Finite-Sample Inference with Monotone Incomplete Multivariate Normal Data, I

Wan-Ying Chang* and Donald St. P. Richards†

July 23, 2008

Abstract

We consider problems in finite-sample inference with two-step, monotone incomplete data drawn from $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a multivariate normal population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We derive a stochastic representation for the exact distribution of $\widehat{\boldsymbol{\mu}}$, the maximum likelihood estimator of $\boldsymbol{\mu}$. We obtain ellipsoidal confidence regions for $\boldsymbol{\mu}$ through T^2, a generalization of Hotelling's statistic. We derive the asymptotic distribution of, and probability inequalities for, T^2 under various assumptions on the sizes of the complete and incomplete samples. Further, we establish an upper bound for the supremum distance between the probability density functions of $\widehat{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}$, a normal approximation to $\widehat{\boldsymbol{\mu}}$.

1 Introduction

During the past eighty years, there has been an enduring interest in multivariate statistical inference with incomplete data. Wilks [30] was one of the earliest contributors to this area of research, the subsequent literature has been voluminous, and we refer to Little and Rubin [23] for an extensive treatment of the field.

In this paper, we consider problems in inference with multivariate, d-dimensional data, drawn from a normal population. We suppose that the data are composed of N mutually independent observations consisting of a random sample of n complete observations on all $d = p + q$ characteristics and an additional $N - n$ incomplete observations on the last q characteristics only. We write the data in the form

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{Y}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_2 \\ \boldsymbol{Y}_2 \end{pmatrix} \cdots \begin{pmatrix} \boldsymbol{X}_n \\ \boldsymbol{Y}_n \end{pmatrix} \boldsymbol{Y}_{n+1} \ \boldsymbol{Y}_{n+2} \ \cdots \ \boldsymbol{Y}_N , \tag{1.1}$$

where each $\boldsymbol{X}_j$ is $p \times 1$, each $\boldsymbol{Y}_j$ is $q \times 1$, the complete observations $(\boldsymbol{X}_j', \boldsymbol{Y}_j')'$, for $j = 1, \ldots, n$, are drawn from $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a multivariate normal population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and the incomplete data $\boldsymbol{Y}_j$, $j = n+1, \ldots, N$, are observations on the last q characteristics of the same population. The data in (1.1) are called *two-step*

*Washington Department of Fish and Wildlife, Olympia, WA 98501, USA.

†Department of Statistics, Penn State University, University Park, PA 16802; and The Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC 27709, USA.

†Supported in part by National Science Foundation grants DMS-0705210 and DMS-0112069.

2000 Mathematics Subject Classification: Primary 62H10; Secondary 60D10, 62E15.

Key words and phrases. Bonferroni inequalities; ellipsoidal confidence regions; Esseen inequality; Hotelling's T^2-statistic; matrix F-distribution; maximum likelihood estimation; missing data; simultaneous confidence intervals; Wishart distribution.

monotone, and have been widely studied; cf. Anderson [1], Bhargava [7], Morrison [24], Eaton and Kariya [11], and Hao and Krishnamoorthy [15].

Given a sample (1.1) from the population $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, it is well-known that closed-form expressions for $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, may be obtained by factoring the likelihood function into a product of likelihoods with non-overlapping sets of parameters; consequently, explicit expressions for various likelihood ratio test statistics may be obtained. We refer to Anderson [1], Little and Rubin [23], Anderson and Olkin [3], and Jinadasa and Tracy [16] for derivations of the explicit formulas for $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$; Bhargava [7, 8], Eaton and Kariya [11], and Andersson and Perlman [4, 5, 6] for other aspects of inference with missing data in which factorization of the likelihood function plays a crucial role; and Morrison [24], Giguère and Styan [14], Little and Rubin [23], Kanda and Fujikoshi [17], for results on the moments and asymptotic distributions of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$.

In the literature on inference for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, it is noticeable that the exact distributions of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ have remained unknown. This problem is basic to inference with incomplete data when large samples are infeasible or impractical such as: in sociological research, where subjects have transient lifestyles and cannot be contacted for further data collection after relocation to new addresses; in panel surveys, where subjects may be available for only part of the study; and in astronomy, where monotone incomplete data arises in the classification of galaxies (Lang [20]). It is noticeable that the area of monotone incomplete multivariate normal inference is not well-endowed with the range of explicit formulas appearing in Anderson [2], Eaton [10], and Muirhead [26]. Thus, this paper initiates a program of research on inference for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, where n and N are fixed, with the goal of deriving explicit results analogous to those existing in the classical complete case. A synopsis of our results is as follows.

We provide in Section 2 some preliminary results needed in the sequel. In Section 3, we derive a stochastic representation for the exact distribution of $\widehat{\boldsymbol{\mu}}$. Then, generalizing results of Morrison [24], we apply the stochastic representation to deduce formulas for all marginal central moments of $\widehat{\mu}_1, \ldots, \widehat{\mu}_{p+q}$, the components of $\widehat{\boldsymbol{\mu}}$.

In Section 4, we list some properties of $\widehat{\boldsymbol{\Sigma}}$ that are needed to analyze the distribution of T^2, an analog of Hotelling's statistic, and for a companion paper [9]. In Section 5, we obtain the asymptotic distribution of T^2 and inequalities for its distribution function; by means of these results, lower and upper bounds may be obtained for the confidence levels of ellipsoidal confidence regions obtained for $\boldsymbol{\mu}$ through T^2. Finally, in Section 6, we derive an upper bound on the supremum distance between the density and distribution functions of $\widehat{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}$, a normal approximation to $\widehat{\boldsymbol{\mu}}$.

2 Preliminary results

Throughout the paper, we write all vectors and matrices in boldface type. We denote by $\mathbf{0}$ any zero vector or matrix, the dimension of which will be clear from the context, and we denote the identity matrix of order d by $\boldsymbol{I}_d$. We write $\boldsymbol{A} > \mathbf{0}$ to denote that a matrix $\boldsymbol{A}$ is positive definite (symmetric), and we write $\boldsymbol{A} \geq \boldsymbol{B}$ to mean that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

Let $\boldsymbol{M}$ be a $p \times q$ matrix, $\boldsymbol{C}$ and $\boldsymbol{D}$ be $p \times p$ and $q \times q$ positive definite (symmetric) matrices, respectively, and denote by $\boldsymbol{C} \otimes \boldsymbol{D}$ the Kronecker product of $\boldsymbol{C}$ and $\boldsymbol{D}$. If $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\boldsymbol{C}$, denote by $\boldsymbol{C}^{1/2}$ the positive definite square root of $\boldsymbol{C}$ whose eigenvalues are $\lambda_1^{1/2}, \ldots, \lambda_p^{1/2}$ [26, p. 588, Theorem A9.3 *infra*], and denote by $\boldsymbol{C}^{-1/2}$ the inverse of $\boldsymbol{C}^{1/2}$.

Following [26, p. 79], we say that a $p \times q$ random matrix $\boldsymbol{B}_{12}$ has a multivariate normal distribution with mean $\boldsymbol{M}$ and covariance matrix $\boldsymbol{C} \otimes \boldsymbol{D}$, denoted $\boldsymbol{B}_{12} \sim \mathrm{N}(\boldsymbol{M}, \boldsymbol{C} \otimes \boldsymbol{D})$, if the probability density function of $\boldsymbol{B}_{12}$ is

$$(2\pi)^{-pq/2}|\boldsymbol{C}|^{-q/2}|\boldsymbol{D}|^{-p/2}\exp\big[-\tfrac{1}{2}\mathrm{tr}\,\boldsymbol{C}^{-1}(\boldsymbol{B}_{12}-\boldsymbol{M})\boldsymbol{D}^{-1}(\boldsymbol{B}_{12}-\boldsymbol{M})'\big],$$

$\boldsymbol{B}_{12}\in\mathbb{R}^{p\times q}$. As noted in [26, *loc. cit.*], this distribution is related to the classical multivariate normal distribution as follows: Let $\boldsymbol{T}$ be a rectangular matrix with columns $\boldsymbol{t}_1,\ldots,\boldsymbol{t}_r$, and define the vector $\mathrm{vec}(\boldsymbol{T})$ as

$$\mathrm{vec}(\boldsymbol{T})=\begin{pmatrix}\boldsymbol{t}_1\\ \vdots\\ \boldsymbol{t}_r\end{pmatrix}.$$

Then $\boldsymbol{B}_{12}\sim \mathrm{N}(\boldsymbol{M},\boldsymbol{C}\otimes\boldsymbol{D})$ is equivalent to $\mathrm{vec}(\boldsymbol{B}_{12}')\sim \mathrm{N}_{pq}(\mathrm{vec}(\boldsymbol{M}'),\boldsymbol{C}\otimes\boldsymbol{D})$.

Lemma 2.1. *Let $\boldsymbol{B}_{12}\sim \mathrm{N}(\boldsymbol{0},\boldsymbol{C}\otimes\boldsymbol{D})$, $\boldsymbol{\Lambda}\geq\boldsymbol{0}$ be $q\times q$, and $\boldsymbol{u}\in\mathbb{R}^p$. Then*

$$E\exp(-\boldsymbol{u}'\boldsymbol{B}_{12}\boldsymbol{D}^{-1}\boldsymbol{\Lambda}\boldsymbol{D}^{-1}\boldsymbol{B}_{12}'\boldsymbol{u})=|\boldsymbol{I}_q+2(\boldsymbol{u}'\boldsymbol{C}\boldsymbol{u})\boldsymbol{\Lambda}\boldsymbol{D}^{-1}|^{-1/2}. \tag{2.1}$$

Proof. Since $\boldsymbol{B}_{12}\sim \mathrm{N}(\boldsymbol{0},\boldsymbol{C}\otimes\boldsymbol{D})$, equivalently, $\mathrm{vec}(\boldsymbol{B}_{12}')\sim \mathrm{N}_{pq}(\boldsymbol{0},\boldsymbol{C}\otimes\boldsymbol{D})$, then $\boldsymbol{D}^{-1/2}\boldsymbol{B}_{12}'\boldsymbol{u}\sim \mathrm{N}_q(\boldsymbol{0},(\boldsymbol{u}'\boldsymbol{C}\boldsymbol{u})\boldsymbol{I}_q)$. Hence,

$$\boldsymbol{D}^{-1/2}\boldsymbol{B}_{12}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{B}_{12}\boldsymbol{D}^{-1/2}\equiv(\boldsymbol{D}^{-1/2}\boldsymbol{B}_{12}'\boldsymbol{u})(\boldsymbol{D}^{-1/2}\boldsymbol{B}_{12}'\boldsymbol{u})'\overset{\mathcal{L}}{=}(\boldsymbol{u}'\boldsymbol{C}\boldsymbol{u})\boldsymbol{W},$$

where $\boldsymbol{W}\sim W_q(1,\boldsymbol{I}_q)$, a Wishart distribution with 1 degree of freedom. Then, (2.1) follows from the well-known formula for the moment-generating function of the Wishart distribution. □

Suppose that $\boldsymbol{W}\sim \mathrm{W}_d(a,\boldsymbol{\Lambda})$, a Wishart distribution, where $a>d-1$ and $\boldsymbol{\Lambda}>\boldsymbol{0}$, i.e., $\boldsymbol{W}$ is a $d\times d$ positive definite random matrix with density function

$$\frac{1}{2^{ad/2}\,|\boldsymbol{\Lambda}|^{a/2}\,\Gamma_d(a/2)}\,|\boldsymbol{W}|^{\frac{1}{2}a-\frac{1}{2}(d+1)}\,\exp(-\tfrac{1}{2}\,\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}\boldsymbol{W}), \tag{2.2}$$

$\boldsymbol{W}>\boldsymbol{0}$, where

$$\Gamma_d(a)=\pi^{d(d-1)/4}\prod_{j=1}^{d}\Gamma(a-\tfrac{1}{2}(j-1)),$$

$\mathrm{Re}(a)>(d-1)/2$, is the multivariate gamma function (Muirhead [26], p. 62).

We will need some well-known properties of the Wishart distribution, all of which follow from results given by Anderson [2, pp. 142–143, 262], Eaton [10, pp. 310–312], or Muirhead [26, pp. 93–96, 117]. For ease of exposition, we collect together these properties. In stating these result, we partition $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ into p and q rows and columns, i.e.,

$$\boldsymbol{W}=\begin{pmatrix}\boldsymbol{W}_{11} & \boldsymbol{W}_{12}\\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22}\end{pmatrix},\qquad \boldsymbol{\Lambda}=\begin{pmatrix}\boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12}\\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22}\end{pmatrix},$$

where $\boldsymbol{W}_{11}$ and $\boldsymbol{\Lambda}_{11}$ are $p\times p$, $\boldsymbol{W}_{12}=\boldsymbol{W}_{21}'$ and $\boldsymbol{\Lambda}_{12}=\boldsymbol{\Lambda}_{21}'$ are $p\times q$, and $\boldsymbol{W}_{22}$ and $\boldsymbol{\Lambda}_{22}$ are $q\times q$. We set $\boldsymbol{W}_{11\cdot 2}=\boldsymbol{W}_{11}-\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{21}$ and define $\boldsymbol{\Lambda}_{11\cdot 2}$ similarly.

Proposition 2.2. *Suppose that* $\boldsymbol{W} \sim \mathrm{W}_d(a, \boldsymbol{\Lambda})$. *Then,*

(i) $\boldsymbol{W}_{11\cdot 2}$ *and* $\{\boldsymbol{W}_{12}, \boldsymbol{W}_{22}\}$ *are mutually independent, and* $\boldsymbol{W}_{11\cdot 2} \sim \mathrm{W}_p(a-q, \boldsymbol{\Lambda}_{11\cdot 2})$.

(ii) $\boldsymbol{W}_{12}|\boldsymbol{W}_{22} \sim \mathrm{N}(\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{W}_{22}, \boldsymbol{\Lambda}_{11\cdot 2} \otimes \boldsymbol{W}_{22})$.

