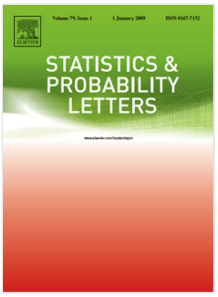
# Estimates of low bias for the multivariate normal

Christopher S. Withers, Saralees Nadarajah

# Accepted Manuscript




Estimates of low bias for the multivariate normal

Christopher S. Withers, Saralees Nadarajah

PII: S0167-7152(11)00216-1
DOI: 10.1016/j.spl.2011.06.014
Reference: STAPRO 6028

To appear in: *Statistics and Probability Letters*

Received date: 17 January 2011
Revised date: 8 June 2011
Accepted date: 9 June 2011

Please cite this article as: Withers, C.S., Nadarajah, S., Estimates of low bias for the multivariate normal. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.06.014

# Estimates of low bias for the multivariate normal

by

Christopher S. Withers
Applied Mathematics Group
Industrial Research Limited
Lower Hutt, NEW ZEALAND

Saralees Nadarajah
School of Mathematics
University of Manchester
Manchester M13 9PL, UK

**Abstract:** Given a sample from a multivariate normal with mean $\boldsymbol{\mu}$, a method is given for obtaining estimates with low bias for a function of the parameters. When the function is a product of positive powers of the parameters an unbiased estimate is available. Estimates of ratios like $\mu_1/\mu_2$ are given with bias $\sim n^{-5}$, where $n$ is the sample size. Simulation studies show superior performance of these estimates versus traditional ones.

**AMS 2000 subject classification:** Primary 62F10; Secondary 62H12.

**Keywords and phrases:** Bias reduction; Moments; Multivariate Hermite polynomials; Multivariate normal; Ratio estimates; Unbiased estimates; Wishart distribution.

## 1 Introduction and summary

Many authors have given unbiased estimates (UEs) for functions $t(\boldsymbol{\theta})$ say of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in terms of independent

$$\widetilde{\boldsymbol{\mu}} \sim \mathcal{N}_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n\right), \ \nu\widetilde{\boldsymbol{\Sigma}} \sim W_p\left(\boldsymbol{\Sigma}, \nu\right),$$

the multivariate normal and Wishart distributions. When $\nu = n-1$ these correspond of course to the usual UEs of $\boldsymbol{\theta}$ given a sample of size $n$ from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, while other $\nu$ occur for regression with normal errors.

For example, Ghurye and Olkin (1969), and Eaton and Morris (1970) give an UE $\widehat{\phi}_{\boldsymbol{\Sigma}}(\mathbf{x}-\boldsymbol{\mu})$ say of the density of $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. So, for $f : \mathbb{R}^p \to \mathbb{R}$ suitably smooth, an UE of $\mathbb{E}[f(\mathbf{X})]$ is $\int f(\mathbf{x})\widehat{\phi}_{\boldsymbol{\Sigma}}(\mathbf{x}-\boldsymbol{\mu})d\mathbf{x}$. They do not give examples, where this can be evaluated in closed form. For recent references, see Hutson (2002), Wan and Zou (2004), Misra *et al.* (2005), Sun and Sun (2005, 2006) and Yanagihara (2006).

Here, we consider obtaining UEs and estimates of low bias for functions of the form

$$t(\boldsymbol{\theta}) = \boldsymbol{\mu}^{\mathbf{I}} f\left(\boldsymbol{\Sigma}\right), \tag{1.1}$$

where

$$\boldsymbol{\mu}^{\mathbf{I}} = \prod_{i=1}^{p} \mu_i^{I_i}$$

and $\mathbf{I}$ is an integer in $\mathbb{R}^p$. The class given (1.1) contains a wide range of parameters of practical interest. For example,

1. ratios of the form $\mu_1/\mu_2$. Such ratios arise most often in biological sciences and related areas, for example, with respect to safety assessment of a substance such as a new pharmaceutical compound or a pesticide relative to a vehicle or negative control. Bose (1942), Roy and Potthoff (1958), and Hannig (2006) consider ratios of means from a bivariate normal.

2. the multivariate coefficient of variation defined by $\Sigma_{i,j}/(\mu_i\mu_j)$, a useful measure of variability (Bennett, 1977).

3. the correlation coefficient defined by $\Sigma_{i,j}/\sqrt{\Sigma_{i,i}\Sigma_{j,j}}$ and others of its kind, for example, the partial correlation coefficient, the multiple correlation coefficient, the coefficient of determination and the canonical correlations (Mardia *et al.*, 1980).

4. the measure of generalized variance, $|\mathbf{\Sigma}|$, and the measure of total variation, trace $\mathbf{\Sigma}$ (Mardia *et al.*, 1980).

5. $\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}$, $\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$ and other quadratic forms involving $\boldsymbol{\mu}$, $\mathbf{\Sigma}^{-1}$ or both. Parameters of this form are common place in every area of multivariate analysis (Mardia *et al.*, 1980). The elements of these parameters are particular cases of (1.1).

6. conditional moments of $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, i.e if $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$, where $\mathbf{X}_1$ is $q \times 1$, and if $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are partitioned similarly then $\mathbb{E}(\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1) = \boldsymbol{\mu}_2 + \mathbf{\Sigma}_{2,1}\mathbf{\Sigma}_{1,1}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ (Mardia *et al.*, 1980).

7. central moments, skewness, kurtosis and cumulants of $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ (Mardia, 1970; Mardia *et al.*, 1980; Holmquist, 1988).

8. moments, central moments and cumulants of the central and non-central Wishart distributions and their variants (Kollo and von Rosen, 1995; Kang and Kim, 1996; Sultan and Tracy, 1996a; Mukhopadhyay, 2009).

9. moments, central moments and cumulants of complex normal distributions and their variants (Sultan and Tracy, 1996b).

10. moments, central moments and cumulants of the central and non-central complex Wishart distributions and their variants (Sultan and Tracy, 1999; Withers and Nadarajah, 2010b).

11. the entropy, a measure of uncertainty, of $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ given by $\sqrt{(2\pi e)^p|\mathbf{\Sigma}|}$ and its variants (Zografos and Nadarajah, 2005).

12. the characteristic function of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ given by $\exp\{-(1/2)\mathbf{t}'\mathbf{\Sigma}\mathbf{t}\}$ (Mardia *et al.*, 1980), the empirical version of which can be used, for example, for estimation.

One could mention many more examples. To the best of our knowledge, we are not aware of a class of parameters wider than (1.1) for which UEs have been sought.

We give a method for obtaining an UE of $t(\boldsymbol{\theta})$ when $\mathbf{I} \geq \mathbf{0}$ (componentwise). For $0 \leq s \leq [I./2]$, an UE $\widetilde{f}_s$ is available for $f_s$, where

$$f_0 = f(\mathbf{\Sigma}), \; f_s = f_{i_1,\ldots,i_{2s}} = \Sigma_{i_1,i_2} \cdots \Sigma_{i_{2s-1},i_{2s}} f(\mathbf{\Sigma}) \tag{1.2}$$

for $s = 1, 2, 3, \ldots, I_\cdot = \sum_{i=1}^{p} I_i$ and $[x]$ is the integral part of $x$. Specific formulas are given for $I_\cdot \leq 8$.

Also for $\mathbf{I}$ not necessarily non-negative, for example $\boldsymbol{\mu}^{\mathbf{I}} = \mu_1/\mu_2$, we provide a method for obtaining an estimate of $t(\boldsymbol{\theta})$ with bias $O(n^{-k})$, where $k \geq 1$. Specific formulas are provided for $k \leq 5$. When $\mathbf{I} \geq \mathbf{0}$ it has form

$$\widetilde{t}_{n,k} = \sum_{i=0}^{k-1} n^{-i} \widetilde{t}_i \left( \widetilde{\boldsymbol{\mu}} \right) \tag{1.3}$$

for $k \geq 1$, where $\widetilde{t}_i(\cdot)$ is linear in $\widetilde{f}_i$ and we now assume

$$\nu/n \text{ is bounded as } n \to \infty \tag{1.4}$$

and for $\{f_s\}$ of (1.2), for $0 \leq s \leq k$,

$$\widetilde{f}_s = f_s + O\left(\nu^{-(k-s)}\right) \tag{1.5}$$

as $\nu \to \infty$. This is based on the general series method of bias reduction given in Withers (1987) which can also be used to obtain $\{\widetilde{f}_s\}$ as in (1.5) for smooth $f$.

For general $\mathbf{I}$ our estimate of $t(\boldsymbol{\theta})$ with bias $O(n^{-k})$ has the form

$$\widetilde{t}_{n,k} L_n,$$

where

$$L_n = \prod_{i=1}^{p} L_{n,i},$$

where

$$L_{n,i} = \begin{cases} I\left(\mu_{o,i} < |\widetilde{\mu}_i|\right), & \text{if } I_i < 0, \\ 1, & \text{if } I_i \geq 0, \end{cases}$$

and

$$I(A) = \begin{cases} 1, & \text{if } A \text{ is true,} \\ 0, & \text{if } A \text{ is false,} \end{cases}$$

where, for $I_i < 0$, we assume

$$\mu_{o,i} \text{ is known such that } 0 < \mu_{o,i} < \mid \mu_i \mid . \tag{1.6}$$

An adaptation like this is necessary since $\mathbb{E}[\widetilde{\boldsymbol{\mu}}^{\mathbf{I}}]$ does not exist unless $\mathbf{I} \geq \mathbf{0}$.

Our main results are given in Section 2. Section 3 gives a method of obtaining UEs of $f_s$ for $f(\boldsymbol{\Sigma}) = 1$. They are given explicitly for $s \leq 3$. Applied to Section 2, they provide UEs for $\boldsymbol{\mu}^{\mathbf{I}}$ if $I_\cdot \leq 6$, for $\boldsymbol{\mu}^{\mathbf{I}}\Sigma_{i,j}$ if $I_\cdot \leq 4$, and for $\boldsymbol{\mu}^{\mathbf{I}}\Sigma_{i,j}\Sigma_{k,l}$ if $I_\cdot \leq 2$, where $\mathbf{I} \geq \mathbf{0}$. For general $\mathbf{I}$ they provide estimates for $\boldsymbol{\mu}^{\mathbf{I}}$ of bias $\sim n^{-5}$, and estimates for $\boldsymbol{\mu}^{\mathbf{I}}\Sigma_{i_1,i_2} \cdots \Sigma_{i_{25-1},i_{25}}$ of bias $\sim n^{-(5-s)}$, $1 \leq s \leq 4$. Section 4 applies the methods of Sections 2 and 3 to obtain estimates for four of the parameters discussed above. A simulation study is performed to show that these estimates outperform traditional ones in terms of both bias and mean square error.

