left\{I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2, I(\Omega_{(j)})(Z-\sqrt{n}\Delta_{(j)})^2\right\}.$$

However, by the computational form of the variance and the definition of $\zeta_{(i)}(k)$, we have

$$\mathbf{Var}\left\{I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2\right\} = \zeta_{(i)}(4) - [\zeta_{(i)}(2)]^2;$$

while since $I(\Omega_{(i)})I(\Omega_{(j)}) = 0$ whenever $i \neq j$, the computational form of the covariance leads to $\mathbf{Cov}\left\{I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2, I(\Omega_{(j)})(Z-\sqrt{n}\Delta_{(j)})^2\right\} = -\zeta_{(i)}(2)\zeta_{(j)}(2)$. Consequently,

$$\begin{aligned}\mathbf{Var}\left\{\sum_{i=1}^p I(\Omega_{(i)})(Z-\sqrt{n}\Delta_{(i)})^2\right\} &= \sum_{i=1}^p\left[\zeta_{(i)}(4) - [\zeta_{(i)}(2)]^2\right] + \sum_{i\neq j}\left[-\zeta_{(i)}(2)\zeta_{(j)}(2)\right] \\ &= \sum_{i=1}^p \zeta_{(i)}(4) - \left\{\sum_{i=1}^p \zeta_{(i)}(2)^2 + \sum_{i\neq j}\zeta_{(i)}(2)\zeta_{(j)}(2)\right\} = \sum_{i=1}^p \zeta_{(i)}(4) - \left(\sum_{i=1}^p \zeta_{(i)}(2)\right)^2. \, \|\end{aligned}$$

When there are only two sub-models we obtain more simplified forms of the mean and variance functions of $\hat{\sigma}^2_{p,MLE}/\sigma^2$.

Corollary 4.2 *If* $p = 2$ *so that* $\mathbf{\Delta} = (0,\Delta)$, *then under the conditions of Theorem 4.1*

$$EpMLE(\Delta) = 1 - \left(\frac{2}{\sqrt{n}}|\Delta|\right)\left\{\phi\left(\frac{\sqrt{n}}{2}|\Delta|\right) - \left(\frac{\sqrt{n}}{2}|\Delta|\right)\left[1-\Phi\left(\frac{\sqrt{n}}{2}|\Delta|\right)\right]\right\};$$

and

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

Proof: By symmetry it suffices to consider $\Delta \geq 0$. For this case, we have $\Omega_{(1)} = (-\infty, \sqrt{n}\Delta/2]$ and $\Omega_{(2)} = (\sqrt{n}\Delta/2, \infty)$. The result for the mean follows immediately from the general formula in Theorem 4.2. To prove the variance result, from (29) we obtain, for $m \geq 1$, that $\zeta_{(1)}(m) = \int_{-\infty}^{\frac{\sqrt{n}}{2}\Delta} z^m \phi(z)dz$ and $\zeta_{(2)}(m) = \int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z^m\phi(z)dz + \sum_{k=0}^{m-1}(-1)^{m-k}\binom{m}{k}(\sqrt{n}\Delta)^{m-k}\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z^k\phi(z)dz$. Consequently, $\sum_{i=1}^{2}\zeta_{(i)}(m) = \int_{-\infty}^{\infty} z^m\phi(z)dz + \sum_{k=0}^{m-1}(-1)^{m-k}\binom{m}{k}(\sqrt{n}\Delta)^{m-k}\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z^k\phi(z)dz$. In particular,

$$\sum_{i=1}^{2}\zeta_{(i)}(2) = 1 + (\sqrt{n}\Delta)^2\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty}\phi(z)dz - 2(\sqrt{n}\Delta)\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z\phi(z)dz;$$

while

$$\sum_{i=1}^{2}\zeta_{(i)}(4) = 3 + (\sqrt{n}\Delta)^4\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty}\phi(z)dz - 4(\sqrt{n}\Delta)^3\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z\phi(z)dz + 6(\sqrt{n}\Delta)^2\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z^2\phi(z)dz - 4(\sqrt{n}\Delta)\int_{\frac{\sqrt{n}}{2}\Delta}^{\infty} z^3\phi(z)dz.$$

By substitution of these expressions into the general formula for VpMLE($\boldsymbol{\Delta}$) in Theorem 4.3, we get the expression given in the statement of the corollary. ‖

We note from the expression in Corollary 4.2 that, for a fixed n, $\lim_{|\Delta|\to 0}$ EpMLE(Δ) $\to 1$ and $\lim_{|\Delta|\to 0}$ VpMLE(Δ) $\to 2/n$, the latter being the variance of the MLE of σ^2 under the true model. Also, for a fixed Δ, we see that $\lim_{n\to\infty}$ EpMLE(Δ) $\to 1$ and $\lim_{n\to\infty}\{n(\text{VpMLE}(\Delta))\} \to 2$.

The next result shows that even though the sub-models MLEs $\hat{\sigma}^2_{i,MLE}$'s are each unbiased for σ^2, the adaptive 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.

Corollary 4.3 *Under the conditions of Corollary 4.2 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.*

Proof: The result will follow from Corollary 4.2 if we could show that the continuous function $g : \Re_+ \to \Re$ defined via $g(u) = \phi(u) - u[1 - \Phi(u)]$ is positive. However, the positivity of this function follows from the facts that $\lim_{u \downarrow 0} h(u) > 0$, $\lim_{u \to \infty} h(u) = 0$, and because $g'(u) = \phi'(u) + u\phi(u) - [1 - \Phi(u)] = -[1 - \Phi(u)] < 0$ since $\phi'(u) = -u\phi(u)$. ‖

Corollary 4.4 *Under $\mathcal{M}_p$ and loss function L_1 in (2), the risk function of $\hat{\sigma}^2_{p,MLE}$ is*

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

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.

Theorem 4.4 *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*

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

and

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

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

Proof: First, recall that

$$\begin{aligned} \text{EpMLE}(\mathbf{\Delta}) \quad = \quad & 1 + \frac{2}{\sqrt{n}} \sum_{i=1}^{p} \Delta_{(i)} \left[\phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i+1)})\right) - \phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i-1)})\right)\right] + \\ & \sum_{i=1}^{p} \Delta^2_{(i)} \left[\Phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i+1)})\right) - \Phi\left(\frac{\sqrt{n}}{2}(\Delta_{(i)} + \Delta_{(i-1)})\right)\right]. \end{aligned}$$

With $\Delta^*_{(i)} = \frac{\sqrt{n}}{2}\Delta_{(i)}$, this becomes

$$\begin{aligned} \text{EpMLE}(\mathbf{\Delta}) \quad = \quad & 1 + \frac{4}{n} \sum_{i=1}^{p} \Delta^*_{(i)} \left[\phi\left((\Delta^*_{(i)} + \Delta^*_{(i+1)})\right) - \phi\left((\Delta^*_{(i)} + \Delta^*_{(i-1)})\right)\right] + \\ & \frac{4}{n} \sum_{i=1}^{p} [\Delta^*_{(i)}]^2 \left[\Phi\left((\Delta^*_{(i)} + \Delta^*_{(i+1)})\right) - \Phi\left((\Delta^*_{(i)} + \Delta^*_{(i-1)})\right)\right]. \end{aligned}$$

Since $\Delta^*_{(0)} = -\infty$ and $\Delta^*_{(p+1)} = \infty$, we have

$$\begin{aligned}
&\sum_{i=1}^{p} \Delta^*_{(i)} \left[\phi(\Delta^*_{(i)} + \Delta^*_{(i+1)}) - \phi(\Delta^*_{(i)} + \Delta^*_{(i-1)})\right] \\
&= \Delta^*_{(1)}\phi(\Delta^*_{(1)} + \Delta^*_{(2)}) - \Delta^*_{(p)}\phi(\Delta^*_{(p)} + \Delta^*_{(p-1)}) + \\
&\quad \sum_{i=2}^{p-1} \Delta^*_{(i)} \left[\phi(\Delta^*_{(i)} + \Delta^*_{(i+1)}) - \phi(\Delta^*_{(i)} + \Delta^*_{(i-1)})\right] \\
&= \Delta^*_{(1)} \left[\phi(2\Delta^*_{(1)} + (\Delta^*_{(2)} - \Delta^*_{(1)})) - \phi(2\Delta^*_{(1)})\right] + \Delta^*_{(1)}\phi(2\Delta^*_{(1)}) - \\
&\quad \Delta^*_{(p)} \left[\phi(2\Delta^*_{(p)} - (\Delta^*_{(p)} - \Delta^*_{(p-1)})) - \phi(2\Delta^*_{(p)})\right] - \Delta^*_{(p)}\phi(2\Delta^*_{(p)}) + \\
&\quad \sum_{i=2}^{p-1} \Delta^*_{(i)} \left[\phi(2\Delta^*_{(i)} + (\Delta^*_{(i+1)} - \Delta^*_{(i)})) - \phi(2\Delta^*_{(i)} - (\Delta^*_{(i)} - \Delta^*_{(i-1)}))\right] \\
&= \Delta^*_{(1)}[2\phi'(2\Delta^*_{(1)}) + o(1)](\Delta^*_{(2)} - \Delta^*_{(1)}) + \Delta^*_{(1)}\phi(2\Delta^*_{(1)}) + \\
&\quad \Delta^*_{(p)}[2\phi'(2\Delta^*_{(p)}) + o(1)](\Delta^*_{(p)} - \Delta^*_{(p-1)}) - \Delta^*_{(p)}\phi(2\Delta^*_{(p)}) + \\
&\quad \sum_{i=2}^{p-1} \Delta^*_{(i)}[2\phi'(2\Delta^*_{(i)}) + o(1)](\Delta^*_{(i+1)} - \Delta^*_{(i-1)})
\end{aligned}$$

with the last equality holding as $p \to \infty$ with $\max_{2 \le i \le p} |\Delta^*_{(i)} - \Delta^*_{(i-1)}| \to 0$ and because for a differentiable function $g(\cdot)$, as $\max\{h_1, h_2\} \to 0$, $g(2w + h_1) - g(2w) = [2g'(2w) + o(1)]h_1$ and $g(2w + h_2) - g(2w - h_1) = [2g'(2w) + o(1)](h_2 + h_1)$. Taking the limit as $p \to \infty$ and using the conditions that $\Delta_{(1)} \to \Delta_{min}$ and $\Delta_{(p)} \to \Delta_{max}$, we therefore find that

$$\begin{aligned}
&\sum_{i=1}^{p} \Delta^*_{(i)} \left[\phi(\Delta^*_{(i)} + \Delta^*_{(i+1)}) - \phi(\Delta^*_{(i)} + \Delta^*_{(i-1)})\right] \\
&\to \int_{\frac{\sqrt{n}}{2}\Delta_{min}}^{\frac{\sqrt{n}}{2}\Delta_{max}} w2\phi'(2w)dw + \left(\frac{\sqrt{n}}{2}\Delta_{min}\right)\phi(\sqrt{n}\Delta_{min}) - \left(\frac{\sqrt{n}}{2}\Delta_{max}\right)\phi(\sqrt{n}\Delta_{max}) \\
&= \frac{1}{2}\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} w\phi'(w)dw + \frac{\sqrt{n}}{2}\{\Delta_{min}\phi(\sqrt{n}\Delta_{min}) - \Delta_{max}\phi(\sqrt{n}\Delta_{max})\} \\
&= -\frac{1}{2}\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} w^2\phi(w)dw + \frac{\sqrt{n}}{2}\{\Delta_{min}\phi(\sqrt{n}\Delta_{min}) - \Delta_{max}\phi(\sqrt{n}\Delta_{max})\}
\end{aligned}$$

since $\int w\phi'(w)dw = -\int w^2\phi(w)dw$.