(iii) If $\boldsymbol{\Lambda}_{12} = \boldsymbol{0}$ *then* $\boldsymbol{W}_{11\cdot 2}$, $\boldsymbol{W}_{22}$, *and* $\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{21}$ *are mutually independent. Moreover,* $\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{21} \sim \mathrm{W}_p(q, \boldsymbol{\Lambda}_{11})$.

(iv) For $k \le d$, *if* $\boldsymbol{M}$ *is* $k \times d$ *of rank* k *then* $(\boldsymbol{M}\boldsymbol{W}^{-1}\boldsymbol{M}')^{-1} \sim \mathrm{W}_k(a-d+k, (\boldsymbol{M}\boldsymbol{\Lambda}^{-1}\boldsymbol{M}')^{-1})$. *In particular, if a* $d \times 1$ *random vector* $\boldsymbol{Y}$ *is independent of* $\boldsymbol{W}$ *and satisfies* $P(\boldsymbol{Y} = \boldsymbol{0}) = 0$ *then* $\boldsymbol{Y}$ *is independent of* $\boldsymbol{Y}'\boldsymbol{\Lambda}^{-1}\boldsymbol{Y}/\boldsymbol{Y}'\boldsymbol{W}^{-1}\boldsymbol{Y} \sim \chi^2_{a-d+1}$.

Lemma 2.3. *Suppose that* $\boldsymbol{B} \sim \mathrm{W}_q(n-1, \boldsymbol{I}_q)$ *and* $t \in \mathbb{C}$, *where* $\mathrm{Re}(t) \ge 0$. *Then*

$$E|\boldsymbol{I}_q + t\boldsymbol{B}^{-1}|^{-1/2} = E\exp(-\tfrac{1}{2}tQ_1^{-1}Q_2), \tag{2.3}$$

where $Q_1 \sim \chi^2_{n-q}$, $Q_2 \sim \chi^2_q$, *and* Q_1 *and* Q_2 *are mutually independent. In addition, if* $\boldsymbol{C}$ *is a* $q \times q$ *positive semidefinite random matrix that is independent of* $\boldsymbol{B}$ *then, for* $t \in \mathbb{R}$,

$$E|\boldsymbol{I}_q - 2it\boldsymbol{C}\boldsymbol{B}^{-1}|^{-1/2} = E\exp(itQ_1^{-1}\boldsymbol{V}'\boldsymbol{C}\boldsymbol{V}), \tag{2.4}$$

where $\boldsymbol{V} \sim \mathrm{N}_q(\boldsymbol{0}, \boldsymbol{I}_q)$, Q_1, *and* $\boldsymbol{C}$ *are mutually independent.*

Proof. Let $\boldsymbol{V} \sim \mathrm{N}_q(\boldsymbol{0}, \boldsymbol{I}_q)$, so that $\boldsymbol{V}\boldsymbol{V}' \sim \mathrm{W}_q(1, \boldsymbol{I}_q)$, and let $\boldsymbol{V}$ be independent of $\boldsymbol{B}$. By the formula for the moment-generating function of a Wishart matrix, $E|\boldsymbol{I}_q + t\boldsymbol{B}^{-1}|^{-1/2} = E\exp(-\frac{1}{2}t\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V})$. By Proposition 2.2(iv), $\boldsymbol{V}'\boldsymbol{V}/\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V} \sim \chi^2_{n-q}$; also, $\boldsymbol{V}'\boldsymbol{V}/\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V}$ is independent of $\boldsymbol{V}$, so we may write $\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V}$ in the form

$$\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V} = (\boldsymbol{V}'\boldsymbol{V}/\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V})^{-1}\boldsymbol{V}'\boldsymbol{V} \overset{\mathcal{L}}{=} Q_1^{-1}Q_2,$$

where $Q_1 = \boldsymbol{V}'\boldsymbol{V}/\boldsymbol{V}'\boldsymbol{B}^{-1}\boldsymbol{V}$, $Q_2 = \boldsymbol{V}'\boldsymbol{V} \sim \chi^2_q$, and Q_1 and Q_2 are independent. This establishes (2.3).

The proof of (2.4) is similar. Note that

$$\begin{aligned} E|\boldsymbol{I}_q - 2it\boldsymbol{C}\boldsymbol{B}^{-1}|^{-1/2} &= E\exp(it\boldsymbol{V}'\boldsymbol{C}^{1/2}\boldsymbol{B}^{-1}\boldsymbol{C}^{1/2}\boldsymbol{V}) \\ &= E\exp(it(\boldsymbol{C}^{1/2}\boldsymbol{V})'\boldsymbol{B}^{-1}(\boldsymbol{C}^{1/2}\boldsymbol{V})). \end{aligned} \tag{2.5}$$

By Proposition 2.2(iv),

$$Q_1 = \frac{(\boldsymbol{C}^{1/2}\boldsymbol{V})'(\boldsymbol{C}^{1/2}\boldsymbol{V})}{(\boldsymbol{C}^{1/2}\boldsymbol{V})'\boldsymbol{B}^{-1}(\boldsymbol{C}^{1/2}\boldsymbol{V})} = \frac{\boldsymbol{V}'\boldsymbol{C}\boldsymbol{V}}{\boldsymbol{V}'\boldsymbol{C}^{1/2}\boldsymbol{B}^{-1}\boldsymbol{C}^{1/2}\boldsymbol{V}} \sim \chi^2_{n-q}$$

and Q_1 is independent of $\boldsymbol{V}$ and $\boldsymbol{C}$. Therefore

$$\boldsymbol{V}'\boldsymbol{C}^{1/2}\boldsymbol{B}^{-1}\boldsymbol{C}^{1/2}\boldsymbol{V} = \Big(\frac{\boldsymbol{V}'\boldsymbol{C}\boldsymbol{V}}{\boldsymbol{V}'\boldsymbol{C}^{1/2}\boldsymbol{B}^{-1}\boldsymbol{C}^{1/2}\boldsymbol{V}}\Big)^{-1}\boldsymbol{V}'\boldsymbol{C}\boldsymbol{V} \overset{\mathcal{L}}{=} \boldsymbol{V}'\boldsymbol{C}\boldsymbol{V}/Q_1^{-1},$$

and in conjunction with (2.5), we now have (2.4). □

3 The distribution of $\widehat{\boldsymbol{\mu}}$

We partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in conformity with (1.1), writing $\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}$ where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are of dimensions p and q, respectively, and $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$, and $\boldsymbol{\Sigma}_{22}$ are of order $p \times p$, $p \times q$, and $q \times q$, respectively. We assume throughout that $n > q + 2$ to ensure that all means and variances are finite and that all integrals encountered later are absolutely convergent. We will use the notation $\tau = n/N$ for the proportion of data which are complete; and we denote $1 - \tau$ by $\bar{\tau}$, so that $\bar{\tau} = (N - n)/N$ is the proportion of incomplete observations.

Define sample means

$$\begin{aligned} \bar{\boldsymbol{X}} &= \tfrac{1}{n}\sum_{j=1}^{n} \boldsymbol{X}_j, & \bar{\boldsymbol{Y}}_1 &= \tfrac{1}{n}\sum_{j=1}^{n} \boldsymbol{Y}_j, \\ \bar{\boldsymbol{Y}}_2 &= \tfrac{1}{N-n}\sum_{j=n+1}^{N} \boldsymbol{Y}_j, & \bar{\boldsymbol{Y}} &= \tfrac{1}{N}\sum_{j=1}^{N} \boldsymbol{Y}_j, \end{aligned} \tag{3.1}$$

and the corresponding matrices of sums of squares and products by

$$\begin{aligned} \boldsymbol{A}_{11} &= \sum_{j=1}^{n} (\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})', & \boldsymbol{A}_{12} = \boldsymbol{A}_{21}' &= \sum_{j=1}^{n} (\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_1)', \\ \boldsymbol{A}_{22,n} &= \sum_{j=1}^{n} (\boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_1)(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_1)', & \boldsymbol{A}_{22,N} &= \sum_{j=1}^{N} (\boldsymbol{Y}_j - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}})'. \end{aligned} \tag{3.2}$$

By Anderson [1] (cf. Morrison [24], Anderson and Olkin [3], Jinadasa and Tracy [16]), the maximum likelihood estimator of $\boldsymbol{\mu}$ is $\widehat{\boldsymbol{\mu}} = \begin{pmatrix}\widehat{\boldsymbol{\mu}}_1\\ \widehat{\boldsymbol{\mu}}_2\end{pmatrix}$, where

$$\widehat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{X}} - \bar{\tau}\boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}(\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2), \qquad \widehat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{Y}}. \tag{3.3}$$

The estimator $\widehat{\boldsymbol{\mu}}_1$ is sometimes called the *regression estimator* of $\boldsymbol{\mu}_1$ (Little [21], p. 594); this terminology stems from a well-known procedure in sampling theory in which additional observations on a subset of variables are used to improve estimation of a parameter.

Introduce the matrix

$$\begin{aligned} \boldsymbol{\Omega} &= \begin{pmatrix} \frac{1}{n}\big(\boldsymbol{\Sigma}_{11} - \bar{\tau}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big) & \frac{1}{N}\boldsymbol{\Sigma}_{12} \\ \frac{1}{N}\boldsymbol{\Sigma}_{21} & \frac{1}{N}\boldsymbol{\Sigma}_{22} \end{pmatrix} \\ &= \frac{1}{N}\boldsymbol{\Sigma} + \frac{\bar{\tau}}{n}\begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \end{aligned} \tag{3.4}$$

where we have applied the standard notation $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Here and throughout the paper, we use the notation "$R_1 \overset{\mathcal{L}}{=} R_2$" whenever two random entities R_1 and R_2 have the same probability distribution. If R_1 is a statistic that depends on a sample size N, then we use the notation "$R_1 \overset{\mathcal{L}}{\to} R_2$ as $N \to \infty$" to denote that R_1 converges in distribution to R_2 as $N \to \infty$. If R_1 and R_2 are real-valued random variables, then we write "$R_1 \overset{\mathcal{L}}{\geq} R_2$" or "$R_2 \overset{\mathcal{L}}{\leq} R_1$" if $P(R_1 \geq t) \geq P(R_2 \geq t)$ for all $t \in \mathbb{R}$.

Theorem 3.1. *The maximum likelihood estimator $\widehat{\boldsymbol{\mu}}$ satisfies the stochastic representation*

$$\widehat{\boldsymbol{\mu}} \overset{\mathcal{L}}{=} \boldsymbol{\mu} + \boldsymbol{V}_1 + \Big(\frac{\bar{\tau}Q_2}{nQ_1}\Big)^{1/2} \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2 \\ \boldsymbol{0} \end{pmatrix}, \tag{3.5}$$

where $\boldsymbol{V}_1 \sim \mathrm{N}_{p+q}(\mathbf{0}, \boldsymbol{\Omega})$, $\boldsymbol{V}_2 \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$, $Q_1 \sim \chi^2_{n-q}$, $Q_2 \sim \chi^2_q$, *and* $\boldsymbol{V}_1$, Q_1, Q_2, *and* $\boldsymbol{V}_2$ *are mutually independent. Further,* $\widehat{\boldsymbol{\mu}}_1$ *and* $\widehat{\boldsymbol{\mu}}_2$ *are mutually independent if and only if* $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

The representation (3.5) will be seen later to provide fundamental insight into the probabilistic behavior of $\widehat{\boldsymbol{\mu}}$ and inference about $\boldsymbol{\mu}$.

We remark that the appearance of stochastic representations in the context of monotone samples is not new; in testing that a monotone sample from a normal population is missing completely at random, Little [22] proposed a test statistic and derived a stochastic representation for its null distribution.

The asymptotic distribution of $\widehat{\boldsymbol{\mu}}$ for large values of N or n can also be deduced from (3.5). For example, if n is fixed and $N \to \infty$ then $\boldsymbol{\Omega} \to n^{-1} \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, a singular matrix, hence $\sqrt{n}(\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) \xrightarrow{\mathcal{L}} \boldsymbol{\Sigma}_{11\cdot 2}^{1/2}(\widetilde{\boldsymbol{V}}_{11} + \sqrt{Q_2/Q_1}\boldsymbol{V}_2)$ where $\widetilde{\boldsymbol{V}}_{11}$ and $\boldsymbol{V}_2$ are i.i.d. $\mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$. As $n, N \to \infty$ with $n/N \to \delta$, $0 < \delta \le 1$, note that $Q_2/Q_1 \sim \chi^2_q/\chi^2_{n-q} \to 0$, almost surely, and, from (3.4), $\mathrm{Cov}(\sqrt{N}\boldsymbol{V}_1) = N\boldsymbol{\Omega} \to \boldsymbol{\Sigma} + (\delta^{-1} - 1) \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$; hence we obtain immediately the following result.

Corollary 3.2. *Suppose* $n, N \to \infty$ *with* $n/N \to \delta$, $0 < \delta \le 1$. *Then*

$$\sqrt{N}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{\mathcal{L}} \mathrm{N}_{p+q}\left(\mathbf{0}, \boldsymbol{\Sigma} + (\delta^{-1} - 1) \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right).$$

We shall also derive from (3.5) some properties of the moments of $\widehat{\boldsymbol{\mu}}$. The following results generalize from the case in which $p = 1$ results of Morrison [24]. Throughout, we use the notation $(a)_j = a(a+1)\cdots(a+j-1)$, where $j = 0, 1, 2, \ldots$, for the shifted factorial.

