Finite-Sample Inference with Monotone Incomplete Multivariate Normal Data, III: Hotelling’s $T^2$-Statistic

Megan M. Romer* and Donald St. P. Richards*

February 15, 2012

Abstract

In the setting of inference with two-step monotone incomplete data drawn from $N_d(\mu, \Sigma)$, a multivariate normal population with mean $\mu$ and covariance matrix $\Sigma$, we derive a stochastic representation for the exact distribution of a generalization of Hotelling’s $T^2$-statistic, thereby enabling the construction of exact-level ellipsoidal confidence regions for $\mu$. By applying the equivariance of $\hat{\mu}$ and $\hat{\Sigma}$, the maximum likelihood estimators of $\mu$ and $\Sigma$, respectively, we show that the $T^2$-statistic is invariant under affine transformations. Further, as a consequence of the exact stochastic representation, we derive upper and lower bounds for the cumulative distribution function of the $T^2$-statistic. We apply these results to construct simultaneous confidence regions for linear combinations of $\mu$, and we apply these results to analyze a data set consisting of cholesterol measurements on a group of Pennsylvania heart-disease patients.

1 Introduction

In statistical inference for the mean of a multivariate normal population, Hotelling’s $T^2$-statistic is a classical cornerstone of multivariate analysis and applications of that statistic commensurately are ubiquitous. We refer to Anderson (2003), Eaton (1983), Johnson and Wichern (2002), and Muirhead (1982) for extensive accounts of the theory and applications of Hotelling’s statistic.

Denote by $N_d(\mu, \Sigma)$ a multivariate normal population with mean $\mu$ and covariance matrix $\Sigma$. For the classical situation in which the observed data are complete, the theory underpinning the $T^2$-statistic rests on the mutual independence of $\hat{\mu}$ and $\hat{\Sigma}$, the maximum likelihood estimators of $\mu$ and $\Sigma$, respectively. By contrast, if the data are incomplete then the estimators $\hat{\mu}$ and $\hat{\Sigma}$ generally are not mutually independent and it becomes exceedingly difficult to ascertain the probabilistic behavior of the natural analog of the $T^2$-statistic.

For the case in which the data are monotone incomplete, explicit formulas are available for $\hat{\mu}$ and $\hat{\Sigma}$; we refer to Anderson (1957), Morrison (1971), Anderson and Olkin (1985), Jinadasa and Tracy (1992), Krishnamoorthy and Pannala (1999), and Chang and Richards (2009, 2010) for derivations or applications of those explicit formulas.

---

*Department of Statistics, Penn State University, University Park, PA 16802.

*Supported in part by National Science Foundation grant DMS-0705210.

Key words and phrases. Asymptotic distribution; Confidence region; Equivariance; Hotelling’s $T^2$-squared statistic; Maximum likelihood estimator; Missing completely at random; Missing data; Monotone incomplete data; Multivariate normal distribution; Orthogonal invariance; Pivotal quantity; Stochastic representation; Simultaneous confidence intervals.
Denote by $\text{Cov} (\hat{\mu})$ the covariance matrix of $\hat{\mu}$ and by $\widehat{\text{Cov}} (\hat{\mu})$ the maximum likelihood estimator of $\text{Cov} (\hat{\mu})$. Following Chang and Richards (2009), we study the pivotal quantity,

$$T^2 = (\hat{\mu} - \mu)'(\widehat{\text{Cov}} (\hat{\mu}))^{-1}(\hat{\mu} - \mu),$$

(1.1)

which generalizes Hotelling’s statistic for inference about $\mu$ in the setting of monotone incomplete data. In this paper, we derive a stochastic representation for the exact distribution of the $T^2$-statistic in (1.1). Our results therefore complement the work of: Krishnamoorthy and Pannala (1999), who obtained an $F$-approximation to a version of the $T^2$-statistic that used a modified form of $\text{Cov} (\hat{\mu})$; Chang and Richards (2009), who obtained upper and lower bounds for the distribution of the statistic (1.1) that lead to conservative ellipsoidal confidence regions for $\mu$, and who derived the asymptotic distribution of the $T^2$-statistic for the cases in which $n$ is fixed or $n, N \to \infty$ with $n/N \to \delta \in (0, 1]$; and Seko, Yamazaki, and Seo (2008), who presented numerical investigations of the distribution function of the statistic (1.1).

In deriving an exact stochastic representation for the statistic (1.1), we were motivated by three main considerations. First, it was proved by Chang and Richards (2009, 2010) that the statistic (1.1) is asymptotically chi-square distributed for various large $n$ and $N$; however, Seo, et al. (2008) noted that the chi-square percentiles do not accurately approximate the percentiles of the $T^2$-statistic when $N$ is not large; therefore, there was a need for accurate calculation of the exact percentiles of the statistic (1.1).

Second, there remains open the problem of generalizing to incomplete data the optimality properties of the classical Hotelling’s $T^2$-statistic. In the classical case, such optimality results are based on the exact distribution of the $T^2$-statistic, and we expect to obtain generalizations by means of the exact stochastic representation derived here. Third, the results of Richards and Yamada (2010) now provide Stein-rule, or shrinkage, methods for estimating $\mu$ when the data are monotone incomplete and $\Sigma$ is unknown; to extend those results to the case when $\Sigma$ is unknown, it is necessary to know the exact distribution function of the statistic (1.1), and we expect to derive such results through the exact stochastic representation obtained here.

This paper continues a series of publications (Chang and Richards (2009, 2010); Richards and Yamada (2010); Romer (2009); Yamada, Romer, and Richards (2012); Romer and Richards (2010)). We consider monotone incomplete data, drawn from a multivariate normal population, consisting of mutually independent observations of the form,

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}, X_{n+1}, X_{n+2}, \ldots, X_N$$

(1.2)

where each $X_j \in \mathbb{R}^p$; each $Y_j \in \mathbb{R}^q$; $(X'_j, Y'_j)'$, $j = 1, \ldots, n$ are observations from $N_{p+q}(\mu, \Sigma)$; and the incomplete data $X_j$, $j = n+1, \ldots, N$, are observations on the first $p$ characteristics of the same population. To ensure that all means and variances are finite and that all integrals encountered later are absolutely convergent, we also assume that $n > p + 2$ and $N > n \geq p + q$; see Chang and Richards (2009). As explained by Yamada, et al. (2012), we also assume that data are missing completely at random; indeed, it is necessary to assume missingness completely at random in order to derive the estimators $\hat{\mu}$ and $\hat{\Sigma}$.

We prove in Section 2 an invariance property of the $T^2$-statistic, and we repeatedly draw on that property in deriving the exact stochastic representation. We will deduce from the stochastic representation upper and lower bounds on the distribution function of the $T^2$-statistic. Further, we will derive simultaneous confidence intervals for linear combinations of $\mu$, and we apply our confidence intervals and those previously available to the Pennsylvania cholesterol data as a numerical example.
2 An invariance property of the $T^2$-statistic

We modify suitably the notation for sample mean vectors, covariance matrices, and related entities from the paper of Chang and Richards (2009). Thus, $\tau = n/N$ denotes the proportion of data in (1.2) which are complete, and we denote $1 - \tau$ by $\bar{\tau}$; also, $\mathbf{0}$ will denote any zero vector, the dimension of which will be clear from the context in which it appears. We decompose $\mu$ and $\Sigma$ in conformity with (1.2),

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$  \hspace{1cm} (2.1)

so that $\mu_1$ and $\mu_2$ are $p \times 1$ and $q \times 1$, respectively; $\Sigma_{11}, \Sigma_{12} = \Sigma_{21}$, and $\Sigma_{22}$ are of orders $p \times p$, $p \times q$, and $q \times q$, respectively. We also define the well-known Schur complement $\Sigma_{22,1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Define the sample mean vectors

$$\bar{X}_1 = \frac{1}{n} \sum_{j=1}^{n} X_j, \quad \bar{X}_2 = \frac{1}{N-n} \sum_{j=n+1}^{N} X_j,$$

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j, \quad \bar{X} = \frac{1}{N} \sum_{j=1}^{N} X_j,$$  \hspace{1cm} (2.2)

and the corresponding matrices of sums of squares and products

$$A_{11,n} = \sum_{j=1}^{n} (X_j - \bar{X}_1)(X_j - \bar{X}_1)', \quad A_{12} = A_{21} = \sum_{j=1}^{n} (X_j - \bar{X}_1)(Y_j - \bar{Y})',$$

$$A_{22} = \sum_{j=1}^{n} (Y_j - \bar{Y})(Y_j - \bar{Y})', \quad A_{11,N} = \sum_{j=1}^{N} (X_j - \bar{X})(X_j - \bar{X})'.$$  \hspace{1cm} (2.3)

By Anderson (1957) (cf. Morrison (1971), Anderson and Olkin (1985), Jinadasa and Tracy (1992)), the maximum likelihood estimators of $\mu$ and $\Sigma$ are, respectively,