By an analogous approach, we also find that under the conditions of the theorem,

$$\begin{aligned}
&\sum_{i=1}^{p} (\Delta^*_{(i)})^2 \left[\Phi(\Delta^*_{(i)} + \Delta^*_{(i+1)}) - \Phi(\Delta^*_{(i)} + \Delta^*_{(i-1)})\right] \\
&\to \frac{1}{4}\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} w^2\phi(w)dw + \frac{n}{4}\left\{(\Delta_{min})^2\Phi(\sqrt{n}\Delta_{min}) + (\Delta_{max})^2[1 - \Phi(\sqrt{n}\Delta_{max})]\right\}.
\end{aligned}$$

Using these two intermediate results, it follows that under the conditions of the theorem,

$$\begin{aligned}\text{EpMLE}(\mathbf{\Delta}) \quad \to \quad & 1-\frac{1}{n}\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} w^2\phi(w)dw + \frac{2}{\sqrt{n}}\left\{\Delta_{min}\phi(\sqrt{n}\Delta_{min}) - \Delta_{max}\phi(\sqrt{n}\Delta_{max})\right\} + \\ & \left\{(\Delta_{min})^2\Phi(\sqrt{n}\Delta_{min}) + (\Delta_{max})^2[1-\Phi(\sqrt{n}\Delta_{max})]\right\}.\end{aligned}$$

This proves the first result about the mean. The second result for the mean when $\Delta_{min} \to -\infty$ and $\Delta_{max} \to \infty$ follows from this expression by noting that $\mathbf{E}(Z^2) = 1$.

To prove the result for the variance, define for $k \in \{0, 1, 2, \ldots\}$, $Q_k(z) = \int_{-\infty}^{z} w^k\phi(w)dw$ so that $Q_k'(z) = z^k\phi(z)$. We may therefore express $\xi(k; \Omega_{(i)})$ via

$$\xi(k;\Omega_{(i)}) = \int_{\Delta^*_{(i)}+\Delta^*_{(i-1)}}^{\Delta^*_{(i)}+\Delta^*_{(i+1)}} z^k\phi(z)dz = Q_k(\Delta^*_{(i)} + \Delta^*_{(i+1)}) - Q_k(\Delta^*_{(i)} + \Delta^*_{(i-1)}).$$

We consider the behavior of $\xi(k; \Omega_{(i)})$ as $p \to \infty$ for the three cases pertaining to $i = 1$, $i = p$, and $1 < i < p$. First, when $i = 1$, we have

$$\begin{aligned}\xi(k;\Omega_{(1)}) &= Q_k(\Delta^*_{(1)} + \Delta^*_{(2)}) - Q_k(-\infty) = Q_k(\Delta^*_{(1)} + \Delta^*_{(2)}) \\ &= [Q_k(2\Delta^*_{(1)} + (\Delta^*_{(2)} - \Delta^*_{(1)})) - Q_k(2\Delta^*_{(1)})] + Q_k(2\Delta^*_{(1)}) \\ &= [2Q_k'(2\Delta^*_{(1)}) + o(1)](\Delta^*_{(2)} - \Delta^*_{(1)}) + Q_k(2\Delta^*_{(1)}).\end{aligned}$$

Next, when $i = p$, we have

$$\begin{aligned}\xi(k;\Omega_{(p)}) &= Q_k(\Delta^*_{(p)} + \Delta^*_{(p+1)}) - Q_k(\Delta^*_{(p)} + \Delta^*_{(p-1)}) \\ &= [Q_k(\infty) - Q_k(2\Delta^*_{(p)})] - [Q_k(2\Delta^*_{(p)} - (\Delta^*_{(p)} - \Delta^*_{(p-1)})) - Q_k(2\Delta^*_{(p)})] \\ &= [\mathbf{E}(Z^k) - Q_k(2\Delta^*_{(p)})] + [2Q_k'(2\Delta^*_{(p)}) + o(1)](\Delta^*_{(p)} - \Delta^*_{(p-1)}).\end{aligned}$$

The final case is when $1 < i < p$. For this case we have

$$\begin{aligned}\xi(k;\Omega_{(i)}) &= Q_k(\Delta^*_{(i)} + \Delta^*_{(i+1)}) - Q_k(\Delta^*_{(i)} + \Delta^*_{(i-1)}) \\ &= Q_k(2\Delta^*_{(i)} + (\Delta^*_{(i+1)} - \Delta^*_{(i)})) - Q_k(2\Delta^*_{(i)} - (\Delta^*_{(i)} - \Delta^*_{(i-1)})) \\ &= [2Q_k'(2\Delta^*_{(i)}) + o(1)](\Delta^*_{(i+1)} - \Delta^*_{(i-1)}).\end{aligned}$$

Consequently, it follows from (29) that

$$\begin{aligned}\sum_{i=1}^{p}\zeta_{(i)}(m) &= \sum_{i=1}^{p}\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}(\sqrt{n}\Delta_{(i)})^{m-k}\xi(k;\Omega_{(i)}) \\ &= \sum_{k=0}^{m}(-1)^{m-k}2^{m-k}\binom{m}{k}\sum_{i=1}^{p}(\Delta^*_{(i)})^{m-k}\xi(k;\Omega_{(i)}).\end{aligned}$$

But now, by using the expressions obtained above for $\xi(k;\Omega_{(i)})$'s as $p\to\infty$, we have

$$\begin{aligned}\sum_{i=1}^{p}(\Delta^*_{(i)})^{m-k}\xi(k;\Omega_{(i)}) &= (\Delta^*_{(1)})^{m-k}\xi(k;\Omega_{(1)})+\sum_{i=2}^{p-1}(\Delta^*_{(i)})^{m-k}\xi(k;\Omega_{(i)})+(\Delta^*_{(p)})^{m-k}\xi(k;\Omega_{(p)})\\ &= (\Delta^*_{(1)})^{m-k}[2Q_k'(2\Delta^*_{(1)})+o(1)](\Delta^*_{(2)}-\Delta^*_{(1)})+(\Delta^*_{(1)})^{m-k}Q_k(2\Delta^*_{(1)})+\\ &\quad \sum_{i=2}^{p-1}(\Delta^*_{(i)})^{m-k}[2Q_k'(2\Delta^*_{(i)})+o(1)](\Delta^*_{(i+1)}-\Delta^*_{(i-1)})+\\ &\quad (\Delta^*_{(p)})^{m-k}[2Q_k'(2\Delta^*_{(p)})+o(1)](\Delta^*_{(p)}-\Delta^*_{(p-1)})+(\Delta^*_{(p)})^{m-k}[\mathbf{E}(Z^k)-Q_k(2\Delta^*_{(p)})].\end{aligned}$$

Under the conditions of the theorem, as $p\to\infty$, we obtain

$$\begin{aligned}\sum_{i=1}^{p}(\Delta^*_{(i)})^{m-k}\xi(k;\Omega_{(i)}) &\to \int_{\frac{\sqrt{n}}{2}\Delta_{min}}^{\frac{\sqrt{n}}{2}\Delta_{max}} w^{m-k}2Q_k'(2w)dw+\\ &\quad \left(\frac{\sqrt{n}}{2}\Delta_{min}\right)^{m-k}Q_k(\sqrt{n}\Delta_{min})+\left(\frac{\sqrt{n}}{2}\Delta_{max}\right)^{m-k}\left[\mathbf{E}(Z^k)-Q_k(\sqrt{n}\Delta_{max})\right]\\ &= \int_{\frac{\sqrt{n}}{2}\Delta_{min}}^{\frac{\sqrt{n}}{2}\Delta_{max}} w^{m-k}2(2w)^k\phi(2w)dw+\left(\frac{\sqrt{n}}{2}\right)^{m-k}\Big\{(\Delta_{min})^{m-k}Q_k(\sqrt{n}\Delta_{min})+\\ &\quad (\Delta_{max})^{m-k}[\mathbf{E}(Z^k)-Q_k(\sqrt{n}\Delta_{max})]\Big\}\\ &= \frac{1}{2^{m-k}}\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} z^m\phi(z)dz+\left(\frac{\sqrt{n}}{2}\right)^{m-k}\left\{(\Delta_{min})^{m-k}\int_{-\infty}^{\sqrt{n}\Delta_{min}} z^k\phi(z)dz+\right.\\ &\quad \left.(\Delta_{max})^{m-k}\int_{\sqrt{n}\Delta_{max}}^{\infty} z^k\phi(z)dz\right\}.\end{aligned}$$

Therefore, as $p\to\infty$,

$$\begin{aligned}\sum_{i=1}^{p}\zeta_{(i)}(m) &\to \left\{\sum_{k=0}^{m}(-1)^{m-k}2^{m-k}\binom{m}{k}2^{-(m-k)}\right\}\left\{\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} z^m\phi(z)dz\right\}+\\ &\quad \sum_{k=0}^{m}(-1)^{m-k}2^{m-k}\binom{m}{k}\left(\frac{\sqrt{n}}{2}\right)^{m-k}\times\\ &\quad \left\{(\Delta_{min})^{m-k}\int_{-\infty}^{\sqrt{n}\Delta_{min}} z^k\phi(z)dz+(\Delta_{max})^{m-k}\int_{\sqrt{n}\Delta_{max}}^{\infty} z^k\phi(z)dz\right\}\\ &= (1-1)^m\int_{\sqrt{n}\Delta_{min}}^{\sqrt{n}\Delta_{max}} z^m\phi(z)dz+\int_{-\infty}^{\sqrt{n}\Delta_{min}}\left[\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}(\sqrt{n}\Delta_{min})^{m-k}z^k\right]\phi(z)dz+\\ &\quad \int_{\sqrt{n}\Delta_{max}}^{\infty}\left[\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}(\sqrt{n}\Delta_{max})^{m-k}z^k\right]\phi(z)dz\\ &= \int_{-\infty}^{\sqrt{n}\Delta_{min}}(z-\sqrt{n}\Delta_{min})^m\phi(z)dz+\int_{\sqrt{n}\Delta_{max}}^{\infty}(z-\sqrt{n}\Delta_{max})^m\phi(z)dz\\ &= \mathbf{E}\left\{(Z-\sqrt{n}\Delta_{min})^m I(Z<\sqrt{n}\Delta_{min})\right\}+\mathbf{E}\left\{(Z-\sqrt{n}\Delta_{max})^m I(Z>\sqrt{n}\Delta_{max})\right\}.\end{aligned}$$

Using the general expression of VpMLE($\mathbf{\Delta}$) in Theorem 4.3 and substituting the limiting result of $\sum_{i=1}^{p} \zeta_{(i)}(m)$ obtained above for $m = 4$ and $m = 2$, the limiting expression given in the statement of the theorem for VpMLE($\mathbf{\Delta}$) is obtained. The second result concerning the variance as $\Delta_{min} \to -\infty$ and $\Delta_{max} \to \infty$ follows since the terms involving expectations all converge to zeros. This completes the proof of the theorem. $\|$

Corollary 4.5 *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 $\hat{\sigma}^2_{UMVU}$.

Proof: From Theorem 4.4,

$$\begin{aligned} R\left(\hat{\sigma}^2_{p,MLE}, (\mu_{i_0}, \sigma^2)\right) &= \text{VpMLE}(\mathbf{\Delta}) + [\text{EpMLE}(\mathbf{\Delta}) - 1]^2 \\ &\to (2/n)(1 - 1/n) + [(1 - 1/n) - 1]^2 = (2n-1)/n^2. \end{aligned}$$

Also, since $\hat{\sigma}^2_{p,MRE} = (n/(n+2))\, \hat{\sigma}^2_{p,MLE}$, we have

$$\begin{aligned} R\left(\hat{\sigma}^2_{p,MRE}, (\mu_{i_0}, \sigma^2)\right) &= \left(\frac{n}{n+2}\right)^2 \text{VpMLE}(\mathbf{\Delta}) + \left[\left(\frac{n}{n+2}\right) \text{EpMLE}(\mathbf{\Delta}) - 1\right]^2 \\ \to \quad & \left(\frac{n}{n+2}\right)^2 \left(\frac{2}{n}\right)\left(1 - \frac{1}{n}\right) + \left[\left(\frac{n}{n+2}\right)\left(1 - \frac{1}{n}\right) - 1\right]^2 = \frac{2n+7}{(n+2)^2}. \end{aligned}$$

The efficiency expressions relative to $\hat{\sigma}^2_{UMVU}$ of $\hat{\sigma}^2_{p,MLE}$ and $\hat{\sigma}^2_{p,MRE}$ are then 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 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 $p \to \infty$, $\hat{\sigma}^2_{MRE}$ 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 $n > 2$. $\|$

The fourth result in Corollary 4.5 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 adaptive 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 4.2, 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 adaptive estimator developed under model $\mathcal{M}_p$. These results are in accordance 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 also seem able to adapt well and their performance seems not degraded by an increase in the number of sub-models. We point out however that it is conceivable that there might exist a better two-step estimator than $\hat{\sigma}^2_{p,MRE}$ derivable under $\mathcal{M}_p$, which dominates $\hat{\sigma}^2_{MRE}$ for *all* p. We shall defer the exploration of this intriguing possibility to future research.