Corollary 3.3. *(i) All odd central moments of* $\widehat{\mu}_{1r}$*, the* r*th component of* $\widehat{\boldsymbol{\mu}}_1$*, are zero. In particular,* $\widehat{\boldsymbol{\mu}}$ *is unbiased.*

(ii) For $n > q + 2$*, the covariance matrix of* $\widehat{\boldsymbol{\mu}}$ *is*

$$\mathrm{Cov}(\widehat{\boldsymbol{\mu}}) = \frac{1}{N}\boldsymbol{\Sigma} + \frac{(n-2)\bar{\tau}}{n(n-q-2)} \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{3.6}$$

(iii) Denote by ω_{ij} *and* $\sigma_{ij\cdot p+1,\ldots,p+q}$ *the* (i,j)*th entries of* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Sigma}_{11\cdot 2}$*, respectively. Then the even central moments of* $\widehat{\mu}_{1r}$ *are*

$$E(\widehat{\mu}_{1r} - \mu_{1r})^{2k} = \frac{(2k)!}{k!} \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^j (\frac{1}{2}q)_j}{(-\frac{1}{2}(n-q)+1)_j} \left(\frac{\bar{\tau}}{n}\right)^{2j} (\sigma_{rr\cdot p+1,\ldots,p+q})^j \omega_{rr}^{k-j} \tag{3.7}$$

for $k < (n-q)/2$*. If* $k \ge (n-q)/2$ *then* $E(\widehat{\mu}_{1r} - \mu_{1r})^{2k}$ *does not exist.*

We note that the unbiasedness, and the odd central moments, of $\widehat{\boldsymbol{\mu}}$ can be derived from (3.3) and the sampling distributions of the means and covariance matrices there. Also, (3.6) was obtained earlier by Kanda and Fujikoshi [17]).

Proof of Theorem 3.1. We shall establish this result through an analysis of ϕ, the characteristic function of $\widehat{\boldsymbol{\mu}}$, simplifying expressions for ϕ until we recognize that we have obtained the characteristic function of the right-hand side of (3.5).

Since $\tau = 1 - \bar{\tau} = n/N$ then, by (3.1),

$$\widehat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{X}} - \bar{\tau} \boldsymbol{A}_{12} \boldsymbol{A}_{22,n}^{-1} (\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2), \qquad \widehat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{Y}} = \tau \bar{\boldsymbol{Y}}_1 + \bar{\tau} \bar{\boldsymbol{Y}}_2.$$

For $\boldsymbol{t} = \begin{pmatrix} \boldsymbol{t}_1 \\ \boldsymbol{t}_2 \end{pmatrix} \in \mathbb{R}^{p+q}$, the joint characteristic function of $\widehat{\boldsymbol{\mu}} = \begin{pmatrix} \widehat{\boldsymbol{\mu}}_1 \\ \widehat{\boldsymbol{\mu}}_2 \end{pmatrix}$ is

$$\begin{aligned} \phi(\boldsymbol{t}) &= E e^{i(\boldsymbol{t}_1' \widehat{\boldsymbol{\mu}}_1 + \boldsymbol{t}_2' \widehat{\boldsymbol{\mu}}_2)} \\ &= E \exp \Big[i \big(\boldsymbol{t}_1' \bar{\boldsymbol{X}} - \bar{\tau} \boldsymbol{t}_1' \boldsymbol{A}_{12} \boldsymbol{A}_{22,n}^{-1} (\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2) + \tau \boldsymbol{t}_2' \bar{\boldsymbol{Y}}_1 + \bar{\tau} \boldsymbol{t}_2' \bar{\boldsymbol{Y}}_2 \big) \Big] \\ &= E \, \exp \Big[i \big(\boldsymbol{t}_1' \bar{\boldsymbol{X}} + (\tau \boldsymbol{t}_2 - \bar{\tau} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{A}_{21} \boldsymbol{t}_1)' \bar{\boldsymbol{Y}}_1 + \bar{\tau} (\boldsymbol{t}_2 + \boldsymbol{A}_{22,n}^{-1} \boldsymbol{A}_{21} \boldsymbol{t}_1)' \bar{\boldsymbol{Y}}_2 \big) \Big]. \end{aligned}$$

Observe that

$$\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix} \equiv \sum_{j=1}^{n} \begin{pmatrix} \boldsymbol{X}_j - \bar{\boldsymbol{X}} \\ \boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_j - \bar{\boldsymbol{X}} \\ \boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_1 \end{pmatrix}' \sim \mathrm{W}_{p+q}(n-1, \boldsymbol{\Sigma}),$$

and that $\begin{pmatrix} \bar{\boldsymbol{X}} \\ \bar{\boldsymbol{Y}}_1 \end{pmatrix}$, $\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}$, and $\bar{\boldsymbol{Y}}_2$ are mutually independent; therefore

$$\begin{aligned} \phi(\boldsymbol{t}) = {} & E_{\{\boldsymbol{A}_{12}, \boldsymbol{A}_{22,n}\}} E_{\{\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}_1\}} \exp \Big[i \big(\boldsymbol{t}_1' \bar{\boldsymbol{X}} + (\tau \boldsymbol{t}_2 - \bar{\tau} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{A}_{21} \boldsymbol{t}_1)' \bar{\boldsymbol{Y}}_1 \big) \Big] \\ & \times E_{\bar{\boldsymbol{Y}}_2} \exp \Big[i \bar{\tau} (\boldsymbol{t}_2 + \boldsymbol{A}_{22,n}^{-1} \boldsymbol{A}_{21} \boldsymbol{t}_1)' \bar{\boldsymbol{Y}}_2 \Big]. \end{aligned}$$

Since $\begin{pmatrix} \bar{\boldsymbol{X}} \\ \bar{\boldsymbol{Y}}_1 \end{pmatrix} \sim \mathrm{N}_{p+q}(\boldsymbol{\mu}, n^{-1} \boldsymbol{\Sigma})$ and $\bar{\boldsymbol{Y}}_2 \sim \mathrm{N}_q(\boldsymbol{\mu}_2, (N-n)^{-1} \boldsymbol{\Sigma}_{22})$ then, on applying the usual formula for the characteristic function of the multivariate normal distribution and simplifying the algebraic expressions, we obtain

$$\begin{aligned} \phi(\boldsymbol{t}) = {} & \exp \Big(i \boldsymbol{t}_1' \boldsymbol{\mu}_1 + i \boldsymbol{t}_2' \boldsymbol{\mu}_2 - \frac{1}{2N} \boldsymbol{t}_2' \boldsymbol{\Sigma}_{22} \boldsymbol{t}_2 \Big) \exp \Big(- \frac{1}{2n} \boldsymbol{t}_1' \boldsymbol{\Sigma}_{11} \boldsymbol{t}_1 - \frac{1}{n} \tau \boldsymbol{t}_2' \boldsymbol{\Sigma}_{21} \boldsymbol{t}_1 \Big) \\ & \times E_{\{\boldsymbol{A}_{12}, \boldsymbol{A}_{22,n}\}} \exp \Big[- \frac{1}{2n} \bar{\tau} \boldsymbol{t}_1' \boldsymbol{A}_{12} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{\Sigma}_{22} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{A}_{21} \boldsymbol{t}_1 \\ & \qquad\qquad + \frac{1}{n} \bar{\tau} \boldsymbol{t}_1' \boldsymbol{A}_{12} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{t}_1 \Big]. \end{aligned} \tag{3.8}$$

By Proposition 2.2(i), (ii), $\boldsymbol{A}_{12} | \boldsymbol{A}_{22,n} \sim \mathrm{N}(\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{A}_{22,n}, \boldsymbol{\Sigma}_{11 \cdot 2} \otimes \boldsymbol{A}_{22,n})$ and $\boldsymbol{A}_{22,n} \sim \mathrm{W}_q(n-1, \boldsymbol{\Sigma}_{22})$. Making the transformation from $\boldsymbol{A}_{12}$ to $\boldsymbol{B}_{12} = \boldsymbol{A}_{12} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{A}_{22,n}$, we have $\boldsymbol{B}_{12} | \boldsymbol{A}_{22,n} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{11 \cdot 2} \otimes \boldsymbol{A}_{22,n})$. After a lengthy, but straightforward, calculation, we find that the expectation in (3.8) equals

$$\exp \Big[\frac{1}{2n} \bar{\tau} \boldsymbol{t}_1' \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{t}_1 \Big] \, E_{\boldsymbol{A}_{22,n}} \, E_{\{\boldsymbol{B}_{12} | \boldsymbol{A}_{22,n}\}} \exp \Big[- \frac{1}{2n} \bar{\tau} \boldsymbol{t}_1' \boldsymbol{B}_{12} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{\Sigma}_{22} \boldsymbol{A}_{22,n}^{-1} \boldsymbol{B}_{12}' \boldsymbol{t}_1 \Big]. \tag{3.9}$$

Applying (2.1) with $\boldsymbol{C} = \boldsymbol{\Sigma}_{11 \cdot 2}$, $\boldsymbol{D} = \boldsymbol{A}_{22,n}$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{22}$, and $\boldsymbol{u} = (\bar{\tau}/2n)^{1/2} \boldsymbol{t}_1$, the inner expectation in (3.9) is seen to equal $\big| \boldsymbol{I}_q + n^{-1} \bar{\tau} (\boldsymbol{t}_1' \boldsymbol{\Sigma}_{11 \cdot 2} \boldsymbol{t}_1) \boldsymbol{\Sigma}_{22} \boldsymbol{A}_{22,n}^{-1} \big|^{-1/2}$; inserting this result in (3.9), substituting the outcome in (3.8), and simplifying the resulting expression, we obtain

$$\phi(\boldsymbol{t}) = \exp(i \boldsymbol{t}' \boldsymbol{\mu} - \tfrac{1}{2} \boldsymbol{t}' \boldsymbol{\Omega} \boldsymbol{t}) \, E_{\boldsymbol{A}_{22,n}} \big| \boldsymbol{I}_q + n^{-1} \bar{\tau} (\boldsymbol{t}_1' \boldsymbol{\Sigma}_{11 \cdot 2} \boldsymbol{t}_1) \boldsymbol{\Sigma}_{22} \boldsymbol{A}_{22,n}^{-1} \big|^{-1/2}, \tag{3.10}$$

where $\boldsymbol{\Omega}$ is defined in (3.4). Since $\boldsymbol{A}_{22,n} \sim \mathrm{W}_q(n-1, \boldsymbol{\Sigma}_{22})$ then $\boldsymbol{B}_{22} := \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{A}_{22,n} \boldsymbol{\Sigma}_{22}^{-1/2} \sim \mathrm{W}_q(n-1, \boldsymbol{I}_q)$; therefore, by (2.3) of Lemma 2.3, the expectation in (3.10) equals

$$E_{\boldsymbol{B}_{22}} |\boldsymbol{I}_q + n^{-1}\bar{\tau}(\boldsymbol{t}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{t}_1) \boldsymbol{B}_{22}^{-1}|^{-1/2} = E \exp\big(-\bar{\tau} Q_1^{-1} Q_2 \, \boldsymbol{t}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{t}_1 / 2n\big), \tag{3.11}$$

where $Q_1 \sim \chi^2_{n-q}$ and $Q_2 \sim \chi^2_q$ are mutually independent. Substituting (3.11) in (3.10), we have

$$\begin{aligned} \phi(\boldsymbol{t}) &= \exp(i\boldsymbol{t}'\boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{t}'\boldsymbol{\Omega}\boldsymbol{t}) \, E \exp\big(-\bar{\tau} Q_1^{-1} Q_2 \, \boldsymbol{t}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{t}_1/2n\big) \\ &= E \exp\big(i\boldsymbol{t}'(\boldsymbol{\mu} + \boldsymbol{V}_1)\big) \, \exp\big(-\bar{\tau} Q_1^{-1} Q_2 \, \boldsymbol{t}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{t}_1/2n\big), \end{aligned} \tag{3.12}$$

where $\boldsymbol{V}_1 \sim \mathrm{N}_{p+q}(\boldsymbol{0}, \boldsymbol{\Omega})$ independently of Q_1 and Q_2. Furthermore, by writing

$$E \exp\big(-\bar{\tau} Q_1^{-1} Q_2 \, \boldsymbol{t}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{t}_1/2n\big) = E \exp\big(i(\bar{\tau} Q_1^{-1} Q_2/n)^{1/2} \; \boldsymbol{t}_1' \boldsymbol{\Sigma}_{11\cdot 2}^{1/2} \boldsymbol{V}_2\big),$$

where $\boldsymbol{V}_2 \sim \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{I}_p)$ independently of $\boldsymbol{V}_1$, Q_1, and Q_2, and substituting this latter result in (3.12), we obtain (3.5).

