Journal of Multivariate Analysis 107 (2012) 232–243

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

James–Stein type estimators of variances

Tiejun Tong [a], Homin Jang [b], Yuedong Wang [b,*]

[a] Department of Mathematics, Hong Kong Baptist University, Hong Kong
[b] Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA

ARTICLE INFO

Article history:
Received 2 August 2011
Available online 31 January 2012

AMS subject classification:
62H12

Keywords:
Inadmissibility
Shrinkage estimation
Shrinkage parameter
Squared loss function
Stein loss function
Variance estimation

ABSTRACT

In this paper we propose James–Stein type estimators for variances raised to a fixed power by shrinking individual variance estimators towards the arithmetic mean. We derive and estimate the optimal choices of shrinkage parameters under both the squared and the Stein loss functions. Asymptotic properties are investigated under two schemes when either the number of degrees of freedom of each individual estimate or the number of individuals approaches to infinity. Simulation studies indicate that the performance of various shrinkage estimators depends on the loss function, and the proposed estimator outperforms existing methods under the squared loss function.

1. Introduction

Let $Z_1, \ldots, Z_p$ be unbiased estimators of the variances $\sigma_1^2, \ldots, \sigma_p^2$. Furthermore, assume that Z_i can be represented as $Z_i = \sigma_i^2 \chi_{\nu,i}^2/\nu$ for $i = 1, \ldots, p$ where $\chi_{\nu,i}^2$ are independent and identically distributed chi-squared random variables with ν degrees of freedom. We considerthe problem of estimating $\boldsymbol{\sigma}^{2t} = (\sigma_1^{2t}, \ldots, \sigma_p^{2t})$ for any $t \neq 0$ under both the squared loss function

$$L_Q(\sigma^2, \hat{\sigma}^2) = \left(\frac{\hat{\sigma}^2}{\sigma^2} - 1\right)^2, \tag{1}$$

and the Stein loss function [7]

$$L_T(\sigma^2, \hat{\sigma}^2) = \frac{\hat{\sigma}^2}{\sigma^2} - \ln\left(\frac{\hat{\sigma}^2}{\sigma^2}\right) - 1. \tag{2}$$

The Stein loss function is also known as the entropy loss or Kullback–Leibler loss function [9]. Note that the estimation of variances σ_i^2 or their reciprocals σ_i^{-2} are special cases with $t = 1$ or $t = -1$.

One of the motivations for the above problem is the detection of differentially expressed genes in microarray experiments. In this case Z_i corresponds to the sample variance of gene i. Typically the number of genes p is large and the number of degrees of freedom ν is small. Therefore, the conventional gene-by-gene t-test has low power [4,15]. This problem is quite common in high-dimensional data, and various methods have been proposed to improve the variance estimation

* Corresponding author.
E-mail address: yuedong@pstat.ucsb.edu (T. Tong).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.01.019

[2,8,14,16,13,4,6,10,15]. It is also known that improved variance estimation can substantially improve classification accuracy for high-dimensional low-sample-size data [12,11].

The basic technique is to borrow information across estimates of variances, an idea originated from the James–Stein estimator of means [7]. In particular, Cui et al. [4] proposed shrinkage estimators by applying the James–Stein estimator to variances on the logarithmic scale. Tong and Wang [15] considered optimal shrinkage estimators within a general family of estimators, also on the logarithmic scale. They showed that the sample variances are inadmissible.

In this paper, rather than on the logarithmic scale, we consider shrinkage estimation on the original scale. Consequently, individual variance estimators are shrunk towards the arithmetic mean rather than the geometric mean. Let $Z_i(t) = h(t)Z_i^t$ where $h(t) = \Gamma(\nu/2)(\nu/2)^t/\Gamma(\nu/2+t)$ and $\Gamma(\cdot)$ is the gamma function. For any nonzero $t > -\nu/2$, $Z_i(t)$ is an unbiased estimator of σ_i^{2t} [15]. When $\sigma_i^2 = \sigma^2$ for all i, $\bar{Z}(t) = \sum_{i=1}^p Z_i(t)/p$ is an unbiased estimator of σ^{2t}. Therefore, we consider the following shrinkage estimator for σ_i^{2t},

$$\hat{\sigma}_i^{2t} = \alpha\bar{Z}(t) + \beta Z_i(t), \quad i = 1, \ldots, p, \tag{3}$$

where $\alpha \geq 0$ and $\beta \geq 0$ are the shrinkage parameters with $\alpha + \beta > 0$. It is clear that the above shrinkage estimator is an extension of the James–Stein estimator for multiple means to multiple variances. There is no shrinkage when $\alpha = 0$ and $\beta = 1$. On the other hand, all variance estimates are shrunken to the bias-corrected arithmetic mean $\bar{Z}(t)$ when $\alpha = 1$ and $\beta = 0$. Note that we do not require $\alpha + \beta = 1$. Nevertheless, it is easy to check that, for any $t > -\nu/2$, the total bias $\sum_{i=1}^p \{\mathrm{E}(\hat{\sigma}_i^{2t}) - \sigma_i^{2t}\}$ equals zero if and only if $\alpha + \beta = 1$. Therefore, we will also consider the special case when $\alpha + \beta = 1$. In what follows, under both the squared and the Stein loss functions, we study the optimal shrinkage estimators and estimate the optimal shrinkage parameters.

The remainder of the paper is organized as follows. In Sections 2 and 3, we derive the optimal shrinkage estimators for variances under the squared and the Stein loss functions, respectively. We also propose estimators for the optimal shrinkage parameters and investigate their asymptotic properties. We then conduct simulations in Section 4 to evaluate the performance of the proposed estimators and compare them to some existing methods.

2. Optimal shrinkage estimator under the squared loss

2.1. Optimal shrinkage estimator

Let $\bar{\sigma}^{2\xi} = \sum_{i=1}^p \sigma_i^{2\xi}/p$ for any ξ and $\hat{\boldsymbol{\sigma}}^{2t} = (\hat{\sigma}_1^{2t}, \ldots, \hat{\sigma}_p^{2t})$. It is straightforward to show that, under the squared loss function (1), the average risk of $\hat{\sigma}_i^{2t}$ is

$$\begin{aligned} R_Q(\alpha, \beta; \boldsymbol{\sigma}^{2t}) &\triangleq \frac{1}{p}\sum_{i=1}^p \mathrm{E}L_Q(\sigma_i^{2t}, \hat{\sigma}_i^{2t}) \\ &= A_2(t)\alpha^2 + A_3(t)\beta^2 + 2A_4(t)\alpha\beta - 2A_1(t)\alpha - 2\beta + 1, \end{aligned} \tag{4}$$

where $A_1(t) = \bar{\sigma}^{2t}\bar{\sigma}^{-2t}$, $A_2(t) = \bar{\sigma}^{-4t}[\bar{\sigma}^{4t}\{A_3(t) - 1\}/p + (\bar{\sigma}^{2t})^2]$, $A_3(t) = \Gamma(\nu/2)\Gamma(\nu/2+2t)/\Gamma^2(\nu/2+t)$, and $A_4(t) = \{A_3(t) - 1\}/p + A_1(t)$. Note that $\bar{\sigma}^{2t}, \bar{\sigma}^{-2t}, \bar{\sigma}^{4t}$ and $\bar{\sigma}^{-4t}$ correspond to $\bar{\sigma}^{2\xi}$ with $\xi = t, \xi = -t, \xi = 2t$ and $\xi = -2t$ respectively. In addition, $R_Q(\alpha, \beta; \boldsymbol{\sigma}^{2t})$ is a positive definite quadratic function of α and β.

Theorem 1. *For any fixed $p \geq 2$, ν, and nonzero $t > -\nu/4$, $R_Q(\alpha, \beta; \boldsymbol{\sigma}^{2t})$ is a strictly convex function of α and β with the unique minimum point at*

$$\alpha_{Q_1}^* = \frac{A_1(t)A_3(t) - A_4(t)}{A_2(t)A_3(t) - A_4^2(t)} \quad \text{and} \quad \beta_{Q_1}^* = \frac{A_2(t) - A_1(t)A_4(t)}{A_2(t)A_3(t) - A_4^2(t)}. \tag{5}$$

The proof of Theorem 1 is omitted since it is straightforward. $\alpha_{Q_1}^*$ and $\beta_{Q_1}^*$ are the optimal shrinkage parameters. Denote the corresponding optimal shrinkage estimator as $\hat{\sigma}_{i,Q_1}^{2t} = \alpha_{Q_1}^*\bar{Z}(t) + \beta_{Q_1}^* Z_i(t)$.

$A_1(t) \geq 1$ by the Cauchy inequality and $A_3(t) > 1$ for any $t > -\nu/4$ [1]. Then $A_1(t)A_3(t) - A_4(t) = \{A_3(t) - 1\}\{A_1(t) - 1/p\} > 0$ for any $p \geq 2$. Furthermore, $A_2(t)A_3(t) - A_4^2(t) > 0$ since it is the determinant of the Hessian matrix of the strictly convex function $R_Q(\alpha, \beta; \boldsymbol{\sigma}^{2t})$. Therefore, $\alpha_{Q_1}^* > 0$ and $\hat{\sigma}_{i,Q_1}^{2t}$ has a smaller average risk than the original estimator $Z_i(t)$. When $\sigma_i^2 = \sigma^2$ for all i, we have $A_1(t) = 1$ and $A_2(t) = A_4(t) = \{A_3(t) - 1\}/p + 1$. Plugging them into (5) leads to $\alpha_{Q_1}^* = [\{A_3(t) - 1\}/p + 1]^{-1}$ and $\beta_{Q_1}^* = 0$. Now since $A_3(t) > 1$ for any $t > -\nu/4$, we have $\alpha_{Q_1}^* < 1$. Therefore, $\hat{\sigma}_{i,Q_1}^{2t}$ also has a smaller average risk than the pooled variance estimator $\bar{Z}(t)$ when $\sigma_i^2 = \sigma^2$ for all i. The following theorem indicates that there is no need to borrow information across estimates of variances when the degrees of freedom ν approaches to infinity.