4.1.2 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

$$\hat{\sigma}^2_{p,LB} \equiv \hat{\sigma}^2_{p,LB}(\alpha) = \sum_{i=1}^{p} \left\{ \frac{(\hat{\sigma}^2_i)^{-\alpha}}{\sum_{j=1}^{p} (\hat{\sigma}^2_j)^{-\alpha}} \right\} \hat{\sigma}^2_i. \tag{31}$$

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 4.5 *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 $\boldsymbol{\Delta}$.

Proof: Starting from (31) and using Corollary 4.1, we obtain

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

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

From the distributional representation in Theorem 4.5, a closed-form expression for the risk function of $\hat{\sigma}^2_{p,LB}$ will be difficult to obtain because of the adaptive nature of the mixing probabilities $\theta_i(\boldsymbol{T})$'s 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

$$\frac{\hat{\sigma}^2_{p,LB}}{\sigma^2} \stackrel{d}{\approx} \frac{W}{n}\left\{1 + H + \boldsymbol{H}^{(1)\prime}(\boldsymbol{T}-\boldsymbol{\nu}) + \frac{1}{2}(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})\right\}. \tag{32}$$

Lemma 4.1 *Under the conditions of Proposition 4.1, (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)* $\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})(\boldsymbol{T}-\boldsymbol{\nu})'\} = \boldsymbol{J} + n\left(1/C_n^2 - 1 + 3/n\right)\boldsymbol{\nu}^{\otimes 2}$, *where the constant* C_n *is given in Corollary 4.1.*

Proof: Since $\mathbf{E}(\boldsymbol{T}|W) = \mathbf{E}(\boldsymbol{V}/\sqrt{W}|W) = \mathbf{E}(\boldsymbol{V})/\sqrt{W} = -\sqrt{n}\boldsymbol{\Delta}/\sqrt{W} = \sqrt{n}\boldsymbol{\nu}/(\sqrt{W}C_n)$, 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

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

$$\mathbf{E}\{\boldsymbol{T}\boldsymbol{\nu}'|W\} = -\frac{\sqrt{n}\boldsymbol{\Delta}\boldsymbol{\nu}'}{\sqrt{W}} = \frac{\sqrt{n}\boldsymbol{\nu}^{\otimes 2}}{\sqrt{W}C_n}.$$

Consequently,

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

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. ‖

Theorem 4.6 *Under $\mathcal{M}_p$, a second-order approximation to the mean of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is*

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

Proof: First, note that by using the fourth result in Lemma 4.1, we have

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

From (32), and using Lemma 4.1 and the preceding result, the approximation to the mean of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is immediately obtained to be

$$\begin{aligned}
E_2(\boldsymbol{\Delta}) &= \frac{1}{n}\left\{(1+H)\mathbf{E}(W)+\boldsymbol{H}^{(1)'}\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})\}+\frac{1}{2}\mathbf{E}\{W(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})\}\right\}\\
&= \frac{1}{n}\left\{(1+H)(n-1)-\boldsymbol{H}^{(1)'}\boldsymbol{\nu}+\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})+\frac{n}{2}\left(\frac{1}{C_n^2}-1+\frac{3}{n}\right)(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\right\},
\end{aligned}$$

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

To obtain second-order approximations 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

$$Q \equiv 1+H+\boldsymbol{H}^{(1)'}(\boldsymbol{T}-\boldsymbol{\nu})+\frac{1}{2}(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu}), \tag{33}$$

which is the second-order Taylor approximation of the function $1+H(\boldsymbol{T})$ (cf., equation (32)).

Lemma 4.2 *Under the conditions of Proposition 4.1,*

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

Proof: From the proof of and results in Lemma 4.1, it is immediate to see that

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

Consequently, from the first result in Lemma 4.1 and the preceding result,

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

Lemma 4.3 *Under the conditions of Proposition 4.1,*

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

Proof: First, we have that

$$\begin{aligned}&\mathbf{Var}\{\boldsymbol{H}^{(1)\prime}(\boldsymbol{T}-\boldsymbol{\nu})|W\} = \boldsymbol{H}^{(1)\prime}\mathbf{Cov}\{\boldsymbol{T}|W\}\boldsymbol{H}^{(1)} \\ &= \frac{1}{W}\boldsymbol{H}^{(1)\prime}\mathbf{Cov}(\boldsymbol{V})\boldsymbol{H}^{(1)} = \frac{1}{W}\boldsymbol{H}^{(1)\prime}\boldsymbol{J}\boldsymbol{H}^{(1)} = \frac{1}{W}(\boldsymbol{H}^{(1)\prime}\mathbf{1})^2.\end{aligned}$$

Next, we have

$$\begin{aligned}&\mathbf{Var}\{(\boldsymbol{T}-\boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T}-\boldsymbol{\nu})|W\} = \mathbf{Var}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T} - 2\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} \\ &= \mathbf{Var}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} + 4\mathbf{Var}\{\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} - 4\mathbf{Cov}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\}.\end{aligned}$$

But

$$\mathbf{Var}\{\boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} = \mathbf{Var}\left\{\frac{\boldsymbol{V}'}{\sqrt{W}}\boldsymbol{H}^{(2)}\frac{\boldsymbol{V}}{\sqrt{W}}|W\right\} = \frac{1}{W^2}\mathbf{Var}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}\}.$$

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

$$\begin{aligned}\mathbf{Var}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}\} &= \mathbf{Var}\{(Z\mathbf{1} - \sqrt{n}\boldsymbol{\Delta})'\boldsymbol{H}^{(2)}(Z\mathbf{1} - \sqrt{n}\boldsymbol{\Delta})\} \\ &= \mathbf{Var}\{(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})Z^2 - 2\sqrt{n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\Delta})Z\} \\ &= (\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2\mathbf{Var}(Z^2) + 4n(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\Delta})^2\mathbf{Var}(Z) - \\ &\quad 4\sqrt{n}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\Delta})\mathbf{Cov}(Z^2, Z) \\ &= 2(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2 + \frac{4n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\end{aligned}$$

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

$$\mathbf{Var}\{\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} = \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\mathbf{Cov}(\boldsymbol{T}|W)\tfrac{1}{W}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{J}\boldsymbol{H}^{(2)}\boldsymbol{\nu} = \tfrac{1}{W}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2$$

and

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

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

$$\begin{aligned}
&\mathbf{Cov}\{\boldsymbol{V}'\boldsymbol{H}^{(2)}\boldsymbol{V}, \boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{V}\} \\
&= \mathbf{Cov}\{(Z\mathbf{1} - \sqrt{n}\boldsymbol{\Delta})'\boldsymbol{H}^{(2)}(Z\mathbf{1} - \sqrt{n}\boldsymbol{\Delta}), \boldsymbol{\nu}'\boldsymbol{H}^{(2)}(Z\mathbf{1} - \sqrt{n}\boldsymbol{\Delta})\} \\
&= \mathbf{Cov}\{Z^2(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1}) - 2\sqrt{n}(\mathbf{1}\boldsymbol{H}^{(2)}\boldsymbol{\Delta})Z, (\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\mathbf{1})Z\} \\
&= (\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\mathbf{Cov}(Z^2, Z) - 2\sqrt{n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\Delta})(\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\mathbf{1})\mathbf{Var}(Z) \\
&= \frac{2\sqrt{n}}{C_n}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2,
\end{aligned}$$

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 intermediate results, we therefore obtain

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

We also have that

$$\begin{aligned}
&\mathbf{Cov}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T} - \boldsymbol{\nu}), (\boldsymbol{T} - \boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T} - \boldsymbol{\nu})|W\} \\
&= \mathbf{Cov}\{\boldsymbol{H}^{(1)'}\boldsymbol{T}, \boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T} - 2\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W\} \\
&= \boldsymbol{H}^{(1)'}\mathbf{Cov}(\boldsymbol{T}, \boldsymbol{T}'\boldsymbol{H}^{(2)}\boldsymbol{T}|W) - 2\boldsymbol{H}^{(1)'}\mathbf{Cov}(\boldsymbol{T}|W)\boldsymbol{H}^{(2)}\boldsymbol{\nu}.
\end{aligned}$$

But, once again,

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

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

$$\begin{aligned}
&\mathbf{Var}(Q|W) = \mathbf{Var}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T} - \boldsymbol{\nu}) + \frac{1}{2}(\boldsymbol{T} - \boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T} - \boldsymbol{\nu})|W\} \\
&= \mathbf{Var}\{\boldsymbol{H}^{(1)'}(\boldsymbol{T} - \boldsymbol{\nu})\} + \frac{1}{4}\mathbf{Var}\{(\boldsymbol{T} - \boldsymbol{\nu})'\boldsymbol{H}^{(2)}(\boldsymbol{T} - \boldsymbol{\nu})|W\} +
\end{aligned}$$

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

Theorem 4.7 *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*

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

and

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

Proof: From (32) and (33), and by the iterated variance rule, an approximate variance of $\hat{\sigma}^2_{p,LB}/\sigma^2$ is $V_2(\boldsymbol{\Delta}) \equiv \mathbf{Var}\left\{\frac{1}{n}WQ\right\} = \frac{1}{n}\left[\mathbf{Var}\left\{\frac{W}{\sqrt{n}}\mathbf{E}(Q|W)\right\}+\mathbf{E}\left\{\frac{W^2}{n}\mathbf{Var}(Q|W)\right\}\right]$. By Lemma 4.2,

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

$$= \frac{1}{n}\left\{\left[1+H-\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}+\frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right]^2\mathbf{Var}(W)+\right.$$
$$\frac{n}{C_n^2}\left(\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}-\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2\mathbf{Var}(\sqrt{W})+\frac{2\sqrt{n}}{C_n}\left[1+H-\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}+\frac{1}{2}\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right]\times$$
$$\left.\left(\boldsymbol{H}^{(1)\prime}\boldsymbol{\nu}-\boldsymbol{\nu}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)\mathbf{Cov}(W,\sqrt{W})\right\},$$

which, since $\mathbf{Var}(W) = 2(n-1)$, $\mathbf{Var}(\sqrt{W}) = (n-1)-(n-2)^2C_n^2/n$, and $\mathbf{Cov}(W,\sqrt{W}) = (n-2)C_n/\sqrt{n}$, yields the expression for $\mathrm{VE}(\boldsymbol{\Delta})$ given in the statement of the theorem. Also, from Lemma 4.3,

$$\mathrm{EV}(\boldsymbol{\Delta}) \equiv \mathbf{E}\left\{\frac{W^2}{n}\mathbf{Var}(Q|W)\right\}\frac{1}{n}\left\{\mathbf{E}(W)\left(\boldsymbol{H}^{(1)\prime}\mathbf{1}-\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)^2+\right.$$
$$\left.\frac{2\sqrt{n}}{C_n}\left(\boldsymbol{H}^{(1)\prime}\mathbf{1}-\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu}\right)(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})\mathbf{E}(\sqrt{W})+\frac{1}{2}(\mathbf{1}'\boldsymbol{H}^{(2)}\mathbf{1})^2+\frac{n}{C_n^2}(\mathbf{1}'\boldsymbol{H}^{(2)}\boldsymbol{\nu})^2\right\},$$

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. ‖

We now have the following second-order approximations to the risk functions of $\hat{\sigma}^2_{p,LB1}$. For the other limiting Bayes and weighted estimators in Table 1 and (21), analogous approximate risk expressions can be obtained similarly as for $\hat{\sigma}^2_{p,LB1}$.