Finally, note that by (3.5), $\widehat{\boldsymbol{\mu}}_1$ and $\widehat{\boldsymbol{\mu}}_2$ are independent if and only if $\boldsymbol{V}_{11}$ and $\boldsymbol{V}_{12}$ are independent, equivalently, $\boldsymbol{\Omega}$ is block-diagonal. However, by (3.4), $\boldsymbol{\Omega}$ is block-diagonal if and only if $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$. □

Remark 3.4. As an application of Theorem 3.1, we consider the problem of deriving a $100(1-\alpha)\%$ confidence interval for a linear combination $\boldsymbol{\nu}'\boldsymbol{\mu}$, where $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$ is specified. Writing $\boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix}$ where $\boldsymbol{\nu}_1 \in \mathbb{R}^p$ and $\boldsymbol{\nu}_2 \in \mathbb{R}^q$, it follows from (3.5) that

$$\boldsymbol{\nu}'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \stackrel{\mathcal{L}}{=} \boldsymbol{\nu}'\boldsymbol{V}_1 + (\bar{\tau} Q_2/nQ_1)^{1/2} \boldsymbol{\nu}_1' \boldsymbol{\Sigma}_{11\cdot 2}^{1/2} \boldsymbol{V}_2. \tag{3.13}$$

To obtain an approximate confidence interval for $\boldsymbol{\nu}'\boldsymbol{\mu}$, we approximate the distribution of $\boldsymbol{\nu}'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ by a normal distribution, $\mathrm{N}(0, \theta^2)$, where $\theta^2 = \mathrm{Var}(\boldsymbol{\nu}'\widehat{\boldsymbol{\mu}})$. By (3.6) and (3.13),

$$\theta^2 = \frac{1}{N} \boldsymbol{\nu}'\boldsymbol{\Sigma}\boldsymbol{\nu} + \frac{(n-2)\bar{\tau}}{n(n-q-2)} \boldsymbol{\nu}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{\nu}_1.$$

Using the approximation $\boldsymbol{\nu}'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \approx \mathrm{N}(0, \theta^2)$, we obtain an approximate $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\nu}'\boldsymbol{\mu}$ as $\boldsymbol{\nu}'\widehat{\boldsymbol{\mu}} \mp z_{\alpha/2}\widehat{\theta}$, where $z_{\alpha/2}$ is the usual percentage point of the standard normal distribution, and

$$\widehat{\theta} = \Big(\frac{1}{N} \boldsymbol{\nu}'\widehat{\boldsymbol{\Sigma}}\boldsymbol{\nu} + \frac{(n-2)\bar{\tau}}{n(n-q-2)} \boldsymbol{\nu}_1' \widehat{\boldsymbol{\Sigma}}_{11\cdot 2} \boldsymbol{\nu}_1\Big)^{1/2},$$

where the estimators $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_{11\cdot 2}$ are defined in (4.1). To obtain a rigorous bound on the error in the above normal approximation, we apply the arguments in Section 6 and deduce that if f_1 and f_2 are the density functions of $\boldsymbol{\nu}'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ and $\mathrm{N}(0, \theta^2)$, respectively, then there exists a universal constant $C > 0$ such that $\sup_{t \in \mathbb{R}} |f_1(t) - f_2(t)| \le C \, \boldsymbol{\nu}_1' \boldsymbol{\Sigma}_{11\cdot 2} \boldsymbol{\nu}_1$.

Proof of Corollary 3.3. In the case of (i), denote by μ_{1r}, V_{1r}, and $(\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2)_r$ the rth components of $\boldsymbol{\mu}_1$, $\boldsymbol{V}_1$, and $\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2$, respectively. By (3.5),

$$\widehat{\mu}_{1r} - \mu_{1r} \stackrel{\mathcal{L}}{=} V_{1r} + (\bar{\tau} Q_2/nQ_1)^{1/2} (\boldsymbol{\Sigma}_{11\cdot 2}^{1/2} \boldsymbol{V}_2)_r. \tag{3.14}$$

Since the distributions of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are symmetric about $\mathbf{0}$ then $\widehat{\mu}_{1r} - \mu_{1r} \stackrel{\mathcal{L}}{=} -(\widehat{\mu}_{1r} - \mu_{1r})$, so it follows that all odd moments of $\widehat{\mu}_{1r} - \mu_{1r}$ are equal to zero. In particular, $E(\widehat{\mu}_{1r}) = \mu_{1r}$ for all r; therefore $\widehat{\boldsymbol{\mu}}$ is unbiased.

The proof of (ii) follows directly from 3.5).

To establish (iii), we apply the binomial theorem to (3.14). Noting that the odd moments of V_{1r} and $(\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2)_r$ are zero, we obtain

$$E(\widehat{\mu}_{1r} - \mu_{1r})^{2k} = E\sum_{j=0}^{k} \binom{2k}{2j} (\bar{\tau}/n)^j V_{1r}^{2(k-j)} Q_2^j Q_1^{-j} \big((\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2)_r\big)^{2j}.$$

Since $V_{1r} \sim \mathrm{N}(0, \omega_{rr})$ and $(\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2)_r \sim \mathrm{N}(0, (\boldsymbol{\Sigma}_{11\cdot 2})_{rr})$ then

$$EV_{1r}^{2(k-j)} = \frac{(2k-2j)!}{(k-j)!\,2^{k-j}}\omega_{rr}^{k-j}$$

and

$$E\big((\boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{V}_2)_r\big)^{2j} = \frac{(2j)!}{j!\,2^j}\big((\boldsymbol{\Sigma}_{11\cdot 2})_{rr}\big)^j.$$

By standard calculations, $E(Q_2^j) = 2^j(\frac{1}{2}q)_j$ and

$$E(Q_1^{-j}) = \begin{cases} (-1)^j/2^j(-\frac{1}{2}(n-q)+1)_j, & \text{if } j < (n-q)/2, \\ \infty, & \text{if } j \geq (n-q)/2. \end{cases}$$

Combining these results and simplifying the resulting sum, we obtain (3.7).

Finally, $E(\widehat{\mu}_{1r} - \mu_{1r})^{2k}$ diverges for $k \geq (n-q)/2$ because $E(Q_1^{-(n-q)/2})$ diverges. □

4 Some properties of $\widehat{\boldsymbol{\Sigma}}$

In this section, we list some properties of $\widehat{\boldsymbol{\Sigma}}$ that are needed in Section 5. The proofs of these properties all are provided in the companion paper [9].

By Anderson [1] or Anderson and Olkin [3] (cf. Morrison [24], Giguère and Styan [14]), the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11} & \widehat{\boldsymbol{\Sigma}}_{12} \\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22} \end{pmatrix}$ where, in the notation of (3.2),

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_{11} &= \frac{1}{n}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{21}) + \frac{1}{N}\boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{22,N}\boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{21}, \\ \widehat{\boldsymbol{\Sigma}}_{12} &= \widehat{\boldsymbol{\Sigma}}_{21}' = \frac{1}{N}\boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{22,N}, \\ \widehat{\boldsymbol{\Sigma}}_{22} &= \frac{1}{N}\boldsymbol{A}_{22,N}. \end{aligned} \tag{4.1}$$

Proposition 4.1. *Define* $\boldsymbol{A}_{11\cdot 2,n} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{21}$, $\boldsymbol{B}_1 = \sum_{j=n+1}^{N}(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_2)(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}}_2)'$, $\boldsymbol{B}_2 = n\bar{\tau}(\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2)(\bar{\boldsymbol{Y}}_1 - \bar{\boldsymbol{Y}}_2)'$, *and* $\boldsymbol{B} = \boldsymbol{B}_1 + \boldsymbol{B}_2$. *Then*

$$n\widehat{\boldsymbol{\Sigma}} \stackrel{\mathcal{L}}{=} \tau\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix} + \bar{\tau}\begin{pmatrix} \boldsymbol{A}_{11\cdot 2,n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \tau\begin{pmatrix} \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_q \end{pmatrix}\begin{pmatrix} \boldsymbol{B} & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{B} \end{pmatrix}\begin{pmatrix} \boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{21} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_q \end{pmatrix}, \tag{4.2}$$

where $\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix} \sim \mathrm{W}_{p+q}(n-1, \boldsymbol{\Sigma})$ *and* $\boldsymbol{B} \sim \mathrm{W}_q(N-n, \boldsymbol{\Sigma}_{22})$ *are mutually independent. Moreover,* $N\widehat{\boldsymbol{\Sigma}}_{22} \sim \mathrm{W}_q(N-1, \boldsymbol{\Sigma}_{22})$.

We will also need some results on the matrix F-distribution. A $q \times q$ random matrix $\boldsymbol{F} \geq \mathbf{0}$ is said to have a *matrix* F-*distribution* with degrees of freedom (a, b), denoted $\boldsymbol{F} \sim \mathrm{F}_{a,b}^{(q)}$, if $\boldsymbol{F} \stackrel{\mathcal{L}}{=} \boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2}$, where $\boldsymbol{A} \sim \mathrm{W}_q(a, \boldsymbol{\Sigma}_{22})$ and $\boldsymbol{B} \sim \mathrm{W}_q(b, \boldsymbol{\Sigma}_{22})$ are mutually independent. Necessarily, we require $b > q - 1$ to ensure that $\boldsymbol{B}$ is nonsingular, almost surely. If both $a, b > q - 1$ then $\boldsymbol{F}$ is nonsingular, almost surely, and its density function is

$$\frac{\Gamma_q\big((a+b)/2\big)}{\Gamma_q(a/2)\Gamma_q(b/2)} |\boldsymbol{F}|^{\frac{1}{2}a - \frac{1}{2}(q+1)/2} |\boldsymbol{I}_q + \boldsymbol{F}|^{-(a+b)/2},$$

$\boldsymbol{F} > \mathbf{0}$. It is well-known (Muirhead [26], pp. 312–313) that if $\boldsymbol{A} \sim \mathrm{W}_q(a, \boldsymbol{\Sigma}_{22})$ and $\boldsymbol{B} \sim \mathrm{W}_q(b, \boldsymbol{\Sigma}_{22})$ are independent with $a, b > q - 1$ then both $\boldsymbol{A}^{1/2}\boldsymbol{B}^{-1}\boldsymbol{A}^{1/2}$ and $\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2}$ have the $\mathrm{F}_{a,b}^{(q)}$ distribution; also, if $\boldsymbol{F} \sim \mathrm{F}_{a,b}^{(q)}$ then $\boldsymbol{F}^{-1} \sim \mathrm{F}_{b,a}^{(q)}$. Note that for $q = 1$, the notation $\mathrm{F}_{a,b}^{(q)}$ is a slight departure from the notation for the classical F-distribution, for $\mathrm{F}_{a,b}^{(1)} \equiv \chi_a^2/\chi_b^2$.

Proposition 4.2. *Suppose that* $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. *Then*

(i) $\boldsymbol{A}_{22,n}$, $\boldsymbol{A}_{11\cdot 2,n}$, $\boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\boldsymbol{A}_{21}$, $\boldsymbol{B}_1$, $\bar{\boldsymbol{X}}$, $\bar{\boldsymbol{Y}}_1$, *and* $\bar{\boldsymbol{Y}}_2$ *are mutually independent. Also,* $\boldsymbol{B}_2$ *and* $\bar{\boldsymbol{Y}}$ *are independent.*

(ii) $\widehat{\boldsymbol{\Sigma}}_{11}$ *has a stochastic representation,*

$$\boldsymbol{\Sigma}_{11}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{11}\boldsymbol{\Sigma}_{11}^{-1/2} \stackrel{\mathcal{L}}{=} \frac{1}{n}\boldsymbol{W}_1 + \frac{1}{N}\boldsymbol{W}_2^{1/2}\big(\boldsymbol{I}_p + \boldsymbol{F}\big)\boldsymbol{W}_2^{1/2}, \tag{4.3}$$

where $\boldsymbol{W}_1 \sim \mathrm{W}_p(n - q - 1, \boldsymbol{I}_p)$, $\boldsymbol{W}_2 \sim \mathrm{W}_p(q, \boldsymbol{I}_p)$, $\boldsymbol{F} \sim \mathrm{F}_{N-n,n-q+p-1}^{(p)}$, *and* $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, *and* $\boldsymbol{F}$ *are mutually independent.*

Let $O(q)$ denote the group of $q \times q$ orthogonal matrices. The Haar measure on $O(q)$ is the unique probability distribution on $O(q)$ that is invariant under the two-sided action of $O(q)$. For $p \leq q$, denote by $S_{p,q}$ the *Stiefel manifold* of all $p \times q$ matrices $\boldsymbol{H}_1$ such that $\boldsymbol{H}_1\boldsymbol{H}_1' = \boldsymbol{I}_p$. It is well-known (cf. Muirhead [26], p. 67) that there exists on $S_{p,q}$ a unique probability distribution which is left-invariant under $O(p)$ and right-invariant under $O(q)$; we refer to this distribution as the *uniform distribution on* $S_{p,q}$.

Let $\boldsymbol{H} \in O(q)$ be a random matrix which is distributed according to Haar measure. Expressing $\boldsymbol{H}$ in the form $\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \end{pmatrix}$ where $\boldsymbol{H}_1 \in S_{p,q}$ then $\boldsymbol{H}_1$ is uniformly distributed on $S_{p,q}$. Conversely, given a uniformly distributed $\boldsymbol{H}_1 \in S_{p,q}$, we may complete $\boldsymbol{H}_1$ to form a random $q \times q$ orthogonal matrix $\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \end{pmatrix}$ having the Haar measure on $O(q)$.

Lemma 4.3. *Let* $p \leq q$, $\boldsymbol{F} \sim \mathrm{F}_{a,b}^{(q)}$, $\boldsymbol{H}_1$ *be uniformly distributed on* $S_{p,q}$, *and* $\boldsymbol{F}$ *and* $\boldsymbol{H}_1$ *be independent. Then* $\boldsymbol{H}_1\boldsymbol{F}\boldsymbol{H}_1' \sim \mathrm{F}_{a,b-q+p}^{(p)}$. *Furthermore,* $\boldsymbol{H}_1\boldsymbol{F}\boldsymbol{H}_1' \stackrel{\mathcal{L}}{=} \boldsymbol{F}_{11}$, *the principal* $p \times p$ *submatrix of* $\boldsymbol{F}$.

We also have a stochastic representation for $\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}$, the estimated regression matrix.