Throughout this note, we set

$$(j)_k = j(j-1)\cdots(j-k+1),\ I_{i,j,\ldots,a,b,\ldots} = (I_i)_a\,(I_j)_b\cdots$$

and $\mathbf{e}_i$ is the $i$th unit vector in $\mathbb{R}^p$. For example, $I_{i,2} = (I_i)_2 = I_i(I_i-1)$, $I_{i,j,1,1} = (I_i)_1(I_j)_1 = I_iI_j$, $I_{i,j,k,2,1,1} = (I_i)_2(I_j)_1(I_k)_1 = I_i(I_i-1)I_jI_k$, $I_{i,j,k,l,1,1,1,1} = (I_i)_1(I_j)_1(I_k)_1(I_l)_1 = I_iI_jI_kI_l$, and so on. We also write $t$ to mean $t = t(\boldsymbol{\theta})$ of (1.1). We also assume throughout that (1.4), (1.5) and (1.6) or the relevant ones of these hold.

## 2 The method

Theorem 2.1 gives the method for obtaining estimates with low bias for $t(\boldsymbol{\theta})$ of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

**Theorem 2.1** *With the notations as above, suppose that (1.1) holds. Then*

$$\widetilde{t}_{n,k} = \widetilde{\boldsymbol{\mu}}^{\mathbf{I}}\left(\widetilde{f}_0 - \sum_{s=1}^{k-1} n^{-3}\widetilde{c}_s^{\mathbf{I}}\right) \tag{2.1}$$

*satisfies*

$$\mathbb{E}\left[\widetilde{t}_{n,k}L_n\right] = t + O\left(n^{-k}\right), \tag{2.2}$$

*where*

$$\widetilde{c}_{r/2}^{\mathbf{I}} = \sum_{i_1,i_2,\ldots,i_r} w_{i_1,\ldots,i_r}^{\mathbf{I}} F\left(i_1,\ldots,i_r\right),\ F\left(i_1,\ldots,i_r\right) = \widetilde{f}_{i_1,\ldots,i_r}\left(\widetilde{\mu}_{i_1}\cdots\widetilde{\mu}_{i_r}\right)^{-1}, \tag{2.3}$$

*and $w_{i_1,\ldots,i_r}^{\mathbf{I}}$ is a constant determined by the proof.*

**Proof:** By Withers and Nadarajah (2010a), $\mathbb{E}[I(\mu_{o,i} \geq |\widetilde{\mu}_i|)] \sim \exp(-\lambda_i n)$ as $n \to \infty$, where $\lambda_i > 0$. So, $1 - \mathbb{E}[L_n] \sim \exp(-\lambda n)$, where $\lambda > 0$. As shown in Withers and Nadarajah (2010a) this implies for $\widetilde{t} = \widetilde{\boldsymbol{\mu}}^{\mathbf{I}}\widetilde{f}_0$ with $\mathbb{E}[\widetilde{f}_0] = f_0$ that

$$\mathbb{E}\ \left[\widetilde{t}L_n\right] \approx \mathbb{E}\ \left[\widetilde{t}\right] = t\mathbb{E}\ \left[\prod_{j=1}^{p}\left(1 + n^{-1/2}X_j\right)^{I_j}\right] = t\sum_{i=0}^{\infty} n^{-i}b_i^{\mathbf{I}} = \sum_{i=0}^{\infty} n^{-i}g_i^{\mathbf{I}}$$

say, where

$$b_i^{\mathbf{I}} = b_i^{\mathbf{I}}(\mathbf{V}) = \sum_{j_\cdot=2i}\begin{pmatrix}\mathbf{I}\\ \mathbf{j}\end{pmatrix}\mathbb{E}\left[\mathbf{X}^{\mathbf{j}}\right],$$

$$\mathbf{j} \in \mathbb{N}^p,\ j_\cdot = \sum_{i=1}^{p} j_i,\ \mathbb{N} = \{0,1,2,\ldots\},\ \mathbf{X} \sim \mathcal{N}_p\left(\mathbf{0},\mathbf{V}\right),$$

$$\begin{pmatrix}\mathbf{I}\\ \mathbf{j}\end{pmatrix} = \prod_{i=1}^{p}\begin{pmatrix}I_i\\ j_i\end{pmatrix},$$

$$V_{i,j} = \left(\mu_i\mu_j\right)^{-1}\Sigma_{i,j}.$$

Here, $A_n \approx \sum_{i=0}^{\infty} n^{-i} g_i$ means $A_n = \sum_{i=0}^{k-1} n^{-i} g_i + O(n^{-k})$ for $k \geq 1$. For $\mathbf{I} \geq \mathbf{0}$,

$$\text{the expansion is exact and finite with } b_i^{\mathbf{I}} = 0 \text{ for } i > I./2, \tag{2.4}$$

but otherwise it diverges. Expressions for $\mathbb{E}[\mathbf{X}^{\mathbf{j}}] = \mathbb{E}[\prod_{i=1}^{p} X_i^{j_i}]$ are given in Appendix A.

Then

$$b_0^{\mathbf{I}} = 1,\ b_1^{\mathbf{I}} = \sum_i I_{i,2} V_{i,i}/2 + \sum_{i<j} I_{i,j,1,1} V_{i,j} = \sum_{i,j} w_{i,j}^{\mathbf{I}} V_{i,j}$$

say, and

$$\begin{aligned} b_2^{\mathbf{I}} \;=\; & \sum_i I_{i,4} V_{i,i}^2/8 + \sum_{i \neq j} I_{i,j,3,1} V_{i,i} V_{i,j}/2 + \sum_{i<j} I_{i,j,2,2} \left(V_{i,i} V_{j,j} + 2V_{i,j}^2\right)/4 \\ & + \sum_{j<k,\ i \neq j \text{ or } k} I_{i,j,k,2,1,1} \left(V_{i,i} V_{j,k} + 2V_{i,j} V_{i,k}\right)/2 \\ & + \sum_{i<j<k<l} I_{i,j,k,l,1,1,1,1} \left(V_{i,j} V_{k,l} + V_{i,k} V_{j,l} + V_{i,l} V_{j,k}\right). \end{aligned}$$

So, $\mathbb{E}[\widetilde{t}] = t + O(n^{-1})$. Set $\widetilde{f}_{i_1,\ldots,i_{2s}} = \widetilde{f}_s$ for $s \geq 1$. Given $\boldsymbol{\mu}$,

$$\widetilde{g}_1^{\mathbf{I}}(\boldsymbol{\mu}) = \boldsymbol{\mu}^{\mathbf{I}} \sum_{i,j} w_{i,j}^{\mathbf{I}} \widetilde{f}_{i,j} \mu_i^{-1} \mu_j^{-1} = \sum_{i,j} w_{i,j}^{\mathbf{I}} \widetilde{f}_{i,j} \boldsymbol{\mu}^{\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j} \tag{2.5}$$

is an UE of $g_1^{\mathbf{I}}$. Also

$$\mathbb{E}\left[\widetilde{g}_1^{\mathbf{I}}(\widetilde{\boldsymbol{\mu}})\, L_n\right] \approx \mathbb{E}\left[\widetilde{g}_1^{\mathbf{I}}(\widetilde{\boldsymbol{\mu}})\right] = \sum_{i,j} w_{i,j}^{\mathbf{I}} f_{i,j} \boldsymbol{\mu}^{\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j} \sum_{r=0}^{\infty} n^{-r} b_r^{\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j} = \sum_{r=0}^{\infty} n^{-r} g_{1\cdot r}^{\mathbf{I}}$$

say. Set

$$\widetilde{t}_{n,2} = \widetilde{t} - n^{-1} \widetilde{g}_1^{\mathbf{I}}(\widetilde{\boldsymbol{\mu}}). \tag{2.6}$$

Then $\widetilde{t}_{n,2} L_n$ estimates $t$ with bias $n^{-2} G_2^{\mathbf{I}} + O(n^{-3})$, where

$$G_2^{\mathbf{I}} = g_2^{\mathbf{I}} - g_{1\cdot 1}^{\mathbf{I}},\ g_{1\cdot 1}^{\mathbf{I}} = t H_2^{\mathbf{I}},\ H_2^{\mathbf{I}} = \sum_{i,j,k,l} w_{i,j}^{\mathbf{I}} w_{k,l}^{\mathbf{I}-\mathbf{e}_i-\mathbf{e}_j} V_{i,j} V_{k,l}.$$

Writing $t \sum_{i,j,k,l} w_{i,j,k,l}^{\mathbf{I}} V_{i,j} V_{k,l} = G_2^{\mathbf{I}}$, an UE for it (given $\boldsymbol{\mu}$) is

$$\widetilde{G}_2^{\mathbf{I}}(\boldsymbol{\mu}) = \sum_{i,j,k,l} w_{i,j,k,l}^{\mathbf{I}} \widetilde{f}_{i,j,k,l} \boldsymbol{\mu}^{\mathbf{I}-\mathbf{e}_i-\cdots-\mathbf{e}_l}.$$

Set

$$\widetilde{t}_{n,3} = \widetilde{t}_{n,2} - n^{-2} \widetilde{G}_2^{\mathbf{I}}(\widetilde{\boldsymbol{\mu}}).$$

Then $\widetilde{t}_{n,3} L_n$ estimates $t$ with bias $n^{-3} G_3^{\mathbf{I}} + O(n^{-4})$, where

$$G_3^{\mathbf{I}} = g_3^{\mathbf{I}} - g_{1\cdot 2}^{\mathbf{I}} - G_{2\cdot 1}^{\mathbf{I}}$$

and $G_{2\cdot 1}^{\mathbf{I}}$ is given by

$$\mathbb{E}\left[\widetilde{G}_2^{\mathbf{I}}\left(\widetilde{\boldsymbol{\mu}}\right)L_n\right] \approx \mathbb{E}\left[\widetilde{G}_2^{\mathbf{I}}\left(\widetilde{\boldsymbol{\mu}}\right)\right] = t\sum_{i,j,k,l} w_{i,j,k,l}^{\mathbf{I}} V_{i,j}V_{k,l}\sum_{r=0}^{\infty} n^{-r} b_r^{\mathbf{I}-\mathbf{e}_i-\cdots-\mathbf{e}_l} = \sum_{r=0}^{\infty} n^{-r} G_{2\cdot r}^{\mathbf{I}}.$$