Theorem 2. *For any fixed $p \geq 2$ and nonzero t, as $\nu \to \infty$, we have*

(i) *$\alpha_{Q_1}^* \to 0$ and $\beta_{Q_1}^* \to 1$ when σ_i^2 are not all the same,*
(ii) *$\alpha_{Q_1}^* \to 1$ and $\beta_{Q_1}^* = 0$ when $\sigma_i^2 = \sigma^2$ for all i.*

The proof of Theorem 2 is given in Appendix A. We now consider the optimal shrinkage under the constraint of $\alpha+\beta=1$. Substituting $\beta=1-\alpha$ in (4), we have $R_Q(\alpha,1-\alpha;\boldsymbol{\sigma}^{2t})=\{A_2(t)+A_3(t)-2A_4(t)\}\alpha^2-2\{A_1(t)+A_3(t)-A_4(t)-1\}\alpha+A_3(t)-1$ for $0\le\alpha\le1$. Then the optimal shrinkage estimator $\hat{\sigma}^{2t}_{i,Q_2}=\alpha^*_{Q_2}\bar{Z}(t)+(1-\alpha^*_{Q_2})Z_i(t)$, where $\alpha^*_{Q_2}=(1-1/p)\{A_3(t)-1\}/\{A_2(t)+A_3(t)-2A_4(t)\}$. It is not difficult to show that $\alpha^*_{Q_2}\le1$. Since $A_3(t)>1$ for any $t>-\nu/4$ and $A_2(t)+A_3(t)-2A_4(t)>0$, then $\alpha^*_{Q_2}>0$ for any $p\ge2$. This implies that $\hat{\sigma}^{2t}_{i,Q_2}$ has a smaller average risk than the original estimator $Z_i(t)$. When $\sigma_i^2=\sigma^2$ for all i, noting that $A_1(t)=1$ and $A_2(t)=A_4(t)$, we have $\alpha^*_{Q_2}=1$ regardless of the value of ν. As a consequence, $\hat{\sigma}^{2t}_{i,Q_2}$ reduces to the pooled variance estimator $\bar{Z}(t)$. This is different from the unconstrained situation where $\alpha^*_{Q_1}<1$. Finally, for any fixed $p\ge2$ and nonzero t, as $\nu\to\infty$, we have $\alpha^*_{Q_2}\to0$ when σ_i^2 are not all the same.

2.2. Estimation of the optimal shrinkage parameters

The optimal shrinkage parameters $\alpha^*_{Q_1}$ and $\beta^*_{Q_1}$ depend on the unknown quantities $\bar{\sigma}^{2\xi}$ where $\xi=t,-t,2t$, and $-2t$. We estimate $\bar{\sigma}^{2\xi}$ by $\bar{Z}(\xi)=\sum_{i=1}^p Z_i(\xi)/p$. Correspondingly, let $\tilde{A}_1(t)=\bar{Z}(t)\bar{Z}(-t)$, $\tilde{A}_2(t)=[\bar{Z}(2t)\{A_3(t)-1\}/p+\bar{Z}^2(t)]\bar{Z}(-2t)$, and $\tilde{A}_4(t)=\{A_3(t)-1\}/p+\bar{Z}(t)\bar{Z}(-t)$ be the estimates of $A_1(t)$, $A_2(t)$, and $A_4(t)$, respectively. Then we estimate the optimal shrinkage estimators by

$$\tilde{\alpha}^*_{Q_1}=\frac{\tilde{A}_1(t)A_3(t)-\tilde{A}_4(t)}{\tilde{A}_2(t)A_3(t)-\tilde{A}_4^2(t)}\quad\text{and}\quad\tilde{\beta}^*_{Q_1}=\frac{\tilde{A}_2(t)-\tilde{A}_1(t)\tilde{A}_4(t)}{\tilde{A}_2(t)A_3(t)-\tilde{A}_4^2(t)}.$$

The estimated optimal shrinkage estimator under the squared loss function is then $\tilde{\sigma}^{2t}_{i,Q_1}=\tilde{\alpha}^*_{Q_1}\bar{Z}(t)+\tilde{\beta}^*_{Q_1}Z_i(t)$.

Similarly, for $\alpha^*_{Q_2}$ under the constraint of $\alpha+\beta=1$, we have $\tilde{\alpha}^*_{Q_2}=(1-1/p)\{A_3(t)-1\}/\{\tilde{A}_2(t)+A_3(t)-2\tilde{A}_4(t)\}$ and $\tilde{\sigma}^{2t}_{i,Q_2}=\tilde{\alpha}^*_{Q_2}\bar{Z}(t)+(1-\tilde{\alpha}^*_{Q_2})Z_i(t)$. The following theorem summarizes the asymptotic behavior of the estimated optimal shrinkage parameters as $\nu\to\infty$.

Theorem 3. *For any fixed $p\ge2$ and nonzero t, as $\nu\to\infty$, we have*

(i) $\tilde{\alpha}^*_{Q_1}\xrightarrow{\text{a.s.}}0$ *and* $\tilde{\beta}^*_{Q_1}\xrightarrow{\text{a.s.}}1$ *when σ_i^2 are not all the same,*
(ii) $\tilde{\alpha}^*_{Q_2}\xrightarrow{\text{a.s.}}0$ *when σ_i^2 are not all the same.*

The proof of Theorem 3 is given in Appendix B. For high-dimensional data, it is common that ν is relatively small and p is large. In what follows we investigate the asymptotic behavior of the estimated optimal shrinkage parameters as $p\to\infty$. We assume that σ_i^2 are i.i.d. random variables from a certain distribution F. Denote $\mu_\xi=\mathrm{E}\sigma_1^{2\xi}$ as the ξ-th moment of F.

Theorem 4. *For any fixed ν and nonzero $|t|<\nu/4$, assume that $\sigma_i^2\overset{\text{i.i.d.}}{\sim}F$ with $0<\mu_\xi<\infty$ for $\xi=2t$ and $-2t$. Then*

(i) $\tilde{\alpha}^*_{Q_1}-\alpha^*_{Q_1}\xrightarrow{\text{a.s.}}0$ *and* $\tilde{\beta}^*_{Q_1}-\beta^*_{Q_1}\xrightarrow{\text{a.s.}}0$ *as $p\to\infty$,*
(ii) $\tilde{\alpha}^*_{Q_2}-\alpha^*_{Q_2}\xrightarrow{\text{a.s.}}0$ *as $p\to\infty$.*

The proof of Theorem 4 is given in Appendix C.

3. Optimal shrinkage estimator under the Stein loss

3.1. Optimal shrinkage estimator

Under the Stein loss function (2), the average risk of $\hat{\sigma}_i^{2t}$ is given as

$$\begin{aligned}R_T(\alpha,\beta;\boldsymbol{\sigma}^{2t})&\triangleq\frac{1}{p}\sum_{i=1}^p\mathrm{E}L_T(\sigma_i^{2t},\hat{\sigma}_i^{2t})\\&=A_1(t)\alpha+\beta-\frac{1}{p}\sum_{i=1}^p\mathrm{E}\ln\{\alpha\bar{Z}(t)+\beta Z_i(t)\}+\ln\sigma_i^{2t}-1.\end{aligned}\tag{6}$$

Note that $\mathrm{E}\ln\{\alpha\bar{Z}(t)+\beta Z_i(t)\}$ does not have a closed form as it involves the expectation of the logarithm of a linear combination of independent but non-identically distributed chi-squared random variables. By Beckenbach and Bellman [3], we have

$$\frac{1}{\alpha+\beta}\left\{\frac{\alpha}{p}\sum_{i=1}^p\ln Z_i(t)+\beta\ln Z_i(t)\right\}\le\ln\{\alpha\bar{Z}(t)+\beta Z_i(t)\}\le\alpha\bar{Z}(t)+\beta Z_i(t).$$

Since both $Z_i(t)$ and $\ln Z_i(t)$ are integrable for $t > -\nu/2$, then $\mathrm{E}\ln\{\alpha\bar{Z}(t)+\beta Z_i(t)\}$ exists for any given α and β. Furthermore, since $-\ln x$ is a strictly convex function of x and $\alpha\bar{Z}(t)+\beta Z_i(t)$ is a linear function of α and β, then $-\mathrm{E}\ln\{\alpha\bar{Z}(t)+\beta Z_i(t)\}$ is a strictly convex function of α and β. Therefore, $R_T(\alpha,\beta;\boldsymbol{\sigma}^{2t})$ is a strictly convex function of α and β.

Lemma 1. *For any $\alpha \geq 0$, $\beta \geq 0$, and $\alpha+\beta > 0$, the minimum of $R_T(\alpha,\beta;\boldsymbol{\sigma}^{2t})$ is obtained on the line $A_1(t)\alpha+\beta = 1$.*

The proof of Lemma 1 is given in Appendix D. Replacing β by $1-A_1(t)\alpha$ in (6), the average risk reduces to

$$R_{T_1}(\alpha;\boldsymbol{\sigma}^{2t}) = -\frac{1}{p}\sum_{i=1}^{p}\mathrm{E}\ln[\alpha\bar{Z}(t)+\{1-A_1(t)\alpha\}Z_i(t)] + \frac{1}{p}\sum_{i=1}^{p}\ln\sigma_i^{2t},$$

where $0 \leq \alpha \leq 1/A_1(t)$.