Theorem 4.4. *For arbitrary* $\boldsymbol{\Sigma}_{12}$,

$$\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1} \stackrel{\mathcal{L}}{=} \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} + \boldsymbol{\Sigma}_{11\cdot 2}^{1/2}\boldsymbol{W}^{-1/2}\boldsymbol{K}\boldsymbol{\Sigma}_{22}^{-1/2},$$

where $\boldsymbol{W}$ *and* $\boldsymbol{K}$ *are independent,* $\boldsymbol{W} \sim \mathrm{W}_p(n - q + p - 1, \boldsymbol{I}_p)$, *and* $\boldsymbol{K} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I}_p \otimes \boldsymbol{I}_q)$. *In particular,* $\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}$ *is an unbiased estimator of* $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$, *and*

$$\boldsymbol{\Sigma}_{11\cdot 2}^{-1/2}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1})\boldsymbol{\Sigma}_{22}(\widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1})'\boldsymbol{\Sigma}_{11\cdot 2}^{-1/2} \sim \mathrm{F}_{q,n-q+p-1}^{(p)}. \tag{4.4}$$

By reparametrizing the space of positive definite matrices ([10], Proposition 8.7), we may write

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} \boldsymbol{I}_p & \widehat{\boldsymbol{\Delta}}_{12} \\ \boldsymbol{0} & \boldsymbol{I}_q \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{\Delta}}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ \widehat{\boldsymbol{\Delta}}_{21} & \boldsymbol{I}_q \end{pmatrix}. \tag{4.5}$$

This defines the positive definite symmetric matrix $\widehat{\boldsymbol{\Delta}} = \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_{11} & \widehat{\boldsymbol{\Delta}}_{12} \\ \widehat{\boldsymbol{\Delta}}_{21} & \widehat{\boldsymbol{\Delta}}_{22} \end{pmatrix}$, and the set of sub-matrices $\{\widehat{\boldsymbol{\Delta}}_{11}, \widehat{\boldsymbol{\Delta}}_{12}, \widehat{\boldsymbol{\Delta}}_{22}\}$ are also called the *partial Iwasawa coordinates* of $\boldsymbol{\Sigma}$ (Fujisawa [12]). Inverting (4.5), we obtain

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}^{-1} &= \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ -\widehat{\boldsymbol{\Delta}}_{21} & \boldsymbol{I}_q \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{\Delta}}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_p & -\widehat{\boldsymbol{\Delta}}_{12} \\ \boldsymbol{0} & \boldsymbol{I}_q \end{pmatrix} \\ &= \begin{pmatrix} \widehat{\boldsymbol{\Delta}}_{11}^{-1} & -\widehat{\boldsymbol{\Delta}}_{11}^{-1}\widehat{\boldsymbol{\Delta}}_{12} \\ -\widehat{\boldsymbol{\Delta}}_{21}\widehat{\boldsymbol{\Delta}}_{11}^{-1} & \widehat{\boldsymbol{\Delta}}_{22}^{-1} + \widehat{\boldsymbol{\Delta}}_{21}\widehat{\boldsymbol{\Delta}}_{11}^{-1}\widehat{\boldsymbol{\Delta}}_{12} \end{pmatrix}. \end{aligned}$$

Therefore the correspondence between $\widehat{\boldsymbol{\Delta}}$ and $\widehat{\boldsymbol{\Sigma}}$ is one-to-one, with inverse transformation

$$\widehat{\boldsymbol{\Sigma}}_{11} = \widehat{\boldsymbol{\Delta}}_{11} + \widehat{\boldsymbol{\Delta}}_{12}\widehat{\boldsymbol{\Delta}}_{22}\widehat{\boldsymbol{\Delta}}_{21}, \quad \widehat{\boldsymbol{\Sigma}}_{12} = \widehat{\boldsymbol{\Delta}}_{12}\widehat{\boldsymbol{\Delta}}_{22}, \quad \widehat{\boldsymbol{\Sigma}}_{22} = \widehat{\boldsymbol{\Delta}}_{22},$$

where, by (4.1),

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_{11} &= \widehat{\boldsymbol{\Sigma}}_{11\cdot 2} = \frac{1}{n}\boldsymbol{A}_{11\cdot 2,n}, \\ \widehat{\boldsymbol{\Delta}}_{12} &= \widehat{\boldsymbol{\Delta}}_{21}' = \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1} = \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}, \\ \widehat{\boldsymbol{\Delta}}_{22} &= \widehat{\boldsymbol{\Sigma}}_{22} = \frac{1}{N}\boldsymbol{A}_{22,N}. \end{aligned} \tag{4.6}$$

5 Ellipsoidal confidence regions for $\boldsymbol{\mu}$

The problem of testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is a specified vector, has been studied extensively for data are of the form (1.1). Bhargava [7, 8] obtained the likelihood ratio statistic for testing H_0 against H_a and derived a stochastic representation for the corresponding null distribution; Morrison and Bhoj [25] studied the power of the likelihood ratio test; Eaton and Kariya [11] obtained invariant tests under data structures even more general than (1.1); and Krishnamoorthy and Pannala [18] provided alternatives to the likelihood ratio test.

On the other hand, confidence regions for $\boldsymbol{\mu}$ have received less attention. Krishnamoorthy and Pannala [19] noted that the likelihood ratio criterion leads to confidence regions which are non-ellipsoidal in shape and raised the problem of constructing ellipsoidal confidence regions for $\boldsymbol{\mu}$. This problem calls for a generalization of Hotelling's T^2-statistic for the case in which the data have the monotone structure (1.1). Following [19], we study the statistic

$$T^2 = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \tag{5.1}$$

where, by (3.6),

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}) = \frac{1}{N}\widehat{\boldsymbol{\Sigma}} + \frac{(n-2)\bar{\tau}}{n(n-q-2)} \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \tag{5.2}$$

is the maximum likelihood estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\mu}})$. Krishnamoorthy and Pannala [19] derived F-approximations to T^2 *via* the method of moments and used simulations to illustrate that (5.1) has good power properties in comparison with the likelihood ratio test statistic.

A more profound motivation for the T^2-statistic in (5.1) is to be found in the results of Eaton and Kariya [11], p. 657. They prove that the problem of testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against $H_a : \boldsymbol{\mu} \neq \mathbf{0}$ is invariant under a certain group of transformations and that a maximal invariant parameter is the pair $(\gamma_{11\cdot 2}, \gamma_{22})$, where

$$\gamma_{11\cdot 2} := (\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2), \quad \gamma_{22} := \boldsymbol{\mu}_2'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2.$$

Therefore, in performing inference for $\boldsymbol{\mu}$, it is natural to utilize the corresponding maximum likelihood estimator $(\widehat{\gamma}_{11\cdot 2}, \widehat{\gamma}_{22})$, where

$$\widehat{\gamma}_{11\cdot 2} := (\widehat{\boldsymbol{\mu}}_1 - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\mu}}_2)'\widehat{\boldsymbol{\Sigma}}_{11\cdot 2}^{-1}(\widehat{\boldsymbol{\mu}}_1 - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\mu}}_2), \quad \widehat{\gamma}_{22} := \widehat{\boldsymbol{\mu}}_2'\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\mu}}_2.$$

By a well-known identity ([2], p. 63, Exercise 2.54),

$$\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = (\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2) + \boldsymbol{\mu}_2'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \tag{5.3}$$

i.e., $\gamma_{11\cdot 2}+\gamma_{22} \equiv \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. Therefore $\widehat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}$ is the sum of the maximum likelihood estimators of the maximal invariant parameters. On replacing $\widehat{\boldsymbol{\Sigma}}$ by $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})$ to adjust standard errors, we see that T^2 may be viewed as a modification of the maximum likelihood estimator $\widehat{\gamma}_{11\cdot 2} + \widehat{\gamma}_{22}$.

We turn now to the distribution of (5.1). By Krishnamoorthy and Pannala [19, p. 398] (cf. Eaton and Kariya [11, pp. 656–657]), the distribution of the T^2-statistic in (5.1) does not depend on $\boldsymbol{\mu}$ or $\boldsymbol{\Sigma}$. *Therefore, in establishing results on the distribution of (5.1) we assume, without loss of generality that* $\boldsymbol{\mu} = \mathbf{0}$ *and* $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$.

Introduce the notation

$$\gamma = 1 + \frac{(n-2)N\bar{\tau}}{n(n-q-2)}; \tag{5.4}$$

then, by (5.2),

$$N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}) = \widehat{\boldsymbol{\Sigma}} + (\gamma - 1)\begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11} + (\gamma-1)\widehat{\boldsymbol{\Sigma}}_{11\cdot 2} & \widehat{\boldsymbol{\Sigma}}_{12} \\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22} \end{pmatrix}.$$

Applying a well-known formula (see [2, p. 638]) for inverting a partitioned matrix, we obtain

$$N^{-1}\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1} = \gamma^{-1}\widehat{\boldsymbol{\Sigma}}^{-1} + (1-\gamma^{-1})\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Sigma}}_{22}^{-1} \end{pmatrix}. \tag{5.5}$$

5.1 The asymptotic distribution of the T^2-statistic

Proposition 5.1. *If* $n, N \to \infty$ *with* $n/N \to \delta \in (0,1]$ *then* $T^2 \xrightarrow{\mathcal{L}} \chi^2_{p+q}$.

Proof. By the Law of Large Numbers, $n^{-1}\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}$ and $N^{-1}\boldsymbol{A}_{22,N}$ each converge to $\boldsymbol{\Sigma}$; hence, by (4.1), $\widehat{\boldsymbol{\Sigma}} \to \boldsymbol{\Sigma}$, almost surely. Therefore, by (5.2),

$$N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}) \to \boldsymbol{\Sigma} + (\delta^{-1} - 1)\begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

almost surely. Then, the result follows from Corollary 3.2 and an application of Slutsky's theorem to $\widehat{\boldsymbol{\mu}}'\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1}\widehat{\boldsymbol{\mu}} \equiv (\sqrt{N}\widehat{\boldsymbol{\mu}})'\big(N\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})\big)^{-1}(\sqrt{N}\widehat{\boldsymbol{\mu}})$. □

Next, we consider the case in which $N \to \infty$ and n is fixed. Here, by (5.2), $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}) \to \frac{n-2}{n(n-q-2)} \begin{pmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, almost surely. This asymptotic value of $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})$ indicates that large-N inference for $\boldsymbol{\mu}_2$ should be performed entirely with $\widehat{\boldsymbol{\mu}}_2$, and such may be done in a straightforward manner using the exact distribution, $\sqrt{N}(\widehat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2) \sim \mathrm{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_{22})$. As regards inference for $\boldsymbol{\mu}_1$, we construct the corresponding T^2-statistic, $T_1^2 = (\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)' \widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_1)^{-1} (\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)$, and we now derive its limiting distribution.

Theorem 5.2. *Suppose that $N \to \infty$ and n is fixed. Then*

$$T_1^2 \xrightarrow{\mathcal{L}} \frac{n(n-q-2)}{n-2} \frac{\chi_p^2}{\chi_{n-p-q}^2} \Big(1 + \frac{\chi_q^2}{\chi_{n-q}^2}\Big), \tag{5.6}$$

where all the chi-squared random variables above are mutually independent. Further, if both $n, N \to \infty$ with $n/N \to 0$ then $T_1^2 \xrightarrow{\mathcal{L}} \chi_p^2$.

Proof. By (5.2),

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_1) = \frac{1}{N} \widehat{\boldsymbol{\Sigma}}_{11} + \frac{\gamma - 1}{N} \widehat{\boldsymbol{\Sigma}}_{11\cdot 2} = \frac{1}{N} \widehat{\boldsymbol{\Sigma}}_{11} + \frac{\gamma - 1}{N} \widehat{\boldsymbol{\Delta}}_{11}.$$

By (4.3), $N^{-1} \widehat{\boldsymbol{\Sigma}}_{11} \to \mathbf{0}$, almost surely, and by (5.4), $(\gamma - 1)/N \to (n-2)/n(n-q-2)$ as $N \to \infty$. Therefore, $(n-2)T_1^2/n(n-q-2)$ is stochastically equivalent to $\check{T}_1^2 \equiv \widehat{\boldsymbol{\mu}}_1' \widehat{\boldsymbol{\Delta}}_{11}^{-1} \widehat{\boldsymbol{\mu}}_1$, so it suffices to find the distribution of the latter.

Since $\{\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}_1, \bar{\boldsymbol{Y}}_2\}$ and $\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}$ are mutually independent then it follows from (3.3) that for fixed N, $\widehat{\boldsymbol{\mu}}_1 | \{\boldsymbol{A}_{11}, \boldsymbol{A}_{12}, \boldsymbol{A}_{22,n}\} \sim \mathrm{N}_p(\mathbf{0}, n^{-1}(\boldsymbol{I}_p + \bar{\tau} \widehat{\boldsymbol{\Delta}}_{12} \widehat{\boldsymbol{\Delta}}_{21}))$. Therefore, for $t \in \mathbb{R}$,

$$\begin{aligned} E \exp(it\check{T}_1^2) &= E \exp\big(it \operatorname{tr} \widehat{\boldsymbol{\Delta}}_{11}^{-1} \widehat{\boldsymbol{\mu}}_1 \widehat{\boldsymbol{\mu}}_1'\big) \\ &= E|\boldsymbol{I}_p - 2in^{-1}t \widehat{\boldsymbol{\Delta}}_{11}^{-1} (\boldsymbol{I}_p + \bar{\tau} \widehat{\boldsymbol{\Delta}}_{12} \widehat{\boldsymbol{\Delta}}_{21})|^{-1/2}. \end{aligned}$$

As noted in the proof of Theorem 3.1, $n\widehat{\boldsymbol{\Delta}}_{11} = \boldsymbol{A}_{11\cdot 2,n} \sim \mathrm{W}_p(n-q-1, \boldsymbol{I}_p)$, and $\widehat{\boldsymbol{\Delta}}_{11}$ and $\widehat{\boldsymbol{\Delta}}_{12}$ are mutually independent. Applying (2.4) of Lemma 2.3, we obtain

$$\begin{aligned} E \exp(it\check{T}_1^2) &= E\Big|\boldsymbol{I}_p - 2it\frac{n(n-q-2)}{n-2} Q_1^{-1} \big(\boldsymbol{I}_p + \widehat{\boldsymbol{\Delta}}_{12} \widehat{\boldsymbol{\Delta}}_{21}\big)\Big|^{-1/2} \\ &= E \exp\Big(it\frac{n(n-q-2)}{n-2} Q_1^{-1} \boldsymbol{V}_2' (\boldsymbol{I}_p + \bar{\tau} \widehat{\boldsymbol{\Delta}}_{12} \widehat{\boldsymbol{\Delta}}_{21}) \boldsymbol{V}_2\Big), \end{aligned}$$

where $Q_1 \sim \chi_{n-p-q}^2$, $\boldsymbol{V}_2 \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$, and Q_1 and $\boldsymbol{V}_2$ are mutually independent. Noting that $\bar{\tau} \to 1$ as $N \to \infty$, it follows that $T_1^2 \xrightarrow{\mathcal{L}} n(n-q-2) Q_1^{-1} \boldsymbol{V}_2' (\boldsymbol{I}_p + \boldsymbol{F}) \boldsymbol{V}_2 / (n-2)$, where $\boldsymbol{F} = \widehat{\boldsymbol{\Delta}}_{12} \widehat{\boldsymbol{\Delta}}_{21}$ is independent of Q_1 and $\boldsymbol{V}_2$.