Writing $t\sum_{i_1,i_2,\ldots,i_6} w_{i_1,\ldots,i_6}^{\mathbf{I}} V_{i_1,i_2}V_{i_3,i_4}V_{i_5,i_6} = G_3^{\mathbf{I}}$, an UE for it (given $\boldsymbol{\mu}$) is

$$\widetilde{G}_3^{\mathbf{I}}(\boldsymbol{\mu}) = \sum_{i_1,i_2,\ldots,i_6} w_{i_1,\ldots,i_6}^{\mathbf{I}} \widetilde{f}_{i_1,\ldots,i_6}\boldsymbol{\mu}^{\mathbf{I}-\mathbf{e}_{i_1}-\cdots-\mathbf{e}_{i_6}}.$$

Set

$$\widetilde{t}_{n,4} = \widetilde{t}_{n,3} - n^{-3}\widetilde{G}_3^{\mathbf{I}}\left(\widetilde{\boldsymbol{\mu}}\right).$$

Then $\widetilde{t}_{n,4}L_n$ estimates $t$ with bias $n^{-4}\overline{g}_4^{\mathbf{I}} + O(n^{-5})$, where

$$\begin{aligned}
\overline{g}_4^{\mathbf{I}} &= \overline{g}_4^{\mathbf{I}}(\boldsymbol{\theta}) = g_4^{\mathbf{I}} - g_{1\cdot 3}^{\mathbf{I}} - G_{2\cdot 2}^{\mathbf{I}} - G_{3\cdot 1}^{\mathbf{I}}, \\
G_{3\cdot 1}^{\mathbf{I}} &= t\sum_{i_1,i_2,\ldots,i_6} w_{i_1,\ldots,i_6}^{\mathbf{I}} V_{i_1,i_2}V_{i_3,i_4}V_{i_5,i_6} b_1^{\mathbf{I}-\mathbf{e}_{i_1}-\cdots-\mathbf{e}_{i_6}}.
\end{aligned} \tag{2.7}$$

This process can be repeated to obtain the result of the theorem. □

From (2.4) it follows that $\mathbb{E}[\widetilde{t}_{n,k}] \equiv t$ if $\mathbf{I} \geq \mathbf{0}$ and $k - 1 \geq I./2$.

**Corollary 2.1** *If (1.1) is weakened to (1.2) then (1.3) holds with $\nu$ replaced by $n$ and (2.2) still holds. In particular, if (1.1) holds and $\{\widetilde{f}_s, 0 \leq s \leq 3\}$ are UEs and*

$$\mathbb{E}\left[f_4\left(\widetilde{\boldsymbol{\Sigma}}\right)\right] = f_4 + O\left(\nu^{-1}\right)$$

*as $\nu \to \infty$ then*

$$\left(\widetilde{t}_{n,4} - n^{-4}\overline{g}_4^{\mathbf{I}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)L_n \tag{2.8}$$

*has bias $\sim n^{-5}$.*

**Corollary 2.2** *Suppose $p = 2$. Set $\sum^2 Q_{1,2} = Q_{1,2} + Q_{2,1}$ and $I_{a,b} = (I_1)_a(I_2)_b$ for one and two arguments of $\mathbf{I}$ or $F(\cdot)$, where $F(\cdots)$ is as defined by (2.3). For example, $\sum^2 I_{a,b} = I_{a,b} + I_{b,a}$. Then*

$$\widetilde{c}_1^{\mathbf{I}} = \sum^2 \left\{I_{2,0}F\left(1^2\right) + I_{1,1}F(1,2)\right\}/2, \tag{2.9}$$

$$\widetilde{c}_2^{\mathbf{I}} = -\sum^2 \left\{I_{4,0}F\left(1^4\right) + 4I_{3,1}F\left(1^3,2\right) + I_{2,2}\left[3F\left(1^2,2^2\right) + 2F(1,2,1,2)\right]\right\}/8, \tag{2.10}$$

$$\begin{aligned}
\widetilde{c}_3^{\mathbf{I}} = \sum^2 \Big\{ &- I_{6,0}F\left(1^6\right) - 6I_{5,1}F\left(1^5,2\right) + 3I_{4,2}\left[5F\left(1^4,2^2\right) + F\left(1^3,2^2,1\right)\right] \\
&+ I_{3,3}\left[15F\left(1^3,2^3\right) + 2F(1,2,1,2,1,2)\right]\Big\}/48,
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
\widetilde{c}_4^{\mathbf{I}} = -\sum^2 \Big\{ &I_{8,0}F\left(1^8\right) + 8I_{7,1}F\left(1^2,2\right) + 4I_{6,2}\left[7F\left(1^6,2^2\right) - 3F\left(1^5,2^2,1\right)\right] \\
&+ 12I_{5,3}\left[9F\left(1^5,2^3\right) - 2F\left(1^3,2,1,2,1,2\right)\right] \\
&+ 12I_{4,4}\left[2F\left(1^4,2^4\right) + F\left(1^3,2,1,2^3\right) + 2F(1,2,1,2,1,2,1,2)\right]\Big\}/384.
\end{aligned} \tag{2.12}$$

*For simplicity, we have abbreviated* $F(1,1)$ *as* $F(1^2)$, $F(1,1,1,1)$ *as* $F(1^4)$, $F(1,1,1,2)$ *as* $F(1^3,2)$, $F(1,1,2,2)$ *as* $F(1^2,2^2)$, $F(1,1,1,1,1,1)$ *as* $F(1^6)$, $F(1,1,1,1,1,2)$ *as* $F(1^5,2)$, $F(1,1,1,1,2,2)$ *as* $F(1^4,2^2)$, $F(1,1,1,2,2,1)$ *as* $F(1^3,2^2,1)$, $F(1,1,1,2,2,2)$ *as* $F(1^3,2^3)$, *and so on.*

**Proof:** See Appendix B. □

## 3 UEs for functions of $\boldsymbol{\Sigma}$

Given a function $f(\boldsymbol{\Sigma})$, set $\widetilde{f}_0 = f(\widetilde{\boldsymbol{\Sigma}})$, set $f^*$ denote an UE of $f$ if one exists. We shall deal mainly with obtaining $f^*$ for $f(\boldsymbol{\Sigma})$ a polynomial. Some other cases such as $f^*$ for $\boldsymbol{\Sigma}^{-1}$ and $\mid \boldsymbol{\Sigma} \mid^{\lambda}$ are dealt with at the end of this section.

Finding $f^*$ for $f(\boldsymbol{\Sigma})$ a polynomial reduces to finding $f^*$ for

$$f(\boldsymbol{\Sigma}) = \Sigma_{1,2}\Sigma_{3,4}\cdots\Sigma_{2r-1,2r} = P_r$$

say, where without loss of generality we can take $p \geq 2r$. For, the suffixes $1, 2, \ldots, 2r$ can be replaced by any other suffixes from $1, \ldots, p$ with repetitions allowed.

We begin with a method to obtain

$$\mu_{1,\ldots,2r} = \mathbb{E}\left[S_{1,2}S_{3,4}\cdots S_{2r-1,2r}\right],$$

where $\mathbf{S} = \nu\widetilde{\boldsymbol{\Sigma}}$. Obtaining this from the characteristic function for $\mathbf{S}$ is prohibitively laborious for $r > 2$. We apply the following identity for $\mathbf{V} = \mathbf{V}(\mathbf{S})$ a $p \times p$ matrix (see equation (2.4) of Haff (1980)):

$$\mathbb{E}\ \left[\text{trace } \boldsymbol{\Sigma}^{-1}\mathbf{V}\right] = \mathbb{E}\left[\sum_{i\leq j}\partial\left(V_{i,j}+V_{j,i}\right)/\partial S_{i,j}\right] + (\nu-p-1)\mathbb{E}\ \left[\text{trace } \mathbf{S}^{-1}\mathbf{V}\right]. \quad (3.1)$$

Set $\mathbf{V} = \{\prod_{i=1}^{r}\mathbf{S}\mathbf{A}_i\}\boldsymbol{\Sigma}$, where $(\mathbf{A}_i)_{2i,2i+1} = 1$ for $1 \leq i \leq r$, $(\mathbf{A}_r)_{2r,1} = 1$ and the other elements of $\{\mathbf{A}_i\} \subset \mathbb{R}^{p\times p}$ are zero. Then (3.1) becomes

$$\mu_{1,\ldots,2r} = \mathbb{E}\left[\left(\prod_{i=2}^{r}S_{2i-1,2i}\right)\left\{\nu\Sigma_{1,2} + \sum_{k=2}^{r}S^{-1}_{2k-1,2k}\left(\Sigma_{1,2k-1}S_{2,2k} + \Sigma_{1,2k}S_{2,2k-1}\right)\right\}\right]. \quad (3.2)$$

Putting $r = 1, 2, 3$ this gives

$$\mu_{1,2} = \nu\Sigma_{1,2},\ \mu_{1,2,3,4} = \nu^2\Sigma_{1,2}\Sigma_{3,4} + \nu A_{2,0}, \quad (3.3)$$

$$\mu_{1,2,3,4,5,6} = \nu^3\Sigma_{1,2}\Sigma_{3,4}\Sigma_{5,6} + \nu^2 A_{3,1} + \nu A_{3,0}, \quad (3.4)$$

where $A_{r,i}$ is defined as follows. Let us call the product of $r$ 'pairs' $(i,j) = \Sigma_{i,j}$ taken without replacement from $1, 2, \ldots, 2r$ an *r-product*. For example, $P_r = (1,2)(3,4)\cdots(2r-1,2r)$. Then $A_{r,i}$ is the sum of the distinct $r$-products having exactly $i$ pairs in common with $P_r$. So, $A_{r,r} = P_r$, $A_{r,r-1} = 0$ and $A_{2,0} = (1,3)(2,4) + (1,4)(2,3)$.

