THE STATISTICAL ANALYSIS OF MONOTONE INCOMPLETE MULTIVARIATE NORMAL DATA

A Dissertation in
Statistics
by
Megan M. Romer

© 2009 Megan M. Romer

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2009
The dissertation of Megan M. Romer was reviewed and approved* by the following:

Steven F. Arnold
Professor of Statistics
Chair of Committee

Vernon M. Chinchilli
Professor of Statistics
Distinguished Professor and Chair of Public Health Sciences

Diane M. Henderson
Professor of Mathematics

Donald St. P. Richards
Professor of Statistics
Dissertation Adviser

Bruce G. Lindsay
Willaman Professor of Statistics
Head of the Department of Statistics

* Signatures on file in the Graduate School.
Abstract

We consider problems in finite-sample inference with monotone incomplete data drawn from $N_d(\mu, \Sigma)$, a multivariate normal population with mean $\mu$ and covariance matrix $\Sigma$. In the case of two-step, monotone incomplete data, we show that $\hat{\mu}$ and $\hat{\Sigma}$, the maximum likelihood estimators of $\mu$ and $\Sigma$, respectively, are equivariant and obtain a new derivation of a stochastic representation for $\hat{\mu}$. Our new derivation allows us to identify explicitly in terms of the data the independent random variables that arise in that stochastic representation. Again, in the case of two-step, monotone incomplete data, we derive a stochastic representation for the exact distribution of a generalization of Hotelling’s $T^2$, and therefore obtain ellipsoidal confidence regions for $\mu$. We then derive probability inequalities for the cumulative distribution function of $T^2$. We apply these results to construct confidence regions for linear combinations of $\mu$, and provide a numerical example in which we analyze a data set consisting of cholesterol measurements on a group of Pennsylvania hospital patients. In the case of three-step, monotone incomplete data, we examine the independence properties and joint distribution of subvectors of $\hat{\mu}$, the maximum likelihood estimator of $\mu$. In our examination of the joint distribution of $\hat{\mu}$, we first establish that $\hat{\mu}$ is equivariant and then identify the distribution of $\hat{\mu}$ up to a certain set of conditioning variables.
# Table of Contents

List of Tables ........................................ v

List of Figures ........................................ vi

Acknowledgments ..................................... vii

Chapter 1. Introduction ............................... 1

Chapter 2. Preliminaries ............................. 9
  2.1 Some matrix algebra ............................ 9
  2.2 Some multivariate distributions ............ 11
     2.2.1 The matrix normal distribution ........ 11
     2.2.2 The Wishart distribution ............... 12
     2.2.3 The multivariate beta distribution .... 13
     2.2.4 The noncentral $\chi^2$ distribution .... 13
  2.3 Some properties of these distributions .... 13

Chapter 3. Two-step Monotone Incomplete
           Multivariate Normal Data .................. 17
  3.1 Notation and maximum likelihood estimators .. 17
  3.2 A new derivation of an exact stochastic representation for $\hat{\mu}$ ... 19
  3.3 An exact stochastic representation for the $T^2$-statistic ... 31
  3.4 Probability inequalities for the $T^2$-statistic ... 49
  3.5 Applications of the $T^2$-statistic .......... 53
     3.5.1 Simultaneous confidence intervals for linear functions of $\hat{\mu}$ ... 53
     3.5.2 Ellipsoidal prediction regions for future observations ... 54
     3.5.3 Analysis of the Pennsylvania cholesterol data ... 65

Chapter 4. Three-Step Monotone Incomplete
           Multivariate Normal Data .................. 72
  4.1 Notation and maximum likelihood estimators .. 72
  4.2 Correlation properties of $\hat{\mu}_1, \hat{\mu}_2,$ and $\hat{\mu}_3$ ..... 77
  4.3 The distribution of $\hat{\mu}_3$ .................. 86
  4.4 The covariance matrix of $\hat{\mu}_3$ .......... 109

Chapter 5. Concluding Remarks .................... 119

Bibliography ........................................... 121
List of Tables

1.1 The Pennsylvania Cholesterol Data ........................................... 2

3.1 95% Confidence Interval for Mean Cholesterol Levels .................. 66
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 2; q = 1; n = 19; N = 28$)</td>
<td>68</td>
</tr>
<tr>
<td>3.2</td>
<td>Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 2; q = 2; n = 10; N = 15$)</td>
<td>68</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 3; q = 3; n = 15; N = 20$)</td>
<td>69</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 4; q = 4; n = 15; N = 20$)</td>
<td>69</td>
</tr>
<tr>
<td>3.5</td>
<td>Simulated Cumulative Distribution Function of the $T^2$-statistic: Approximation ($p = 2; q = 2; n = 10; N = 15$)</td>
<td>70</td>
</tr>
<tr>
<td>3.6</td>
<td>Simulated Cumulative Distribution Function of the $T^2$-statistic: Approximation ($p = 4; q = 4; n = 15; N = 20$)</td>
<td>70</td>
</tr>
<tr>
<td>3.7</td>
<td>The Pennsylvania Cholesterol Data: Assessment of Multivariate Normality</td>
<td>71</td>
</tr>
</tbody>
</table>
Acknowledgments