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \bar{Y} - \tau A_{21}A_{11,n}^{-1}(\bar{X}_1 - \bar{X}_2) \end{pmatrix},$$  \hspace{1cm} (2.4)

and

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} \\ \hat{\Sigma}_{21} \\ \hat{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} A_{11,n}^{-1} & \frac{1}{n} A_{11,n}^{-1} & \frac{1}{n} A_{11,n}^{-1} \\ \\ \frac{1}{N} A_{21}A_{11,n}^{-1} & \frac{1}{N} A_{21}A_{11,n}^{-1} & \frac{1}{N} A_{21}A_{11,n}^{-1} \end{pmatrix}.$$  \hspace{1cm} (2.5)

Let

$$\gamma = 1 + \frac{(n-2)N\bar{\tau}}{n(n-p-2)}.$$  \hspace{1cm} (2.6)

As shown by Chang and Richards (2009), the maximum likelihood estimator of $\text{Cov}(\hat{\mu})$ is

$$\widehat{\text{Cov}}(\hat{\mu}) = \frac{1}{N} \hat{\Sigma} + \frac{(\gamma - 1)}{N} \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$  \hspace{1cm} (2.7)

Similar to the approach of Chang and Richards (2009, Section 5), we write the $T^2$-statistic (1.1) as a sum of two terms. Define

$$T_1^2 = n(\bar{Y} - A_{21}A_{11,n}^{-1}\bar{X}_1 - \mu_2 + A_{21}A_{11,n}^{-1}\mu_1)'$$

$$\cdot A_{22,1,n}^{-1}(\bar{Y} - A_{21}A_{11,n}^{-1}\bar{X}_1 - \mu_2 + A_{21}A_{11,n}^{-1}\mu_1),$$  \hspace{1cm} (2.8)
and
\[ T_2^2 = N(\overline{X} - \mu_1)'A_{11,N}^{-1}(\overline{X} - \mu_1) \]  

(2.9)

On applying to (1.1) a well-known quadratic identity (Anderson (2003), p. 63, Exercise 2.54), we obtain
\[ T^2 = N(\hat{\mu} - \mu)'(N\hat{\text{Cov}}(\hat{\mu}))^{-1}(\hat{\mu} - \mu) \]
\[ = N(\hat{\mu}_2 - \mu_2 - A_{21}A_{11,n}^{-1}(\hat{\mu}_1 - \mu_1))' \]
\[ \cdot ((N\hat{\text{Cov}}(\hat{\mu})))_{22,1}^{-1}(\hat{\mu}_2 - \mu_2 - A_{21}A_{11,n}^{-1}(\hat{\mu}_1 - \mu_1)) \]
\[ + N(\hat{\mu}_1 - \mu_1)'((N\text{Cov}(\hat{\mu})))_{11}^{-1}(\hat{\mu}_1 - \mu_1). \]  

(2.10)

By (2.4), we have \( \hat{\mu}_2 - A_{21}A_{11,n}^{-1}\hat{\mu}_1 = \overline{Y} - A_{21}A_{11,n}^{-1}\overline{X} \); also, by (2.5) and (2.7), we obtain \((N\hat{\text{Cov}}(\hat{\mu})))_{22,1} = n^{-1}\gamma A_{22,1,n} \) and \((N\text{Cov}(\hat{\mu})))_{11} = A_{11,N} \). Therefore, by (2.10),
\[ T^2 = N(\gamma^{-1}T_1^2 + T_2^2). \]  

(2.11)

Let \( A_{11} \) and \( A_{22} \) be \( p \times p \) and \( q \times q \) positive definite (symmetric) matrices, respectively. Let \( A_{12} \) be a \( p \times q \) matrix and let \( \nu_1 \) and \( \nu_2 \) be \( p \times 1 \) and \( q \times 1 \) vectors, respectively. Denoting by \( I_d \) the identity matrix of order \( d \), we define
\[ \Lambda = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}, \quad C = \begin{pmatrix} I_p & 0 \\ A_{21} & I_q \end{pmatrix}, \quad \nu = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \]  

(2.12)

and consider the set of affine transformations of the data (1.2) of the form
\[ \begin{pmatrix} X_j^* \\ Y_j^* \end{pmatrix} = \Lambda C \begin{pmatrix} X_j \\ Y_j \end{pmatrix} + \nu, \quad j = 1, \ldots, n, \]
\[ X_j^* = \Lambda_{11}X_j + \nu_1, \quad j = n + 1, \ldots, N; \]  

(2.13)

equivalently,
\[ Y_j^* = \Lambda_{22}A_{21}X_j + \Lambda_{22}Y_j + \nu_2, \quad j = 1, \ldots, n. \]

Romer and Richards (2010) also considered the transformation (2.13) and noted that as \( A_{11}, \Lambda_{21}, \Lambda_{22}, \) and \( \nu \) vary over their respective parameter spaces, the set of all transformations (2.13) forms a group; in particular, each such transformation is invertible.

The proof of the following result follows from the equivariance property of maximum likelihood estimators and algebraic calculations similar to those given by Yamada, et al. (2011), and we provide the details so as to make the paper self-contained.

**Proposition 2.1.** The statistics \( T_1^2 \) in (2.8) and \( T_2^2 \) in (2.9) each are algebraically invariant under the transformation (2.13). Consequently, the same holds for the \( T^2 \)-statistic in (1.1).

**Proof.** Let \( \mu^* = \Lambda C\mu + \nu \) and \( \Sigma^* = \Lambda C\Sigma C'\Lambda' \) be the mean and covariance matrix of the transformed data. Because \( \hat{\mu} \) and \( \hat{\Sigma} \) are maximum likelihood estimators then they are
equivariant under (2.13); that is, \( \hat{\mu} \) and \( \hat{\Sigma} \) are transformed under (2.13) to \( \hat{\mu}' = \Lambda C \hat{\mu} + \nu \) and \( \hat{\Sigma}' = \Lambda C \hat{\Sigma} C' \Lambda' \), respectively. Therefore \( T^2_2 \) is transformed to

\[
(\mu^*_2 - \mu^*_1, 11^{-1} (\mu^*_1 - \mu^*_1)) = (\Lambda_{11} \hat{\mu}_1 + \nu_1 - \Lambda_{11} \mu_1 - \nu_1) (\Lambda_{11} \hat{\Sigma}_{11} \Lambda_{11})^{-1} (\Lambda_{11} \hat{\mu}_1 + \nu_1 - \Lambda_{11} \mu_1 - \nu_1)
\]

\[
\check{\Sigma}_{22} \check{\Sigma}_{11}^{-1} (\mu^*_1 - \mu^*_1) = \Lambda_{22} (\Lambda_{21} + \check{\check{\Sigma}}_{21} \check{\check{\Sigma}}_{11}^{-1} (\hat{\mu}_1 - \mu_1)),
\]

hence,

\[
(\hat{\mu}_2 - \mu_2, 22^{-1} (\hat{\mu}_1 - \mu_1)) = \Lambda_{22} (\Lambda_{21} + \check{\check{\Sigma}}_{21} \check{\check{\Sigma}}_{11}^{-1} (\hat{\mu}_1 - \mu_1)),
\]

which is \( T^2_1 \), identically. Therefore, \( T^2_1 \) also is invariant under (2.13).

By direct algebraic computations, we deduce that \( \Sigma_{22,1} = \Lambda_{22} \check{\check{\Sigma}}_{22,1} \Lambda_{22} \) and that

\[
(\hat{\mu}_2 - \mu_2, 22^{-1} (\hat{\mu}_1 - \mu_1)) = \Lambda_{22} (\Lambda_{21} + \check{\check{\Sigma}}_{21} \check{\check{\Sigma}}_{11}^{-1} (\hat{\mu}_1 - \mu_1)),
\]

and then \( T^2_2 \) is transformed to

\[
(\hat{\mu}_2 - \mu_2 - \check{\Sigma}_{21} \check{\Sigma}_{11}^{-1} (\hat{\mu}_1 - \mu_1)) = \Lambda_{22} (\Lambda_{21} + \check{\check{\Sigma}}_{21} \check{\check{\Sigma}}_{11}^{-1} (\hat{\mu}_1 - \mu_1)),
\]

and then \( T^2_1 \) and \( T^2_2 \) each are invariant then, by (2.11), \( T^2 \) also is invariant. □

By taking \( \Lambda_{11} = \Sigma_{11}^{-1/2}, \Lambda_{22} = \Sigma_{22}^{-1/2}, \) and \( \Lambda_{12} = -\Sigma_{12} \Sigma_{11}^{-1} \), the covariance matrix of \( (X', Y')' \) under this transformation is \( \Lambda_C \Sigma C' \Lambda' = I_{p+q} \). Furthermore, by choosing \( \nu = -\Lambda C \mu \) we may assume \( \mu = 0 \). Therefore, in deriving the distribution of the \( T^2 \)-statistic, we assume, without loss of generality, that \( \mu = 0 \) and \( \Sigma = I_{p+q} \).