We also need $a_{r,i}$, the number of terms in $A_{r,i}$. We have $a_{r,r} = 1$, $a_{r,r-1} = 0$, $a_{2,0} = 2$, $a_{r,0} = (2r-2)(a_{r-1,0} + a_{r-2,0})$, and $a_{r,i} = \binom{r}{i}a_{r-i,0}$ for $r - i \geq 1$. As a check $\sum_{i=0}^{r}A_{r,i}$, the sum of all distinct $r$-products, has $a_r = 1\cdot 3\cdot 5\cdots(2r-1)$ terms so that $\sum_{i=0}^{r}a_{r,i} = a_r$.

Given an $r$-product we can link 1 with 2, 3 with 4, and so forth; if this links all pairs, we call the $r$-products *inseparable*, and otherwise *separable*. So, 13.24.57.68 is separable but not 13.25.47.68.

Let $A_{4,0\cdot1}$ denote the sum of the 12 separable terms in $A_{4,0}$ and $A_{4,0\cdot2}$ the sum of the 48 inseparable terms. Then $r = 4$ in (3.2) yields $\mu_{1,\ldots,8} = \nu^4 A_{4,4} + \nu^3 A_{4,2} + \nu^2(A_{4,1} + A_{4,0\cdot1}) + \nu A_{4,0\cdot2}$, and $a_{4,2} = 12$, $a_{4,1} = 32$. Higher order $\mu_{1,\ldots,2r}$ can be expressed similarly.

Another check is provided by replacing $1,\ldots,2r$ by $1,\ldots,1$ giving $\mathbb{E}[S_{1,1}^r] = \Sigma_{1,1}^r\, \mathbb{E}[\chi_\nu^{2r}]$ $= \Sigma_{1,1}^r \nu(\nu+2)\cdots(\nu+2r-2)$.

We now show how to obtain an UE for $P_r$ for $r = 2, 3$. Higher cases can be handled similarly but with greater labor.

**Theorem 3.1** *With the notation as above, $\boldsymbol{\theta}^* = \mathbf{D}_2^{-1}\widetilde{\boldsymbol{\theta}}$ is an UE of $\boldsymbol{\theta}' = (P_2, A_{2,0})$, where*

$$\mathbf{D}_2 = \begin{bmatrix} 1 & \nu^{-1} \\ 2\nu^{-1} & 1+\nu^{-1} \end{bmatrix}.$$

*Also $\boldsymbol{\theta}^* = \mathbf{D}_3^{-1}\widetilde{\boldsymbol{\theta}}$ is an UE of $\boldsymbol{\theta}' = (P_3, A_{3,1}, A_{3,0})$, where*

$$\mathbf{D}_3 = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ 6\alpha & 1+\alpha+4\alpha^2 & 3\alpha+3\alpha^2 \\ 8\alpha^2 & 4\alpha+4\alpha^2 & 1+3\alpha+4\alpha^2 \end{bmatrix}$$

*and $\alpha = \nu^{-1}$.*

**Proof:** Write (3.3), (3.4) as

$$\mathbb{E}\left[\widetilde{P}_2\right] = P_2 + \nu^{-1} A_{2,0}, \tag{3.5}$$

$$\mathbb{E}\left[\widetilde{P}_3\right] = P_3 + \nu^{-1} A_{3,1} + \nu^{-2} A_{3,0}. \tag{3.6}$$

Let $C_{r,i}$ be any of the $a_{r,i}$ terms in $A_{r,i}$. Then (3.5) implies $\mathbb{E}[\widetilde{C}_{2,0}] = C_{2,0} + \nu^{-1}(P_2 +$ one term from $A_{2,0})$. Summing over $C_{2,0}$ gives

$$\mathbb{E}\left[\widetilde{A}_{2,0}\right] = A_{2,0} + \nu^{-1}\left(2P_2 + A_{2,0}\right). \tag{3.7}$$

Writing (3.5), (3.7) as $\mathbb{E}[\widetilde{\boldsymbol{\theta}}] = \mathbf{D}_2\boldsymbol{\theta}$ proves the first statement of the theorem.

Similarly, (3.6) implies

$$\begin{aligned}
\mathbb{E}\left[\widetilde{C}_{3,1}\right] &= C_{3,1} + \nu^{-1}\left(P_3 + \text{ one term from } A_{3,1} + \text{ four from } A_{3,0}\right) \\
&\quad + \nu^{-2}\left(\text{ four from } A_{3,1} + \text{ four from } A_{3,0}\right), \\
\mathbb{E}\left[\widetilde{C}_{3,0}\right] &= C_{3,0} + \nu^{-1}\left(\text{ three from } A_{3,1} + \text{ three from } A_{3,0}\right) \\
&\quad + \nu^{-2}\left(P_3 + \text{ three from } A_{3,1} + \text{ four from } A_{3,0}\right).
\end{aligned}$$

Summing these and using $a_{3,0} = 8$, $a_{3,1} = 6$, we obtain

$$\begin{aligned}
\mathbb{E}\left[\widetilde{A}_{3,1}\right] &= A_{3,1} + \nu^{-1}\left(6P_3 + A_{3,1} + 3A_{3,0}\right) + \nu^{-2}\left(4A_{3,1} + 3A_{3,0}\right), \\
\mathbb{E}\left[\widetilde{A}_{3,0}\right] &= A_{3,0} + \nu^{-1}\left(4A_{3,1} + 3A_{3,0}\right) + \nu^{-2}\left(8P_3 + 4A_{3,1} + 4A_{3,0}\right).
\end{aligned}$$

So, the second statement follows and the proof is complete. □

Replacing $1, \ldots, 2r$ by $1, \ldots, 1$ reduces $P_r^*$ to $\Sigma_{1,1}^r \nu^r / \{\nu(\nu+2)(\nu+4)\cdots(\nu+2r-2)\}$. So, $(\nu-1)(\nu+2)$ must be a factor of $\mid \mathbf{D}_2 \mid$ and $(\nu-1)(\nu+2)(\nu+4)$ must be a factor of $\mid \mathbf{D}_3 \mid$ and so on. One finds $\mid \mathbf{D}_2 \mid = (1-\alpha)(1+2\alpha)$ and $\mid \mathbf{D}_3 \mid = (1-\alpha)^2(1-2\alpha)(1+2\alpha)^2(1+4\alpha)$ so $P_2^* = \{(1+\alpha)\widetilde{P}_2 - \alpha \widetilde{A}_{2,0}\} / \mid \mathbf{D}_2 \mid$ and $P_3^* = \{(1-\alpha)(1+2\alpha)(1+3\alpha-2\alpha^2)\widetilde{P}_3 + (1-\alpha)\alpha(1+2\alpha)^2 \widetilde{A}_{3,1} - 12\alpha^4 \widetilde{A}_{3,0}\} / \mid \mathbf{D}_3 \mid$.

The identity (3.1) can also be used to obtain other UEs, for example

$$\mathbf{f}^* = (\nu - p - 1)\nu^{-1}\boldsymbol{\Sigma}^{-1}$$

for $\mathbf{f} = \boldsymbol{\Sigma}^{-1}$ as given by equation (19), page 164 of Johnson and Kotz (1972). Note that $i$ on the right hand side of their equation (18.2) should be $-2i$.

Theorem 3.2 derives UEs for functions of $\boldsymbol{\Sigma}$. One of the functions considered is the power of the determinant of $\boldsymbol{\Sigma}$. Other possible functions could include the trace of $\boldsymbol{\Sigma}$, correlation coefficient, partial correlation coefficient, multiple correlation coefficient, coefficient of determination, canonical correlations, entropy of multivariate normal, and the characteristic function of multivariate normal with zero means.

**Theorem 3.2** *Consider the notation above. If $\nu + 2\lambda > p - 3$ then $\nu^{p\lambda} f(\widetilde{\boldsymbol{\Sigma}})/a_{\nu,p}$ is an UE of $f(\boldsymbol{\Sigma}) = \mid \boldsymbol{\Sigma} \mid^{\lambda}$, where*

$$a_{\nu,p} = 2^{\lambda p} \prod_{i=1}^{p} \{\Gamma(b_{\nu-i,\lambda}) / \Gamma(b_{\nu-i,0})\}$$

*and $b_{\nu,\lambda} = (\nu+1)/2 + \lambda$. More generally, if $g(\mathbf{B}) = f(\boldsymbol{\Sigma}^{1/2}\mathbf{B}\boldsymbol{\Sigma}^{1/2})/f(\boldsymbol{\Sigma})$ does not depend on $\boldsymbol{\Sigma}$ for $\mathbf{B}$ symmetric and positive semi-definite then $\nu^{p\lambda} f(\widetilde{\boldsymbol{\Sigma}})/a_{\nu,p}$ is an UE of $f(\boldsymbol{\Sigma})$, where*

$$a_{\nu,p} = b_{\nu,p}(g)/c_{\nu,p},\ c_{\nu,p} = 2^{\nu p/2} \pi^{(p^2-p)/4} \prod_{i=1}^{p} \Gamma(b_{\nu-i,0})$$

*and*

$$b_{\nu,p}(g) = \int_{\mathbf{B}>\mathbf{0}} \mid \mathbf{B} \mid^{(\nu-p-1)/2} g(\mathbf{B}) \exp\ (-\mathit{trace}\ \mathbf{B}/2)\, d\mathbf{B}$$

*when $b_{\nu,p}(g)$ exists and $\nu > p - 3$ (so that $c_{\nu,p}$ exists).*

**Proof:** This follows from page 154 and the lemma on page 156 of Anderson (1958). □

## 4 A simulation study

Here, we perform simulations to compare the estimates proposed in Theorem 2.1 and Theorem 3.2 with traditional ones. We consider four of the twelve parameters discussed in Section 1. Throughout, we assume $(X_1, X_2)$ is a random vector having the bivariate normal distribution with means, $(\mu_1, \mu_2)$, variances, $(\Sigma_{1,1}, \Sigma_{2,2})$ and correlation coefficient $\Sigma_{1,2}/\sqrt{\Sigma_{1,1}\Sigma_{2,2}} = \rho$. Let $\{(x_{1,i}, x_{2,i}), i = 1, 2, \ldots, n\}$ denote a random sample from that

distribution for $\mu_1 = \mu_2 = 1$ and $\Sigma_{1,1} = \Sigma_{2,2} = 1$. Let $\overline{x}_1$ and $\overline{x}_2$ denote the sample means, and let $\mathbf{S}$ denote the sample covariance matrix.