Apply the polar coordinates decomposition $\boldsymbol{V}_2 = Q_2^{1/2} \boldsymbol{U}$, where $Q_2 = \boldsymbol{V}_2' \boldsymbol{V}_2$ and $\boldsymbol{U} = \boldsymbol{V}_2 / (\boldsymbol{V}_2' \boldsymbol{V}_2)^{1/2}$. Since $\boldsymbol{V}_2 \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{I}_p)$ then $Q_2 \sim \chi_p^2$, $\boldsymbol{U}$ is uniformly distributed on S^{p-1}, the unit sphere in $\mathbb{R}^p$, and Q_2 and $\boldsymbol{V}_2$ are mutually independent. Then $\boldsymbol{V}_2' (\boldsymbol{I}_p + \boldsymbol{F}) \boldsymbol{V}_2 \overset{\mathcal{L}}{\equiv} Q_2 \boldsymbol{U}' (\boldsymbol{I}_p + \boldsymbol{F}) \boldsymbol{U} = Q_2 (1 + \boldsymbol{U}' \boldsymbol{F} \boldsymbol{U})$. By (4.4), $\boldsymbol{F} \sim \mathrm{F}_{q,n-q+p-1}^{(p)}$ and, by Lemma 4.3, $\boldsymbol{U}' \boldsymbol{F} \boldsymbol{U} \sim \boldsymbol{F}_{q,n-q}^{(1)} \equiv \chi_q^2 / \chi_{n-q}^2$. Therefore

$$T_1^2 \xrightarrow{\mathcal{L}} \frac{n(n-q-2)}{n-2} Q_1^{-1} Q_2 \Big(1 + \frac{\chi_q^2}{\chi_{n-q}^2}\Big) \overset{\mathcal{L}}{\equiv} \frac{n(n-q-2)}{n-2} \frac{\chi_p^2}{\chi_{n-p-q}^2} \Big(1 + \frac{\chi_q^2}{\chi_{n-q}^2}\Big),$$

and all four chi-square variables are mutually independent, so the proof of (5.6) is complete.

As regards the case in which $n, N \to \infty$, it follows from the Central Limit Theorem and (3.5) that $n^{1/2}\widehat{\boldsymbol{\mu}}_1 \xrightarrow{\mathcal{L}} \boldsymbol{V}_2 \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{11\cdot 2})$. Also, by applying the Law of Large Numbers to (4.1) and by (5.2), we obtain $n\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_1) \to \boldsymbol{\Sigma}_{11\cdot 2}$, almost surely. Therefore, as $n, N \to \infty$, $T_1^2 \equiv (n^{1/2}\widehat{\boldsymbol{\mu}}_1)'(n\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}}_1))^{-1}(n^{1/2}\widehat{\boldsymbol{\mu}}_1) \xrightarrow{\mathcal{L}} \boldsymbol{V}_2'\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\boldsymbol{V}_2 \sim \chi_p^2$. □

5.2 Probability inequalities for the T^2-statistic

We now study the small-sample behavior of the T^2-statistic, deriving probability inequalities which lead to conservative confidence levels for ellipsoidal confidence regions for $\boldsymbol{\mu}$. For the case in which N is large, we can already derive from (5.6) a stochastic inequality for T_1^2. That limiting random variable clearly is stochastically greater than

$$\frac{n(n-q-2)}{n-2}\frac{\chi_p^2}{\chi_{n-p-q}^2},$$

a multiple of an F-distributed random variable. Therefore, for $t \ge 0$,

$$\begin{aligned}\lim_{N\to\infty} P(T_1^2 \le t) &\le P\Big(\frac{n(n-q-2)}{n-2}\frac{\chi_p^2}{\chi_{n-p-q}^2} \le t\Big) \\ &= P\Big(\mathrm{F}_{p,n-p-q} \le \frac{(n-2)(n-p-q)}{n(n-q-2)p}t\Big).\end{aligned}$$

This provides an upper bound on the large-N distribution of T_1^2.

We now derive an upper bound on the distribution function of T^2 for the case in which both n and N are fixed.

Proposition 5.3. *For $t \ge 0$, $P(T^2 \le t) \le P\big(\mathrm{F}_{q,N-q} \le (N-q)t/Nq\big)$.*

Proof. By (5.5) and (5.9), we have

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1} \ge \frac{N}{\gamma}\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Sigma}}_{22}^{-1}\end{pmatrix} + \frac{N(\gamma-1)}{\gamma}\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Sigma}}_{22}^{-1}\end{pmatrix} = N\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Sigma}}_{22}^{-1}\end{pmatrix};$$

therefore $T^2 \overset{\mathcal{L}}{\ge} N\widehat{\boldsymbol{\mu}}_2'\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\mu}}_2$, and then the conclusion follows from the distribution of the classical Hotelling's T^2-statistic. □

The derivation of lower bounds on the distribution function of the T^2 statistic requires much more effort. We shall prove the following result.

Theorem 5.4. *For $t \ge 0$,*

$$P(T^2 \le t) \ge P\Big(N^2 n^{-1}\frac{Q_2}{Q_1}\big(1 + \frac{qQ_3}{Q_5}\big) + q\frac{\big(n^{1/2}Q_3^{1/2} + (N-n)^{1/2}Q_4^{1/2}\big)^2}{Q_5} \le t\Big), \tag{5.7}$$

where $Q_1 \sim \chi_{n-p-q}^2$, $Q_2 \sim \chi_p^2$, $Q_3 \sim \chi_q^2$, $Q_4 \sim \chi_q^2$, $Q_5 \sim \chi_2^2$, and $Q_1, \ldots, Q_5$ are mutually independent.

Remark 5.5. In practice, the right-hand side of (5.7) can be calculated by numerical simulation, and this is simpler than simulating the distribution of T^2 directly from its definition.

Lemma 5.6. *Define the modified T^2-statistic,*

$$\widetilde{T}^2 = N(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}^{-1} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \tag{5.8}$$

Then $T^2 \le N\widetilde{T}^2$.

Proof. By (5.5),

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1} = \frac{N}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1} + \frac{N(\gamma-1)}{\gamma} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{\Sigma}}_{22}^{-1} \end{pmatrix}.$$

Also, by [2], p. 63, Exercise 2.54,

$$\widehat{\boldsymbol{\Sigma}}^{-1} \ge \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{\Sigma}}_{22}^{-1} \end{pmatrix}, \tag{5.9}$$

where the ordering is in the sense of positive semidefiniteness; therefore

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1} \le \frac{N}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1} + \frac{N(\gamma-1)}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1} = N\widehat{\boldsymbol{\Sigma}}^{-1}.$$

By (4.2),

$$\widehat{\boldsymbol{\Sigma}} \ge \frac{1}{n}\Big[\tau \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix} + \bar{\tau} \begin{pmatrix} \boldsymbol{A}_{11\cdot 2,n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\Big].$$

Inverting this inequality using a well-known formula ([2], p. 638) for inverting a partitioned matrix, we obtain

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}^{-1} &\le n\left[\tau \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix} + \bar{\tau} \begin{pmatrix} \boldsymbol{A}_{11\cdot 2,n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\right]^{-1} \\ &= n \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}^{-1} + N\bar{\tau} \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22,n}^{-1} \end{pmatrix}. \end{aligned} \tag{5.10}$$

Applying (5.9) to the last term in (5.10) we obtain

$$\widehat{\boldsymbol{\Sigma}}^{-1} \le n \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}^{-1} + N\bar{\tau} \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}^{-1} = N \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}^{-1}.$$

Therefore

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\mu}})^{-1} \le N^2 \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22,n} \end{pmatrix}^{-1},$$

and now the conclusion follows immediately. □

Proof of Theorem 5.4. As shown in Lemma 5.6, $T^2 \le N\widetilde{T}^2$; therefore $T^2 \overset{\mathcal{L}}{\le} N\widetilde{T}^2$, so it suffices to derive a lower bound on the distribution function of $\widetilde{T}^2$. To that end, we assume, without loss of generality, that $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+q}$.

We apply to (5.8) the quadratic identity (5.3), obtaining

$$N^{-1}\widetilde{T}^2 = (\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\widehat{\boldsymbol{\mu}}_2)'\boldsymbol{A}_{11\cdot 2,n}^{-1}(\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\widehat{\boldsymbol{\mu}}_2) + \widehat{\boldsymbol{\mu}}_2'\boldsymbol{A}_{22,n}^{-1}\widehat{\boldsymbol{\mu}}_2.$$

By (3.3), $\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\widehat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1$; therefore

$$N^{-1}\widetilde{T}^2 = (\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)'\boldsymbol{A}_{11\cdot 2,n}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1) + \widehat{\boldsymbol{\mu}}_2'\boldsymbol{A}_{22,n}^{-1}\widehat{\boldsymbol{\mu}}_2.$$

Recall that $\boldsymbol{A}_{11\cdot 2,n} \sim W_{p+q}(n-q-1, \boldsymbol{I}_{p+q})$ and is independent of $\{\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}_1, \bar{\boldsymbol{Y}}_2, \boldsymbol{A}_{12}, \boldsymbol{A}_{22,n}\}$ (see Proposition 4.2(i)). Hence, by Proposition 2.2(iv),

$$Q_1 \equiv \frac{(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)'(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)}{(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)'\boldsymbol{A}_{11\cdot 2,n}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)} \sim \chi^2_{n-p-q},$$

and Q_1 is independent of $\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1$. Therefore

$$N^{-1}\widetilde{T}^2 \stackrel{\mathcal{L}}{=} Q_1^{-1}\,(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)'(\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1) + \bar{\boldsymbol{Y}}'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}},$$

where Q_1 is independent of $\{\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}_1, \bar{\boldsymbol{Y}}_2, \boldsymbol{A}_{12}, \boldsymbol{A}_{22,n}\}$.

By Proposition 2.2(ii), $\boldsymbol{A}_{12}|\boldsymbol{A}_{22,n} \sim N(\boldsymbol{0}, \boldsymbol{I}_p \otimes \boldsymbol{A}_{22,n})$. Let $\boldsymbol{B}_{12} = \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1/2}$, so that $\boldsymbol{B}_{12}|\boldsymbol{A}_{22,n} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_p \otimes \boldsymbol{I}_q)$; since this conditional distribution does not depend on $\boldsymbol{A}_{22,n}$ then $\boldsymbol{B}_{12}$ also is independent of $\boldsymbol{A}_{22,n}$. Conditional on $\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}$, $\bar{\boldsymbol{X}} - \boldsymbol{A}_{12}\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1 = \bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1$, viewed as a linear function of $\bar{\boldsymbol{X}}$ and $\boldsymbol{B}_{12}$, is multivariate normally distributed with conditional mean

$$E(\bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}) = E(\bar{\boldsymbol{X}}) - E(\boldsymbol{B}_{12})\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1 = \boldsymbol{0},$$

and conditional covariance matrix

$$\begin{aligned}\mathrm{Cov}(\bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1)|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}) &= \mathrm{Cov}(\bar{\boldsymbol{X}}) + E(\boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1\bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1/2}\boldsymbol{B}_{12}'|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}) \\ &= n^{-1}\boldsymbol{I}_p + E(\boldsymbol{B}_{12}\boldsymbol{C}\boldsymbol{B}_{12}'|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}),\end{aligned}$$

where $\boldsymbol{C} = \boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1\bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1/2}$. Since $\boldsymbol{B}_{12} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_p \otimes \boldsymbol{I}_q)$ then the entries of $\boldsymbol{B}_{12}$ are i.i.d. $N(0,1)$ variables. By examining the explicit form of the entries of the matrix $\boldsymbol{B}_{12}\boldsymbol{C}\boldsymbol{B}_{12}'$, it is straightforward to show that if $\boldsymbol{C}$ is a fixed symmetric $q \times q$ matrix then $E(\boldsymbol{B}_{12}\boldsymbol{C}\boldsymbol{B}_{12}') = (\mathrm{tr}\,\boldsymbol{C})\boldsymbol{I}_p$. Therefore, $\mathrm{Cov}(\bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1)|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}) = (n^{-1} + \bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)\boldsymbol{I}_p$. Also, having shown that $\bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\} \sim \mathrm{N}_p(\boldsymbol{0}, (n^{-1} + \bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)\boldsymbol{I}_p)$, we obtain

$$(\bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1)'(\bar{\boldsymbol{X}} - \boldsymbol{B}_{12}\boldsymbol{A}_{22,n}^{-1/2}\bar{\boldsymbol{Y}}_1)|\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\} \stackrel{\mathcal{L}}{=} (n^{-1} + \bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)Q_2,$$

where $Q_2 \sim \chi^2_p$ independently of $\{\bar{\boldsymbol{Y}}_1, \boldsymbol{A}_{22,n}\}$.