In the sequel, we will need to decompose the matrix \( A_{11,N} \) in (2.3) as follows (see Chang and Richards (2010)):

\[
A_{11,N} = A_{11,n} + B_1 + B_2,
\]

where

\[
B_1 = \sum_{j=n+1}^N (X_j - \bar{X}_j)(X_j - \bar{X}_j)',
\]

\[
B_2 = \frac{n(N-n)}{N} (X_1 - \bar{X}_2)(X_1 - \bar{X}_2)',
\]

and \( A_{11,n} \sim W_p(n-1, \Sigma_{11}), B_1 \sim W_p(N-n-1, \Sigma_{11}), \) and \( B_2 \sim W_p(1, \Sigma_{11}) \) are mutually independent Wishart matrices. This decomposition leads to the following result due to Chang and Richards (2009).

**Lemma 2.2.** (Chang and Richards (2009)) Suppose that \( \Sigma_{12} = 0 \). Then the random matrices and vectors \( A_{22,1,n}, A_{21}^{-1}, A_{11,n}, X, Y, B_1, \) and \( B_2 \) are mutually independent.
3 A stochastic representation for the $T^2$-statistic

In this section, we establish a stochastic representation for the exact distribution of the $T^2$-statistic \( \{1,1\} \). To prove this result, we will utilize the method of characteristic functions together with repeated applications of the powerful method of orthogonal invariance. The resulting stochastic representation, remarkably, involves only chi-squared and Beta random variables and a $2 \times 2$ Wishart matrix, all mutually independent; therefore, it is a straightforward matter to simulate this representation.

In the sequel, we use the standard notation (cf. Muirhead, 1982, Chapter 3) Beta\( (a,b) \) for the scalar beta distribution; Beta\( _d(a,b) \), for the multivariate beta distribution; and $W_d(m, \Sigma)$, for the Wishart distribution. Also, we use the usual notation, \( \overset{\sim}{=}, \) to denote equality in distribution; that is, if X and Y are random entities, then $X \overset{\sim}{=} Y$ signifies that X and Y have the same probability distribution.

**Theorem 3.1.** Let $Q_1 \sim \chi^2_{n-p-q}$, $Q_2 \sim \chi^2_q$, $Q_3 \sim \chi^2_p$, $Q_4 \sim \chi^2_p$, $\cos^2 \theta \sim \text{Beta}(1/2, (p-1)/2)$, $\beta \sim \text{Beta}((n-p-2)/2, (N-n-1)/2)$, and $W \sim W_2(N-p, I_2)$ be mutually independent. Set $e_{1,2} = (1,0)'$, $e_{2,2} = (0,1)'$, and

$$u = (Q_3^{1/2} + Q_4^{1/2} \cos \theta)e_{1,2} + Q_4^{1/2} \sin \theta e_{2,2} \equiv \begin{pmatrix} Q_3^{1/2} + Q_4^{1/2} \cos \theta \\ Q_4^{1/2} \sin \theta \end{pmatrix}.$$ 

Then,

$$N^{-1}T^2 \overset{\sim}{=} \frac{Q_2}{\gamma Q_1} \left(1 + \frac{Q_3}{\beta} e_{1,2}'W^{-1}e_{1,2}\right) + u'W^{-1}u - \frac{\tau Q_3 + \tau Q_4 - 2(\tau \bar{\tau} Q_3 Q_4)^{1/2} \cos \theta}{1 + (\tau Q_3 + \tau Q_4 - 2(\tau \bar{\tau} Q_3 Q_4)^{1/2} \cos \theta)} e_{1,2}'W^{-1}e_{1,2} (e_{1,2}'W^{-1}u)^2.$$

**3.1 Some preliminary results**

Let $M$ be a $p \times q$ matrix and denote by vec\( (M) \) the $pq \times 1$ column vector formed by stacking the columns of $M$. Let $C$ and $D$ be, respectively, $p \times p$ and $q \times q$ positive definite (symmetric) matrices. Denote by $|C|$ the determinant of $C$, and denote by $C \otimes D$ the Kronecker product of $C$ and $D$. Muirhead (1982, pp. 73–78) provides numerous properties of these matrix operations and we state here those properties that are needed in the sequel.

**Proposition 3.2.**

(i) $(C \otimes D)' = C' \otimes D'$, $\text{tr} (C \otimes D) = (\text{tr} C) (\text{tr} D)$, $(C \otimes D)^{-1} = C^{-1} \otimes D^{-1}$, and $|C \otimes D| = |C|^q |D|^p$.

(ii) If $A$ is $m \times p$ and $B$ is $r \times q$, then $(A \otimes B)(C \otimes D) = AC \otimes BD$.

(iii) If $A$ is $m \times p$ and $B$ is $q \times m$, then

$$\text{vec}(BAM) = (M' \otimes B)\text{vec}(A),$$

$$\text{tr}(BAM) = (\text{vec}(B'))'(I \otimes A)\text{vec}(M),$$

and

$$\text{tr}(AM'CMB) = (\text{vec}(M'))'(BA \otimes C')\text{vec}(M).$$
(iv) For any vector \( \mathbf{v} \in \mathbb{R}^p \),
\[
(C + \mathbf{vv}')^{-1} = C^{-1} - (1 + \mathbf{v'}C^{-1}\mathbf{v})^{-1}C^{-1}\mathbf{vv'}C^{-1}.
\] (3.1)

We now state a result on the characteristic function of a quadratic form in multivariate normal variables. Results of this type appear in various forms in the literature; notably, they often follow from a result of Khatri (1980, p. 446, equation (3.4)).

**Lemma 3.3.** (Khatri, 1980) Let \( C \) be a real, symmetric \( p \times p \) matrix, \( t \in \mathbb{R} \), \( \mathbf{v} \in \mathbb{R}^p \), and \( \mathbf{Z} \sim N_p(0, \Sigma) \), and let \( i \equiv \sqrt{-1} \). Then
\[
E e^{it(Z'\mathbf{CZ} + \mathbf{v}'\mathbf{v})} = |I_p - 2it\mathbf{C}\Sigma|^{-1/2} \exp \left( -\frac{1}{2}t^2\mathbf{v}'\Sigma(I_p - 2it\mathbf{C}\Sigma)^{-1}\mathbf{v} \right). \tag{3.2}
\]

Moreover, (3.2) remains valid if \( \mathbf{C} \) is a complex symmetric matrix whose imaginary part is positive definite and \( \mathbf{v} \) is a complex vector.

The following result extends Lemma 2.2 of Chang and Richards (2009).

**Lemma 3.4.** Let \( \mathbf{A} \) be a \( q \times q \) positive definite matrix and \( \mathbf{U} \) be a \( p \times p \) positive semidefinite matrix. If \( \mathbf{B}_{12} \sim N(0, \mathbf{C} \otimes \mathbf{D}) \), then
\[
E \exp(-\text{tr} \mathbf{U} \mathbf{B}_{12} \mathbf{D}^{-1} \mathbf{A} \mathbf{B}_{12}' \mathbf{D}^{-1}) = |I_p + 2q \mathbf{C} \mathbf{D}^{-1} - I_p|^{-1/2}. \tag{3.3}
\]
This result remains valid if \( \mathbf{U} \) is a symmetric complex matrix whose real part is positive semidefinite.

**Proof.** We attribute the following proof of (3.3) to an anonymous referee of Chang and Richards (2009). First, recall that if \( \mathbf{X} \sim N_d(0, \mathbf{I}_d) \), then \( \mathbf{XX}' \sim W_d(1, \mathbf{I}_d) \). Therefore, for any \( p \times p \) positive definite matrix \( \mathbf{A} \) and \( t > 0 \),
\[
E \exp \left( -t\mathbf{X}'\mathbf{AX} \right) = E \exp \left( -t \text{tr} \mathbf{AXX}' \right) = |I + 2t\mathbf{A}|^{-1/2}, \tag{3.4}
\]

Define \( \mathbf{K} = \mathbf{D}^{-1/2} \mathbf{B}_{12}' \mathbf{C} \mathbf{D}^{-1/2} \), \( \phi = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} \), and \( \psi = \mathbf{C}^{-1/2} \mathbf{U} \mathbf{C}^{-1/2} \). By Proposition 3.2(ii), \( \text{vec}(\mathbf{K}) = (\mathbf{C}^{-1/2} \otimes \mathbf{D}^{-1/2}) \text{vec}(\mathbf{B}_{12}) \). Because \( \mathbf{B}_{12} \sim N(0, \mathbf{C} \otimes \mathbf{D}) \) then \( \text{vec}(\mathbf{B}_{12}) \sim N(0, \mathbf{I} \otimes \mathbf{D}) \), so it follows that
\[
\text{vec}(\mathbf{K}) \sim N_{pq}(0, (\mathbf{C}^{-1/2} \otimes \mathbf{D}^{-1/2})((\mathbf{C} \otimes \mathbf{D}))(\mathbf{C}^{-1/2} \otimes \mathbf{D}^{-1/2})).
\]

By Proposition 3.2(ii),
\[
(C^{-1/2} \otimes \mathbf{D}^{-1/2})(\mathbf{C} \otimes \mathbf{D})(C^{-1/2} \otimes \mathbf{D}^{-1/2}) = C^{-1/2} \mathbf{C} \mathbf{D}^{-1/2} \otimes \mathbf{D}^{-1/2} \mathbf{D} \mathbf{D}^{-1/2} = I_p \otimes I_q = I_{pq};
\]

hence, \( \text{vec}(\mathbf{K}) \sim N_{pq}(0, I_{pq}) \). Moreover, by Proposition 3.2(iii),
\[
(\text{vec} \mathbf{K})'(\psi \otimes \phi)(\text{vec} \mathbf{K}) = (\text{vec} \mathbf{K})'\text{vec}(\phi \mathbf{K} \psi) = \text{tr}(\mathbf{K}'\phi \mathbf{K} \psi) = \text{tr}(\psi \mathbf{K}'\phi \mathbf{K}),
\]
and from the definitions of ψ, K, and φ, we have ψK′φK = UBD−1ΛD−1B′. Because vec(K) ∼ Npq(0, Ipq), then we obtain the desired result from the moment-generating function stated above, (3.4), with t = 1 and A = ψ′ ⊗ φ. □

Chang and Richards (2009) gathered together a collection of properties of the Wishart distribution that we will also need here, all of which are available from Anderson (2003), Eaton (1983), or Muirhead (1982).