First, consider the ratio, $\mu_1/\mu_2$, the first example discussed in Section 1. An estimate of bias $O(1/n^2)$ can be obtained from (2.1) by setting $k = 2$:

$$(\overline{x}_1/\overline{x}_2)\left(1 - n^{-3}\widetilde{c}_1^{\mathbf{I}}\right), \tag{4.1}$$

where $\mathbf{I} = (1, -1)$ and $\widetilde{c}_1^{\mathbf{I}}$ is given by Corollary 2.2. A traditional estimate for $\mu_1/\mu_2$ is

$$\overline{x}_1/\overline{x}_2. \tag{4.2}$$

Next, consider the multivariate coefficient of variation defined by $\Sigma_{1,2}/(\mu_1\mu_2)$, the second example discussed in Section 1. An estimate of bias $O(1/n^2)$ from (2.1) is:

$$(\overline{x}_1\overline{x}_2)^{-1}\left(S_{1,2} - n^{-3}\widetilde{c}_1^{\mathbf{I}}\right), \tag{4.3}$$

where $\mathbf{I} = (-1, -1)$ and $\widetilde{c}_1^{\mathbf{I}}$ is given by Corollary 2.2. A traditional estimate for $\Sigma_{1,2}/(\mu_1\mu_2)$ is

$$(\overline{x}_1\overline{x}_2)^{-1} S_{1,2}. \tag{4.4}$$

Next, consider the conditional moment, $\mathbb{E}(X_2 \mid X_1 = 0) = \mu_2 - \Sigma_{2,1}\Sigma_{1,1}^{-1}\mu_1$, the sixth example discussed in Section 1. An estimate of bias $O(1/n^2)$ obtained from (2.1) by noting that $\overline{x}_2$ is an unbiased estimate of $\mu_2$ is:

$$\overline{x}_2 - \overline{x}_1\left(S_{2,1}S_{1,1}^{-1} - n^{-3}\widetilde{c}_1^{\mathbf{I}}\right), \tag{4.5}$$

where $\mathbf{I} = (1, 0)$ and $\widetilde{c}_1^{\mathbf{I}}$ is given by Corollary 2.2. A traditional estimate for $\mathbb{E}(X_2 \mid X_1 = 0)$ is

$$\overline{x}_2 - S_{2,1}S_{1,1}^{-1}\overline{x}_1. \tag{4.6}$$

Finally, consider the entropy given by $(2\pi e)\sqrt{|\mathbf{\Sigma}|}$, the eleventh example discussed in Section 1. By Theorem 3.2, an unbiased estimate is:

$$2\pi e(n-1)\,|\mathbf{S}|^{1/2}/(n-2). \tag{4.7}$$

A traditional estimate for $(2\pi e)\sqrt{|\mathbf{\Sigma}|}$ is

$$2\pi e\,|\mathbf{S}|^{1/2}. \tag{4.8}$$

We now compare the performances of (4.1) versus (4.2), (4.3) versus (4.4), (4.5) versus (4.6), and (4.7) versus (4.8) by simulation. We use two criteria for comparing the two estimates: bias and mean squared error. These criteria for each of (4.1)-(4.8) and for each of $\rho = 0.2, 0.4, 0.6, 0.8$ and $n = 2, 3, \ldots, 1000$ are computed by simulating 10000 replications of $\{(x_{1,i}, x_{2,i}), i = 1, 2, \ldots, n\}$.

The plots of the mean squared error and the bias versus $n$ for $\rho = 0.2, 0.4, 0.6, 0.8$ and for the eight estimates are shown in Figures 4.1 to 4.8. The $x$ axes are plotted on log scale. In Figures 4.7 and 4.8, $n$ ranges from 3 to 1000 because (4.7) is not defined for $n = 2$. In Figures 4.1, 4.3, 4.5 and 4.7, the line of zero bias is drawn in green.

The proposed estimates appear better with respect to both bias and mean squared error for all values of $\rho$ and $n$. The reduction in bias and mean squared error appears significant for small to moderate values of $n$. For all sufficiently large $n$, the proposed and the traditional estimates appear indistinguishable. The biases and mean squared errors for both estimates approach zero as $n$ increases, but those for the proposed estimate approach zero at a faster rate. The biases for the proposed estimates appear to be almost zero for all $n \geq 50$. The biases for both estimates for the conditional moment and entropy appear negative most of the time.

Because of lack of space, we have only considered four of the twelve examples discussed in Section 1. But simulations not reported here show that the results are similar for each of the twelve examples. For each example, the proposed estimates reduce both bias and mean squared error for all values of $\rho$ and $n$.

Although the results in Sections 2 and 3 are tailored for the multivariate normal distribution, they can be applied for non-normal distributions too provided $n$ is sufficiently large. We performed a simulation study as above for a range of non-normal distributions, including the bivariate log-normal, bivariate $t$, bivariate exponential and bivariate gamma distributions. We observed significant reductions in both bias and mean squared error for $n$ sufficiently large. The results are not presented here again because of lack of space.

## Appendix A

For $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{V})$ and $\mathbf{j}$ a non-negative integer in $\mathbb{R}^p$ we give expressions for $\mathbb{E}[\mathbf{X}^{\mathbf{j}}]$, where $\mathbf{X}^{\mathbf{j}} = \prod_{i=1}^p X_i^{j_i}$.

Set $j_\cdot = \sum_{i=1}^p j_i$, $\lambda_i = \mathbb{E}[\mathcal{N}_1(0,1)^{2i}] = 1 \cdot 3 \cdot 5 \cdots (2i-1)$, $\partial_i = \partial/\partial t_i$, $\partial^{\mathbf{j}} = \prod_{i=1}^p \partial_i^{j_i}$ and $v = \mathbf{t}'\mathbf{V}\mathbf{t}$ for $\mathbf{t}$ in $\mathbb{R}^p$. Then $\mathbb{E}[\mathbf{X}^{\mathbf{j}}] = \partial^{\mathbf{j}} \mathbb{E}\left[\exp\left(\mathbf{t}'\mathbf{X}\right)|_{\mathbf{t}=\mathbf{0}}\right] = (-1)^{\mathbf{j}} He_{\mathbf{j}}(\mathbf{0}, -\mathbf{V})$ as noted in Withers (1985), where $He_{\mathbf{j}}(\mathbf{t}, \mathbf{V}) = \exp\left(v/2\right)(-\partial)^{\mathbf{j}} \exp\left(-v/2\right)$, the multivariate Hermite polynomial. More simply $\mathbf{X}^{\mathbf{j}} = \partial^{\mathbf{j}}(\mathbf{t}'\mathbf{X})^k/k!$ for $k = j_\cdot$ so that $\mathbb{E}[\mathbf{X}^{\mathbf{j}}] = a_i \partial^{\mathbf{j}} v^i$ for $j_\cdot = 2i$, where $a_i = \lambda_i/(2i)! = 2^{-i}/i!$.

The dimension can be reduced by one by noting that $\partial^{\mathbf{j}}(\mathbf{t}'\mathbf{X})^k = (k)_{j_\cdot} \mathbf{X}^{\mathbf{j}} (\mathbf{t}'\mathbf{X})^{k-j_\cdot}$, where $(k)_i = k!/(k-i)!$ and setting $\mathbf{t} = \mathbf{e}_p$, the $p$th unit vector. This gives

$$\mathbb{E}\left[\mathbf{X}^{\mathbf{j}} X_p^{2i-j_\cdot}\right] = \partial^{\mathbf{j}} v^i \Big|_{\mathbf{t}=\mathbf{e}_p} a_{i,j_\cdot}, \tag{A.1}$$

where $a_{i,k} = \lambda_i/(2i)_k = (2i-k)! a_i$. For example,

$$\mathbb{E}\left[X_1^{j_1} X_2^{2i-j_1}\right] = \partial_1^{j_1} v^i \Big|_{\mathbf{t}=\mathbf{e}_p} a_{i,j_1}. \tag{A.2}$$

From Withers (1984) it follows that

$$\partial_{\alpha_1} \cdots \partial_{\alpha_r} v^i = \sum_{l+2m=r} 2^{l+m} (i)_{l+m} v^{i-l-m} \sum^M t_{a_1} \cdots t_{a_l} V_{b_1,b_2} \cdots V_{b_{2m-1},b_{2m}}, \tag{A.3}$$

where $l \geq 0$, $m \geq 0$ and $\sum^M$ sums over all $M = r!/(l!m!2^m)$ partitions of $(\alpha_1, \ldots, \alpha_r)$ giving distinct terms allowing for the symmetry of $\mathbf{V}$. For example,

$$\partial_1^r v^i = r! \sum_{l+2m=r} 2^l (i)_{l+m} v^{i-l-m} t_1^l V_{1,1}^m / \left(l!m!\right). \tag{A.4}$$

Note that (A.3) can now be used to obtain the left hand side of (A.1). For example, (A.2), (A.4) imply for $j_1 + j_2 = 2i$

$$\mathbb{E}\left[X_1^{j_1} X_2^{j_2}\right] = 2^{-i} j_1! j_2! \sum_k V_{1,1}^{k_1} \left(2V_{1,2}\right)^k V_{2,2}^{k_2} / \left(k_1! k! k_2!\right),$$

where $k_i = (j_i - k)/2$ and $\sum_k$ sums over $0 \le k \le \min(j_1, j_2)$ with $k$ odd or even according as $(j_1, j_2)$ are odd or even. In particular, for $i \ge 0$

$$\mathbb{E}\left[X_1^{2i}\right] = 2^{-i}(2i)! V_{1,1}^i / i! = \lambda_1 V_{1,1}^i, \tag{A.5}$$

$$\mathbb{E}\left[X_1^{2i-1} X_2\right] = 2^{1-i}(2i-1)! V_{1,1}^{i-1} V_{1,2} / (i-1)! = \lambda_i V_{1,1}^{i-1} V_{1,2}, \tag{A.6}$$

$$\mathbb{E}\left[X_1^{2i-2} X_2^2\right] = 2^{1-i}(2i-2)! V_{1,1}^{i-2} \left\{V_{1,1}V_{2,2} + 2(i-1)V_{1,2}^2\right\} / (i-1)!, \tag{A.7}$$

$$\mathbb{E}\left[X_1^{2i-3} X_2^3\right] = 2^{2-i}(2i-3)! V_{1,1}^{i-3} V_{1,2} \left\{3V_{1,1}V_{2,2} + 4(i-2)V_{1,2}^2\right\} / (i-2)!, \tag{A.8}$$

$$\mathbb{E}\left[\left(X_1 X_2\right)^4 / (4!)^2\right] = V_{1,1}^2 V_{2,2}^2 / 32 + V_{1,1} V_{1,2}^2 V_{2,2} / 8 + V_{1,2}^4 / 4!. \tag{A.9}$$