Corollary 4.6 *Under $\mathcal{M}_p$ and the loss function L_1 in (2), a second-order approximations to the risk functions of $\hat{\sigma}^2_{p,LB1}$ is*

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

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 4.6 and Theorem 4.7, respectively.

Proof: Immediate from the quadratic nature of the loss function, Theorems 4.6 and 4.7, and the relationship of $\hat{\sigma}^2_{p,LB1}$ with $\hat{\sigma}^2_{p,LB}(\alpha)$. ‖

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 $\theta_i(\boldsymbol{T})$'s in Corollary 4.1, we find that for $j,k \in \{1,2,\ldots,p\}$,

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

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

$$\theta^{(2)}_{ikl}(\boldsymbol{T}) = (2\alpha)\left\{ I\{k = l\}\left(\frac{1 - T_k^2}{(1+T_k^2)^2}\right)\theta_k(\boldsymbol{T})[\theta_i(\boldsymbol{T}) - I\{k = i\}] + \left(\frac{T_k}{(1+T_k^2)^2}\right)\left[\theta^{(1)}_{kl}(\boldsymbol{T})[\theta_i(\boldsymbol{T}) - I\{k = i\}] + \theta_k(\boldsymbol{T})\theta^{(1)}_{il}(\boldsymbol{T})\right]\right\}.$$

4.1.3 Assessing the Second-Order Approximations via Simulation

In terms of practicality, it is expected that the approximations will not perform very well when n is small. To establish the appropriateness of the approximations for various values of n and choices of $\boldsymbol{\Delta}$, we compared the values for 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 approximation counterparts. The results are shown in Table 3. As expected, a similar pattern is present 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 worse-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. A very similar behavior is shown by variances and risks also.

One other observation from Table 3 is worth mentioning: 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. This is also supported by Figure 1, which compares the accuracy of the approximations for $n \in \{3, 10, 30\}$ in models with $p = 2$, as a function of the separation between the two means. When $n = 3$ the approximation performs badly regardless of the magnitude of the separation. 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.

4.2 Comparison of Relative Efficiencies of σ^2 Estimators

We now carry out the comparisons of the relative efficiencies of the variance estimators with respect to $\hat{\sigma}^2_{UMVU}$, using simulated datasets with a variety of $\boldsymbol{\Delta}$ and for $n \in \{3, 10, 30\}$. The results are summarized in Tables 4 and 5 and Figures 2 and 3.

Table 4 focuses in particular on the differences in the estimator relative efficiencies between symmetric and asymmetric $\boldsymbol{\Delta}$ cases. As can be seen from the table, there does not seem to be a visible effect of the asymmetry of $\mathbf{\Delta}$ on the estimators. In all $\mathbf{\Delta}$ 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 $\hat{\sigma}^2_{p,MRE}$ also performing decently. 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 5 is designed for the purpose of examining the impact of an increasing number of sub-models. Examining this table, we can see that there clearly is a negative effect on the two-step estimator $\hat{\sigma}^2_{p,MRE}$ to the extent that when p is large, this estimator becomes less efficient than the estimator $\hat{\sigma}^2_{MRE}$, an estimator developed under a wider and more general model. This result is consistent with the theoretical result of Corollary 4.5, which shows that the estimator $\hat{\sigma}^2_{p,MRE}$ becomes less efficient than the estimator $\hat{\sigma}^2_{MRE}$ when $p \to \infty$. We also point out that even for p as low as 33, the ratios of the relative efficiency values from Table 5 start to agree (to the third decimal) with the limiting relative efficiencies theoretically predicted by Corollary 4.5. The performances of the weighted estimators appear not to be affected much by the increasing p, and it seems that the weighted estimators are able to adjust better to an increase in the number of sub-models than the two-step estimator, a possible intrinsic advantage of weighted estimators. The range of the values in $\mathbf{\Delta}$ also matters, as will become more apparent in Figure 2.

Figure 2 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 $\mathbf{\Delta}$ vector. In the top and bottom contour plots symmetric and asymmetric $\mathbf{\Delta}$ cases are considered separately. The top contour plot reveals a structure that is consistent with Corollary 4.5, 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 4.5. The bottom contour plot is constructed using $\mathbf{\Delta}$ that are not symmetric (all $\Delta_{min} = 0$), and therefore could not be compared to the predictions of Corollary 4.5. From this plot we can see however that $\hat{\sigma}^2_{p,MRE}$ seems to dominate $\hat{\sigma}^2_{MRE}$ everywhere. Note that this plot employs $\mathbf{\Delta}$'s which are extremely asymmetric, with $\Delta_{min} = 0$.

Figure 3 is designed to explore the case when $p = 2$ only, so $\mathbf{\Delta} = (0, \Delta)$. This figure depicts the efficiency of all estimators (except $\hat{\sigma}^2_{p,ALB}$ whose performance is quite similar to $\hat{\sigma}^2_{p,LB1}$ and not competitive with the others) as a function of the magnitude of 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 to be better to average over the sub-models' estimators. Finally, 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.

Overall, based on the results of the risk comparisons, 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 may result in improved performance.

5 Comparisons of Estimators of $\tau(t)$

In this section we compare the performances of estimators of $\tau(t) = \Phi((t-\mu)/\sigma)$ for a given t using the loss function in (5). To perform global comparisons of the estimators over the whole region $\Re$, we utilize the global loss function

$$L_3(\hat{F}, F) = \int_{-\infty}^{+\infty} \{\hat{F}(t) - F(t)\}^2 dF(t) = \int_{-\infty}^{+\infty} L_2(\hat{F}(t), F(t)) dF(t). \tag{34}$$

Denoting by $R_2(\hat{F}(t), F(t))$ the risk function of $\hat{F}(t)$ using the loss function L_2, the risk function associated with L_3 is given by $R_3(\hat{F}, F) = \int_{-\infty}^{+\infty} R_2(\hat{F}(t), F(t)) dF(t)$. In the numerical comparison of the estimators, we shall approximate this global risk via

$$\hat{R}_3(\hat{F}, F) = \sum_{k=1}^{M} \hat{R}_2(\hat{F}(t_k), F(t_k))\{F(t_k) - F(t_{k-1})\}, \tag{35}$$

where $\hat{R}_2(\hat{F}(t), F(t))$ represents an approximation based on simulation to $R_2(\hat{F}(t), F(t))$. In particular, in our computations, we have chosen t_k's such that $F(t_k) - F(t_{k-1}) = .01$ for all k's.

We recall the competing estimators. First we have the UMVUE under model $\mathcal{M}$ given by

$$\hat{\tau}_{UMVU}(t) = \mathcal{T}\left(\frac{\sqrt{n-2}z_1(t)}{\sqrt{1-z_1^2(t)}}; n-2\right) I\{|z_1(t)| \le 1\} + I\{z_1(t) > 1\}$$

where $z_1(t) = (\sqrt{n}/(n-1))\left((t-\bar{X})/S\right)$. The second estimator, formed by first selecting the sub-model using the selector $\hat{M}$ and using the selected sub-model's UMVUE is

$$\hat{\tau}_{p,UMVU}(t) = \sum_{i=1}^{p} I\{\hat{M} = i\}\hat{\tau}_{UMVU,i}(t)$$

where, for $i = 1, 2, \ldots, p$, we have

$$\hat{\tau}_{UMVU,i}(t) = \mathcal{T}\left(\frac{\sqrt{n-1}z_{3i}(t)}{\sqrt{1-z_{3i}^2(t)}}; n-1\right) I\{|z_{3i}(t)| \le 1\} + I\{z_{3i}(t) > 1\}$$

with $z_{3i}(t) = 1/\sqrt{n}\left((t-\mu_i)/\hat{\sigma}_i\right)$. The next estimator is similar to the second estimator, except it utilizes a limiting Bayes estimator associated with $\kappa = 1$ at the chosen sub-model. It is

$$\hat{\tau}_{p,ALB}(t) = \sum_{i=1}^{p} I\{\hat{M} = i\}\mathcal{T}\left(\frac{t-\mu_i}{\hat{\sigma}_i}; n\right).$$

The four limiting Bayes estimators $\hat{\tau}_{p,LBk}(t)$, $k = 1, 2, 3, 4$ which are associated with $\kappa \in \{1, 3/2, 2, 3\}$ are enumerated in Table 2.

All of these estimators have distributional representations which indicate that their distributions depend on $(\mu_1, \ldots, \mu_p, \sigma^2)$ only through $\xi(t) = (t - \mu_{i_0})/\sigma$ and $\boldsymbol{\Delta} = (\boldsymbol{\mu} - \mu_{i_0}\mathbf{1})/\sigma$. To demonstrate this for the estimator $\hat{\tau}_{UMVU}(t)$, let $Z \sim N(0,1)$ and $W \sim \chi^2_{n-1}$ with Z and W independent. Then,

$$z_1(t) = \left(\frac{\sqrt{n}}{n-1}\right)\left(\frac{t-\bar{X}}{S}\right) \stackrel{d}{=} \frac{1}{\sqrt{n-1}}\left(\frac{\sqrt{n}\xi(t) - Z}{\sqrt{W}}\right) \equiv z_1^*(t; Z, W).$$

As a consequence, a representation of $\hat{\tau}_{UMVU}(t)$ is

$$\hat{\tau}_{UMVU}(t) \stackrel{d}{=} \mathcal{T}\left(\frac{\sqrt{n-2}z_1^*(t;Z,W)}{\sqrt{1-(z_1^*(t;Z,W))^2}}; n-2\right) I\{|z_1^*(t;Z,W)| \le 1\} + I\{z_1^*(t;Z,W) > 1\}. \quad (36)$$

We also have $z_{3i}(t) \stackrel{d}{=} (\xi(t) - \Delta_i)/\sqrt{W + (Z - \sqrt{n}\Delta_i)^2} \equiv z_{3i}^*(t; Z, W)$, so with $z_{3(i)}^*(t; Z, W) = (\xi(t) - \Delta_{(i)})/\sqrt{W + (Z - \sqrt{n}\Delta_{(i)})^2}$ and

$$\begin{aligned} \hat{\tau}_{UMVU,(i)}(t; Z, W) \quad &\equiv \quad \mathcal{T}\left(\frac{\sqrt{n-1}z_{3(i)}^*(t;Z,W)}{\sqrt{1-(z_{3(i)}^*(t;Z,W))^2}}; n-1\right) I\{|z_{3(i)}^*(t;Z,W)| \le 1\} \\ &\quad + I\{z_{3(i)}^*(t;Z,W) > 1\}, \end{aligned}$$

we obtain the representation

$$\hat{\tau}_{p,UMVU}(t) \stackrel{d}{=} \sum_{i=1}^{p} I\{L(\Delta_{(i)}, \mathbf{\Delta}) < Z < U(\Delta_{(i)}, \mathbf{\Delta})\}\hat{\tau}_{UMVU,(i)}(t; Z, W). \tag{37}$$

The proof of this is analogous to the proof of Theorem 4.1, and omitted. The limiting Bayes estimators have similar representations. For instance, with $\boldsymbol{T} = (Z\mathbf{1} - \sqrt{n}\mathbf{\Delta})/\sqrt{W}$ (cf., Cor. 4.1), we have

$$\hat{\tau}_{p,LB1}(t) \stackrel{d}{=} \sum_{i=1}^{p} \left\{ \frac{(1+T_i^2)^{-n/2}}{\sum_{j=1}^{p}(1+T_j^2)^{-n/2}} \right\} \mathcal{T}\left(\sqrt{n}\, z_{3i}^*(t; Z, W); n\right) \tag{38}$$

with similar representations for the other limiting Bayes estimators. Unfortunately, even with these representations, obtaining closed-form expressions of the risk functions is still difficult, if not impossible. The risk comparisons were therefore facilitated through a simulation study.