To Donald, Jenn, Chuck, and Baby:
For reasons unique to each of you,
I thank you.
Chapter 1

Introduction

In all areas of research, incomplete data sets are ubiquitous. To describe a few of the many situations in which such data occur: in clinical trials, participants often drop out of studies; in engineering research, machines often fail; in scientific laboratories, beakers often break; in astronomy, cloudy weather interferes with data collection. Because of the omnipresence of data sets with missing values, there now exists an extensive literature on the analysis of such data. We refer to Giri [13], Johnson and Wichern [16], Little and Rubin [23], Schafer [30], and Srivastava [31] for treatments of statistical inference with incomplete data and a wide range of applications including astronomy, biology, clinical trials, and sample surveys.

A well-known example of an incomplete data set was provided by Ryan and Joiner [29]. 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 at 2 and at 4 days of follow-up, and 19 patients were measured on day 14. The data are displayed in Table I.1.
Table 1.1 The Pennsylvania Cholesterol Data

<table>
<thead>
<tr>
<th>Day</th>
<th>Cholesterol Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>160 270 186 282 276 278 318 236 200 266 210 226 272</td>
</tr>
<tr>
<td>4</td>
<td>146 218 190 186 220 248 258 242 202 182 236 214 238 276</td>
</tr>
<tr>
<td>14</td>
<td>142 156 168 182 188 198 200 204 214 216 236 242 248 256</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day</th>
<th>Cholesterol Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>288 294 234 244 360 236 142 280 242 206 282 224 280 288</td>
</tr>
<tr>
<td>4</td>
<td>248 240 220 270 352 234 116 200 288 244 294 200 218 278</td>
</tr>
<tr>
<td>14</td>
<td>256 264 264 280 294 * * * * * * * * *</td>
</tr>
</tbody>
</table>

The cholesterol data provide an example of a two-step monotone incomplete pattern. A random sample $X_\alpha = (X_{1\alpha}, \ldots, X_{d\alpha})'$, $\alpha = 1, \ldots, N$, from a $d$-dimensional multivariate population is said to be monotone incomplete if whenever $X_{l\alpha}$ is missing then, for all $j > l$ and $\beta > \alpha$, $X_{j\beta}$ also is missing. More generally, we can conceive of $k$-step monotone incomplete data sets as visualized in Figure 1.
A random vector $X \in \mathbb{R}^d$ is said to have a multivariate normal distribution with mean $\mu \in \mathbb{R}^d$ and positive definite (symmetric) $d \times d$ covariance matrix $\Sigma$, denoted $X \sim N_d(\mu, \Sigma)$, if its density function is

$$
(2\pi)^{-d/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right],
$$

(1.0.1)

$x \in \mathbb{R}^d$. The multivariate normal distribution is undoubtedly the most important distribution in statistics, and many extensive, classical treatments of statistical inference for multivariate normally distributed data are available, e.g. Anderson [2], Eaton [9], and Muirhead [25].

In this dissertation, we consider problems arising in the statistical analysis of monotone incomplete data drawn from a multivariate normal population. In general, $\mu$ and $\Sigma$ are unknown, so it is of interest to perform statistical inference for them. Throughout the dissertation, we focus on maximum likelihood estimators of $\mu$ and $\Sigma$. When the data are monotone incomplete, the maximum likelihood estimators for $\mu$ and $\Sigma$, denoted by $\hat{\mu}$ and $\hat{\Sigma}$, respectively, are well-known [1], [3], [15], [24]; however, the exact distributions of $\hat{\mu}$ and $\hat{\Sigma}$ are far more complicated than in the case of complete data sets. In fact, until recently, neither distribution was known in the case of two-step data and, for general $k$, both distributions still remain unknown.

Morrison [24] and Kanda and Fujikoshi [17] examined the exact means and variances of $\hat{\mu}$ in great detail. Kanda and Fujikoshi also went on to find asymptotic results. Statistical inference for $\mu$ was first developed without knowledge of the exact distribution of $\hat{\mu}$. There has been much research in the area of hypothesis testing for $\mu$ e.g.
Bhargava [4], [5], Eaton and Kariya [10], Giri [13], Hao and Krishnamoorthy [14]. A drawback of hypothesis testing alone is that each element of $\mu$ must be specified in the null hypothesis, therefore it is preferable that the results of a hypothesis test be accompanied by a confidence region. Confidence regions for $\mu$ may be based on the likelihood ratio test statistic; however the resulting regions are not ellipsoidal and, in fact, have rather counterintuitive shapes.