By the Cauchy-Schwartz inequality,

$$\bar{\boldsymbol{Y}}'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}} \equiv (\tau\bar{\boldsymbol{Y}}_1 + \bar{\tau}\bar{\boldsymbol{Y}}_2)'\boldsymbol{A}_{22,n}^{-1}(\tau\bar{\boldsymbol{Y}}_1 + \bar{\tau}\bar{\boldsymbol{Y}}_2) \le \big(\tau(\bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)^{1/2} + \bar{\tau}(\bar{\boldsymbol{Y}}_2'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_2)^{1/2}\big)^2.$$

Therefore

$$N^{-1}\widetilde{T}^2 \stackrel{\mathcal{L}}{\le} \frac{Q_2}{Q_1}(n^{-1} + \bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1) + (\tau(\bar{\boldsymbol{Y}}_1'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_1)^{1/2} + \bar{\tau}(\bar{\boldsymbol{Y}}_2'\boldsymbol{A}_{22,n}^{-1}\bar{\boldsymbol{Y}}_2)^{1/2})^2. \qquad (5.11)$$

Denote by $\lambda_{\max}(\boldsymbol{A}_{22,n}^{-1})$ the largest eigenvalue of $\boldsymbol{A}_{22,n}^{-1}$; by the definition of $\lambda_{\max}(\boldsymbol{A}_{22,n}^{-1})$, we have $\boldsymbol{Y}_j'\boldsymbol{A}_{22,n}^{-1}\boldsymbol{Y}_j \le \lambda_{\max}(\boldsymbol{A}_{22,n}^{-1})\,\boldsymbol{Y}_j'\boldsymbol{Y}_j$ for $j = 1, 2$. On applying these inequalities to (5.11), noting that $\lambda_{\max}(\boldsymbol{A}_{22,n}^{-1}) = 1/\lambda_{\min}(\boldsymbol{A}_{22,n})$, where $\lambda_{\min}(\boldsymbol{A}_{22,n})$ is the smallest eigenvalue of $\boldsymbol{A}_{22,n}$, we obtain

$$N^{-1}\widetilde{T}^2 \overset{\mathcal{L}}{\leq} \frac{Q_2}{Q_1}\Big(n^{-1} + \frac{\bar{\boldsymbol{Y}}_1'\bar{\boldsymbol{Y}}_1}{\lambda_{\min}(\boldsymbol{A}_{22,n})}\Big) + \frac{\big(\tau(\bar{\boldsymbol{Y}}_1'\bar{\boldsymbol{Y}}_1)^{1/2} + \bar{\tau}(\bar{\boldsymbol{Y}}_2'\bar{\boldsymbol{Y}}_2)^{1/2}\big)^2}{\lambda_{\min}(\boldsymbol{A}_{22,n})}.$$

Since $\bar{\boldsymbol{Y}}_1 \sim \mathrm{N}_q(\boldsymbol{0}, n^{-1}\boldsymbol{I}_q)$ then $\bar{\boldsymbol{Y}}_1'\bar{\boldsymbol{Y}}_1 \overset{\mathcal{L}}{=} n^{-1}Q_3$ where $Q_3 \sim \chi^2_q$; similarly, $\bar{\boldsymbol{Y}}_2'\bar{\boldsymbol{Y}}_2 \overset{\mathcal{L}}{=} (N-n)^{-1}Q_4$ where $Q_4 \sim \chi^2_q$. Therefore

$$N^{-1}\widetilde{T}^2 \overset{\mathcal{L}}{\leq} n^{-1}\frac{Q_2}{Q_1}\Big(1 + \frac{Q_3}{\lambda_{\min}(\boldsymbol{A}_{22,n})}\Big) + \frac{\big(n^{1/2}Q_3^{1/2} + (N-n)^{1/2}Q_4^{1/2}\big)^2}{N^2\lambda_{\min}(\boldsymbol{A}_{22,n})}. \tag{5.12}$$

Finally, we obtain a stochastic lower bound on $\lambda_{\min}(\boldsymbol{A}_{22,n})$. For any $t \geq 0$, it is simple to see that the inequality $\{\lambda_{\min}(\boldsymbol{A}_{22,n}) > t\}$ is equivalent to $\{\boldsymbol{A}_{22,n} > t\boldsymbol{I}_q\}$. Therefore, applying the density function (2.2) of $\boldsymbol{A}_{22,n} \sim W_q(n-1, \boldsymbol{I}_q)$, we obtain

$$\begin{aligned} P(\lambda_{\min}(\boldsymbol{A}_{22,n}) > t) &= \int_{\boldsymbol{W} > t\boldsymbol{I}_q} \frac{|\boldsymbol{W}|^{(n-q-2)/2}\exp(-\frac{1}{2}\mathrm{tr}\,\boldsymbol{W})}{2^{(n-1)q/2}\Gamma_q((n-1)/2)}\,d\boldsymbol{W} \\ &= e^{-qt/2}\int_{\boldsymbol{W} > \boldsymbol{0}} \frac{|\boldsymbol{W} + t\boldsymbol{I}_q|^{(n-q-2)/2}\exp(-\frac{1}{2}\mathrm{tr}\,\boldsymbol{W})}{2^{(n-1)q/2}\Gamma_q((n-1)/2)}\,d\boldsymbol{W}, \end{aligned}$$

where the latter equality is obtained by making the transformation $\boldsymbol{W} \to \boldsymbol{W} + t\boldsymbol{I}_q$. Since $n > q+2$ then $|\boldsymbol{W} + t\boldsymbol{I}_q|^{(n-q-2)/2} \geq |\boldsymbol{W}|^{(n-q-2)/2}$ for all $\boldsymbol{W}$ and t; applying this inequality to the integrand above, then the remaining integral equals 1.

Therefore $P(\lambda_{\min}(\boldsymbol{A}_{22,n}) > t) \geq e^{-qt/2}$ for all $t \geq 0$, hence $\lambda_{\min}(\boldsymbol{A}_{22,n}) \overset{\mathcal{L}}{\geq} q^{-1}Q_5$, where $Q_5 \sim \chi^2_2$; equivalently, $1/\lambda_{\min}(\boldsymbol{A}_{22,n}) \overset{\mathcal{L}}{\leq} qQ_5^{-1}$. Substituting this result in (5.12), we obtain

$$T^2 \overset{\mathcal{L}}{\leq} N\widetilde{T}^2 \overset{\mathcal{L}}{\leq} N^2 n^{-1}\frac{Q_2}{Q_1}\Big(1 + \frac{qQ_3}{Q_5}\Big) + q\frac{\big(n^{1/2}Q_3^{1/2} + (N-n)^{1/2}Q_4^{1/2}\big)^2}{Q_5}.$$

The proof of (5.7) now is complete. □

6 A normal approximation to $\widehat{\boldsymbol{\mu}}$

It would be useful to approximate the distribution of $\widehat{\boldsymbol{\mu}}$ by a normal distribution for, in data analysis, such an approximation would make the distribution theory tractable. One approximation arises from discarding the last term in (3.5), so that $\widehat{\boldsymbol{\mu}} \approx \mathrm{N}_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Omega})$. A second, and more accurate, normal approximation is $\widehat{\boldsymbol{\mu}} \approx \mathrm{N}_{p+q}(\boldsymbol{\mu}, \widetilde{\boldsymbol{\Omega}})$, where $\widetilde{\boldsymbol{\Omega}} = \mathrm{Cov}(\widehat{\boldsymbol{\mu}})$ is given in (3.6). Both approximations are easy to apply and are accurate if $\tau \simeq 1$. However, the second approximation is generally more accurate because it utilizes information arising from the second term in the expression for $\widehat{\boldsymbol{\mu}}_1$ in (3.3), whereas the first approximation discards that information. Therefore, we restrict our attention to the second approximation.

To quantify the accuracy of this approximation, we obtain an upper bound on the L^∞ distance between the density and distribution functions of $\widehat{\boldsymbol{\mu}}$ and its approximant. This will be done by applying an extension of the classical Esseen inequality.

Proposition 6.1. *For $k = 1, 2$, let $\boldsymbol{V}_k \sim \mathrm{N}_d(\boldsymbol{\nu}, \boldsymbol{\Lambda}_k)$. Denote by $f_k(\cdot)$ the density function of $\boldsymbol{V}_k$ and let $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 - \boldsymbol{\Lambda}_2$. Then there exists an absolute constant C_0 such that*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d} |f_1(\boldsymbol{x}) - f_2(\boldsymbol{x})| \le \frac{d(d+3)}{(6\pi^d)^{1/(d+3)}} \Big(\frac{C_0}{d+2}\Big)^{(d+2)/(d+3)} (\mathrm{tr}\, \boldsymbol{\Lambda}^2)^{1/2(d+3)}. \tag{6.1}$$

Proof. The characteristic function of $\boldsymbol{V}_k$ is $\phi_k(\boldsymbol{t}) = \exp(i\boldsymbol{t}'\boldsymbol{\nu} - \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Lambda}_k\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^d$. Therefore, by the elementary inequality, $|e^{-a} - e^{-b}| \le |a - b|$, $a, b \ge 0$, which is a consequence of the Taylor expansion of e^{-x}, $x > 0$, we have

$$\begin{aligned} |\phi_1(\boldsymbol{t}) - \phi_2(\boldsymbol{t})| &= |e^{i\boldsymbol{t}'\boldsymbol{\nu}}(e^{-\boldsymbol{t}'\boldsymbol{\Lambda}_1\boldsymbol{t}/2} - e^{-\boldsymbol{t}'\boldsymbol{\Lambda}_2\boldsymbol{t}/2})| \\ &= |e^{-\boldsymbol{t}'\boldsymbol{\Lambda}_1\boldsymbol{t}/2} - e^{-\boldsymbol{t}'\boldsymbol{\Lambda}_2\boldsymbol{t}/2}| \le \tfrac{1}{2}|\boldsymbol{t}'\boldsymbol{\Lambda}\boldsymbol{t}| \le (\mathrm{tr}\, \boldsymbol{\Lambda}^2)^{1/2}\boldsymbol{t}'\boldsymbol{t}, \end{aligned}$$

where the last inequality follows from the Cauchy-Schwartz inequality. It follows that, for $h_1, \dots, h_d > 0$,

$$\begin{aligned} \int_{-h_1}^{h_1} \cdots \int_{-h_d}^{h_d} |\phi_1(\boldsymbol{t}) - \phi_2(\boldsymbol{t})| \, d\boldsymbol{t} &\le \tfrac{1}{2}(\mathrm{tr}\, \boldsymbol{\Lambda}^2)^{1/2} \int_{-h_1}^{h_1} \cdots \int_{-h_d}^{h_d} \boldsymbol{t}'\boldsymbol{t}\, d\boldsymbol{t} \\ &= \frac{2^{d-1}}{3}(\mathrm{tr}\, \boldsymbol{\Lambda}^2)^{1/2} h_1 \cdots h_d(h_1^2 + \cdots + h_d^2). \end{aligned}$$

On applying Theorem 3.1 of Roussas [27], we obtain

$$\begin{aligned} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |f_1(\boldsymbol{x}) - f_2(\boldsymbol{x})| &\le C_0(h_1^{-1} + \cdots + h_d^{-1}) + (2\pi)^{-d} \int_{-h_1}^{h_1} \cdots \int_{-h_d}^{h_d} |\phi_1(\boldsymbol{t}) - \phi_2(\boldsymbol{t})| \, d\boldsymbol{t} \\ &\le C_0(h_1^{-1} + \cdots + h_d^{-1}) + C_1 h_1 \cdots h_d(h_1^2 + \cdots + h_d^2), \end{aligned} \tag{6.2}$$

where $C_1 = \pi^{-d}(\mathrm{tr}\, \boldsymbol{\Lambda}^2)^{1/2}/6$, and C_0 is an absolute positive constant, i.e., not dependent on d, f_1, or f_2. It is simple to show that (6.2), as a function of $h_1, \dots, h_d > 0$, is minimized at $(h_0, \dots, h_0)$, where $h_0 = (C_0/(d+2)C_1)^{1/(d+3)}$, therefore (6.2) has minimum value $d(C_1 h_0^{d+2} + C_0 h_0^{-1})$. Simplifying this expression for the minimum value, we obtain (6.1). □

We now obtain a bound for the L^∞-distance between $f_{\widehat{\boldsymbol{\mu}}}$ and $f_{\widetilde{\boldsymbol{\mu}}}$, the density functions of $\widehat{\boldsymbol{\mu}}$ and its normal approximation $\widetilde{\boldsymbol{\mu}} \sim \mathrm{N}_{p+q}(\boldsymbol{\mu}, \mathrm{Cov}(\widehat{\boldsymbol{\mu}}))$.