5.1 Simulated Comparisons of Estimators of $\tau(t)$

We perform simulated comparisons of the estimators of $\tau(t)$ by taking specific combinations of n, p, $\mathbf{\Delta}$, and $\xi(t)$, since these are the quantities on which the distributions depend. The set of combinations of $(n, p, \mathbf{\Delta})$ are given in Table 6. For $\xi(t)$, we choose t-values ranging from the 1st to the 99th percentile of the true underlying distribution. Equivalently, the values of $\xi(t)$ range from the 1st to the 99th percentiles of the standard normal distribution.

We first examine the pointwise simulated risks of the estimators to ascertain their bias and variance behaviors. Figure 4 presents the simulated bias, variance, risk, and relative efficiency curves of the estimators for $n = 10$ and two sets of $\mathbf{\Delta}$: $\mathbf{\Delta} = (-1, 0, 1)$ and $\mathbf{\Delta} = (0, .5, 1)$. For each value of t or $\xi(t)$, 10000 replications were performed. As expected, the simulated bias curve of $\hat{\tau}_{UMVU}$ is almost zero; whereas the other estimators developed under $\mathcal{M}_p$ all have non-negligible bias. $\hat{\tau}_{p,UMVU}$ has less bias than the other adaptive estimators, with $\hat{\tau}_{p,LB1}$ and $\hat{\tau}_{p,LB4}$ estimators seeming to be more biased than other adaptive estimators. Looking at the simulated variance curves, we now observe that these adaptive estimators have considerably lower variances than the $\hat{\tau}_{UMVU}$ estimator in the central portion of the domain of the distribution, which in this case contains most of the probability. Combining the variance and bias curves to obtain risk curves, we see that the adaptive estimators are better than the $\hat{\tau}_{UMVU}$. Translating these risks into relative efficiencies, we see that there is considerable improvement by utilizing an adaptive estimator relative to the UMVU estimator under $\mathcal{M}$. However, just by considering these two sets of pointwise curves, it is not clear whether it is preferable to utilize a limiting Bayes or a two-step estimator.

Table 6 presents the simulated relative global efficiencies of the two-step estimators $\hat{\tau}_{p,UMVU}(t)$, $\hat{\tau}_{p,ALB}(t)$, and the limiting Bayes estimators $\hat{\tau}_{p,LBk}(t), k = 1, 2, 3, 4$. The limiting Bayes estimator

$\hat{\tau}_{p,LB1}(t)$, which is associated with the improper prior as $\kappa \to 1$, outperforms consistently all other estimators in all the cases considered. (Interestingly, recall that the corresponding limiting Bayes estimator of σ^2, $\hat{\sigma}^2_{p,LB1}$, did not even have competitive efficiency!) This is followed by the $\hat{\tau}_{p,LB2}(t)$ estimator. Between the two-step estimators, $\hat{\tau}_{p,ALB}(t)$ is better than the $\hat{\tau}_{p,UMVU}(t)$ estimator. Considering the impact of the range of Δ values, which is done by examining $\boldsymbol{\Delta}$ sets $\{1,2,3\}$, $\{4,5,6\}$, and $\{7,8\}$, we note that the advantage of the $\tau_{p,LB1}(t)$ estimator over $\hat{\tau}_{UMVU}(t)$ diminishes as the range of the Δ values increases, except that there is a sudden increase in advantage for the case $\boldsymbol{\Delta} = (-1,0,1)$ at $n = 30$, indicating that the behavior is not always monotone.

Intuitively, one would expect that with the increasing sample size n, the relative global efficiency of $\hat{\tau}_{p,LB1}(t)$ should decrease, but this is not so. This counter-intuitive behavior is also visible in Figure 5, which shows two contour plots of the relative global efficiencies of $\hat{\tau}_{p,LB1}(t)$ with respect to $\hat{\tau}_{UMVU}(t)$ as a function of n and Δ_{max}. The top contour plot is for asymmetric $\boldsymbol{\Delta}$ with $p = 2$, while the bottom plot is for symmetric $\boldsymbol{\Delta}$ with $p = 3$. In both plots it is clear that the relationship between the relative global efficiency and Δ_{max} is not monotone in n. The highest relative efficiency regions are observed for the smallest and largest Δ_{max} values, but the *valley* that separates these *high planes* varies remarkably with n, being widest for small n and narrowing as n increases.

An apparent explanation of this fascinating behavior is as follows. For a fixed n, when $|\Delta_{max}|$ is large, the model selector is able to select the correct sub-model with high probability, rendering $\mathcal{M}_p$-based estimators efficient relative to the $\mathcal{M}$-based UMVU estimator. When $|\Delta_{max}|$ is small (close to zero), though the model selector may have difficulty selecting the correct model, the negative impact of erroneously chosen sub-model is not so pronounced as the bias is negligible. So with a smaller variance, $\mathcal{M}_p$-based estimators still predominate. The story changes when $|\Delta_{max}|$ is neither too small nor too large, because now when the model selector chooses a wrong sub-model the bias of the $\mathcal{M}_p$-based estimators becomes magnified. So in spite of the variances being small, the overall improvement over the $\mathcal{M}$-based estimators is attenuated. This explains the observed non-monotone behavior for a given n in Figure 5 contour plots. To explain the behavior as n changes, we simply need to note that it is easier for the model selector to select the correct sub-model when n is large. If $n_1 < n_2$ are two samples sizes, and if $|\Delta_{max}(n,c)|$ is the value of $|\Delta_{max}|$ for sample size n that will yield the relative efficiency level c, then $|\Delta_{max}(n_1,c)| > |\Delta_{max}(n_2,c)|$.

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. This type of situation arises in many contexts (regression analysis, reliability and survival analysis, goodness-of-fit testing, to name a few). It is of interest to determine which of the three possible strategies is preferable: (i) to utilize a wider model, possibly nonparametric, that encompasses all competing sub-models; (ii) to adopt a two-step approach, where the first step is to select the sub-model, and having selected the sub-model, the second step is to use an inference procedure within this chosen sub-model, but with both steps utilizing the same sample data; (iii) to adopt 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 specimen with a finite number of Gaussian (normal) sub-models with common variance but different means, we examined the strategies in the problem of estimating the common variance and the distribution function. Based on the theoretical and simulated comparisons of the different types of estimators, and with the estimator performance evaluated through risk functions based on quadratic loss, the following conclusions can be made:

- there could be considerable improvement in using adaptive estimators, developed by exploiting the structure of the sub-models, over the strategy of simply using estimators from a wider model;

- however, the properties of the resulting adaptive estimators may be extremely 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;

- based on the theoretical and simulated results for the two parameters (σ^2 and τ) 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, which could be safeguarding against the worst scenario;

- 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 attenuated; and

- finally, when developing weighted estimators through the Bayesian framework, care must be observed in assigning prior parameters: a particular parameter specification may lead to poor estimators for estimating one parameter, but produce excellent estimators for other parameters.

Lastly, we point out that this study does not in any way answer conclusively many other issues pertaining to the situation of a model with a finite number of sub-models, calling for studies of more complicated situations in varied settings. For example, when adopting 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 for making inference in the chosen sub-model, an issue alluded for instance in Hastie, Tibshirani and Friedman ([6])? Furthermore, this paper did not address the important issues of admissibility or minimaxity of estimators, which clearly will be hard issues to deal with for the adaptive estimators, but which nevertheless should be studied in the future. For instance, are the estimators of the variance $\hat{\sigma}^2_{p,MRE}$ and $\hat{\sigma}^2_{p,PLB4}$ admissible?

References

[1] Akaike, H. (1973). Information theory and the maximum likelihood principle. In *2nd International Symposium on Information Theory*, pp. 610–624, (B. N. Petrov and F. Csaki, eds.). Akademiai Kiado, Budapest.

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

[3] Brewster, J. and Zidek, J. (1974). Improving on equivariant estimators. *Ann. Statist.*, **2**, 21–38.

[4] Brown, L. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.*, **39**, 29–48.

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

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

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

[8] Kallenberg, W. and Ledwina, T. (1995). Consistency and Monte Carlo simulations of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.*, **23**, 1594–1608/

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

[10] Ledwina, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.*, **89**, 1000-1005.

[11] Lehmann, E. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition. Springer: New York.

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

[13] Neyman, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuarietidskr.*, **20**, 149–199.

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

[15] Pal, N., Ling, C. and Lin, J. (1998). Estimation of a normal variance — a critical review. *Statist. Papers*, **39**, 389–404.

[16] Peña, E. (1998). Smooth goodness-of-fit tests for composite hypothesis in hazard based models. *Ann. Statist.*, **26**, 1935–1971.

[17] Rayner, J. and Best, D. (1989). *Smooth Tests of Goodness of Fit.* Oxford University Press.

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

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

[20] Schwartz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.

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

[22] Strawderman, W. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.*, **2**, 190–198.

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

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

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

[26] Zacks, S. (1971). *The Theory of Statistical Inference.* John Wiley & Sons: New York.

Table 3: Second-order approximation and simulation results for the mean and variance functions of $\hat{\sigma}^2_{p,LB}(\alpha = n/2)/\sigma^2$, and the risk function of $\hat{\sigma}^2_{p,LB}(\alpha = n/2)$ for different combinations of p, $\boldsymbol{\Delta}$, and n. For each combination, 10000 simulation replications were performed.

Combinations of p and $\boldsymbol{\Delta}$	n	Mean Appr.	Mean Sim.	Variance Appr.	Variance Sim.	Risk Appr.	Risk Sim.
	3	0.80	0.96	0.45	0.65	0.49	0.65
$\boldsymbol{\Delta}$=(-0.25,0,0.25)	10	0.96	0.99	0.18	0.20	0.18	0.20
p=3	30	0.99	0.99	0.07	0.07	0.07	0.07
	3	0.84	0.94	0.40	0.65	0.43	0.66
$\boldsymbol{\Delta}$=(-0.5,0,0.5)	10	0.99	0.99	0.20	0.21	0.20	0.21
p=3	30	1.02	1.00	0.07	0.07	0.07	0.07
	3	0.95	1.04	0.60	0.75	0.60	0.75
$\boldsymbol{\Delta}$=(0,0.25, 0.50)	10	1.01	1.01	0.20	0.21	0.20	0.21
p=3	30	1.00	1.00	0.07	0.07	0.07	0.07
	3	1.07	1.11	0.73	0.89	0.73	0.90
$\boldsymbol{\Delta}$=(0,0.5,1)	10	1.03	1.02	0.22	0.22	0.22	0.22
p=3	30	1.01	1.00	0.07	0.07	0.07	0.07
	3	0.86	0.98	0.50	0.65	0.52	0.65
$\boldsymbol{\Delta}$=(-0.25:0.0625:0.25)	10	0.97	0.98	0.19	0.20	0.19	0.20
p=9	30	0.99	0.99	0.07	0.07	0.07	0.07
	3	0.87	0.97	0.50	0.67	0.52	0.67
$\boldsymbol{\Delta}$=(-0.25:0.03125:0.25)	10	0.97	0.99	0.19	0.20	0.19	0.20
p=17	30	0.99	0.99	0.07	0.07	0.07	0.07
	3	0.99	1.03	0.64	0.74	0.64	0.74
$\boldsymbol{\Delta}$=(0:0.0625:0.5)	10	1.02	1.03	0.21	0.22	0.21	0.22
p=9	30	1.01	1.01	0.07	0.07	0.07	0.07
	3	1.00	1.04	0.65	0.76	0.65	0.76
$\boldsymbol{\Delta}$=(0:0.03125:0.5)	10	1.02	1.03	0.21	0.22	0.21	0.22
p=17	30	1.02	1.02	0.07	0.07	0.07	0.07