**Proposition 3.5.** For \( W \sim W_d(a, \Lambda) \), suppose that

\[
W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}
\]

are partitioned as in (2.1), and set \( \Lambda_{22-1} = \Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12} \). Then,

(i) \( W_{22-1} \) and \( \{ W_{21}, W_{11} \} \) are mutually independent, and \( W_{22-1} \sim W_q(a - p, \Lambda_{22-1}) \).
(ii) \( W_{21} | W_{11} \sim N(\Lambda_{21}\Lambda_{11}^{-1}W_{11}, \Lambda_{22-1} \otimes W_{11}) \).
(iii) If \( \Lambda_{12} = 0 \), then \( W_{22-1}, W_{11} \), and \( W_{21} W_{11}^{-1} W_{12} \) are mutually independent. Moreover, \( W_{21} W_{11}^{-1} W_{12} \sim W_q(p, \Lambda_{22}) \).
(iv) Suppose that \( M \) is a \( k \times d \) matrix of rank \( k \) where \( k \leq d \). Then \( (MW^{-1}M')^{-1} \sim W_k(a - d + k, (M\Lambda^{-1}M')^{-1}) \). In particular, if \( Y \) is a \( d \times 1 \) random vector which is independent of \( W \) and is such that \( P(Y = 0) = 0 \), then \( Y \) is independent of \( Y'\Lambda^{-1}Y / Y'W^{-1}Y \), and \( Y'\Lambda^{-1}Y / Y'W^{-1}Y \sim \chi^2_{a - d + 1} \).
(v) Let \( G \sim W_d(a, \Sigma) \), \( H \sim W_d(b, \Phi) \), where \( G \) and \( H \) are independent. Then \( L = (G + H)^{-1/2}G(G + H)^{-1/2} \sim Beta_d(a/2, b/2) \) and \( L \) is independent of \( G + H \).

### 3.2 The proof of Theorem 3.1

As noted earlier we assume, without loss of generality, that \( \mu = 0 \) and \( \Sigma = I_{p+q} \). Recall from (2.11) that

\[
N^{-1}T^2 = \gamma^{-1}T_1^2 + T_2^2,
\]

where

\[
T_1^2 = n(Y - A_{21}A_{11}^{-1}\bar{X}_1)'A_{22-1}^{-1}(Y - A_{21}A_{11}^{-1}\bar{X}_1)
\]

and

\[
T_2^2 = N\bar{X}'(A_{11}^{-1}B_1 + B_2)^{-1}\bar{X}.
\]

By elementary properties of the multivariate normal distribution,

\[
n^{1/2}(Y - A_{21}A_{11}^{-1}\bar{X}_1) | \{ \bar{X}_1, \bar{X}_2, A_{21}, A_{11}, B_1 \} \sim N_q(-n^{1/2}A_{21}A_{11}^{-1}\bar{X}_1, I_q), \quad (3.5)
\]

and by Proposition 3.5(i), \( A_{22-1} \sim W_q(n - p - 1, I_q) \) and is independent of \( \{ A_{12}, A_{11} \} \).

Define

\[
Q_1 = \frac{n(Y - A_{21}A_{11}^{-1}\bar{X}_1)'(Y - A_{21}A_{11}^{-1}\bar{X}_1)}{T_1^2};
\]

then by Proposition 3.5(iv), \( Q_1 | \{ A_{12}, A_{11}, \bar{X}_1 \} \sim \chi^2_{n - p - q} \) and \( Q_1 \) is independent of \( Y - A_{21}A_{11}^{-1}\bar{X}_1 \). Because this distribution does not depend on \( \{ A_{12}, A_{11}, \bar{X}_1 \} \), then \( Q_1 \) is also independent of \( \{ A_{12}, A_{11}, \bar{X}_1 \} \). Therefore,

\[
T_1^2 \overset{\text{iid}}{\sim} \frac{n(Y - A_{21}A_{11}^{-1}\bar{X}_1)'(Y - A_{21}A_{11}^{-1}\bar{X}_1)}{Q_1},
\]

8
where $Q_1 \sim \chi^2_{n-p-q}$ and the numerator and denominator are mutually independent. By (3.5),

$$n(\bar{Y} - A_{21} A_{11, n}^{-1} \bar{X}_1)'(\bar{Y} - A_{21} A_{11, n}^{-1} \bar{X}_1) \sim \chi^2_q(n \bar{X}_1' A_{11, n}^{-1} A_{12} A_{21} A_{11, n}^{-1} \bar{X}_1),$$

(3.6)
a noncentral chi-squared distribution with $q$ degrees of freedom and noncentrality parameter $n \bar{X}_1' A_{11, n}^{-1} A_{12} A_{21} A_{11, n}^{-1} \bar{X}_1$.

Let $t \in \mathbb{R}$. By Lemma 2.2, the characteristic function of $T^2/N$ is given by

$$E \exp(itN^{-1}T^2) = E \exp \left[ \frac{it}{\gamma Q_1} n(\bar{Y} - A_{21} A_{11, n}^{-1} \bar{X}_1)'(\bar{Y} - A_{21} A_{11, n}^{-1} \bar{X}_1) + N \bar{X}'(A_{11, n} + B_1 + B_2)^{-1} \bar{X} \right]$$

$$= E_{Q_1} E_{X_1, X_2} E_{B_1, E_{A_{21}, A_{11, n}}} \exp \left[ itN \bar{X}'(A_{11, n} + B_1 + B_2)^{-1} \bar{X} \right]$$

$$\cdot E_{X_1, X_2, A_{21}, A_{11, n}, B_1} \exp \left[ \frac{it}{\gamma Q_1} n(\bar{Y} - A_{21} A_{11, n}^{-1} \bar{X}_1)'(\bar{Y} - A_{21} A_{11, n}^{-1} \bar{X}_1) \right] \quad (3.7)$$

Applying to (3.6) the well-known formula for the characteristic function of the noncentral chi-squared distribution, and inserting the result into (3.7) yields

$$E \exp(itN^{-1}T^2) = E_{Q_1} E_{X_1, X_2} E_{B_1, E_{A_{21}, A_{11, n}}} \exp \left[ itN \bar{X}'(A_{11, n} + B_1 + B_2)^{-1} \bar{X} \right]$$

$$\cdot (1 - \frac{2it}{\gamma Q_1})^{-q/2} \exp \left[ itn(\gamma Q_1 - 2it)^{-1} \bar{X}_1' A_{11, n}^{-1} A_{12} A_{21} A_{11, n}^{-1} \bar{X}_1 \right]. \quad (3.8)$$

By Proposition 3.5 (ii), $A_{21} | A_{11, n} \sim N(0, I_q \otimes A_{11, n}$); therefore (3.8) equals

$$E_{Q_1} \left( 1 - \frac{2it}{\gamma Q_1} \right)^{-q/2} E_{X_1, X_2} E_{B_1, E_{A_{11, n}}} \exp \left[ itn(\gamma Q_1 - 2it)^{-1} \bar{X}_1' A_{11, n}^{-1} A_{12} A_{21} A_{11, n}^{-1} \bar{X}_1 \right]$$

$$\cdot E_{A_{21} | A_{11, n}} \exp \left[ itn(\gamma Q_1 - 2it)^{-1} \bar{X}_1' A_{11, n}^{-1} A_{12} A_{21} A_{11, n}^{-1} \bar{X}_1 \right]. \quad (3.9)$$

By Lemma 3.4 with $U = -itn(\gamma Q_1 - 2it)^{-1} I_q$, $B_{12} = A_{21}$, $D = A_{11, n}$, $C = I_q$, and $\Lambda = \bar{X}_1 \bar{X}_1'$, we have

$$E_{A_{21} | A_{11, n}} \exp \left[ itn(\gamma Q_1 - 2it)^{-1} \bar{X}_1' A_{11, n}^{-1} A_{12} A_{21} A_{11, n}^{-1} \bar{X}_1 \right]$$

$$= E_{A_{21} | A_{11, n}} \exp \left[ itn(\gamma Q_1 - 2it)^{-1} \text{tr} \left( A_{21} A_{11, n}^{-1} \bar{X}_1 \bar{X}_1' A_{11, n}^{-1} A_{12} \right) \right]$$

$$= |I_p - 2itn(\gamma Q_1 - 2it)^{-1} I_q \otimes A_{11, n}^{-1/2} \bar{X}_1 \bar{X}_1' A_{11, n}^{-1/2}|^{-q/2}$$

$$= (1 - 2itn(\gamma Q_1 - 2it)^{-1} \bar{X}_1' A_{11, n}^{-1} \bar{X}_1)^{-q/2}.$$