The noncentral moments of $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{V})$ are now obtainable from

$$\mathbb{E}\left[\mathbf{Y}^{\mathbf{j}}\right] = \mathbb{E}\left[(\boldsymbol{\mu} + \mathbf{X})^{\mathbf{j}}\right] = \sum_{\mathbf{l}=\mathbf{0}}^{\mathbf{j}} \binom{\mathbf{j}}{\mathbf{l}} \boldsymbol{\mu}^{\mathbf{j}-\mathbf{l}} \mathbb{E}\left[\mathbf{X}^{\mathbf{l}}\right]. \tag{A.10}$$

From Withers (1985), $\mathbb{E}[\mathbf{Y}^{\mathbf{j}}] = (-1)^{j\cdot} He_{\mathbf{j}}(\mathbf{V}^{-1}\boldsymbol{\mu}, -\boldsymbol{\Sigma})$, so (A.10) gives an expression for the general multivariate Hermite polynomial.

## Appendix B

Here, we give $\overline{g}_4^{\mathbf{I}} = \overline{g}_4^{\mathbf{I}}(\boldsymbol{\theta})$ needed for (2.8) and details of the proof of (2.9)-(2.12). If we drop the superscript $\mathbf{I}$ then (2.5) implies (2.9). Note also that (2.6) implies

$$\begin{aligned}(g_2, g_{1\cdot 1}, G_2) \;=\; & t \sum^{2} \Big\{ \left(I_{4,0} V_{1,1}^2 + 4 I_{3,1} V_{1,1} V_{1,2}\right)(1, 2, -1) \\ & + I_{2,2} \Big[ V_{1,1} V_{2,2} (1, 4, -3) + 2 V_{1,2}^2 (1, 2, -1) \Big] \Big\} / 8,\end{aligned}$$

which implies (2.10). Note also that (A.5)-(A.8) and (2.1) imply

$$\begin{aligned}(g_3, g_{1\cdot 2}, G_{2\cdot 1}, G_3) \;=\; & t \sum^{2} \Big\{ \left(I_{6,0} V_{1,1}^3 + 6 I_{5,1} V_{1,1}^2 V_{1,2}\right)(1, 3, -3, -1) \\ & + 3 I_{4,2} \Big[ V_{1,1}^2 V_{2,2} (1, 3, -7, 5) + V_{1,1} V_{1,2}^2 (4, 13, -10, 1) \Big] \\ & + I_{3,3} \Big[ 3 V_{1,1} V_{1,2} V_{2,2} (2, 4, -7, 5) + 2 V_{1,2}^3 (4, 6, -3, 1) \Big] \Big\} / 48,\end{aligned}$$

which implies (2.11). Note also that (A.5)-(A.9) and (2.7) imply

$$\begin{aligned}
& (g_4, g_{1\cdot 3}, G_{2\cdot 2}, G_{3\cdot 1}, \overline{g}_4) \\
= \; & t\sum^{2}\Big\{ \left(I_{8,0}V_{1,1}^4 + 8I_{7,1}V_{1,1}^3V_{1,2}\right)(1,2,-6,4,-1) \\
& +4I_{6,2}\Big[V_{1,1}^3V_{2,2}(1,4,-12,16,-7) + 3V_{1,1}^2V_{1,2}(2,8,-12,5,1)\Big] \\
& +4I_{5,3}\Big[3V_{1,1}^2V_{1,2}V_{2,2}(1,8,-20,22,-9) + 2V_{1,1}V_{1,2}^3(4,20,-24,5,3)\Big] \\
& +2I_{4,4}\Big[3V_{1,1}^2V_{2,2}^2(1,2,-7,10,-4) + 6V_{1,1}V_{1,2}^2V_{2,2}(2,8,-16,11,-1) \\
& +4V_{1,2}^4(1,8,-6,2,-3)\Big]\Big\}/384,
\end{aligned}$$

which implies (2.12).

## Acknowledgments

The authors would like to thank the Editor and the Associate Editor for careful reading and for their comments which greatly improved the paper.

## References

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Analysis*. John Wiley and Sons, New York.

[2] Bennett, B. M. (1977). On multivariate coefficients of variation. *Statistical Papers*, **18**, 123-128.

[3] Bose, P. (1942). On the exact distribution of the ratio of two means belonging to samples drawn from a given correlated bivariate normal population. *Bulletin of Calcutta Mathematical Society*, **34**, 139-141.

[4] Eaton, M. L. and Morris, C. N. (1970). The application of invariance to unbiased estimation. *Annals of Mathematical Statistics*, **41**, 1708-1716.

[5] Ghurye, S. G. and Olkin, I. (1969). Unbiased estimation of some multivariate probability densities and related functions. *Annals of Mathematical Statistics*, **40**, 1261-1271.

[6] Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Annals of Mathematical Statistics*, **8**, 586-597.

[7] Hannig, J. (2006). Asymptotic bounds for coverage probabilities for a class of confidence intervals for ratio of means in a bivariate normal distribution. *Journal of Probability and Statistical Science*, **4**, 41-49.

[8] Holmquist, B. (1988). Moments and cumulants of the multivariate normal distribution. *Stochastic Analysis and Applications*, **6**, 273-278.

[9] Hutson, A. D. (2002). Quasi-medians are robust and relatively efficient estimators of a common mean given multivariate normality. *Statistics and Probability Letters*, **57**, 403-408.

[10] Johnson, N. L. and Kotz, S. (1972). *Distribution in Statistics: Continuous Multivariate Distributions*. Wiley, New York.

[11] Kang, C. and Kim, B. C. (1996). The general moment of non-central Wishart distribution. *Journal of the Korean Statistical Society*, **25**, 393-406.

[12] Kollo, T. and von Rosen, D. (1995). Minimal moments and cumulants of symmetric matrices: an application to the Wishart distribution. *Journal of Multivariate Analysis*, **55**, 149-164.

[13] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519-530.

[14] Mardia, K. V., Kent, J. T. and Bibby, J. M. (1980). *Multivariate Analysis*. Academic Press, London.

[15] Misra, N., Singh, H. and Demchuk, E. (2005). Estimation of the entropy of a multivariate normal distribution. *Journal of Multivariate Analysis*, **92**, 324-342.

[16] Mukhopadhyay, P. (2009). *Multivariate Statistical Analysis*. World Scientific Publishing Company, Hackensack, New Jersey.

[17] Roy, S. N. and Potthoff, R. F. (1958). Confidence bounds on vector analogues of the "ratio of means" and the "ratio of variances" for two correlated normal variates and some associated tests. *Annals of Mathematical Statistics*, **29**, 829-841.

[18] Sultan, S. A. and Tracy, D. S. (1996a). Moments of Wishart distribution. *Stochastic Analysis and Applications*, **14**, 237-243.

[19] Sultan, S. A. and Tracy, D. S. (1996b). Moments of the complex multivariate normal distribution. *Linear Algebra and Its Applications*, **237-238**, 191-204.

[20] Sultan, S. A. and Tracy, D. S. (1999). Moments of the complex Wishart distribution. *International Journal of Mathematical and Statistical Sciences*, **8**, 51-62.

[21] Sun, D. and Sun, X. (2005). Estimation of the multivariate normal precision and covariance matrices in a star-shape model. *Annals of the Institute of Statistical Mathematics*, **57**, 455-484.

[22] Sun, D. and Sun, X. (2006). Estimation of multivariate normal covariance and precision matrices in a star-shape model with missing data. *Journal of Multivariate Analysis*, **97**, 698-719.

[23] Wan, A. T. K. and Zou, G. (2004). On unbiased and improved loss estimation for the mean of a multivariate normal distribution with unknown variance. *Journal of Statistical Planning and Inference*, **119**, 17-22.

[24] Withers, C. S. (1984). A chain rule for differentiation with applications to multivariate Hermite polynomials. *Bulletin of the Australian Mathematical Society*, **30**, 247-250.

[25] Withers, C. S. (1985). The moments of the multivariate normal. *Bulletin of the Australian Mathematical Society*, **32**, 103-108.

[26] Withers, C. S. (1987). Bias reduction by Taylor series. *Communications in Statistics—Theory and Methods*, **16**, 2369-2384.

[27] Withers, C. S. and Nadarajah, S. (2010a). Estimates for the inverse powers of a normal mean. *Technical Report*, Applied Mathematics Group, Industrial Research Ltd., Lower Hutt, New Zealand.

[28] Withers, C. S. and Nadarajah, S. (2010b). Moments and cumulants for the complex Wishart. *Utilitas Mathematica*, to appear.

[29] Yanagihara, H. (2006). Corrected version of AIC for selecting multivariate normal linear regression models in a general nonnormal case. *Journal of Multivariate Analysis*, **97**, 1070-1089.

[30] Zografos, K. and Nadarajah, S. (2005). Expressions for Rényi and Shannon entropies for multivariate distributions. *Statistics and Probability Letters*, **71**, 71-84.