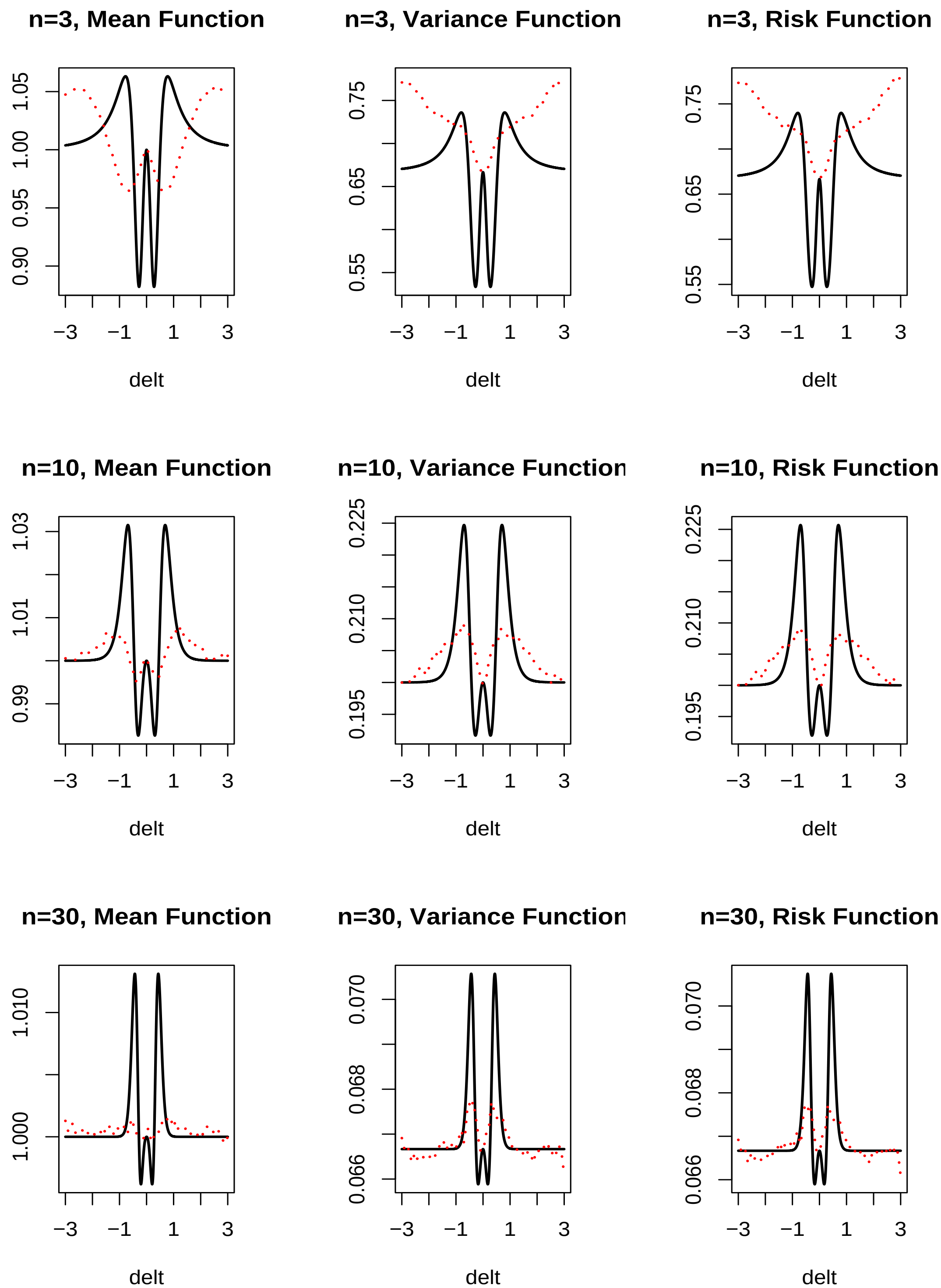


Figure 1: Second-order approximations (solid lines) and lowess curves (dotted lines) for the simulated mean and variance functions of $\hat{\sigma}^2_{p,LB}(\alpha = n/2)/\sigma^2$, and the risk function of $\hat{\sigma}^2_{p,LB}(\alpha = n/2)$ when $p = 2$, $\mathbf{\Delta} = (0, \Delta \equiv \text{delt})$, and three values of n. The number of simulation replications for each value of Δ was 10000.

Table 4: Efficiencies (relative to of the UMVU estimator $\hat{\sigma}^2_{UMVU}$) of the different variance estimators for different combinations of p, $\boldsymbol{\Delta}$, and n. For the limiting Bayes and weighted estimators, the values are based on simulation studies with 10000 replications for each combination.

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

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

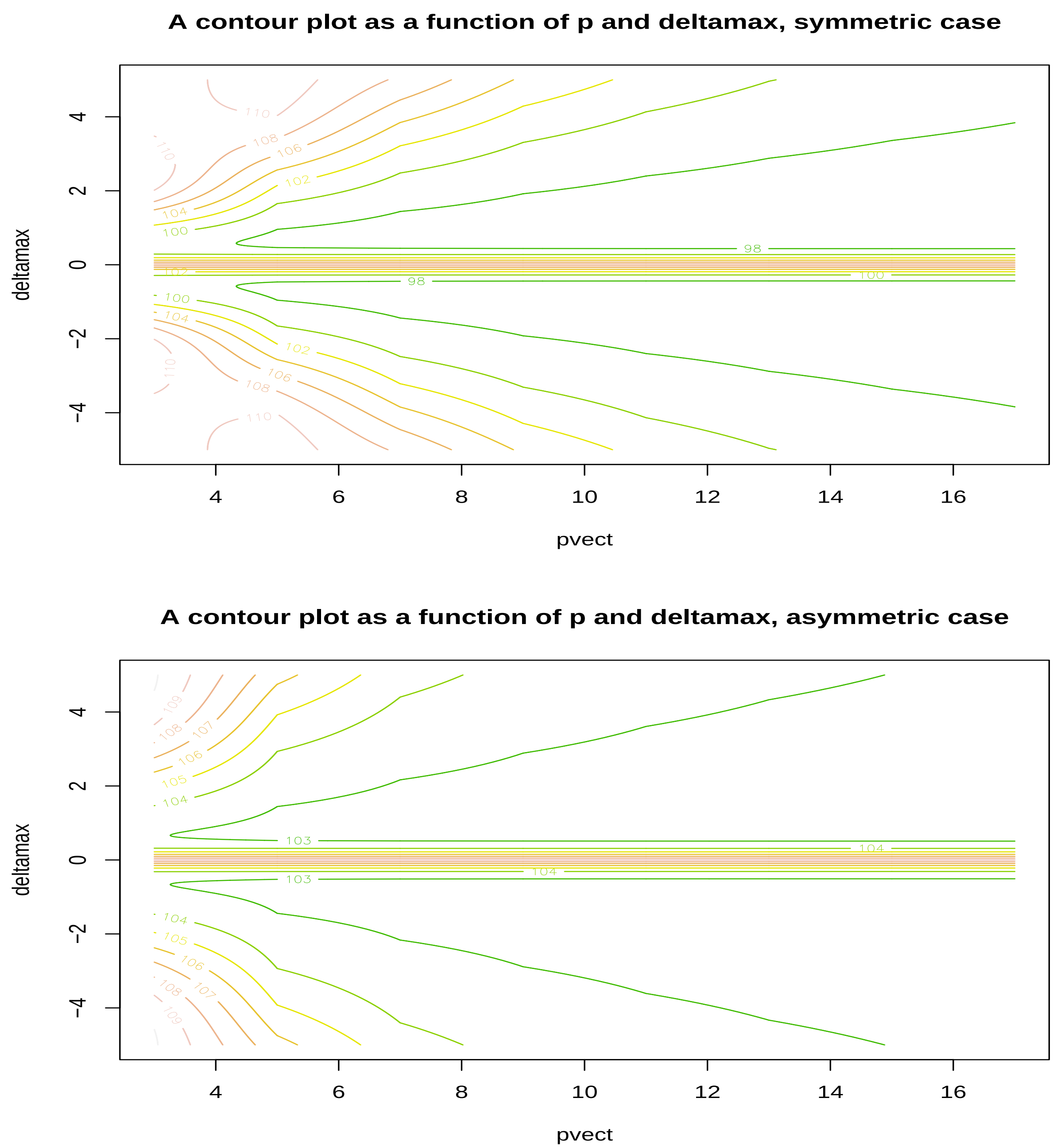


Figure 2: Relative efficiencies of pMRE with respect to MRE in a symmetric and asymmetric Δ cases, as a function of $\Delta_{\max}$ and number of sub-models p for sample size of $n = 10$. The symmetric case is of form $\mathbf{\Delta} = [-\Delta_{max} : \Delta_{max}/(p-1) : \Delta_{max}]$, while the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

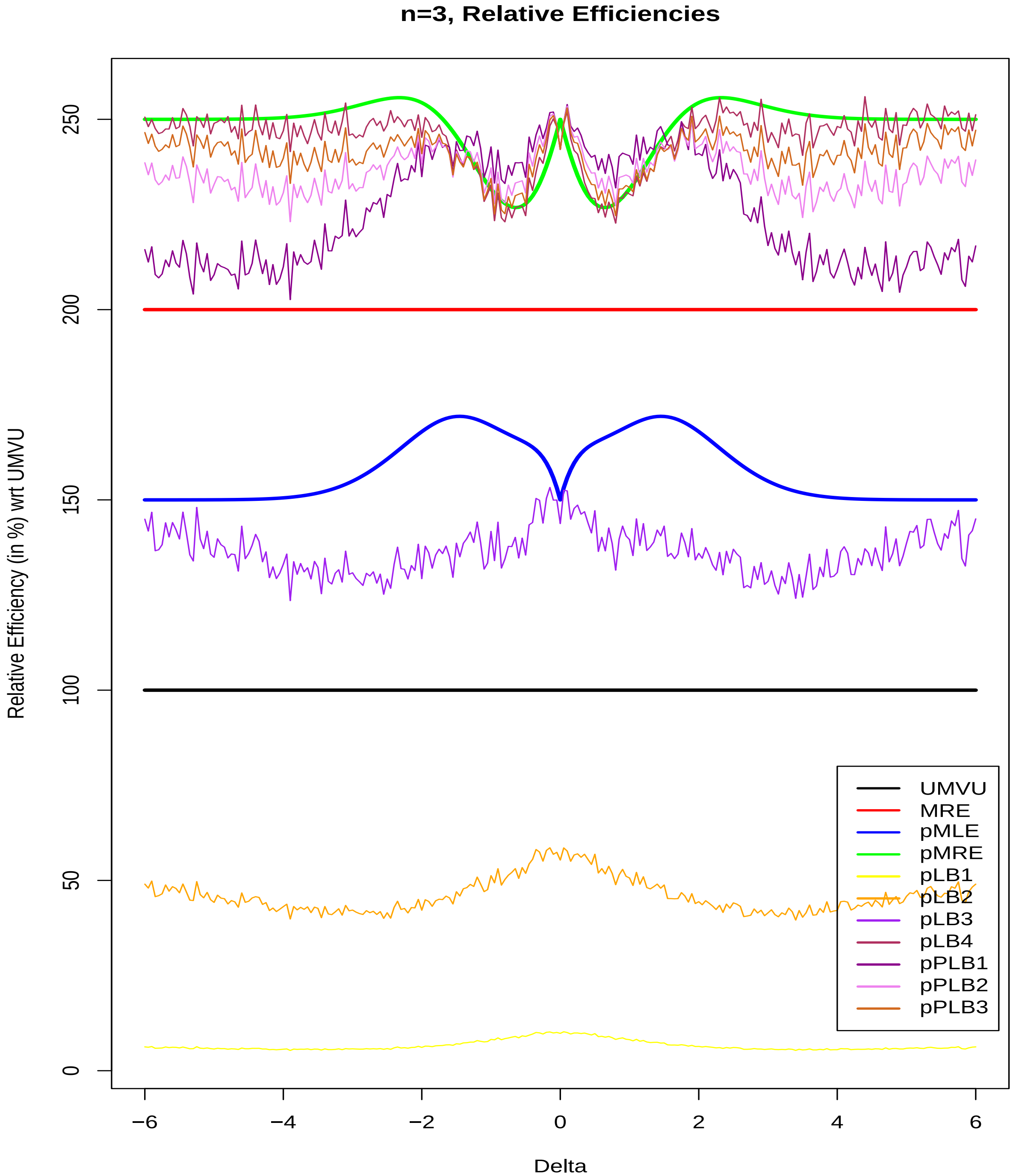


Figure 3: Relative efficiencies of the 11 estimators of σ^2 relative to the UMVUE $\hat{\sigma}^2_{UMVUE}$ for $p = 2$ and $\boldsymbol{\Delta} = (0, \Delta)$, with the Δ varying, for $n = 3$. For the limiting Bayes (pLBk) and weighted (pPLBk) the (connected) scatterplot represents the simulated estimates of relative efficiency based on 10000 replications for each Δ.