Theorem 5. *For any fixed $p \geq 2$, ν, and nonzero $|t| < \nu/2$, $R_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})$ is a strictly convex function of α on $[0, 1/A_1(t)]$ that satisfies*

(i) $R'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=0} < 0$,
(ii) $R'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=1/A_1(t)} \geq 0$ *where the equality holds if and only if $\sigma_i^2 = \sigma^2$ for all i.*

The proof of Theorem 5 is given in Appendix E. Let $\alpha^*_{T_1} = \mathrm{argmin}_{\alpha\in[0,1/A_1(t)]}R_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})$ be the optimal shrinkage parameter under the Stein loss function. By Theorem 5, there exists a unique $\alpha^*_{T_1}$ in $(0, 1/A_1(t)]$ that satisfies $R'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t}) = 0$. The optimal shrinkage estimator under the Stein loss function is then $\hat{\sigma}^{2t}_{i,T_1} = \alpha^*_{T_1}\bar{Z}(t)+\{1-A_1(t)\alpha^*_{T_1}\}Z_i(t)$. Since $\alpha^*_{T_1} > 0$, then $\hat{\sigma}^{2t}_{i,T_1}$ has a smaller average risk than $Z_i(t)$. When $\sigma_i^2 = \sigma^2$ for all i, it is seen that $\alpha^*_{T_1} = 1/A_1(t) = 1$. Consequently $\hat{\sigma}^{2t}_{i,T_1}$ reduces to the pooled variance estimator $\bar{Z}(t)$. The following theorem indicates that there is no need to borrow information when ν is large.

Theorem 6. *For any fixed $p \geq 2$ and nonzero t, as $\nu \to \infty$, we have*

(i) $\alpha^*_{T_1} \to 0$ *when σ_i^2 are not all the same,*
(ii) $R_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})$ *tends to a constant function of α when $\sigma_i^2 = \sigma^2$ for all i.*

The proof of Theorem 6 is omitted due to its simplicity. We now consider the optimal shrinkage under the constraint of $\alpha+\beta = 1$. Substituting $\beta = 1-\alpha$ in (6), we have

$$R_{T_2}(\alpha;\boldsymbol{\sigma}^{2t}) = \alpha\{A_1(t)-1\} - \frac{1}{p}\sum_{i=1}^{p}\mathrm{E}\ln\{\alpha\bar{Z}(t)+(1-\alpha)Z_i(t)\} + \frac{1}{p}\sum_{i=1}^{p}\ln\sigma_i^{2t}$$

for $0 \leq \alpha \leq 1$. Denote the optimal shrinkage parameter under the constraint of $\alpha+\beta = 1$ as $\alpha^*_{T_2} = \mathrm{argmin}_{\alpha\in[0,1]}R_{T_2}(\alpha;\boldsymbol{\sigma}^{2t})$ and the corresponding optimal shrinkage estimator as $\hat{\sigma}^{2t}_{i,T_2} = \alpha^*_{T_2}\bar{Z}(t)+(1-\alpha^*_{T_2})Z_i(t)$. Similarly it can be shown that for any fixed $p \geq 2$, ν, and nonzero $|t| < \nu/2$, $R_{T_2}(\alpha;\boldsymbol{\sigma}^{2t})$ is a strictly convex function of α on $[0, 1]$ that satisfies (i) $R'_{T_2}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=0} < 0$, and (ii) $R'_{T_2}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=1} \geq 0$ where the equality holds if and only if $\sigma_i^2 = \sigma^2$ for all i. Therefore, there exists a unique $\alpha^*_{T_2}$ in $(0, 1]$ that satisfies $R'_{T_2}(\alpha;\boldsymbol{\sigma}^{2t}) = 0$. Since $\alpha^*_{T_2} > 0$, $\hat{\sigma}^{2t}_{i,T_2}$ has a smaller average risk than the original estimator $Z_i(t)$. When $\sigma_i^2 = \sigma^2$ for all i, it can be shown that $\alpha^*_{T_2} = 1$. Consequently $\hat{\sigma}^{2t}_{i,T_2}$ reduces to the pooled variance estimator $\bar{Z}(t)$. Finally, for any fixed $p \geq 2$ and nonzero t, as $\nu \to \infty$ we have (i) $\alpha^*_{T_2} \to 0$ when σ_i^2 are not all the same, and (ii) $R_{T_2}(\alpha;\boldsymbol{\sigma}^{2t})$ tends to a constant function of α when $\sigma_i^2 = \sigma^2$ for all i.

3.2. Estimation of the optimal shrinkage parameters

The optimal shrinkage parameter $\alpha^*_{T_1}$ is the unique solution to

$$R'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t}) = -\frac{1}{p}\sum_{i=1}^{p}\mathrm{E}\left[\frac{\bar{Z}(t)-A_1(t)Z_i(t)}{\alpha\bar{Z}(t)+\{1-A_1(t)\alpha\}Z_i(t)}\right] = 0$$

in $(0, 1/A_1(t)]$. We estimate $A_1(t)$ by $\tilde{A}_1(t) = \bar{Z}(t)\bar{Z}(-t)$, and $R'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})$ by

$$\hat{R}'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t}) = -\frac{1}{p}\sum_{i=1}^{p}\frac{\bar{Z}(t)-\tilde{A}_1(t)Z_i(t)}{\alpha\bar{Z}(t)+\{1-\tilde{A}_1(t)\alpha\}Z_i(t)}.$$

It is easy to verify that $\hat{R}'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=0} = \bar{Z}(t)\sum_{i=1}^{p}Z_i^{-t}\{h(-t)-1/h(t)\}/p < 0$, $\hat{R}'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=1/\tilde{A}_1(t)} = \tilde{A}_1(t)\{\tilde{A}_1(t)-1\}$, and $\hat{R}''_{T_2}(\alpha;\boldsymbol{\sigma}^{2t}) > 0$ for any $\alpha \in (0, 1/\tilde{A}_1(t)]$. Note that $\hat{R}'_{T_1}(\alpha;\boldsymbol{\sigma}^{2t})|_{\alpha=1/\tilde{A}_1(t)}$ is not guaranteed to be non-negative. Nevertheless,

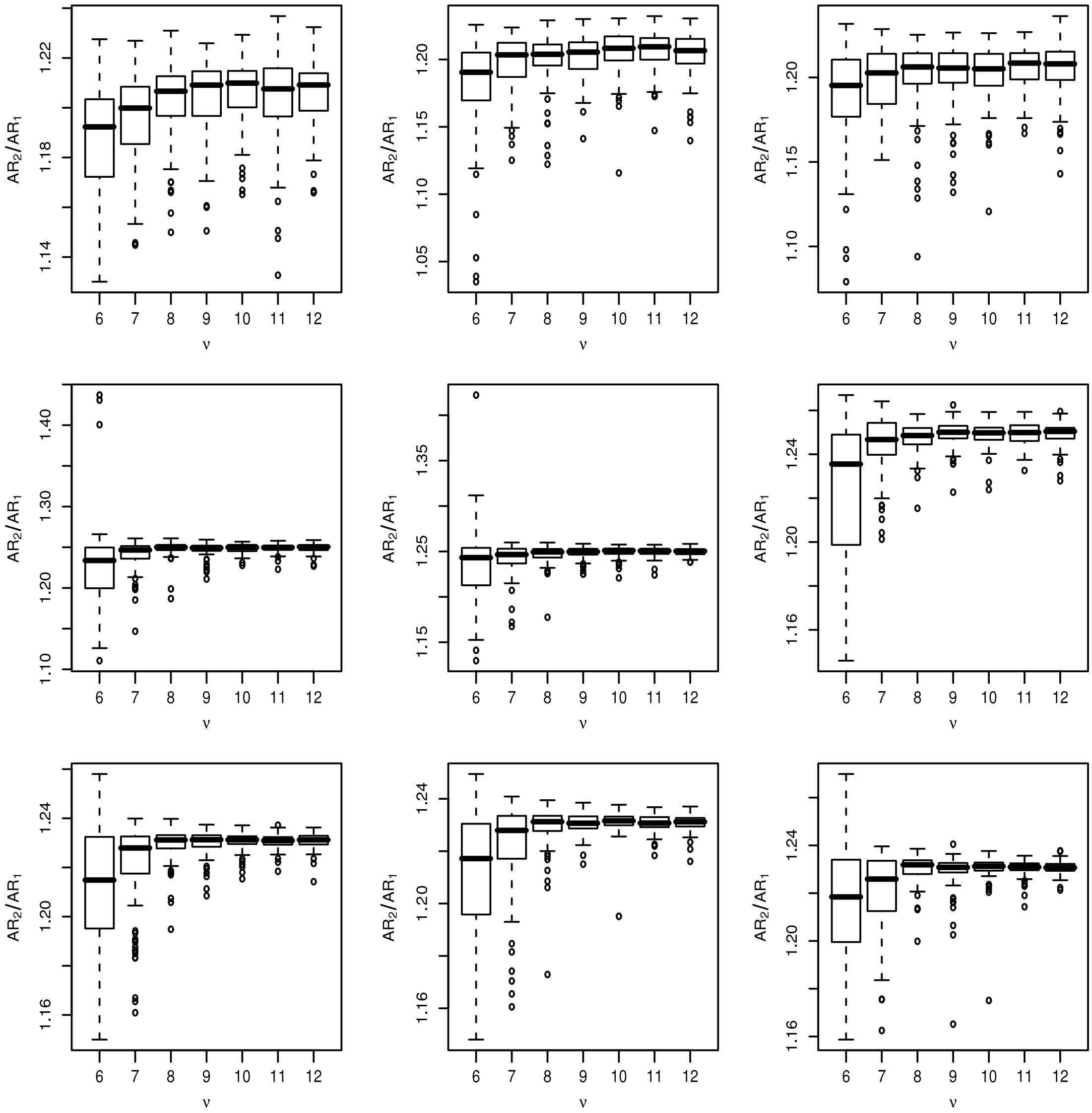

Fig. 1. Boxplots of the ratios of $AR(\tilde{\alpha}^*_{Q_2})/AR(\tilde{\alpha}^*_{Q_1})$ under the squared loss function. Rows from top to bottom correspond to the shape parameter γ with values 3, 6 and 9, respectively. Columns from left to right correspond to the mean parameter μ with values 1/3, 1 and 3, respectively.

we have $\tilde{A}_1(t) \xrightarrow{\text{a.s.}} A_1(t) \geq 1$ as $\nu \to \infty$ and $\tilde{A}_1(t) \xrightarrow{\text{a.s.}} \mu_t\mu_{-t} \geq 1$ as $p \to \infty$. Therefore, $\hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=1/\tilde{A}_1(t)}$ is non-negative when either ν or p is large. Consequently, there is a unique α in $(0, 1/\tilde{A}_1(t)]$ that satisfies $\hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t}) = 0$ when ν or p is large. We denote the solution as $\tilde{\alpha}^*_{T_1}$, and the estimated optimal shrinkage estimator by $\tilde{\sigma}^{2t}_{i,T_1} = \tilde{\alpha}^*_{T_1}\bar{Z}(t) + \{1 - \tilde{A}_1(t)\tilde{\alpha}^*_{T_1}\}Z_i(t)$.