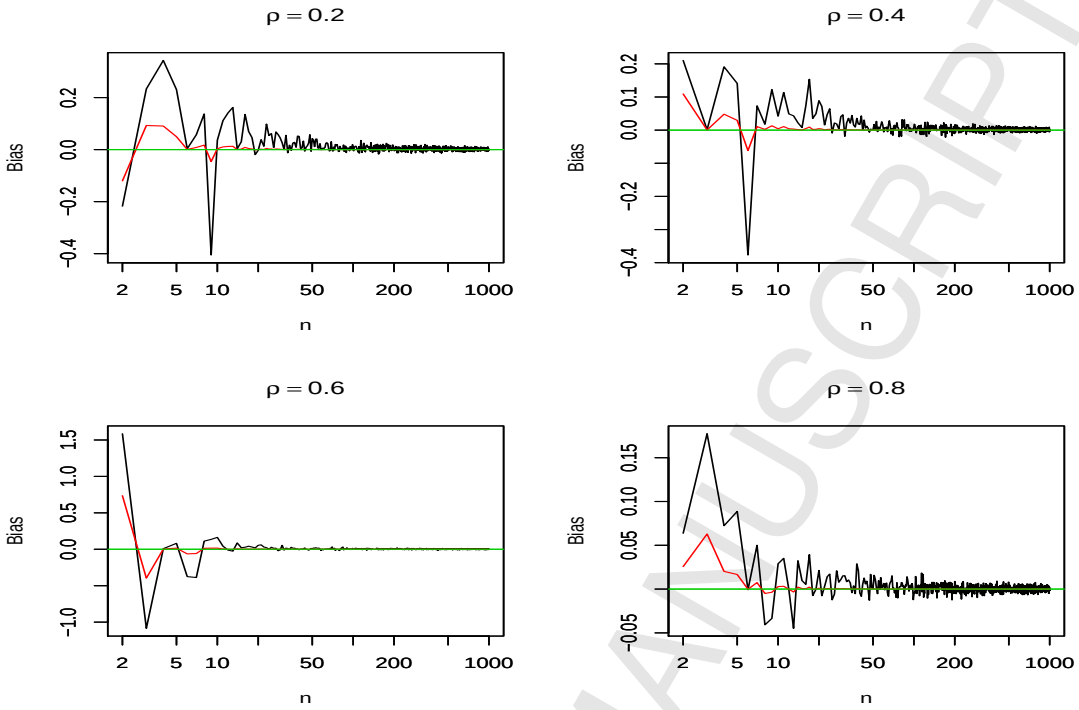



**Figure 4.1** Biases of the estimates given by (4.1) (in red) and (4.2) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

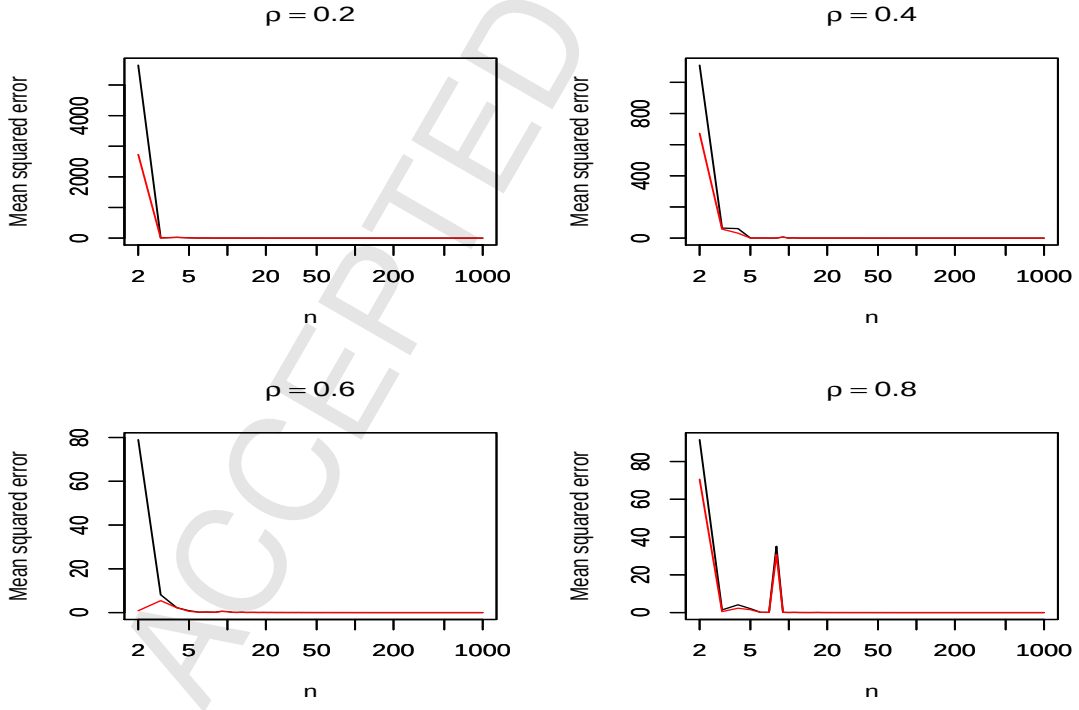



**Figure 4.2** Mean squared errors of the estimates given by (4.1) (in red) and (4.2) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

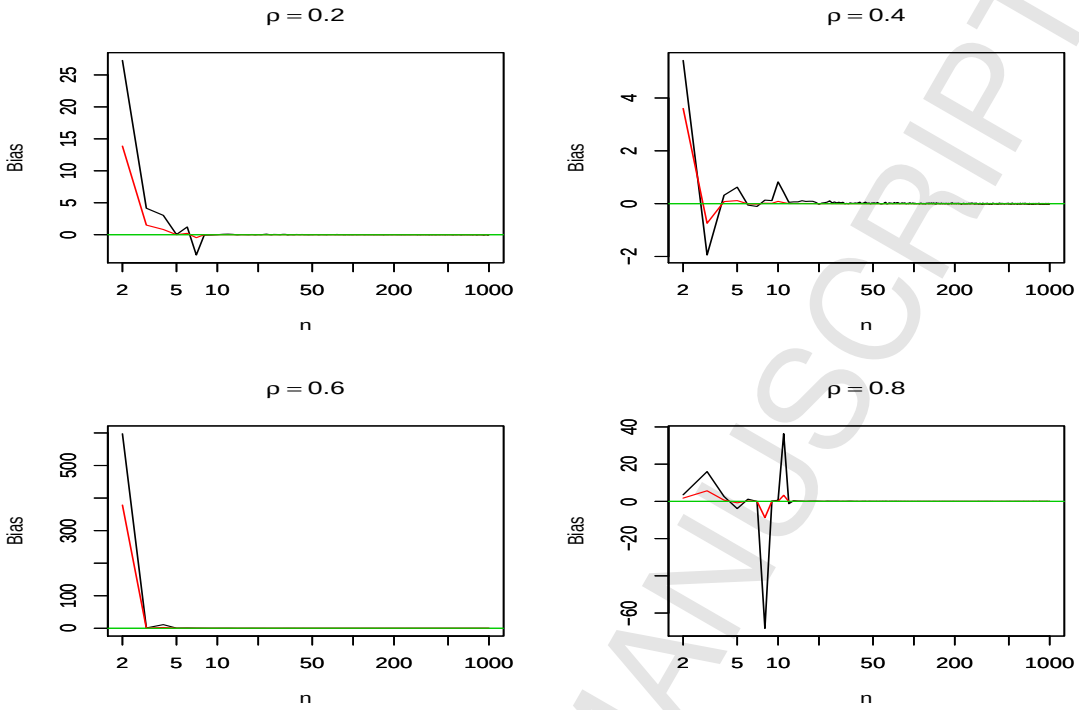



**Figure 4.3** Biases of the estimates given by (4.3) (in red) and (4.4) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

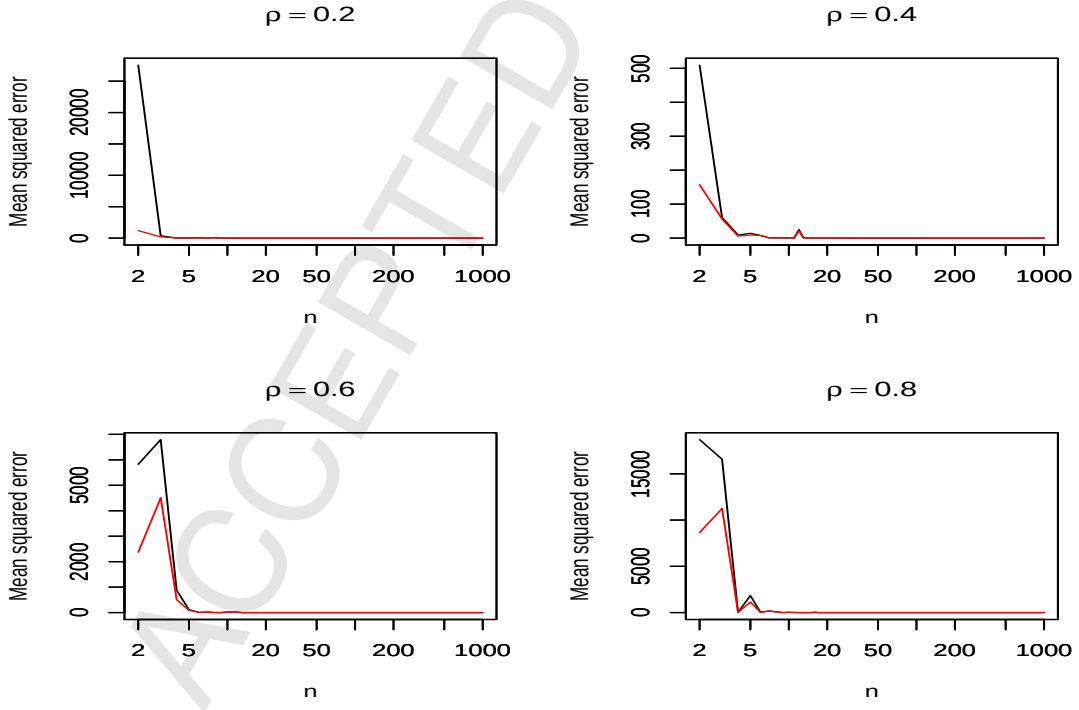



**Figure 4.4** Mean squared errors of the estimates given by (4.3) (in red) and (4.4) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