Substituting this result into (3.9) yields

$$E \exp(itN^{-1}T^2)$$

$$= E_{Q_1} \left( 1 - \frac{2it}{\gamma Q_1} \right)^{-q/2} E_{X_1, X_2} E_{B_1, E_{A_{11, n}}} \exp \left[ itn \bar{X}'(A_{11, n} + B_1 + B_2)^{-1} \bar{X} \right]$$

$$\cdot \left( 1 - 2itn(\gamma Q_1 - 2it)^{-1} \bar{X}_1' A_{11, n}^{-1} \bar{X}_1 \right)^{-q/2}. \quad (3.10)$$
Because
\[
\left(1 - \frac{2i\tau}{\gamma Q_1}\right) (1 - 2i\tau n(\gamma Q_1 - 2i\tau)^{-1} \tilde{X}_1 A_{11,n}^{-1} \tilde{X}_1) = 1 - \frac{2i\tau}{\gamma Q_1} (1 + n\tilde{X}'_1 A_{11,n}^{-1} \tilde{X}_1),
\]
it follows that
\[
E \exp(i\tau N^{-1}T^2) = E Q_1, E_{X_1, X_2, E B_1, E A_{11,n}} \exp \left[ i\tau N \tilde{X}'(A_{11,n} + B_1 + B_2)^{-1} \tilde{X} \right]
\cdot \left(1 - \frac{2i\tau}{\gamma Q_1} (1 + n\tilde{X}'_1 A_{11,n}^{-1} \tilde{X}_1)\right)^{-g/2}
\equiv E Q_1, E_{X_1, X_2, E B_1, E A_{11,n}} \exp \left[ i\tau N \tilde{X}'(A_{11,n} + B_1 + B_2)^{-1} \tilde{X} \right]
\cdot E_{Q_2} \exp \left( \frac{i\tau Q_2}{\gamma Q_1} (1 + n\tilde{X}'_1 A_{11,n}^{-1} \tilde{X}_1)\right),
\]
where $Q_2 \sim \chi^2$ and $Q_2$ is independent of $\{Q_1, \tilde{X}_1, \tilde{X}_2, B_1, A_{11,n}\}$. Hence,
\[
E \exp(i\tau N^{-1}T^2) = E Q_1, E_{Q_2} \exp \left( \frac{i\tau Q_2}{\gamma Q_1} \right)
\cdot \left[ E_{X_1, X_2, E B_1, E A_{11,n}} \exp \left[ i\tau N \tilde{X}'(A_{11,n} + B_1 + B_2)^{-1} \tilde{X} \right]
\exp \left( \frac{i\tau Q_2}{\gamma Q_1} \tilde{X}'_1 A_{11,n}^{-1} \tilde{X}_1 \right) \right].
\]

We now apply the method of orthogonal invariance to simplify this characteristic function. For fixed $Q_1$ and $Q_2$, define the function
\[
f(\tilde{X}_1, \tilde{X}_2) = E_{B_1, E A_{11,n}} \exp \left[ i\tau N \tilde{X}'(A_{11,n} + B_1 + B_2)^{-1} \tilde{X} \right] \exp \left( \frac{i\tau Q_2}{\gamma Q_1} \tilde{X}'_1 A_{11,n}^{-1} \tilde{X}_1 \right).
\]
(Here, it must be kept in mind that $\tilde{X}$ and $B_2$ are functions of $\tilde{X}_1$ and $\tilde{X}_2$; see (2.2) and (2.16).) We now show that $f(\tilde{X}_1, \tilde{X}_2)$ is invariant under the transformation $(\tilde{X}_1, \tilde{X}_2) \rightarrow (H \tilde{X}_1, H \tilde{X}_2)$, where $H \in O(p)$, the set of $p \times p$ orthogonal matrices.

Suppose $(\tilde{X}_1, \tilde{X}_2)$ is transformed to $(H \tilde{X}_1, H \tilde{X}_2)$. Then by (2.2), $\tilde{X}$ is transformed to $H \tilde{X} H'$ and, by (2.16), $B_2$ is transformed to $H B_1 H'$. Since $A_{11,n} \sim W_p(n - 1, I_p)$ and $B_1 \sim W_p(N - n - 1, I_p)$ then $H A_{11,n} H' \equiv A_{11,n}$ and $H B_1 H' \equiv B_1$. Therefore,
\[
f(H \tilde{X}_1, H \tilde{X}_2) = E_{B_1, E A_{11,n}} \exp \left[ i\tau N (H \tilde{X})'(H A_{11,n} H' + H B_1 H' + H B_2 H')^{-1} H \tilde{X} \right]
\cdot \exp \left( \frac{i\tau Q_2}{\gamma Q_1} (H \tilde{X}_1)'(H A_{11,n} H')^{-1} H \tilde{X}_1 \right).
\]

On simplifying this expression, we find that all terms involving $H$ are cancelled, hence $f(H \tilde{X}_1, H \tilde{X}_2) \equiv f(\tilde{X}_1, \tilde{X}_2)$. Therefore, $f(\tilde{X}_1, \tilde{X}_2)$ is invariant under the transformation $(\tilde{X}_1, \tilde{X}_2) \rightarrow (H \tilde{X}_1, H \tilde{X}_2)$ and the simplification of (3.10) will follow from a judicious choice of $H$, as follows.

Because $(\tilde{X}_1 - \tilde{X}_2)(\tilde{X}_1 - \tilde{X}_2)'$ is of rank 1, there exists $H_1 \in O(p)$ such that
\[
H_1(\tilde{X}_1 - \tilde{X}_2)(\tilde{X}_1 - \tilde{X}_2)' H_1' = \|\tilde{X}_1 - \tilde{X}_2\|^2 e_{1,p} e_{1,p}',
\]
where $e_{1,p} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$. So we can replace $B_2 = n\tilde{\tau} (\tilde{X}_1 - \tilde{X}_2)(\tilde{X}_1 - \tilde{X}_2)'$ in (3.10) by $n\tilde{\tau} \|\tilde{X}_1 - \tilde{X}_2\|^2 H_1' e_{1,p} e_{1,p}' H_1$. Also, by replacing $(\tilde{X}_1, \tilde{X}_2)$ with $(H_1' \tilde{X}_1, H_1' \tilde{X}_2)$ then by
orthogonal invariance, (3.10) becomes

\[ E \exp(itN^{-1}T^2) = E_{Q_1} E_{Q_2} \exp \left( \frac{i\gamma Q_2}{Q_1} \right) \tilde{E}_{X_1,X_2} E_{B_1} \]

\[ \cdot E_{A_{1,n}} \exp \left[ i\gamma n \tilde{X}' (A_{11,n} + B_1 + n\bar{\tau} ||\tilde{X}_1 - \tilde{X}_2||^2 e_{1,p} e_{1,p}')^{-1} \tilde{X} \right] \]

\[ \cdot \exp \left( \frac{i\gamma Q_2}{Q_1} \tilde{X}' A_{11,n}^{-1} \tilde{X} \right). \]

(3.11)

We now make a second orthogonal transformation. There exists an orthogonal matrix \( H_2 \in O(p) \) with first row \( \tilde{X}_1' / ||\tilde{X}_1|| \) and for which the remaining rows of \( H_2 \) may be constructed using the standard Gram-Schmidt orthogonalization process. We transform \( \tilde{X}_1 \) to \( H_2 \tilde{X}_1 = ||\tilde{X}_1|| e_{1,p} \) and \( \tilde{X}_2 \) to \( H_2 \tilde{X}_2 = \alpha_1 e_{1,p} + \alpha_2 e_{2,p} \) where \( e_{1,p} \) is defined as before, \( e_{2,p} = (0,1,0,\ldots,0)' \in \mathbb{R}^p \), and \( \alpha_1 \) and \( \alpha_2 \) are such that

\[ \alpha_1 ||\tilde{X}_1|| = ||\tilde{X}_1|| \cos \theta, \quad \alpha_2^2 + \alpha_2 = ||\tilde{X}_2||. \]

Let \( \theta \) be the angle between \( \tilde{X}_1 \) and \( \tilde{X}_2 \), i.e., \( \cos \theta = \tilde{X}_1' \tilde{X}_2 / ||\tilde{X}_1|| \cdot ||\tilde{X}_2|| \). Then,

\[ \alpha_1 = \frac{||\tilde{X}_1'|| \tilde{X}_2}{||\tilde{X}_1'||} = ||\tilde{X}_2|| \cos \theta \]

and

\[ \alpha_2 = \left( ||\tilde{X}_2||^2 - \alpha_1^2 \right)^{1/2} = \left( ||\tilde{X}_2||^2 - ||\tilde{X}_2||^2 \cos^2 \theta \right)^{1/2} = ||\tilde{X}_2|| \sin \theta. \]

(3.13)