Similarly, for $\alpha^*_{T_2}$ under the constraint of $\alpha + \beta = 1$, the estimated optimal shrinkage parameter $\tilde{\alpha}^*_{T_2}$ is derived by solving the equation $\hat{R}'_{T_2}(\alpha; \boldsymbol{\sigma}^{2t}) = 0$ where

$$\hat{R}'_{T_2}(\alpha; \boldsymbol{\sigma}^{2t}) = \tilde{A}_1(t) - 1 - \frac{1}{p}\sum_{i=1}^{p}\left\{\frac{\bar{Z}(t) - Z_i(t)}{\alpha\bar{Z}(t) + (1-\alpha)Z_i(t)}\right\}.$$

The corresponding optimal shrinkage estimator is denoted as $\tilde{\sigma}^{2t}_{i,T_2} = \tilde{\alpha}^*_{T_2}\bar{Z}(t) + (1 - \tilde{\alpha}^*_{T_2})Z_i(t)$. The following theorems summarize the asymptotic properties of the estimated optimal shrinkage parameters under the Stein loss function when $\nu \to \infty$ or $p \to \infty$.

Theorem 7. *For any fixed $p \geq 2$ and nonzero t, as $\nu \to \infty$, we have $\tilde{\alpha}^*_{T_1} \xrightarrow{\text{a.s.}} 0$ and $\tilde{\alpha}^*_{T_2} \xrightarrow{\text{a.s.}} 0$ when σ_i^2 are not all the same.*

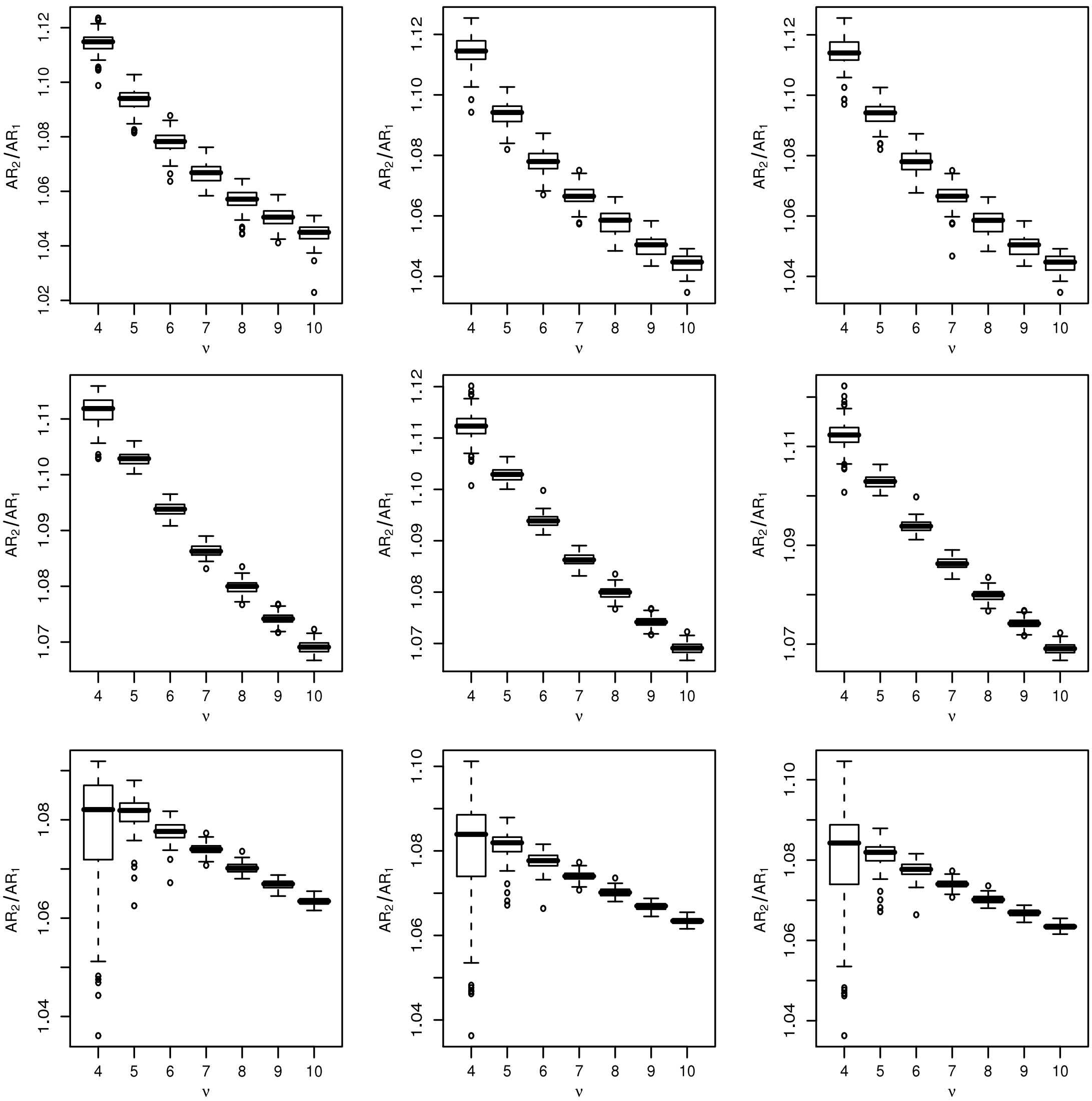

Fig. 2. Boxplots of the ratios of $AR(\tilde{\alpha}^*_{T_2})/AR(\tilde{\alpha}^*_{T_1})$ under the Stein loss function. Rows from top to bottom correspond to the shape parameter γ with values 3, 6 and 9, respectively. Columns from left to right correspond to the mean parameter μ with values 1/3, 1 and 3, respectively.

Theorem 8. *For any fixed* ν *and nonzero* $|t| < \nu/2$*, assume that* $\sigma_i^2 \overset{\text{i.i.d.}}{\sim} F$ *with* $0 < \mu_\xi < \infty$ *for* $\xi = 2t$ *and* $-2t$*. In addition, assume that* $\mathrm{E}[\sum_{i=1}^p Z_i^2(t)/\{p\bar{Z}(t)\}]^2 < \infty$ *and* $\mathrm{E}\{A_1(t) - \mu_t\mu_{-t}\}^2 \to 0$ *as* $p \to \infty$*. Then, as* $p \to \infty$*,*

(i) $\hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t}) - R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t}) \overset{\text{a.s.}}{\to} 0$ *uniformly for* $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ *with any* $c > 1$*. In addition,* $\tilde{\alpha}^*_{T_1} - \alpha^*_{T_1} \overset{\text{a.s.}}{\to} 0$ *as* $p \to \infty$*.*

(ii) $\hat{R}'_{T_2}(\alpha; \boldsymbol{\sigma}^{2t}) - R'_{T_2}(\alpha; \boldsymbol{\sigma}^{2t}) \overset{\text{a.s.}}{\to} 0$ *uniformly for* $\alpha \in (0, 1]$*. In addition,* $\tilde{\alpha}^*_{T_2} - \alpha^*_{T_2} \overset{\text{a.s.}}{\to} 0$ *as* $p \to \infty$*.*

The proofs of Theorems 7 and 8 are given in Appendices F and G, respectively.

4. Simulation study

As in [15], we set $p = 5000$ in this section. We simulate σ_i^2 for $i = 1, \ldots, p$ from a gamma distribution with shape parameter γ and mean parameter μ. We consider all nine combinations of $\gamma = 3, 6$ and 9 and $\mu = 1/3, 1$ and 3. For each given σ_i^2, we simulate $\nu + 1$ observations from $\mathrm{N}(\theta_i, \sigma_i^2)$ where θ_i is a random sample from $\mathrm{N}(0, 1)$, and then compute Z_i as the sample variance based on these $\nu + 1$ observations. Based on Theorems 4 and 8, we need $\nu > 4|t|$ for the squared loss and $\nu > 2|t|$ for the Stein loss. For $t = 1$ we need $\nu > 4$ and $\nu > 2$ for the squared and the Stein losses respectively. The estimates of shrinkage parameters are unstable when $\nu = 5$ under the squared loss and when $\nu = 3$ under the Stein loss. Therefore, we consider $\nu = 6, 7, \ldots, 12$ for the estimation under the squared loss function, and $\nu = 4, 5, \ldots, 10$ for the estimation under the Stein loss function.