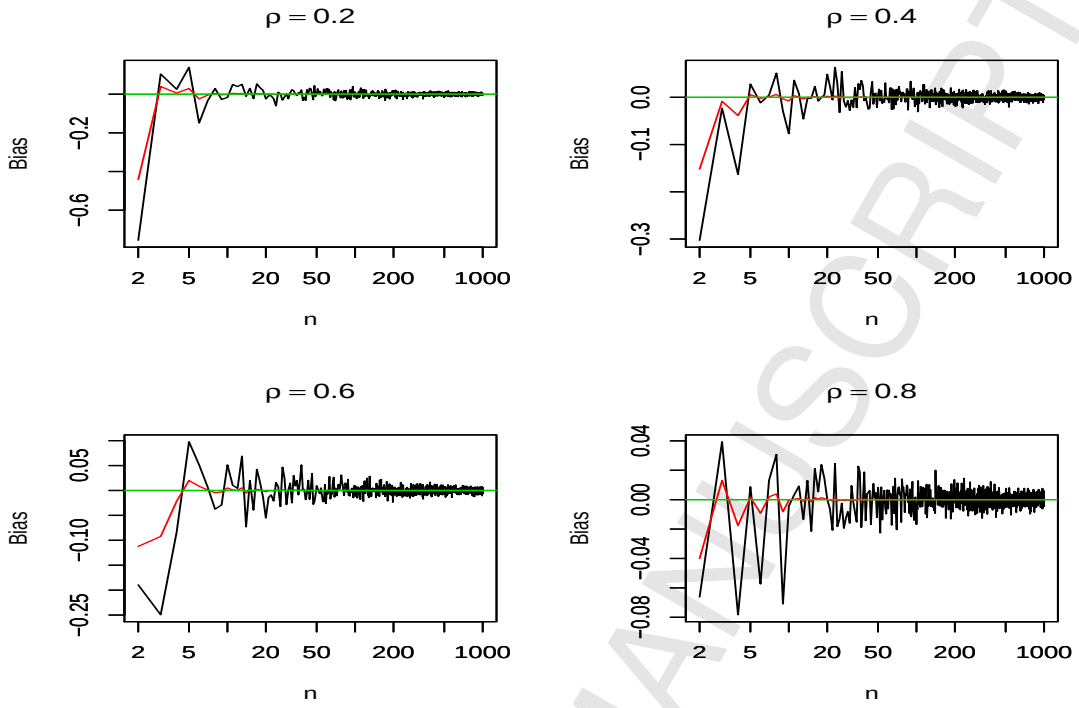



**Figure 4.5** Biases of the estimates given by (4.5) (in red) and (4.6) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

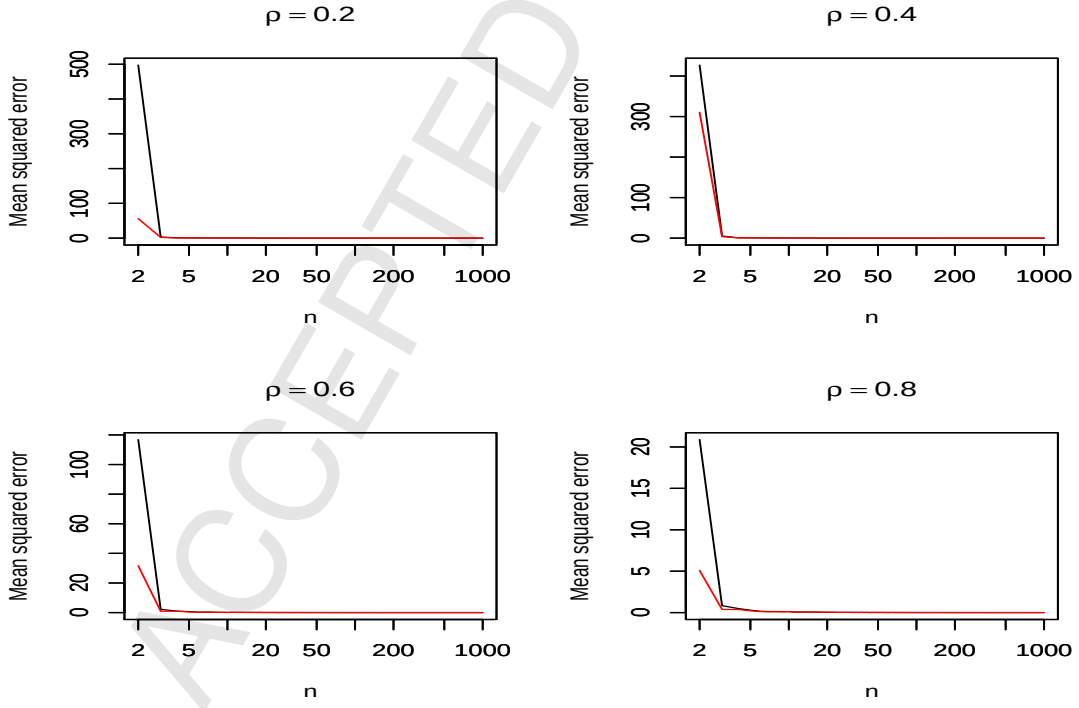



**Figure 4.6** Mean squared errors of the estimates given by (4.5) (in red) and (4.6) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

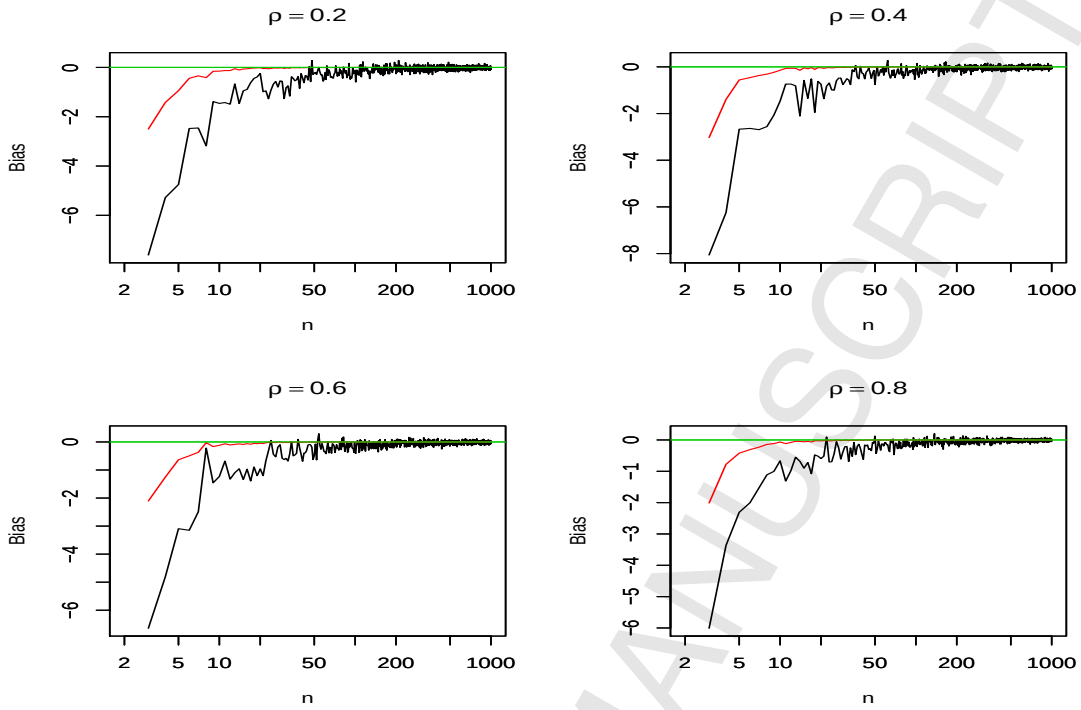



**Figure 4.7** Biases of the estimates given by (4.7) (in red) and (4.8) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.

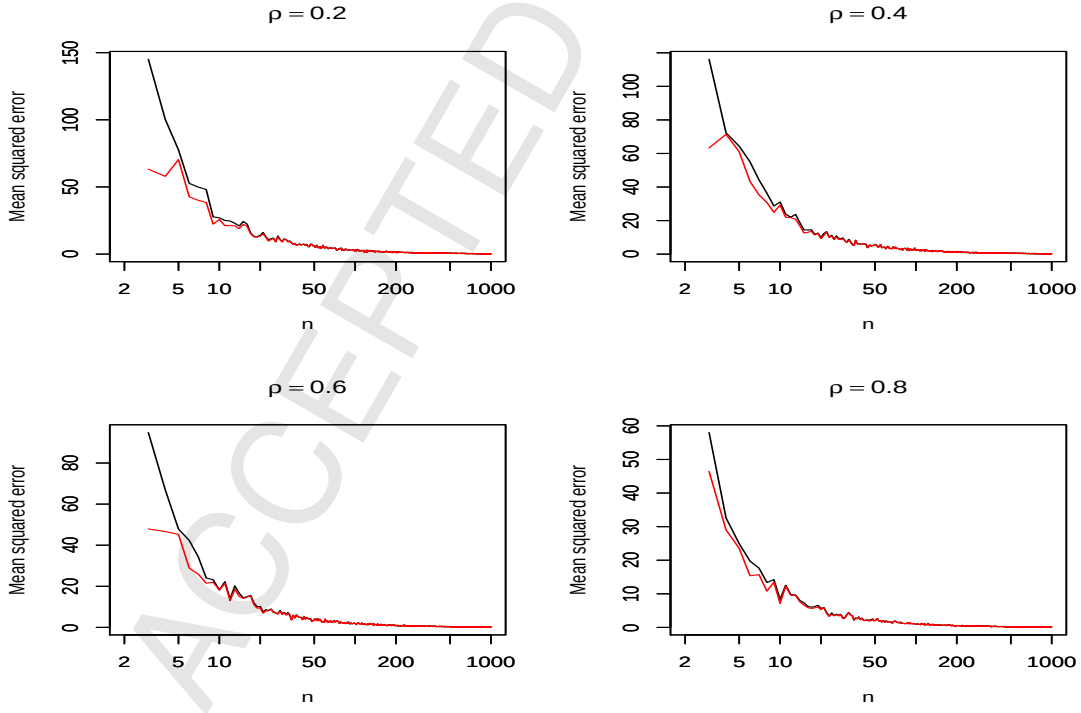



**Figure 4.8** Mean squared errors of the estimates given by (4.7) (in red) and (4.8) (in black) for $\mu_1 = \mu_2 = 1$, $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\rho = 0.2, 0.4, 0.6, 0.8$.