Therefore

\[ \tilde{X} = \tau \tilde{X}_1 + \bar{\tau} \tilde{X}_2 = \tau ||\tilde{X}_1|| e_{1,p} + \bar{\tau} ||\tilde{X}_2|| (\cos \theta e_{1,p} + \sin \theta e_{2,p}) \]

and

\[ ||\tilde{X}_1 - \tilde{X}_2||^2 = ||\tilde{X}_1||^2 + ||\tilde{X}_2||^2 - 2 ||\tilde{X}_1|| ||\tilde{X}_2|| \cos \theta. \]

Because \( n^{1/2} \tilde{X}_1 \sim N_p(0, I_p) \) and \( (N-n)^{1/2} \tilde{X}_2 \sim N_p(0, I_p) \), then \( \tilde{X}_1 \) and \( \tilde{X}_2 \) are orthogonally invariant random vectors. Therefore \( \tilde{X}_1' / ||\tilde{X}_1|| \) and \( \tilde{X}_2' / ||\tilde{X}_2|| \) are mutually independent and uniformly distributed on \( S^{p-1} \), the unit sphere in \( \mathbb{R}^p \). Hence \( \cos \theta \overset{\text{iid}}{\sim} U_1 U_2 \), where \( U_1 \) and \( U_2 \) are independent and uniformly distributed on \( S^{p-1} \). By Muirhead (1982, p. 38), we deduce that \( \tilde{X}_1, \tilde{X}_2, \) and \( \theta \) are mutually independent and also that \( \cos^2 \theta \sim \text{Beta}(1/2, (p-1)/2) \).

By (3.11),

\[ N^{-1/2} \tilde{X}_1 \sim N_p(0, I_p) \quad \text{and} \quad (N-n)^{-1/2} \tilde{X}_2 \sim N_p(0, I_p) \]

and therefore \( Q_3 \equiv (N-n) ||\tilde{X}_2||^2 \sim \chi_p^2 \). Similarly, \( (N-n)^{-1/2} \tilde{X}_2 \sim N_p(0, I_p) \) and therefore \( Q_4 \equiv (N-n) ||\tilde{X}_2||^2 \sim \chi_p^2 \). In addition, \( A_{11,n}, Q_1, Q_2, ||\tilde{X}_1||, ||\tilde{X}_2||, \theta, \) and \( B_1 \) are mutually independent. Thus, we have mutual independence between \( Q_1, Q_2, Q_3, Q_4, \theta, \)
We conclude, therefore, that

\[
N^{-1}T^2 \equiv \frac{Q_2}{Q_1} (1 + Q_3\beta_1 e_{1,p} A_{11,n}^{-1} e_{1,p}) \\
+ \left[ (Q_3^{1/2} + Q_4^{1/2} \cos \theta) e_{1,p} + Q_4^{1/2} e_{2,p} \sin \theta \right]^T \\
\cdot \left( A_{11,n} + B_1 + (\tau Q_3 + \tau Q_4 - 2(\tau \tau Q_3 Q_4)^{1/2} \cos \theta) e_{1,p} e_{1,p}^T \right)^{-1} \\
\cdot \left[ (Q_3^{1/2} + Q_4^{1/2} \cos \theta) e_{1,p} + Q_4^{1/2} \sin \theta e_{2,p} \right].
\]

(3.15)

We note that, up to this point, it is possible to set \( n = N \) in the foregoing, and we will do so later in Remark 3.7 to derive the distribution of the classical \( T^2 \)-statistic from (3.15).

Define \( L = (A_{11,n} + B_1)^{-1/2} A_{11,n} (A_{11,n} + B_1)^{1/2} \) and \( P = A_{11,n} + B_1 \). By Proposition 3.5(v), \( L \sim \text{Beta}_p((n-1)/2, (N-n-1)/2) \), \( P \sim \mathcal{W}_p(N-2, I_p) \), and \( L \) and \( P \) are mutually independent. Therefore, (3.15) reduces to

\[
N^{-1}T^2 \leq \frac{Q_2}{Q_1} (1 + Q_3(P^{-1/2}e_{1,p})^T L^{-1}(P^{-1/2}e_{1,p})) \\
+ \left[ (Q_3^{1/2} + Q_4^{1/2} \cos \theta) e_{1,p} + Q_4^{1/2} \sin \theta e_{2,p} \right]^T \\
\cdot \left( P + (\tau Q_3 + \tau Q_4 - 2(\tau \tau Q_3 Q_4)^{1/2} \cos \theta) e_{1,p} e_{1,p}^T \right)^{-1} \\
\cdot \left[ (Q_3^{1/2} + Q_4^{1/2} \cos \theta) e_{1,p} + Q_4^{1/2} \sin \theta e_{2,p} \right].
\]

(3.16)

Because the distribution of \( L \) is invariant under orthogonal transformations then we may replace \( L \) by \( H' LH \) for any \( H \in O(p) \). We therefore choose \( H \), judiciously, as an orthogonal matrix with first row \((P^{-1/2}e_{1,p})^T/\|P^{-1/2}e_{1,p}\|\); then, all other rows of \( H \) are orthogonal to \( P^{-1/2}e_{1,p} \), so

\[
HP^{-1/2}e_{1,p} = \|P^{-1/2}e_{1,p}\| e_{1,p}.
\]

Therefore, when \( L \mapsto H' LH \) we obtain

\[
(P^{-1/2}e_{1,p})^T L^{-1}(P^{-1/2}e_{1,p}) \mapsto (P^{-1/2}e_{1,p})^T (H' LH)^{-1}(P^{-1/2}e_{1,p}) \\
= (HP^{-1/2}e_{1,p})^T L^{-1}(HP^{-1/2}e_{1,p}) \\
= \|P^{-1/2}e_{1,p}\|^2 \cdot e_{1,p}^T L^{-1} e_{1,p} \\
= e_{1,p}^T P^{-1/2} e_{1,p}/\beta,
\]

(3.17)

where \( \beta \equiv 1/e_{1,p}^T L^{-1} e_{1,p} \). We note that \( \beta \sim \text{Beta}((n-p-2)/2, (N-n-1)/2) \); cf. Muirhead (1982, p. 120, Problem 3.22(e)).

Next, we set \( U = U_1 e_{1,1} + U_2 e_{1,2} \) where

\[
U_1 = Q_3^{1/2} + Q_4^{1/2} \cos \theta, \quad U_2 = Q_4^{1/2} \sin \theta,
\]

(3.18)

and let

\[
V = \bar{e} Q_3 + \tau Q_4 - 2(\tau \bar{e} Q_3 Q_4)^{1/2} \cos \theta.
\]

(3.19)

Applying (3.17), the representation (3.16) then becomes

\[
N^{-1}T^2 \leq \frac{Q_2}{Q_1} (1 + \frac{Q_3}{\beta} e_{1,p} P^{-1} e_{1,p}) + (U_1 e_{1,p} + U_2 e_{2,p})^T (P + V e_{1,p} e_{1,p}^T)^{-1} (U_1 e_{1,p} + U_2 e_{2,p}).
\]

(3.20)
By \((3.1)\),
\[
(P + Ve_{1,p}e'_{1,p})^{-1} = P^{-1} - \frac{V}{1 + Ve'_{1,p}P^{-1}e_{1,p}}P^{-1}e_{1,p}e'_{1,p}P^{-1};
\]
on inserting this expression in \((3.20)\) we obtain
\[
N^{-1}T^2 \leq \frac{Q_2}{\gamma Q_1} \left(1 + \frac{Q_3}{\beta} e'_{1,p}P^{-1}e_{1,p}\right) + (U_1e_{1,p} + U_2e_{2,p})'P^{-1}(U_1e_{1,p} + U_2e_{2,p})
- \frac{V}{1 + Ve'_{1,p}P^{-1}e_{1,p}}[e'_{1,p}P^{-1}(U_1e_{1,p} + U_2e_{2,p})]^2. \tag{3.21}
\]
Although this representation involves \(P\), only the random variables \(e'_{1,p}P^{-1}e_{1,p}\), \(e'_{1,p}P^{-1}e_{2,p}\), and \(e'_{2,p}P^{-1}e_{2,p}\) are germane. Therefore, to conclude the proof, we need to derive the joint distribution of those three terms.

To that end, recall that \(P \sim W_p(N - 2, I_p)\). Set \(M = (e_{1,p}, e_{2,p})'\) and let \(W = (MP^{-1}M')^{-1}\); noting that \(MM' = I_2\), it follows from Proposition \((3.5)\,\text{iv})\) that \(W \sim W_2(N - p, I_2)\). Because
\[
W^{-1} = MP^{-1}M' \equiv \begin{pmatrix} e'_{1,p}P^{-1}e_{1,p} & e'_{1,p}P^{-1}e_{2,p} \\ e'_{2,p}P^{-1}e_{1,p} & e'_{2,p}P^{-1}e_{2,p} \end{pmatrix}
\]
then \(e'_{1,2}W^{-1}e_{1,2} \equiv e'_{1,p}P^{-1}e_{1,p}\) and \(e'_{1,2}W^{-1}e_{2,2} \equiv e'_{1,p}P^{-1}e_{2,p}\). Therefore, \((3.21)\) reduces to
\[
N^{-1}T^2 \leq \frac{Q_2}{\gamma Q_1} \left(1 + \frac{Q_3}{\beta} e'_{1,2}W^{-1}e_{1,2}\right) + u'W^{-1}u - \frac{V}{1 + Ve'_{1,2}W^{-1}e_{1,2}}(e'_{1,2}W^{-1}u)^2, \tag{3.22}
\]
where \(W \sim W_2(N - p, I_2)\), and the proof now is complete. \(\blacksquare\)

**Remark 3.6.** It is a straightforward matter to simulate a \(2 \times 2\) Wishart matrix and hence to simulate the stochastic representation of the \(T^2\)-statistic derived above. Nevertheless, we now show that the representation \((3.22)\) can be reduced further to involve only scalar-valued, mutually independent random variables, and we use this result later to derive an upper bound on the distribution function of the \(T^2\)-statistic.