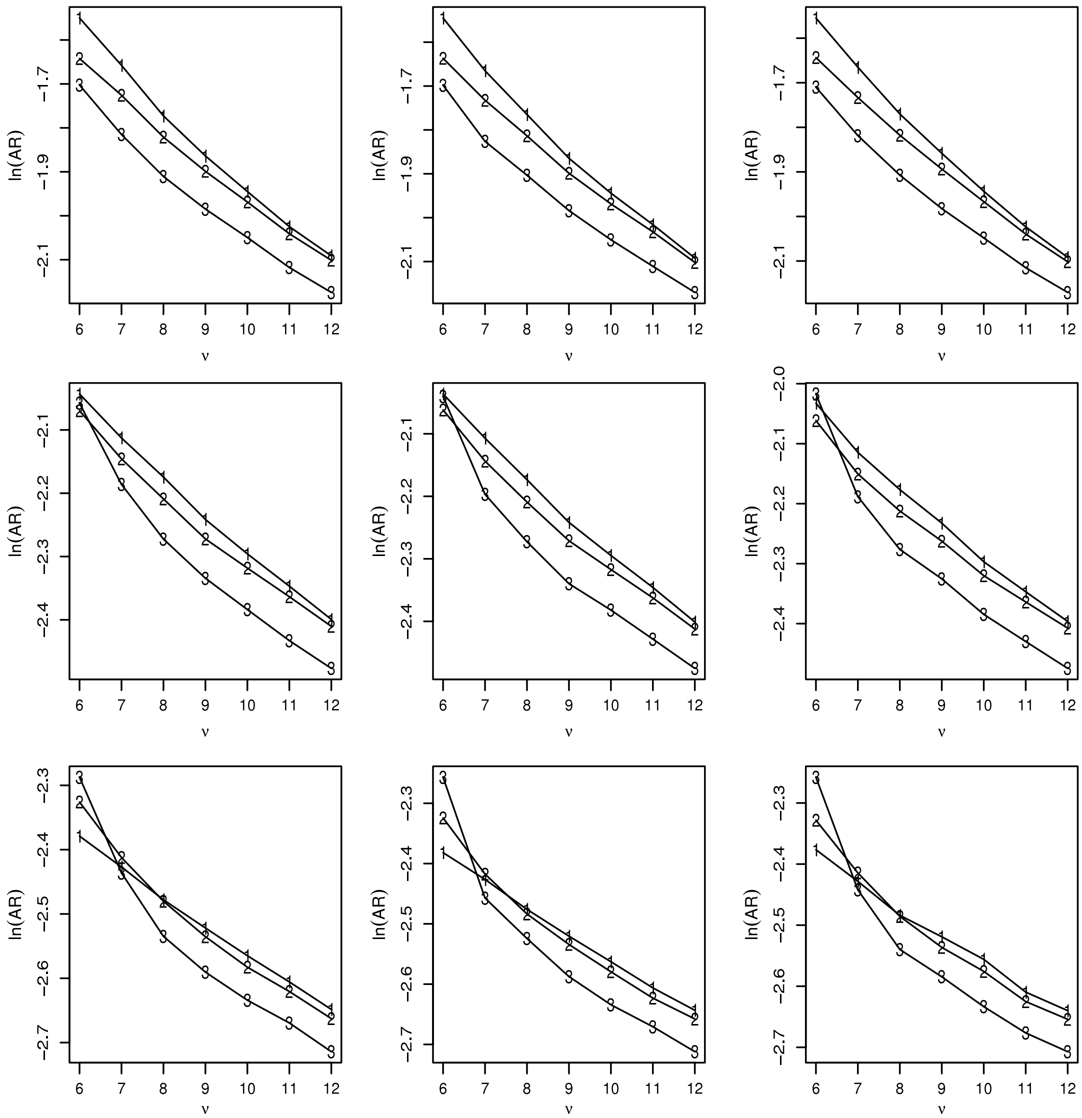

Fig. 3. Plots of log average risks for estimating σ_i^2 under the squared loss function. Lines marked with "1", "2" and "3" correspond to the CHQBC estimator, the TW estimator under the squared loss function, and $\tilde{\sigma}^2_{i,Q_1}$, respectively. Rows from top to bottom correspond to the shape parameter γ with values 3, 6 and 9, respectively. Columns from left to right correspond to the mean parameter μ with values 1/3, 1 and 3, respectively.

For each setting, we repeat the simulation 100 times and then compute the following average risk

$$AR = \frac{1}{100p} \sum_{r=1}^{100} \sum_{i=1}^{p} L(\sigma_{ir}^2, \hat{\sigma}_{ir}^2), \tag{7}$$

where r represents simulation replications and L can be either the squared or the Stein loss function. To save space, we present the comparison results for estimating σ_i^2 (i.e. $t = 1$) only. Comparative results for other values of t are similar.

We present results from two simulations. The purpose of our first simulation is to evaluate the loss of efficiency due to the constraint $\alpha + \beta = 1$. Let $AR(\tilde{\alpha}^*_{Q_1})$, $AR(\tilde{\alpha}^*_{Q_2})$, $AR(\tilde{\alpha}^*_{T_1})$ and $AR(\tilde{\alpha}^*_{T_2})$ be the average risks computed using (7) for the estimated optimal estimators $\tilde{\sigma}^2_{i,Q_1}$, $\tilde{\sigma}^2_{i,Q_2}$, $\tilde{\sigma}^2_{i,T_1}$ and $\tilde{\sigma}^2_{i,T_2}$ respectively. To evaluate the amount of efficiency loss under the constraint of $\alpha + \beta = 1$, we plot the ratios of $AR(\tilde{\alpha}^*_{Q_2})/AR(\tilde{\alpha}^*_{Q_1})$ and $AR(\tilde{\alpha}^*_{T_2})/AR(\tilde{\alpha}^*_{T_1})$ in Figs. 1 and 2 for the squared and the Stein loss functions, respectively. We note that the ratios are all greater than 1 as expected. That is, the shrinkage variance estimators without the constraint have smaller risks than those with the constraint. Under the squared loss, the median efficiency loss of $AR(\tilde{\alpha}^*_{Q_2})$ over $AR(\tilde{\alpha}^*_{Q_1})$ has a range between 19% and 25%. Under the Stein loss, the median efficiency loss of $AR(\tilde{\alpha}^*_{T_2})$ over $AR(\tilde{\alpha}^*_{T_1})$ has a range between 4% and 12%. It is interesting to note that the median efficiency loss under the squared loss

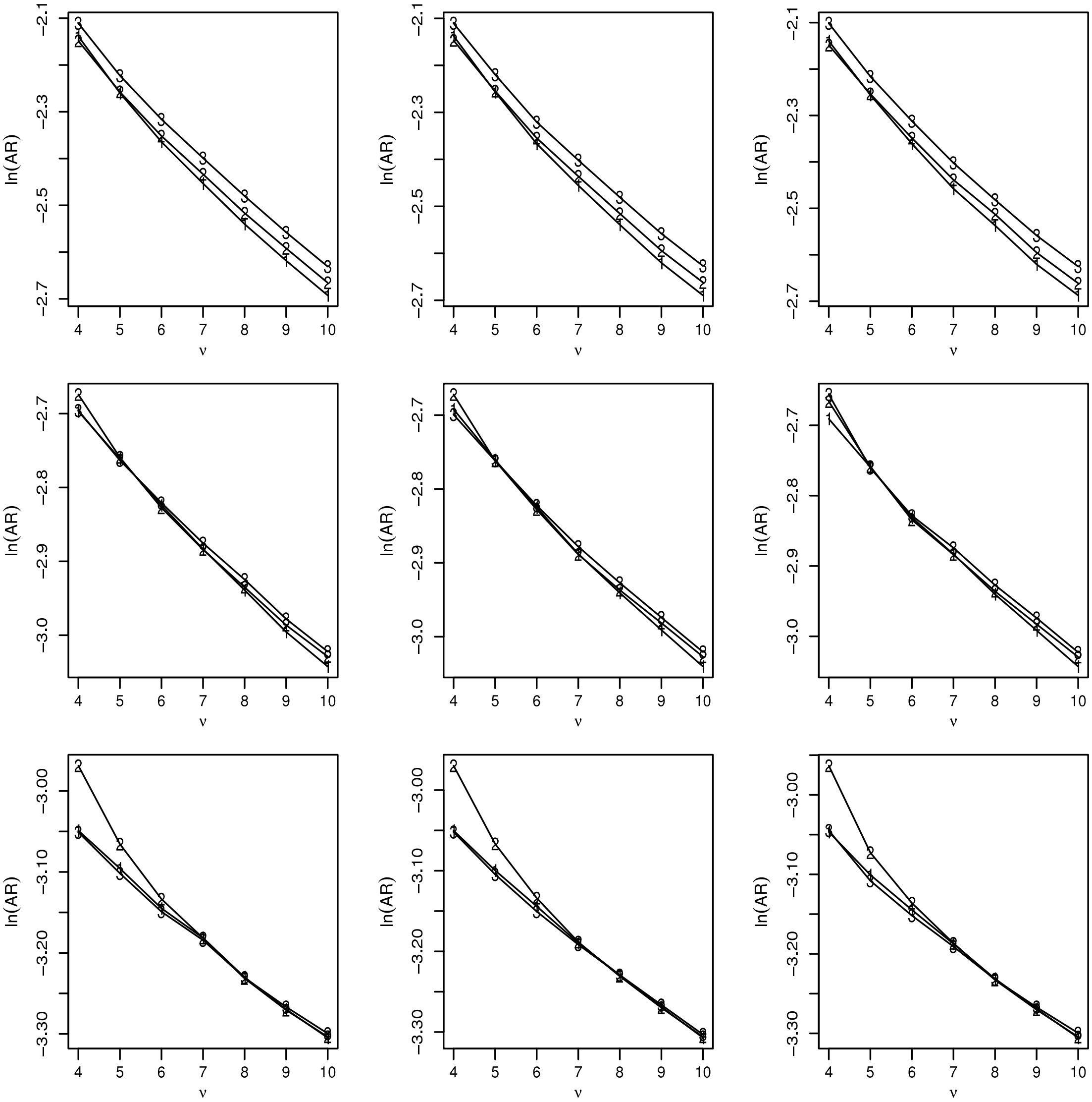

Fig. 4. Plots of log average risks for estimating σ_i^2 under the Stein loss function. Lines marked with "1", "2" and "3" correspond to the CHQBC estimator, the TW estimator under the Stein loss function, and $\tilde{\sigma}^2_{i,T_1}$, respectively. Rows from top to bottom correspond to the shape parameter γ with values 3, 6 and 9, respectively. Columns from left to right correspond to the mean parameter μ with values 1/3, 1 and 3, respectively.

function remains constant over different values of ν while the median efficiency loss under the Stein loss function decreases as ν increases.