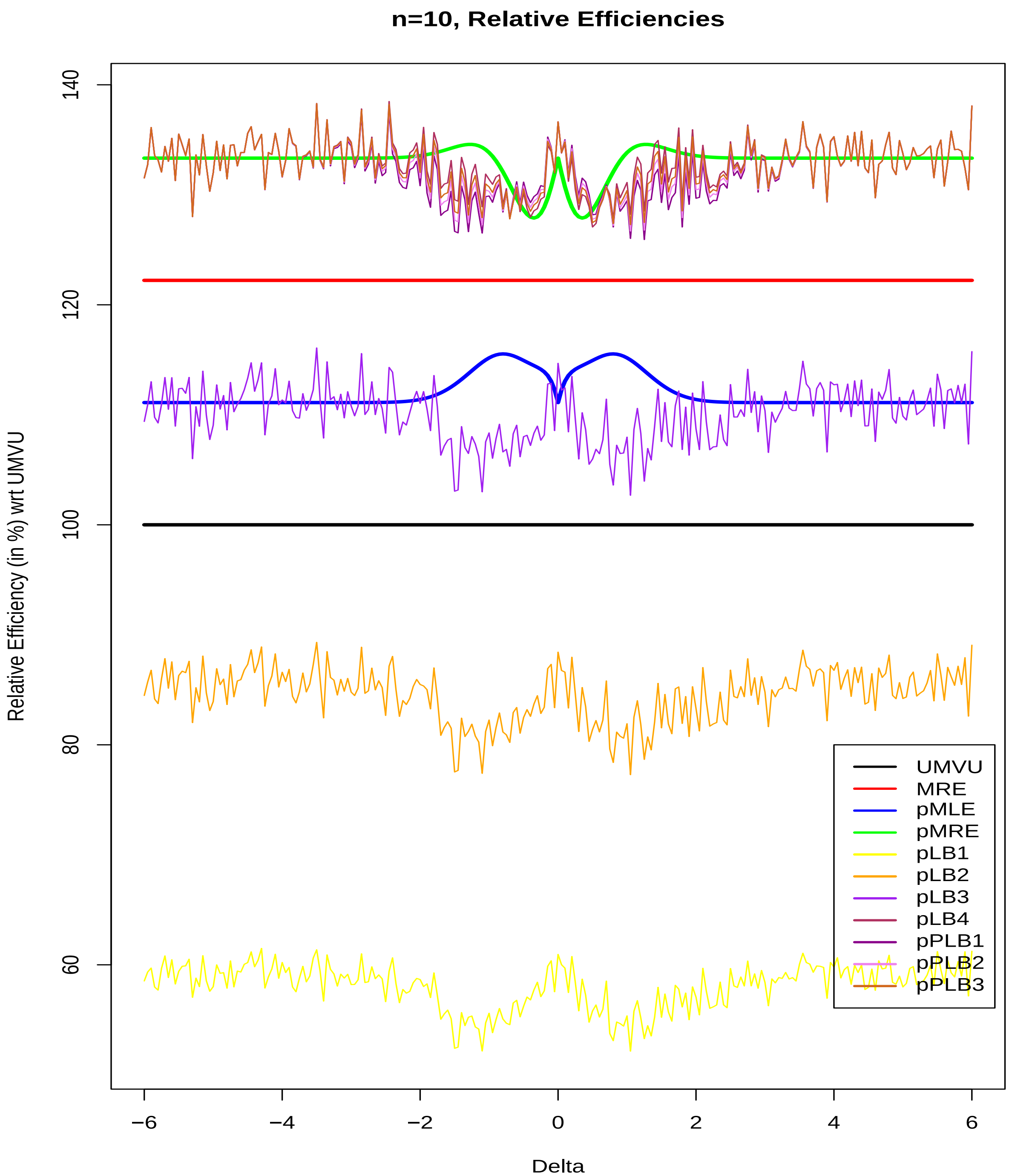


Figure 3 continued. Case of $n = 10$.

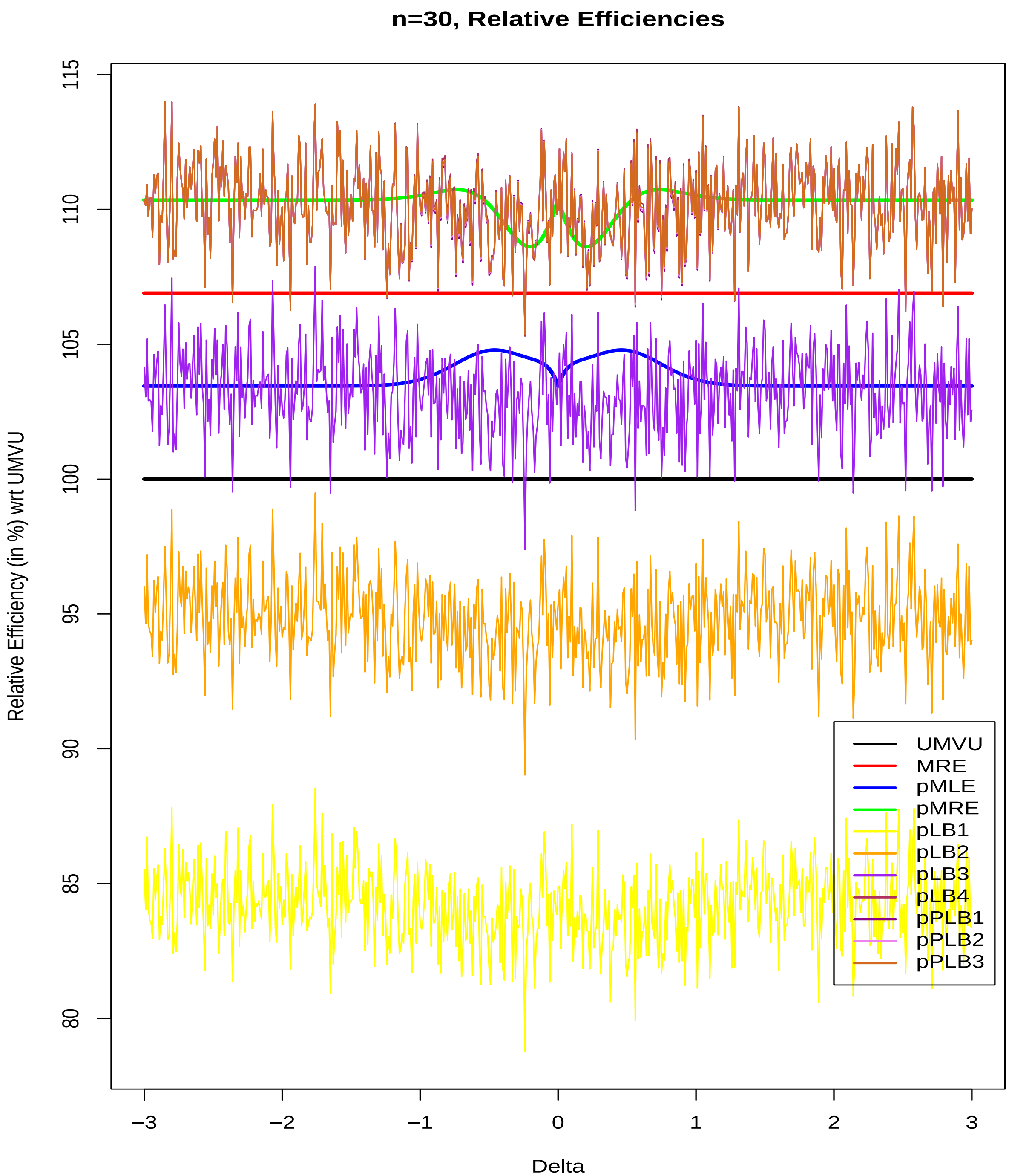


Figure 3 continued. Case of $n = 30$.

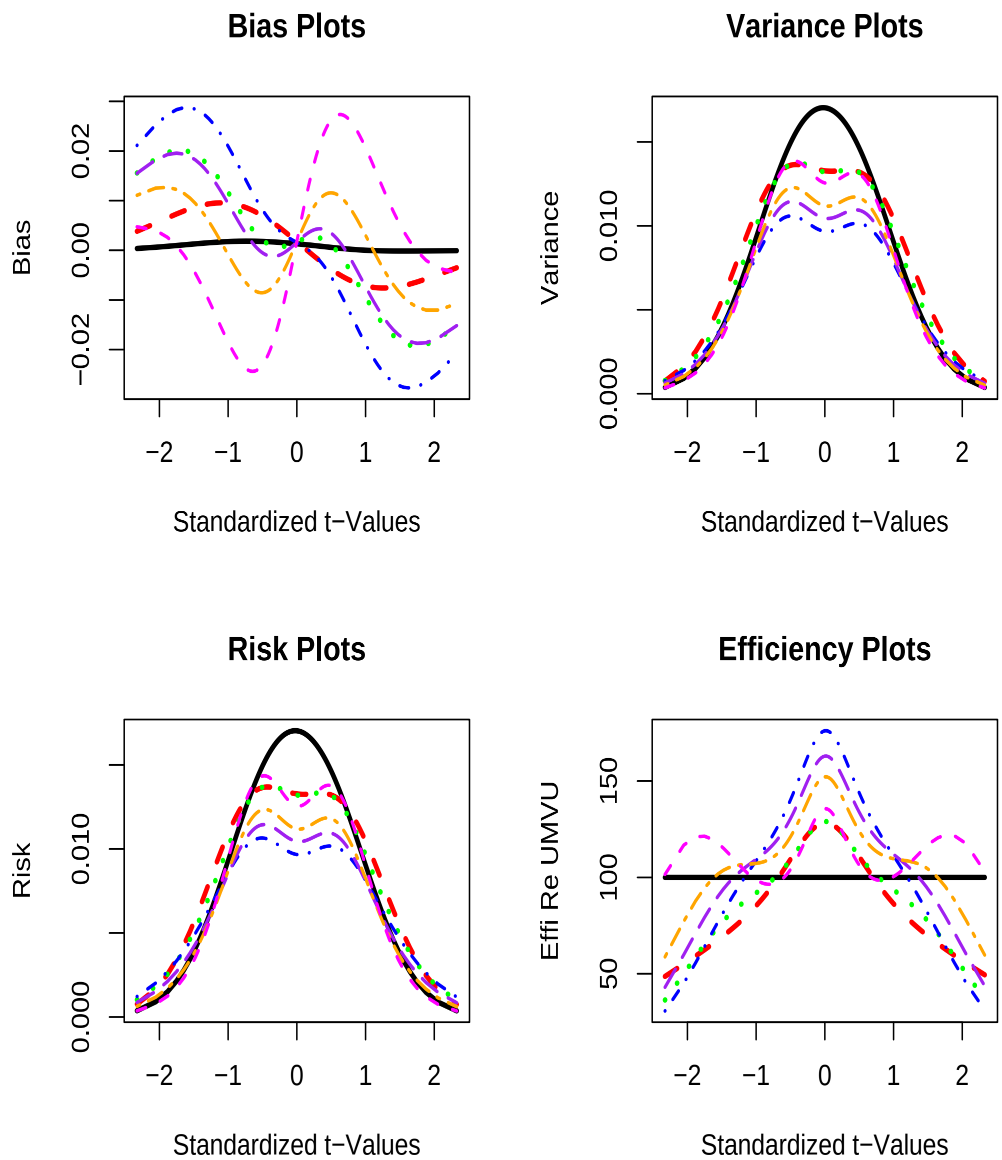


Figure 4: Pointwise biases, variances, risks, and relative efficiencies of the distribution estimators $\hat{\tau}$, for $\mathbf{\Delta} = (-1, 0, 1)$ and sample size $n = 10$. For each time point t, 10000 simulation replications were performed. **Legend:** — (black) = UMVU; $- - -$ (red) = pUMVU; $\cdots$ (green) = pALB; $- \cdot - \cdot -$ (blue) = pLB1; — — — (purple) = pLB2; - - - - - (orange) = pLB3; - - - (magenta) = pLB4.