Theorem 6.2. *There exists a positive constant $C_{p,q,n}$ such that*

$$\sup_{\boldsymbol{x}\in\mathbb{R}^{p+q}} |f_{\widehat{\boldsymbol{\mu}}}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x})| \le C_{p,q,n} \left(\bar{\tau}\, \mathrm{tr}\, \boldsymbol{\Sigma}_{11\cdot 2}^2\right)^{1/2(p+q+3)}. \tag{6.3}$$

Proof. Denote by Q the random variable $(\bar{\tau}Q_2/nQ_1)^{1/2}$ in (3.5); then,

$$\widehat{\boldsymbol{\mu}}|Q \sim \mathrm{N}_{p+q}\Big(\boldsymbol{\mu}, \boldsymbol{\Omega} + Q^2\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\Big).$$

Since $\widetilde{\boldsymbol{\mu}} \sim \mathrm{N}_{p+q}(\boldsymbol{\mu}, \mathrm{Cov}(\widehat{\boldsymbol{\mu}}))$ then, by (6.1), the L^∞-distance between $f_{\widehat{\boldsymbol{\mu}}|Q}$, the conditional density of $\widehat{\boldsymbol{\mu}}$ given Q, and $f_{\widetilde{\boldsymbol{\mu}}}$ satisfies

$$\sup_{\boldsymbol{x}\in\mathbb{R}^{p+q}} |f_{\widehat{\boldsymbol{\mu}}|Q}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x})| \le \frac{(p+q)(p+q+3)}{(6\pi^{p+q})^{1/(p+q+3)}}\Big(\frac{C_0}{p+q+2}\Big)(\mathrm{tr}\, \boldsymbol{\Lambda}_Q^2)^{1/2(p+q+3)}, \tag{6.4}$$

where

$$\boldsymbol{\Lambda}_Q = \mathrm{Cov}(\widehat{\boldsymbol{\mu}}|Q) - \mathrm{Cov}(\widetilde{\boldsymbol{\mu}}) = \big(Q^2 - E(Q^2)\big)\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}.$$

Noting that $f_{\widehat{\boldsymbol{\mu}}}(\boldsymbol{x}) = E_Q f_{\widehat{\boldsymbol{\mu}}|Q}(\boldsymbol{x})$ for all $x \in \mathbb{R}^{p+q}$, we have

$$\begin{aligned}\sup_{\boldsymbol{x}\in\mathbb{R}^{p+q}} |f_{\widehat{\boldsymbol{\mu}}}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x})| &= \sup_{\boldsymbol{x}\in\mathbb{R}^{p+q}} |E(f_{\widehat{\boldsymbol{\mu}}|Q}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x}))|\\ &\le \sup_{\boldsymbol{x}\in\mathbb{R}^{p+q}} E|f_{\widehat{\boldsymbol{\mu}}|Q}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x})|\\ &\le C_{p,q}(\mathrm{tr}\,\boldsymbol{\Sigma}_{11\cdot 2}^2)^{1/2(p+q+3)} E|Q^2 - E(Q^2)|^{1/(p+q+3)},\end{aligned}$$

where $C_{p,q}$ is the constant in (6.4). By Jensen's inequality,

$$\begin{aligned}E|Q^2 - E(Q^2)|^{1/(p+q+3)} &\equiv E(|Q^2 - E(Q^2)|^2)^{1/2(p+q+3)}\\ &\le (E|Q^2 - E(Q^2)|^2)^{1/2(p+q+3)} = (\mathrm{Var}(Q^2))^{1/2(p+q+3)}.\end{aligned}$$

Since $\mathrm{Var}(Q^2) = n^{-1}\bar{\tau}\mathrm{Var}(Q_2/Q_1)$ and

$$\begin{aligned}\mathrm{Var}(Q_2/Q_1) &= E(Q_2^2)E(Q_1^{-2}) - (E(Q_2)E(Q_1^{-1}))^2\\ &= \frac{2q}{(n-q-2)(n-q-4)} - \frac{q^2}{(n-q-2)^2},\end{aligned}$$

then we obtain (6.3) with $C_{p,q,n} = (n^{-1}\mathrm{Var}(Q_2/Q_1))^{1/2(p+q+3)} C_{p,q}$. □

Corollary 6.3. *For $t_1, \dots, t_{p+q} > 0$,*

$$\Big|P\Big(\bigcap_{j=1}^{p+q}\{|\widehat{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big) - P\Big(\bigcap_{j=1}^{p+q}\{|\widetilde{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big)\Big| \le C_{p,q,n}\Big(\prod_{j=1}^{p+q} t_j\Big)(\bar{\tau}\mathrm{tr}\,\boldsymbol{\Sigma}_{11\cdot 2}^2)^{1/2(p+q+3)}. \quad (6.5)$$

Further,

$$P\Big(\bigcap_{j=1}^{p+q}\{|\widehat{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big) \ge \prod_{j=1}^{p+q}\Big[2\Phi\Big(\frac{t_j}{2\sqrt{\mathrm{Var}(\widehat{\mu}_j)}}\Big) - 1\Big] - C_{p,q,n}\Big(\prod_{j=1}^{p+q} t_j\Big)(\bar{\tau}\mathrm{tr}\,\boldsymbol{\Sigma}_{11\cdot 2}^2)^{1/2(p+q+3)}. \quad (6.6)$$

Proof. Let $\mathcal{R}$ denote the rectangle $[-t_1/2, t_1/2] \times \cdots \times [-t_{p+q}/2, t_{p+q}/2]$. Then

$$\begin{aligned}\Big|P\Big(\bigcap_{j=1}^{p+q}\{|\widehat{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big) - P\Big(\bigcap_{j=1}^{p+q}\{|\widetilde{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big)\Big| &= \Big|\int_{\mathcal{R}} (f_{\widehat{\boldsymbol{\mu}}}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x}))d\boldsymbol{x}\Big|\\ &\le \int_{\mathcal{R}} |f_{\widehat{\boldsymbol{\mu}}}(\boldsymbol{x}) - f_{\widetilde{\boldsymbol{\mu}}}(\boldsymbol{x})|d\boldsymbol{x}\\ &\le \|f_{\widehat{\boldsymbol{\mu}}} - f_{\widetilde{\boldsymbol{\mu}}}\|_\infty \mathrm{Vol}(\mathcal{R}).\end{aligned}$$

Then (6.5) follows from (6.3) and the fact that $\mathrm{Vol}(\mathcal{R}) = \prod_{j=1}^{p+q} t_j$. Next, by (6.5),

$$P\Big(\bigcap_{j=1}^{p+q} \{|\widehat{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big) \ge P\Big(\bigcap_{j=1}^{p+q} \{|\widetilde{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big) - C_{p,q,n}\Big(\prod_{j=1}^{p+q} t_j\Big)(\bar{\tau}\,\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_{11\cdot 2}^2)^{1/2(p+q+3)}. \quad (6.7)$$

Since $\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu} \sim \mathrm{N}_{p+q}(\mathbf{0}, \mathrm{Cov}(\widehat{\boldsymbol{\mu}}))$ then, by an inequality of Šidák [29],

$$\begin{aligned} P\Big(\bigcap_{j=1}^{p+q}\{|\widetilde{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\}\Big) &\ge \prod_{j=1}^{p+q} P\big(|\widetilde{\mu}_j - \mu_j| \le \tfrac{1}{2}t_j\big) \\ &= \prod_{j=1}^{p+q} \Big[2\Phi\Big(\frac{t_j}{2\sqrt{\mathrm{Var}(\widehat{\mu}_j)}}\Big) - 1\Big] \end{aligned}$$

Substituting this lower bound at (6.7), we obtain (6.6). □

Remark 6.4. For a given data set, we may apply (6.6) to obtain an estimated lower bound on the confidence level of simultaneous confidence intervals for $\mu_1, \ldots, \mu_{p+q}$. Replacing each unknown parameter on the right-hand side of (6.6) by its corresponding maximum likelihood estimator, we obtain

$$\prod_{j=1}^{p+q} \Big[2\Phi\big(t_j/2\sqrt{\widehat{\mathrm{Var}}(\widehat{\mu}_j)}\big) - 1\Big] - C_{p,q,n}\Big(\prod_{j=1}^{p+q} t_j\Big)(\bar{\tau}\,\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_{11\cdot 2}^2)^{1/2(p+q+3)},$$

which is an estimated lower bound on the confidence level.

We can also obtain bounds on the supremum distance between the cumulative distribution functions of $\widehat{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}$. In the case of lower-orthant unbounded rectangles, we may apply the results of Sadikova [28] and Gamkrelidze [13] on generalizations of Esseen's inequality to derive an analog for cumulative distribution functions of Proposition 6.1. As an indication of these results, we state without proof an analog of Proposition 6.1 for distribution functions for the case in which $d = 2$. As before, suppose that $\boldsymbol{V}_k \sim \mathrm{N}_2(\boldsymbol{\nu}, \boldsymbol{\Lambda}_k)$, $k = 1, 2$, and denote by F_k the distribution function of $\boldsymbol{V}_k$. Further, let $\Lambda_{ij}^{(k)}$ denote the (i,j)th element of $\boldsymbol{\Lambda}_k$ and define $\widetilde{\boldsymbol{\Lambda}}_k = \frac{1}{2}(\boldsymbol{\Lambda}_k + \mathrm{diag}(\boldsymbol{\Lambda}_k))$.

Proposition 6.5. *There exist constants $c_1, c_2 > 0$ such that, for $T > 0$,*

$$\begin{aligned} \sup_{\boldsymbol{x} \in \mathbb{R}^2} |F_1(\boldsymbol{x}) - F_2(\boldsymbol{x})| \le \frac{1}{\pi^2}\Bigg[&\tfrac{4}{3}T^2 \sinh(|\Lambda_{12}^{(1)}|T^2)(\mathrm{tr}\,(\widetilde{\boldsymbol{\Lambda}}_1 - \widetilde{\boldsymbol{\Lambda}}_2)^2)^{1/2} \\ &+ 16\,\frac{\cosh\big(T^2 \max(\Lambda_{12}^{(1)}, \Lambda_{12}^{(2)})\big) - \cosh\big(T^2 \min(\Lambda_{12}^{(1)}, \Lambda_{12}^{(2)})\big)}{(\Lambda_{12}^{(1)} + \Lambda_{12}^{(2)})T^2}\Bigg] \\ &+ c_1\Big(\sum_{j=1}^{2} |\Lambda_{jj}^{(1)} - \Lambda_{jj}^{(2)}|\Big)T + c_2 T^{-1}. \end{aligned}$$

By applying this result to $\widehat{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}$, we obtain an analog of Theorem 6.2 for the case in which $p = q = 1$.

Acknowledgments. Steven Arnold, in providing references to the literature and holding discussions on all aspects of the results derived here, was a veritable fountain of knowledge for Richards as this manuscript was being prepared; consequently, the authors are deeply grateful to Professor Arnold for his constant encouragement and advice. The authors also are very grateful to the referees and to Megan Romer for providing valuable comments on earlier versions of this article.

References

[1] T. W. Anderson, Maximum likelihood estimators for a multivariate normal distribution when some observations are missing, *J. Amer. Statist. Assoc.*, **52** (1957), 200–203.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, third edition, Wiley, New York, 2003.

[3] T. W. Anderson and I. Olkin, Maximum likelihood estimation of the parameters of a multivariate normal distribution. *Linear Algebra Appl.*, **70** (1985), 147–171.

[4] S. A. Andersson and M. D. Perlman, Lattice-ordered conditional independence models for missing data, *Statist. Probab. Lett.*, **12** (1991), 465–486.

[5] S. A. Andersson and M. D. Perlman, Unbiasedness of the likelihood ratio test for lattice conditional independence models, *J. Multivariate Anal.*, **53** (1995), 1–17.

[6] S. A. Andersson and M. D. Perlman, Testing lattice conditional independence models, *J. Multivariate Anal.*, **53** (1995), 18–38.

[7] R. Bhargava, Multivariate tests of hypotheses with incomplete data, Technical Report No. 3, Applied Math. and Statistics Laboratories, Stanford University, Stanford, CA, 1962.

[8] R. P. Bhargava, Some one-sample hypothesis testing problems when there is a monotone sample from a multivariate normal population, *Ann. Inst. Statist. Math.*, **27** (1975), 329–339.

[9] W.-Y. Chang and D. St. P. Richards, Finite-sample inference with monotone incomplete multivariate normal data, II, 2007, preprint, Penn State University.

[10] M. L. Eaton, *Multivariate Statistics. A Vector Space Approach*, Wiley, New York, 1983.

[11] M. L. Eaton and T. Kariya, Multivariate tests with incomplete data. *Ann. Statist.*, **11** (1983), 654–665.

[12] H. Fujisawa, A note on the maximum likelihood estimators for multivariate normal distribution with monotone data, *Comm. Statist. Theory Methods*, **24** (1995), 1377–1382.

[13] N. G. Gamkrelidze, A multidimensional generalization of Esseen's inequality for distribution functions, *Theor. Probability Appl.*, **22** (1977) 877–880.

[14] M. A. Giguère and G. P. H. Styan, Multivariate normal estimation with missing data on several variates, In: *Trans. Seventh Prague Conference on Information Theory, Statistical Decision Functions*, pp. 129–139, Academia, Prague, 1978.

[15] J. Hao and K. Krishnamoorthy, Inferences on a normal covariance matrix and generalized variance with monotone missing data, *J. Multivariate Anal.*, **78** (2001), 62–82.

[16] K. G. Jinadasa and D. S. Tracy, Maximum likelihood estimation for multivariate normal distribution with monotone sample, *Comm. Statist. Theory Methods*, **21** (1992), 41–50.

[17] T. Kanda and Y. Fujikoshi, Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data, *Amer. J. Math. Management Sci.*, **18** (1988), 161–190.

[18] K. Krishnamoorthy and M. K. Pannala, Some simple test procedures for normal mean vector with incomplete data, *Ann. Inst. Statist. Math.*, **50** (1998), 531–542.

[19] K. Krishnamoorthy and M. K. Pannala, Confidence estimation of a normal mean vector with incomplete data, *Canad. J. Statist.*, **27** (1999), 395–407.

[20] K. R. Lang, *Astrophysical Formulae: Volume II: Space, Time, Matter and Cosmology*, third edition, Springer, New York, 2006.

[21] R. J. A. Little, Inference about means from incomplete multivariate data, *Biometrika*, **63** (1976), 593–604.

[22] R. J. A. Little, A test of missing completely at random for multivariate data with missing values, *J. Amer. Statist. Assoc.*, **83** (1988), 1198–1202.

[23] R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data*, 2nd edition, Wiley-Interscience, Hoboken, NJ, 2002.

[24] D. F. Morrison, Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data, *J. Amer. Statist. Assoc.*, **66** (1971), 602–604.

[25] D. F. Morrison and D. S. Bhoj, Power of the likelihood ratio test on the mean vector of the multivariate normal distribution with missing observations, *Biometrika*, **60** (1973), 365–368.

[26] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.

[27] G. G. Roussas, An Esseen-type inequality for probability density functions, with an application, *Statist. Probab. Lett.* **51** (2001), 397–408; **54** (2001), 449.

[28] S. M. Sadikova, On two-dimensional analogs of an inequality of Esseen and their application to the central limit theorem, *Theor. Probability Appl.*, **11** (1966), 325–335.

[29] Z. Šidák, Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.*, **62** (1967), 626–633.

[30] S. S. Wilks, Moments and distributions of estimates of population parameters from fragmentary samples, *Ann. Math. Statist.*, **3** (1932), 163–195.