For \(i, j = 1, 2\) let \(\omega_{i,j} = e'_{i,2}W^{-1}e_{j,2}\), the \((i, j)\)th entry of \(W^{-1}\). Then \((3.22)\) becomes
\[
N^{-1}T^2 \leq \frac{Q_2}{\gamma Q_1} \left(1 + \frac{Q_3}{\beta} \omega_{1,1}\right) + U_1^2\omega_{1,1} + U_2^2\omega_{2,2} + 2U_1U_2\omega_{2,1}
- \frac{V}{1 + V\omega_{11}}(U_1\omega_{1,1} + U_2\omega_{2,1})^2. \tag{3.23}
\]
By the Bartlett decomposition of the Wishart matrix (Anderson, 2003, Chapter 7), we obtain
\[
W = \begin{pmatrix} t_{1,1} & 0 \\ -t_{2,1} & t_{2,2} \end{pmatrix}
\begin{pmatrix} t_{1,1} & -t_{1,1}t_{2,1} \\ 0 & t_{2,2} \end{pmatrix}
= \begin{pmatrix} t_{1,1}^2 & -t_{1,1}t_{2,1} \\ -t_{1,1}t_{2,1} & t_{2,1}^2 + t_{2,2}^2 \end{pmatrix},
\]
where \(t_{1,1}^2 \sim \chi^2_{N - p}\), \(t_{2,2}^2 \sim \chi^2_{N - p - 1}\), \(t_{2,1} \sim N(0, 1)\), and \(t_{1,1}\), \(t_{2,2}\), and \(t_{2,1}\) are mutually independent. By a direct calculation,
\[
W^{-1} = \frac{1}{t_{1,1}^2 t_{2,2}^2}
\begin{pmatrix} t_{2,1}^2 + t_{2,2}^2 & t_{1,1}t_{2,1} \\ t_{1,1}t_{2,1} & t_{1,1}^2 \end{pmatrix};
\]
where $Q_5 = \ell_{1,1}^2 \sim \chi_5^2$, $Q_6 = \ell_{2,2}^2 \sim \chi_6^2$, and $Q_7 = \ell_{1,1}^2 \sim \chi_1^2$. Substituting these $Q_j$ in (3.23) we obtain a stochastic representation that involves only the scalar-valued, mutually independent random variables $Q_1, Q_2, Q_3, Q_5, Q_6,$ and $Q_7$.

**Remark 3.7.** We can recover from the previous calculations the distribution of the classical $T^2$-statistic. Setting $n = N$ at (3.15) we obtain $\tau = \gamma = 1, \tau = 0, B_1 \equiv 0, Q_4 \equiv 0, and then (3.15) reduces to

$$n^{-1}T^2 \equiv \frac{Q_2}{Q_1}(1 + Q_3e'_{1,p}A_{11,n}^{-1}e_{1,p}) + Q_3e'_{1,p}A_{11,n}^{-1}e_{1,p}.$$ 

Because $A_{11,n} \sim W_p(n - 1, I_p)$ then, by Proposition 3.5(iv), $Q_0 \equiv 1/e'_{1,p}A_{11,n}^{-1}e_{1,p} \sim \chi^2_{n-p}$; hence,

$$n^{-1}T^2 \equiv \frac{Q_2}{Q_1}(1 + \frac{Q_3}{Q_0}) + \frac{Q_3}{Q_0} = \left(1 + \frac{Q_2}{Q_1}\right)\left(1 + \frac{Q_3}{Q_0}\right) - 1,$$

where $Q_0 \sim \chi^2_{n-p}$, $Q_3 \sim \chi^2_{p}$, $Q_2 \sim \chi^2_{q}$, and $Q_1 \sim \chi^2_{n-p-q}$. Noting that

$$\left(1 + \frac{Q_2}{Q_1}\right)^{-1} = \frac{Q_1}{Q_1 + Q_2} \sim Beta\left((n - p - q)/2, q/2\right),$$

and, similarly, that

$$\left(1 + \frac{Q_3}{Q_0}\right)^{-1} = \frac{Q_0}{Q_0 + Q_3} \sim Beta\left((n - p)/2, p/2\right),$$

it follows from the standard method-of-moments approach (Anderson, 2003) to the distribution of products of independent beta-distributed random variables that

$$\frac{Q_1}{Q_1 + Q_2} \cdot \frac{Q_0}{Q_0 + Q_3} \equiv \frac{R_1}{R_1 + R_2},$$

where $R_1 \sim \chi^2_{n-p-q}$, $R_2 \sim \chi^2_{p+q}$, and $R_1$ and $R_2$ are mutually independent. Therefore,

$$n^{-1}T^2 \equiv \left(1 + \frac{Q_2}{Q_1}\right)\left(1 + \frac{Q_3}{Q_0}\right) - 1 \equiv \frac{R_1 + R_2}{R_1} - 1 = \frac{R_2}{R_1},$$

and so we have obtained the well-known result for the distribution of the classical Hotelling’s $T^2$-statistic in terms of a ratio of independent chi-squared random variables.
4 Probability inequalities for the $T^2$-statistic

The exact distribution of the $T^2$-statistic, although completely amenable to numerical simulation, is analytically non-trivial. Therefore, it may be useful to find simpler upper and lower bounds on its distribution.

To that end, Chang and Richards (2009) found upper and lower bounds for the distribution function of the $T^2$-statistic. Because our bounds are based on a stochastic representation of the exact distribution of the $T^2$-statistic, it can be expected that our bounds will lead to more precise confidence regions than those given by Chang and Richards (2009).

We will use standard notation for stochastic orderings: If $X$ and $Y$ are scalar-valued random variables, then $X \preceq Y$ or $Y \succeq X$ denotes that $P(X \geq t) \geq P(Y \geq t)$, $t \in \mathbb{R}$.

**Proposition 4.1.** Let $Q_1 \sim \chi^2_{n-p-q}$, $Q_2 \sim \chi^2_q$, $Q_3 \sim \chi^2_p$, $Q_8 \sim \chi^2_{N-p-1}$, and $\beta \sim \text{Beta}((n-p-2)/2, (N-n-1)/2)$ be mutually independent. Then,

$$N^{-1}T^2 \preceq \frac{Q_2}{\gamma Q_1} \left( 1 + \frac{Q_3}{Q_8 \beta} \right).$$

(4.1)

**Proof.** By omitting from the right-hand side of the stochastic representation (3.20) the positive quadratic form $(U_1 e_{1,p} + U_2 e_{2,p})'(P + V e_{1,p} e_{1,p}')^{-1}(U_1 e_{1,p} + U_2 e_{2,p})$, we obtain

$$N^{-1}T^2 \succeq \frac{Q_2}{\gamma Q_1} \left( 1 + \frac{Q_3}{\beta} e_{1,p}' P^{-1} e_{1,p} \right).$$

By Proposition 3.5(iv), $Q_8 \equiv 1/e_{1,p}' P^{-1} e_{1,p} \sim \chi^2_{N-p-1}$, and so we obtain (4.1). $\square$

In applications, we will use the above stochastic lower bound in an equivalent formulation in terms of an upper bound on the exact cumulative distribution function of the $T^2$-statistic: For $t \geq 0$,

$$P(T^2 \leq t) \leq P \left( \frac{Q_2}{Q_1} \left( 1 + \frac{Q_3}{Q_8 \beta} \right) \leq \frac{\gamma t}{N} \right).$$

(4.2)

We remark also that a weaker stochastic inequality, $T^2 \preceq NQ_2/\gamma Q_1$, which follows immediately from (4.1), leads to the result,

$$P(T^2 \leq t) \leq P \left( \frac{Q_2}{Q_1} \leq \frac{\gamma t}{N} \right),$$

and this bounds the distribution function of $T^2$ by the distribution function of an $F$-statistic.

The proof of the following result shows, in contrast to the derivation of a stochastic lower bound, that it is more complex to obtain a stochastic upper bound for $T^2$. Throughout the proof, the random variables $Q_1, \ldots, Q_4$, $\beta$, $\omega_{1,1}$, $\omega_{2,2}$, and $\omega_{2,1}$ are those encountered in Theorem 3.1 and in (3.24).