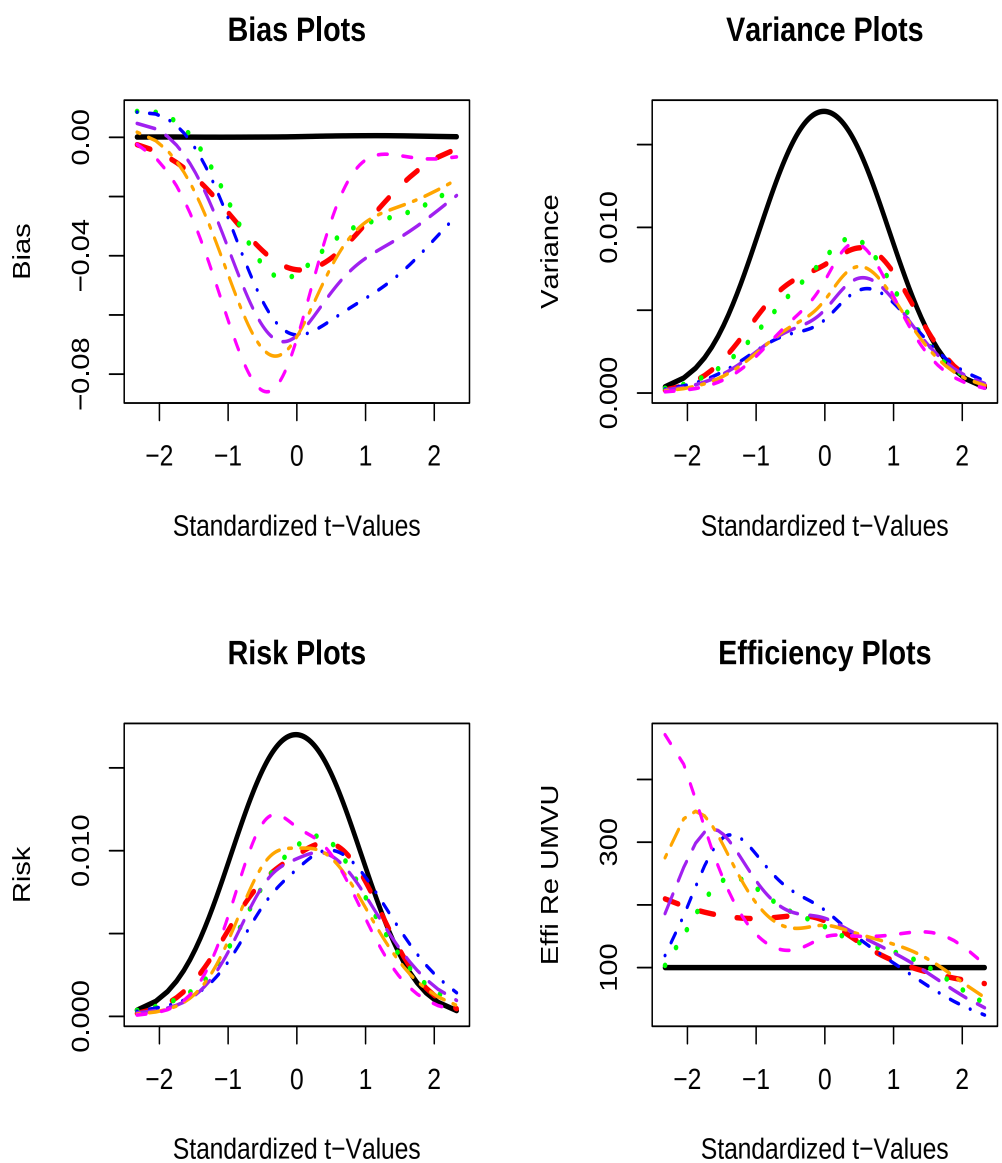
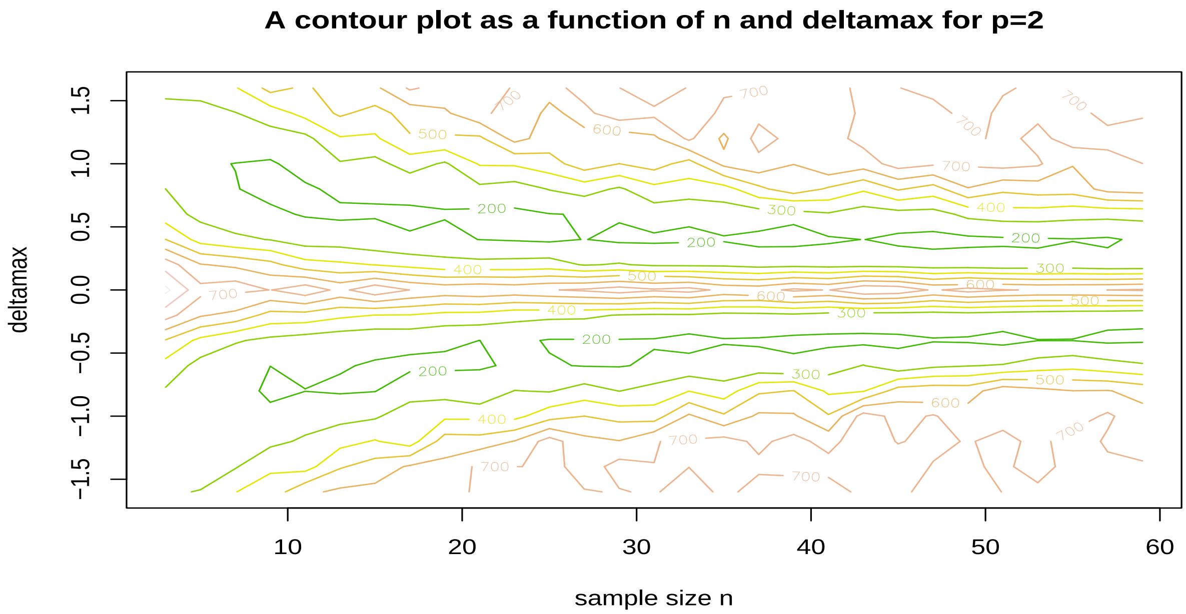


Figure 4 continued. Case of $\mathbf{\Delta} = (0, .5, 1)$.
Legend: — (black) = UMVU; – – – (red) = pUMVU; · · · (green) = pALB; – · – · – (blue) = pLB1; — — — (purple) = pLB2; - - - - - (orange) = pLB3; - - - (magenta) = pLB4.

Table 6: Relative global efficiencies (relative to the UMVU estimator $\hat{\tau}^2_{UMVU}$) of estimators of the distribution fucntion for different combinations of p, $\mathbf{\Delta}$, and n. 10000 simulation replications were performed for each combination.

Combinations of p and $\mathbf{\Delta}$	n	Efficiency % UMVU	pUMVU	pALB	pLB1	pLB2	pLB3	pLB4
$\mathbf{\Delta}$=(0, .25) p=2	3	100	411	542	664	572	465	329
	10	100	295	311	404	387	357	289
	30	100	188	190	248	245	239	222
$\mathbf{\Delta}$=(0,.50) p=3	3	100	264	319	419	358	300	226
	10	100	170	176	225	216	203	175
	30	100	152	154	191	189	186	175
$\mathbf{\Delta}$=(0,1) p=3	3	100	181	211	262	230	200	160
	10	100	187	194	233	226	215	186
	30	100	441	452	499	490	472	418
$\mathbf{\Delta}$=(-.25,0,.25) p=3	3	100	315	393	693	561	438	298
	10	100	192	199	430	393	348	266
	30	100	108	109	195	189	182	166
$\mathbf{\Delta}$=(-.5,0,.5) p=3	3	100	166	191	357	275	220	162
	10	100	96	98	159	147	136	115
	30	100	82	83	110	109	107	102
$\mathbf{\Delta}$=(-1,0,1) p=3	3	100	106	119	182	147	125	100
	10	100	104	107	136	131	125	111
	30	100	360	367	408	403	392	355
$\mathbf{\Delta}$=(0,.25,.50) p=3	3	100	272	332	423	365	308	233
	10	100	188	195	220	213	201	175
	30	100	155	157	160	160	158	151
$\mathbf{\Delta}$=(0,.5,1) p=3	3	100	181	211	242	215	190	156
	10	100	151	156	168	165	159	142
	30	100	150	151	186	184	181	171
$\mathbf{\Delta}$=(-1: 2^{-2} :1) p=9	3	100	111	125	216	169	140	110
	10	100	96	99	120	114	108	95
	30	100	89	90	103	102	101	96
$\mathbf{\Delta}$=(-1: 2^{-4} :1) p=33	3	100	111	125	230	180	148	115
	10	100	100	102	126	119	111	98
	30	100	99	100	104	102	101	96



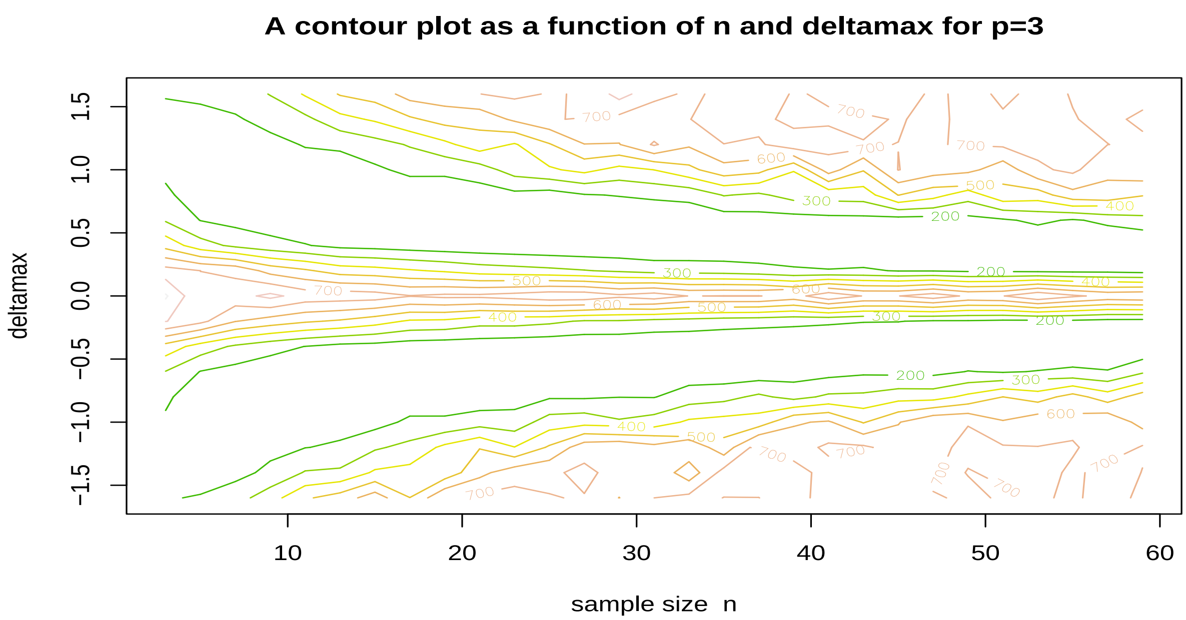


Figure 5: Relative global efficiencies of $\hat{\tau}_{p,LB1}$ with respect to $\hat{\tau}_{UMVU}$ in an asymmetric ($p = 2$) and symmetric ($p = 3$) $\boldsymbol{\Delta}$ cases, as a function of $\Delta_{\max}$ and sample size n. The first scenario corresponds to the asymmetric case of the form $\boldsymbol{\Delta} = (0, \Delta_{max})$, while the second scenario corresponds to the symmetric case of form $\boldsymbol{\Delta} = (-\Delta_{max}, 0, \Delta_{max})$. For each combination of $(n, \boldsymbol{\Delta})$, 10000 simulation replications were performed.