The purpose of our second simulation is to compare the performance of the proposed estimators $\tilde{\sigma}^2_{i,Q_1}$ and $\tilde{\sigma}^2_{i,T_1}$ with the shrinkage estimators in [4,15] which are referred to as the CHQBC and TW estimators, respectively. Note that both the CHQBC and TW estimators are based on the logarithmic scale which shrink towards the geometric mean. All the shrinkage estimators perform considerably better than the original estimator Z_i. For simplicity, we will not present the average risk of Z_i. Figs. 3 and 4 show the average risks on the logarithmic scale under the squared and the Stein loss functions, respectively. We note that, except for very small ν, the proposed estimator $\tilde{\sigma}^2_{i,Q_1}$ has the smallest average risk in most settings under the squared loss. Under the Stein loss function, all three estimators have similar performance except for the case when $\gamma = 3$ where the proposed estimator $\tilde{\sigma}^2_{i,T_1}$ performs slightly worse than the CHQBC and TW estimators.

Acknowledgments

TT's research was supported by a grant from Hong Kong Research Grants Council (HKBU 202711) and two grants from Hong Kong Baptist University (FRG1/10-11/031 and FRG2/10-11/020). YW's research was supported by a grant from National Science Foundation (DMS 0706886). The authors thank the editor, the associate editor, and two referees for their constructive comments and suggestions that have led to a substantial improvement in the article.

Appendix A. Proof of Theorem 2

For any fixed $p \geq 2$ and nonzero t, as $\nu \to \infty$, we have $A_3(t) \to 1, A_2(t) \to (\bar{\sigma}^{2t})^2\bar{\sigma}^{-4t}$, and $A_4(t) \to A_1(t)$. This leads to $A_1(t)A_3(t) - A_4(t) \to 0$ and $\{A_2(t) - A_1(t)A_4(t)\} - \{A_2(t)A_3(t) - A_4^2(t)\} \to 0$ as $\nu \to \infty$. In addition, $A_2(t)A_3(t) - A_4^2(t) \to C$ as $\nu \to \infty$, where $C = (\bar{\sigma}^{2t})^2 \sum_{i=1}^p \sum_{j>i}(\sigma_i^{-2t} - \sigma_j^{-2t})^2/p^2$ with $p \geq 2$. It is seen that $C > 0$ as long as σ_i^2 are not all the same. Therefore, $\alpha_{Q_1}^* \to 0$ and $\beta_{Q_1}^* \to 1$ when σ_i^2 are not all the same. When $\sigma_i^2 = \sigma^2$ for all i, we have $\alpha_{Q_1}^* = [\{A_3(t) - 1\}/p + 1]^{-1} \to 1$ and $\beta_{Q_1}^* = 0$ as $\nu \to \infty$.

Appendix B. Proof of Theorem 3

We first prove Theorem 3(i). For any fixed $p \geq 2$ and nonzero t, as $\nu \to \infty$, we have $Z_i(t) \overset{\text{a.s.}}{\to} \sigma_i^{2t}$ and $Z_i(-t) \overset{\text{a.s.}}{\to} \sigma_i^{-2t}$. This leads to $\bar{Z}(t) \overset{\text{a.s.}}{\to} \bar{\sigma}^{2t}, \bar{Z}(-t) \overset{\text{a.s.}}{\to} \bar{\sigma}^{-2t}$, and $\tilde{A}_1(t) \overset{\text{a.s.}}{\to} A_1(t)$ as $\nu \to \infty$. Similarly, we have $\tilde{A}_2(t) \overset{\text{a.s.}}{\to} (\bar{\sigma}^{2t})^2\bar{\sigma}^{-4t}, A_3(t) \to 1$, and $\tilde{A}_4(t) \overset{\text{a.s.}}{\to} A_1(t)$ as $\nu \to \infty$. Therefore, $\tilde{A}_1(t)A_3(t) - \tilde{A}_4(t) \overset{\text{a.s.}}{\to} 0$ and $\tilde{A}_2(t)A_3(t) - \tilde{A}_4^2(t) \overset{\text{a.s.}}{\to} C$ as $\nu \to \infty$ where $C = (\bar{\sigma}^{2t})^2 \sum_{i=1}^p \sum_{j>i}(\sigma_i^{-2t} - \sigma_j^{-2t})^2/p^2 \geq 0$. Note that $C = 0$ if and only if $\sigma_i^2 = \sigma^2$ for all i. Therefore, $\tilde{\alpha}_{Q_1}^* \overset{\text{a.s.}}{\to} 0$ when σ_i^2 are not all the same. In addition, noting that $\{\tilde{A}_2(t) - \tilde{A}_1(t)\tilde{A}_4(t)\} - \{\tilde{A}_2(t)A_3(t) - \tilde{A}_4^2(t)\} = \tilde{A}_2(t)\{1 - A_3(t)\} - \tilde{A}_4(t)\{\tilde{A}_1(t) - \tilde{A}_4(t)\} \overset{\text{a.s.}}{\to} 0$ as $\nu \to \infty$, we have $\tilde{\beta}_{Q_1}^* \overset{\text{a.s.}}{\to} 1$ when σ_i^2 are not all the same.

The proof of Theorem 3(ii) is similar and thus is omitted.

Appendix C. Proof of Theorem 4

We prove Theorem 4(i) only. By the strong law of large numbers (SLLN), as $p \to \infty$, we have $A_1(t) \overset{\text{a.s.}}{\to} \mu_t\mu_{-t}, A_2(t) \overset{\text{a.s.}}{\to} \mu_t^2\mu_{-2t}$, and $A_4(t) \overset{\text{a.s.}}{\to} \mu_t\mu_{-t}$. Then for any fixed ν and nonzero t, as $p \to \infty$,

$$\alpha_{Q_1}^* \overset{\text{a.s.}}{\to} \frac{\mu_t\mu_{-t}A_3(t) - \mu_t\mu_{-t}}{\mu_t^2\mu_{-2t}A_3(t) - \mu_t^2\mu_{-t}^2} = \frac{A_3(t) - 1}{\mu_t\mu_{-t}\{A_3(t)\mu_{-2t}/\mu_{-t}^2 - 1\}}.$$

Noting that $\mu_{-2t}/\mu_{-t}^2 \geq 1$ and $A_3(t) > 1$, we have $0 < \alpha_{Q_1}^* < 1/(\mu_t\mu_{-t})$.

For any $\xi > -\nu/2$, $Z_i(\xi)$ are i.i.d. random variables with $\mathrm{E}Z_i(\xi) = \mathrm{E}[\mathrm{E}\{Z_i(\xi)|\sigma_i^2\}] = \mathrm{E}\sigma_i^{2\xi} = \mu_\xi$. By SLLN, $\bar{Z}(\xi) \overset{\text{a.s.}}{\to} \mu_\xi$ as $p \to \infty$. Therefore, for any fixed ν and nonzero $|t| < \nu/4$, we have $\tilde{A}_1(t) \overset{\text{a.s.}}{\to} \mu_t\mu_{-t}, \tilde{A}_2(t) \overset{\text{a.s.}}{\to} \mu_t^2\mu_{-2t}$, and $\tilde{A}_4(t) \overset{\text{a.s.}}{\to} \mu_t\mu_{-t}$ as $p \to \infty$. This implies that $\alpha_{Q_1}^*$ and $\tilde{\alpha}_{Q_1}^*$ have the same limit. Thus, by the Slutsky theorem, we have $\tilde{\alpha}_{Q_1}^* - \alpha_{Q_1}^* \overset{\text{a.s.}}{\to} 0$ as $p \to \infty$. Finally, by the same arguments we have $\tilde{\beta}_{Q_1}^* - \beta_{Q_1}^* \overset{\text{a.s.}}{\to} 0$ as $p \to \infty$.

Appendix D. Proof of Lemma 1

Taking the first partial derivatives with respect to α and β yields

$$\frac{\partial}{\partial\alpha}R_T(\alpha, \beta; \boldsymbol{\sigma}^{2t}) = A_1(t) - \frac{1}{p}\sum_{i=1}^p \mathrm{E}\left\{\frac{\bar{Z}(t)}{\alpha\bar{Z}(t) + \beta Z_i(t)}\right\} \overset{\text{set}}{=} 0, \tag{8}$$

$$\frac{\partial}{\partial\beta}R_T(\alpha, \beta; \boldsymbol{\sigma}^{2t}) = 1 - \frac{1}{p}\sum_{i=1}^p \mathrm{E}\left\{\frac{Z_i(t)}{\alpha\bar{Z}(t) + \beta Z_i(t)}\right\} \overset{\text{set}}{=} 0. \tag{9}$$

Multiplying (8) and (9) by α and β respectively, and then adding them together, we have $A_1(t)\alpha + \beta = 1$. Note that $R_T(\alpha, \beta; \boldsymbol{\sigma}^{2t})$ is a strictly convex function of α and β. Thus if the minimum value of $R_T(\alpha, \beta; \boldsymbol{\sigma}^{2t})$ is inside the open set $\{(\alpha, \beta) : \alpha > 0, \beta > 0\}$, it must satisfy $A_1(t)\alpha + \beta = 1$.