**Theorem 4.2.** Let $Q_1 \sim \chi^2_{n-p-q}$, $Q_2 \sim \chi^2_q$, $Q_3 \sim \chi^2_p$, $Q_4 \sim \chi^2_p$, $Q_5 \sim \chi^2_{N-p}$, $Q_6 \sim \chi^2_{N-p}$, $Q_7 \sim \chi^2_1$, and $\beta \sim \text{Beta}((n-p-2)/2, (N-n-1)/2)$ be mutually independent, and let $\omega_{2,2}$, and $\omega_{2,1}$ be defined as in (3.24). Then,

$$N^{-1}T^2 \preceq \frac{Q_2}{\gamma Q_1} \left( 1 + \frac{Q_3}{\beta} \omega_{1,1} \right) + Q_3 \omega_{1,1} + \frac{1}{2} Q_4 (\omega_{2,1} + \omega_{2,2})$$

$$+ Q_4 \left( \frac{1}{4} (\omega_{1,1} - \omega_{2,2})^2 + \omega_{2,1}^2 \right)^{1/2} + 2(Q_3 Q_4)^{1/2}(\omega_{1,1}^2 + \omega_{2,1}^2)^{1/2}. \quad (4.3)$$
Proof. Consider the random variable $V$ in (3.19). Because $\cos \theta \leq 1$, it follows that
\[
V = \bar{\tau}Q_3 + \tau Q_4 - 2(\bar{\tau}Q_3Q_4)^{1/2}\cos \theta \\
\geq \bar{\tau}Q_3 + \tau Q_4 - 2(\bar{\tau}Q_3Q_4)^{1/2} \\
= ((\bar{\tau}Q_3)^{1/2} - (\tau Q_4)^{1/2})^2 \geq 0.
\]

Therefore, by the stochastic representation for $T^2$ at (3.22), we have
\[
N^{-1}T^2 \leq \frac{Q_2}{\gamma Q_1} \left(1 + \frac{Q_3}{\beta} \epsilon_{1,2}' \epsilon_{1,2} \right) + (U_1e_{1,2} + U_2e_{2,2})'W^{-1}(U_1e_{1,2} + U_2e_{2,2}).
\] (4.4)

Recalling from Remark 3.6 that $\omega_{i,j} = \epsilon_{i,2}'W^{-1}\epsilon_{j,2}$, and applying the definition of $U_1$ and $U_2$ in (3.18), we obtain
\[
(U_1e_{1,2} + U_2e_{2,2})'W^{-1}(U_1e_{1,2} + U_2e_{2,2}) \equiv U_1^2\omega_{1,1} + U_2^2\omega_{2,2} + 2U_1U_2\omega_{2,1} \\
= (Q_3^{1/2} + Q_4^{1/2}\cos \theta)^2\omega_{1,1} + Q_4\omega_{2,2}\sin^2 \theta + 2(Q_3^{1/2} + Q_4^{1/2}\cos \theta)Q_4^{1/2}\omega_{1,2}\sin \theta.
\]

Substituting this result in (4.4) and applying the trigonometric identities $\sin 2\theta = 2\cos \theta \sin \theta$ and $\cos 2\theta = 2\cos^2 \theta - 1 = 1 - 2\sin^2 \theta$ we obtain
\[
N^{-1}T^2 \leq \frac{Q_2}{\gamma Q_1} \left(1 + \frac{Q_3}{\beta} \omega_{1,1} \right) + Q_3\omega_{1,1} + \frac{1}{2}Q_4(\omega_{2,1} + \omega_{2,2}) + f(\theta),
\] (4.5)

where
\[
f(\theta) = Q_4\left(\frac{1}{2}(\omega_{1,1} - \omega_{2,2})\cos 2\theta + \omega_{2,2}\sin 2\theta \right) + 2(Q_3Q_4)^{1/2}(\omega_{1,1}\cos \theta + \omega_{2,1}\sin \theta).
\]

Applying the well-known inequality,
\[
a \cos \theta + b \sin \theta \leq (a^2 + b^2)^{1/2}
\]
for all $\theta$, we obtain
\[
f(\theta) \leq Q_4\left(\frac{1}{2}(\omega_{1,1} - \omega_{2,2})^2 + \omega_{2,2}^2 \right)^{1/2} + 2(Q_3Q_4)^{1/2}(\omega_{1,1}^2 + \omega_{2,1}^2)^{1/2}.
\]

Replacing $f(\theta)$ in (4.5) by this upper bound, we obtain (4.3). $\square$

Because the stochastic inequalities obtained here are derived from the exact distribution, it is to be expected that the resulting bounds on the cumulative distribution function of the $T^2$-statistic will be sharper than previously-obtained bounds. Using the statistical package R (R Development Core Team, 2005), we simulated the bounds arising from Proposition 4.1 and Theorem 4.2, the bounds obtained by Chang and Richards (2009), and the exact cumulative distribution function for various values of $p$, $q$, $n$, and $N$. According to these numerical investigations, the bounds obtained from Proposition 4.1 and Theorem 4.2 often are sharper than previously obtained bounds; consequently, those bounds often will lead to less conservative ellipsoidal confidence regions for $\mu$.

We have provided in Figures 1 and 2 graphs in two instances, each over the range $0 \leq t \leq 100$, of the exact cumulative distribution function, the bounds obtained from Proposition 4.1 and Theorem 4.2, and the bounds due to Chang and Richards (2009). It is evident from these figures that the bounds derived from Proposition 4.1 and Theorem 4.2 perform better overall, being in all cases either closer to the exact distribution function or similar to previously obtained bounds. These simulations indicate also that, as $t \to \infty$, the bounds in (4.1) and (4.3) converge more rapidly to the exact cumulative distribution function than the bounds obtained by Chang and Richards (2009).
5 Inference for linear combinations of the components of $\mu$

In the classical setting where the data are complete, the $T^2$-statistic has been used to construct ellipsoidal confidence regions for linear combinations, $v'\mu$ where $v \in \mathbb{R}^{p+q}$, of the components of $\mu$; cf. Anderson (2003, Section 5.3.3). Similar results have also been obtained in the monotone incomplete data setting for the likelihood ratio test; cf. Little (1976), Srivastava (2002).

Proceeding as in the classical case (Anderson, 2003, loc. cit.), we apply the generalized Cauchy-Schwarz inequality,

$$|v'(\hat{\mu} - \mu)|^2 \leq v'\widehat{\text{Cov}}(\hat{\mu})v \cdot (\hat{\mu} - \mu)'(\text{Cov}(\hat{\mu}))^{-1}(\hat{\mu} - \mu) = v'\widehat{\text{Cov}}(\hat{\mu})v \cdot T^2,$$

To obtain

$$\sqrt{T^2} \geq \frac{|v'\hat{\mu} - v'\mu|}{(v'\widehat{\text{Cov}}(\hat{\mu})v)^{1/2}}.$$

Solving this inequality for $v'\mu$, it follows that if $T^2_\alpha$ is a 100(1-$\alpha$)% percentage point for the
The $T^2$-statistic then $\mu$ satisfies for all $v \in \mathbb{R}^{p+q}$ the simultaneous inequalities,
\[ v'\hat{\mu} - \sqrt{T^2_\alpha} \cdot (v'\text{Cov}(\hat{\mu})v)^{1/2} < v'\mu < v'\hat{\mu} + \sqrt{T^2_\alpha} \cdot (v'\text{Cov}(\hat{\mu})v)^{1/2}, \]
with confidence at least $1 - \alpha$. To apply this result in practice we may use either simulated values of the exact distribution function, as obtained from Theorem 3.1, or bounds for the percentage points $T^2_\alpha$ that can be derived from the results in Section 4.

### 5.1 The Pennsylvania cholesterol data set

A well-known example of a monotone incomplete data set was provided by Ryan and Joiner (1985). Researchers at a medical center in Pennsylvania monitored the cholesterol levels of 28 patients over a period of 14 days immediately following a heart attack. All 28 patients subsequently had their cholesterol levels measured on day 2 and on day 4 days of follow-up, and 19 patients were measured on day 14. This data set is an example of a longitudinal study; nevertheless, for the purposes of providing a numerical example to illustrate applications our results, we suppose that the data consist of mutually independent observations satisfying the MCAR assumption.
We first assess the assumption of multivariate normality for this data set. Romer (2009) constructed quantile-quantile, or Q-Q, plots for each characteristic and deduced that the assumption of normality in each case is reasonable. Yamada, et al. (2012) recently applied to this data a statistic for testing kurtosis with monotone incomplete data; that kurtosis statistic also failed to reject the hypothesis of multivariate normality of the full data set at the 5% level of significance.

We simulated 1,000 values of the $T^2$-statistic and our upper and lower bounds in Section 4 to estimate the 97.5th percentile of each distribution. We then calculated 95% confidence intervals for various linear combinations of $\mu$, viz., the mean cholesterol levels for Day 2, Day 4, and Day 14, and the difference in mean cholesterol levels between Days 14 and 2. As expected, in Table 1 the lower bound provides a wider and more-conservative confidence interval, and the upper bound provides a narrower and less-conservative, confidence interval.

**Acknowledgments.** We are grateful to Javier Rojo for the invitation to participate in the Lehmann Symposium.

**References**