We now consider the case when the minimum value of $R_T(\alpha, \beta; \boldsymbol{\sigma}^{2t})$ locates on the boundary. First consider the case when $\alpha = 0$ and $\beta > 0$. Then $R_T(0, \beta; \boldsymbol{\sigma}^{2t}) = \beta - \ln\beta - \sum_{i=1}^p \mathrm{E}\ln Z_i(t)/p + \sum_{i=1}^p \ln\sigma_i^{2t}/p - 1$. Taking the first derivative, we have $(\partial/\partial\beta)R_T(0, \beta; \boldsymbol{\sigma}^{2t}) = 1 - 1/\beta \overset{\text{set}}{=} 0$. This leads to $\beta = 1$. Since $(\partial^2/\partial\beta^2)R_T(0, \beta; \boldsymbol{\sigma}^{2t}) = 1/\beta^2 > 0$, then $R_T(0, \beta; \boldsymbol{\sigma}^{2t})$ is minimized at $(\alpha, \beta) = (0, 1)$. Next consider the case when $\alpha > 0$ and $\beta = 0$, $R_T(\alpha, 0; \boldsymbol{\sigma}^{2t}) = A_1(t)\alpha - \ln\alpha - \mathrm{E}\ln\bar{Z}(t) + \sum_{i=1}^p \ln\sigma_i^{2t}/p - 1$. Taking the first derivative, we have $(\partial/\partial\alpha)R_T(\alpha, 0; \boldsymbol{\sigma}^{2t}) = A_1(t) - 1/\alpha \overset{\text{set}}{=} 0$. This leads to $\alpha = 1/A_1(t)$. Since $(\partial^2/\partial\alpha^2)R_T(\alpha, 0; \boldsymbol{\sigma}^{2t}) = 1/\alpha^2 > 0$, then $R_T(\alpha, 0; \boldsymbol{\sigma}^{2t})$ is minimized at $(\alpha, \beta) = (1/A_1(t), 0)$. We note that both $(\alpha, \beta) = (0, 1)$ and $(\alpha, \beta) = (1/A_1(t), 0)$ satisfy $A_1(t)\alpha + \beta = 1$.

Appendix E. Proof of Theorem 5

We first prove Theorem 5(i). The first derivative of $R_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})$ evaluated at $\alpha = 0$

$$R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=0} = -\frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left\{\frac{\bar{Z}(t) - A_1(t)Z_i(t)}{Z_i(t)}\right\} = A_1(t) - \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left\{\frac{\bar{Z}(t)}{Z_i(t)}\right\}.$$

Noting that $Z_i(t)$ are independent of each other, we have

$$\begin{aligned}\mathrm{E}\left\{\frac{\bar{Z}(t)}{Z_i(t)}\right\} &= \frac{1}{p}\sum_{j\neq i} \mathrm{E}Z_j(t)\mathrm{E}Z_j^{-1}(t) + \frac{1}{p} \\ &= \left(\frac{1}{p}\sum_{j=1}^{p}\sigma_j^{2t}\right)\left\{\frac{K_1(t)}{\sigma_i^{2t}}\right\} + \frac{1}{p}\{1 - K_1(t)\},\end{aligned}$$

where $\mathrm{E}Z_i^{-1}(t) = K_1(t)/\sigma_i^{2t}$ and $K_1(t) = \Gamma(\nu/2 + t)\,\Gamma(\nu/2 - t)\,/\Gamma^2(\nu/2)$. Then

$$\frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left\{\frac{\bar{Z}(t)}{Z_i(t)}\right\} = K_1(t)A_1(t) + \frac{1}{p}\left\{1 - K_1(t)\right\}.$$

Consequently $R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=0} = \{1 - K_1(t)\}\{A_1(t) - 1/p\}$. Finally, noting that $A_1(t) \geq 1$ and $K_1(t) > 1$ for any nonzero $|t| < \nu/2$, we have $R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=0} < 0$ for any $p \geq 2$.

We now prove Theorem 5(ii). The first derivative of $R_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})$ evaluated at $\alpha = 1/A_1(t)$

$$R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=1/A_1(t)} = -\frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left\{\frac{\bar{Z}(t) - A_1(t)Z_i(t)}{\bar{Z}(t)/A_1(t)}\right\} = A_1^2(t) - A_1(t).$$

Note that $A_1(t) \geq 1$. We have $R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=1/A_1(t)} \geq 0$ where the equality holds if and only if $\sigma_i^2 = \sigma^2$ for all g.

Appendix F. Proof of Theorem 7

It has been shown in Appendix B that, for any fixed $p \geq 2$ and nonzero $|t| < \nu/2$, $Z_i(t) \xrightarrow{\text{a.s.}} \sigma_i^{2t}$, $\bar{Z}(t) \xrightarrow{\text{a.s.}} \bar{\sigma}^{2t}$, and $\tilde{A}_1(t) \xrightarrow{\text{a.s.}} A_1(t)$ as $\nu \to \infty$. Noting that $K_1(t) \to 1$ as $\nu \to \infty$, we have

$$\hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})|_{\alpha=0} = -\frac{1}{p}\sum_{i=1}^{p} \frac{\bar{Z}(t) - \tilde{A}_1(t)Z_i(t)}{Z_i(t)} = \tilde{A}_1(t)\{1 - K_1(t)\} \xrightarrow{\text{a.s.}} 0,$$

and

$$\begin{aligned}\hat{R}''_{T_1}(\alpha; \boldsymbol{\sigma}^{2t}) &= \frac{1}{p}\sum_{i=1}^{p}\left[\frac{\bar{Z}(t) - \tilde{A}_1(t)Z_i(t)}{\alpha\bar{Z}(t) + \{1 - \tilde{A}_1(t)\alpha\}Z_i(t)}\right]^2 \\ &\xrightarrow{\text{a.s.}} \frac{1}{p}\sum_{i=1}^{p}\left[\frac{\bar{\sigma}^{2t} - A_1(t)\sigma_i^{2t}}{\alpha\bar{\sigma}^{2t} + \{1 - \alpha A_1(t)\}\sigma_i^{2t}}\right]^2 \\ &\geq 0,\end{aligned}$$

where the equality holds if and only if $\sigma_i^2 = \sigma^2$ for all i. This implies that, as $\nu \to \infty$, $\hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})$ is a strictly increasing function of α with a minimum value at $\alpha = 0$ when σ_i^2 are not all the same. Therefore, $\tilde{\alpha}^*_{T_1} \xrightarrow{\text{a.s.}} 0$ as $\nu \to \infty$ when σ_i^2 are not all the same.

The proof of $\tilde{\alpha}^*_{T_2} \xrightarrow{\text{a.s.}} 0$ is similar and thus is omitted.

Appendix G. Proof of Theorem 8

We prove Theorem 8(i) only. Let

$$\begin{aligned}H_i(\bar{Z}(t), \alpha, \tilde{A}_1(t)) &= -\frac{\bar{Z}(t) - \tilde{A}_1(t)Z_i(t)}{\alpha\bar{Z}(t) + \{1 - \tilde{A}_1(t)\alpha\}Z_i(t)}, \\ H_i(\bar{Z}(t), \alpha, A_1(t)) &= -\frac{\bar{Z}(t) - A_1(t)Z_i(t)}{\alpha\bar{Z}(t) + \{1 - A_1(t)\alpha\}Z_i(t)}, \\ H_i(\mu_t, \alpha, \mu_t\mu_{-t}) &= -\frac{\mu_t - \mu_t\mu_{-t}Z_i(t)}{\alpha\mu_t + \{1 - \mu_t\mu_{-t}\alpha\}Z_i(t)}.\end{aligned}$$

For a fixed ν and $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$, we have

$$\left|\hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t}) - R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})\right| \le \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where

$$\mathrm{I} = \left|\frac{1}{p}\sum_{i=1}^{p} H_i(\bar{Z}(t), \alpha, \tilde{A}_1(t)) - \frac{1}{p}\sum_{i=1}^{p} H_i(\mu_t, \alpha, \mu_t\mu_{-t})\right|,$$

$$\mathrm{II} = \left|\frac{1}{p}\sum_{i=1}^{p} H_i(\mu_t, \alpha, \mu_t\mu_{-t}) - \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}H_i(\mu_t, \alpha, \mu_t\mu_{-t})\right|,$$

$$\mathrm{III} = \left|\frac{1}{p}\sum_{i=1}^{p} \mathrm{E}H_i(\mu_t, \alpha, \mu_t\mu_{-t}) - \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}H_i(\bar{Z}(t), \alpha, A_1(t))\right|.$$

It suffices to show that $\mathrm{I} \xrightarrow{\text{a.s.}} 0$, $\mathrm{II} \xrightarrow{\text{a.s.}} 0$, and $\mathrm{III} \xrightarrow{\text{a.s.}} 0$ uniformly for $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ as $p \to \infty$.

For I, we have

$$\begin{aligned}
\mathrm{I} &= \left|\frac{1}{p}\sum_{i=1}^{p} H_i(\bar{Z}(t), \alpha, \tilde{A}_1(t)) - \frac{1}{p}\sum_{i=1}^{p} H_i(\mu_t, \alpha, \mu_t\mu_{-t})\right| \\
&= \left|\frac{1}{p}\sum_{i=1}^{p} \frac{\{\tilde{A}_1(t) - \mu_t\mu_{-t}\}Z_i^2(t) + \{\bar{Z}(t) - \mu_t\}Z_i(t)}{D_1D_2}\right| \\
&\le |\tilde{A}_1(t) - \mu_t\mu_{-t}|\frac{1}{p}\sum_{i=1}^{p}\frac{Z_i^2(t)}{D_1D_2} + \left|\bar{Z}(t) - \mu_t\right| \frac{1}{p}\sum_{i=1}^{p}\frac{Z_i(t)}{D_1D_2},
\end{aligned} \tag{10}$$

where $D_1 = \alpha\bar{Z}(t) + \{1 - \tilde{A}_1(t)\alpha\}Z_i(t)$ and $D_2 = \alpha\mu_t + \{1 - \mu_t\mu_{-t}\alpha\}Z_i(t)$. Note that $\tilde{A}_1(t) \xrightarrow{\text{a.s.}} \mu_t\mu_{-t}$ as $p \to \infty$. There exists an $N_1 > 0$ such that for any $p > N_1$, $\tilde{A}_1(t) < c\mu_t\mu_{-t}$ for any given $c > 1$. When $\alpha \in (1/2c\mu_t\mu_{-t}, 1/c\mu_t\mu_{-t}]$, we have $D_1 \ge \alpha\bar{Z}(t)$ and $D_2 \ge \alpha\mu_t$ for any $p > N_1$. Consequently, as $p \to \infty$,

$$\frac{1}{p}\sum_{i=1}^{p}\frac{Z_i^2(t)}{D_1D_2} \le \frac{(2c\mu_t\mu_{-t})^2}{\mu_t\bar{Z}(t)}\frac{1}{p}\sum_{i=1}^{p} Z_i^2(t) \xrightarrow{\text{a.s.}} \frac{h^2(t)}{h(2t)}4c^2\mu_{2t}\mu_{-t}^2 < \infty.$$

When $\alpha \in (0, 1/2c\mu_t\mu_{-t}]$, we have $D_1 \ge Z_i(t)/2$ and $D_2 \ge (1 - 1/2c)Z_i(t)$ for any $p > N_1$. Consequently,

$$\frac{1}{p}\sum_{i=1}^{p}\frac{Z_i^2(t)}{D_1D_2} \le \frac{4c}{2c-1} < \infty.$$

Therefore, for any $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$, $\sum_{i=1}^{p} Z_i^2(t)/(pD_1D_2)$ is almost surely bounded as $p \to \infty$. Similar arguments show that $\sum_{i=1}^{p} Z_i(t)/(pD_1D_2)$ is almost surely bounded as $p \to \infty$. Then, together with the facts that $\tilde{A}_1(t) \xrightarrow{\text{a.s.}} \mu_t\mu_{-t}$ and $\bar{Z}(t) \xrightarrow{\text{a.s.}} \mu_t$ as $p \to \infty$, by Slutsky's theorem, we have $\mathrm{I} \xrightarrow{\text{a.s.}} 0$ uniformly for $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$.

For II, noting that $\mathrm{E}\{H_i(\mu_t, \alpha, \mu_t\mu_{-t})\}^2 < \infty$ for any nonzero $|t| < \nu/2$ and $\alpha \in [0, 1/\mu_t\mu_{-t}]$, by Theorem 16 in [5], we have $\mathrm{II} \xrightarrow{\text{a.s.}} 0$ uniformly for $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$.

For III, we have

$$\begin{aligned}
\mathrm{III} &= \left|\frac{1}{p}\sum_{i=1}^{p} \mathrm{E}H_i(\mu_t, \alpha, \mu_t\mu_{-t}) - \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}H_i(\bar{Z}(t), \alpha, A_1(t))\right| \\
&\le \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left|H_i(\mu_t, \alpha, \mu_t\mu_{-t}) - H_i(\bar{Z}(t), \alpha, A_1(t))\right| \\
&\le \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left|\frac{\{A_1(t) - \mu_t\mu_{-t}\}Z_i^2(t)}{D_2D_3}\right| + \frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left|\frac{\{\bar{Z}(t) - \mu_t\}Z_i(t)}{D_2D_3}\right|,
\end{aligned} \tag{11}$$

where D_2 is defined above and $D_3 = \alpha\bar{Z}(t) + \{1 - A_1(t)\alpha\}Z_i(t)$. Following a similar argument as above it can be shown that for $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$,

$$\frac{1}{p}\sum_{i=1}^{p} \mathrm{E}\left|\frac{\{A_1(t) - \mu_t\mu_{-t}\}Z_i^2(t)}{D_2D_3}\right| \le C_1\left[\mathrm{E}\{A_1(t) - \mu_t\mu_{-t}\}^2\right]^{1/2},$$

where C_1 is a finite number independent of α. Now since $\mathrm{E}\{A_1(t) - \mu_t\mu_{-t}\}^2 \to 0$ as $p \to \infty$, the first term in (11) converges to 0 uniformly for $\alpha \in [0, 1/c\mu_t\mu_{-t}]$ with $c > 1$. Similarly, it can be shown that the second term in (11) converges to 0 uniformly for $\alpha \in [0, 1/c\mu_t\mu_{-t}]$ with $c > 1$. Therefore, III $\overset{\text{a.s.}}{\to} 0$ uniformly for $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$.

Now we show that $\tilde{\alpha}^*_{T_1} - \alpha^*_{T_1} \overset{\text{a.s.}}{\to} 0$ as $p \to \infty$. For ease of notation, denote $f_1(\alpha) = \hat{R}'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})$ and $f_2(\alpha) = R'_{T_1}(\alpha; \boldsymbol{\sigma}^{2t})$. When $p > N_1$ such that $\tilde{A}_1(t) < c\mu_t\mu_{-t}$, we have for any $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$,

$$
\begin{aligned}
f_1'(\alpha) &= \frac{1}{p}\sum_{i=1}^{p}\left[\frac{\bar{Z}(t) - \tilde{A}_1(t)Z_i(t)}{\alpha\bar{Z}(t) + \{1 - \tilde{A}_1(t)\alpha\}Z_i(t)}\right]^2 \\
&\geq \frac{1}{p}\sum_{i=1}^{p}\left\{\frac{\bar{Z}(t) - \tilde{A}_1(t)Z_i(t)}{\bar{Z}(t)/c\mu_t\mu_{-t} + Z_i(t)}\right\}^2 \\
&\overset{\text{a.s.}}{\to} (\mu_t\mu_{-t})^2\mathrm{E}\left\{\frac{Z_1(t) - 1/\mu_{-t}}{Z_1(t) + 1/c\mu_{-t}}\right\}^2, \quad \text{as } p \to \infty.
\end{aligned} \tag{12}
$$

Now since $Z_1(t)$ is a non-trivial random variable, we have $\min_{\alpha\in(0,1/c\mu_t\mu_{-t}]}\lim_{p\to\infty} f_1'(\alpha) \geq M > 0$ where M is the limit in (12). Thus, there exists an $N_2 > 0$ such that for any $p > N_2$, $|f_1(\alpha^*_{T_1}) - f_1(\tilde{\alpha}^*_{T_1})|/|\tilde{\alpha}^*_{T_1} - \alpha^*_{T_1}| \geq M/2$. Note that, by definition, $f_1(\tilde{\alpha}^*_{T_1}) = f_2(\alpha^*_{T_1}) = 0$. This leads to $|f_1(\alpha^*_{T_1}) - f_2(\alpha^*_{T_1})| \geq (M/2)|\tilde{\alpha}^*_{T_1} - \alpha^*_{T_1}|$ for any $p > N_2$. Note that $f_1(\alpha) - f_2(\alpha) \overset{\text{a.s.}}{\to} 0$ uniformly for $\alpha \in (0, 1/c\mu_t\mu_{-t}]$ with $c > 1$. By letting $c \to 1$ such that $\alpha^*_{T_1} \in (0, 1/c\mu_t\mu_{-t}]$, we have $f_1(\alpha^*_{T_1}) - f_2(\alpha^*_{T_1}) \to 0$ as $p \to \infty$. Finally, as $M > 0$, we have $\tilde{\alpha}^*_{T_1} - \alpha^*_{T_1} \to 0$ as $p \to \infty$.

The proof of (ii) is similar and thus is omitted.

References

[1] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, Dover, New York, 1972.
[2] P. Baldi, A.D. Long, A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes, Bioinformatics 17 (2001) 509–519.
[3] E. Beckenbach, R. Bellman, Inequalities, Springer-Verlag, New York, 1970.
[4] X. Cui, J.T.G. Hwang, J. Qiu, N.J. Blades, G.A. Churchill, Improved statistical tests for differential gene expression by shrinking variance components estimates, Biostatistics 6 (2005) 59–75.
[5] T.S. Ferguson, A Course in Large Sample Theory, Chapman and Hall, London, 1996.
[6] G. Fox, M. Dimmic, A two-sample Bayesian t-test for microarray data, BMC Bioinformatics 7 (2006) 126.
[7] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1961, pp. 361–379.
[8] A. Kamb, A. Ramaswami, A simple method for statistical analysis of intensity differences in microarray-derived gene expression data, BMC Biotechnology 1 (2001) 1–8.
[9] T. Kubokawa, Shrinkage and modification techniques in estimation of variance and the related problems: a review, Communications in Statistics: Theory and Methods 28 (1999) 613–650.
[10] R. Opgen-Rhein, K. Strimmer, Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach, Statistical Applications in Genetics and Molecular Biology 6 (2007) 9.
[11] H. Pang, K. Ebisu, E. Watanabe, L. Sue, T. Tong, Analyzing breast cancer microarrays from African Americans using shrinkage-based discriminant analysis, Human Genomics 5 (2009) 5–16.
[12] H. Pang, T. Tong, H. Zhao, Shrinkage-based diagonal discriminant analysis and its applications in high-dimensional data, Biometrics 65 (2009) 1021–1029.
[13] G.K. Smyth, Linear models and empirical Bayes methods for assessing differential expression in microarray experiment, Statistical Applications in Genetics and Molecular Biology 3 (2004) 1.
[14] J.D. Storey, R. Tibshirani, SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays, in: G. Parmigiani, E.S. Garrett, R.A. Irizarry, S.L. Zeger (Eds.), The Analysis of Gene Expression Data: Methods and Software, Springer, New York, 2003.
[15] T. Tong, Y. Wang, Optimal shrinkage estimation of variances with applications to microarray data analysis, Journal of the American Statistical Association 102 (2007) 113–122.
[16] G.W. Wright, R.M. Simon, A random variance model for detection of differential gene expression in small microarray experiments, Bioinformatics 19 (2003) 2448–2455.