For the case in which the data are two-step monotone incomplete, the first derivation of ellipsoidal confidence regions for $\mu$ was obtained by Krishnamoorthy and Pannala [20]. The estimator given in [20] of $\text{Cov}(\hat{\mu})$, the covariance matrix of $\hat{\mu}$, is, in retrospect, not identical to $\tilde{\text{Cov}}(\hat{\mu})$, the maximum likelihood estimator of $\text{Cov}(\hat{\mu})$; however, it is asymptotically equivalent. Therefore we denote their estimated covariance matrix by $\tilde{\text{Cov}}(\hat{\mu})$. Krishnamoorthy and Pannala [20] obtained ellipsoidal confidence regions for $\mu$ by means of

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

a generalization of the classical Hotelling’s $T^2$-statistic. Krishnamoorthy and Pannala approximated the distribution of $\tilde{T}^2$ with an $F$-distribution and we shall show that, for small dimensions, their approximation is very close to the exact distribution. Moreover, they also extended this method to general $k$-step monotone incomplete data.

Chang and Richards [6], [7] derived stochastic representations for the exact distributions of $\hat{\mu}$ and $\hat{\Sigma}$ in the case of two-step monotone incomplete data. These stochastic representations are important because the asymptotic distributions hold only for large sample sizes, which often are unavailable, especially for high-dimensional data. Chang
and Richards [6] also derived \( \hat{\text{Cov}}(\hat{\mu}) \), the maximum likelihood estimator of \( \text{Cov}(\hat{\mu}) \) and therefore generalized the classical Hotelling’s \( T^2 \)-statistic to

\[
T^2 = (\hat{\mu} - \mu)'\hat{\text{Cov}}(\hat{\mu})^{-1}(\hat{\mu} - \mu).
\]

(1.0.2)

Chang and Richards based their ellipsoidal confidence regions for \( \mu \) on probability inequalities for \( T^2 \).

In this dissertation, we begin by considering the case of two-step monotone incomplete data, \( i.e., \ k = 2 \). Let \( X \in \mathbb{R}^p \) and \( Y \in \mathbb{R}^q \). Suppose \( \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q}(\mu, \Sigma) \). In the two-step setting, we observe \( n \) mutually independent observations on \( \begin{pmatrix} X \\ Y \end{pmatrix} \) and an additional \( N - n \) independent observations on \( X \) only. Therefore, the data are mutually independent vectors of the form

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

In this setting, we provide an alternative derivation of the stochastic representation for \( \hat{\mu} \) found by Chang and Richards [6]. We first prove that both \( \hat{\mu} \) and \( \hat{\Sigma} \) are equivariant, a result that greatly simplifies the examination of both statistics as we then may assume \( \mu = 0 \) and \( \Sigma = I_{p+q} \), the identity matrix. This new derivation is based on the conditional distribution of the incomplete data given the complete data and it identifies explicitly the independent random variables that appear in the stochastic representation.
We derive next a stochastic representation for the exact distribution of the $T^2$-statistic. We first prove an invariance property of the $T^2$-statistic, and we rely heavily on that property in our derivation of the exact stochastic representation. This stochastic representation allows us to construct ellipsoidal confidence regions for $\mu$, or to perform related tests of hypotheses, with exact confidence levels or levels of significance, respectively. As our confidence regions are based on the exact distribution of the $T^2$-statistic, it follows that our confidence regions are of exact level and hence are preferable to those of Chang and Richards [6] or Krishnamoorthy and Pannala [20]. From this stochastic representation, we also 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. Because our stochastic representation is quite complex, we also provide in Chapter 3 probability inequalities for the $T^2$-statistic. The last application we explore is the construction of prediction regions for new observations; although we are unable to find the exact distribution for our proposed statistic, we are confident that $F$-approximations can be obtained.

In the second part of the dissertation, which appears in Chapter 4, we address several research problems related to three-step monotone incomplete data, i.e., $k = 3$. Let $X_1 \in \mathbb{R}^{p_1}, X_2 \in \mathbb{R}^{p_2}$, and $X_3 \in \mathbb{R}^{p_3}$. Suppose $(X'_1, X'_2, X'_3)' \sim N_{p_1+p_2+p_3}(\mu, \Sigma)$. In the three-step setting, we observe $n_1$ mutually independent observations on $(X'_1, X'_2, X'_3)'$, 


an additional $n_2$ independent observations on $(X_1', X_2')'$, and an additional $n_3$ observations on $X_1$ only. Therefore, the data are mutually independent vectors of the form

\[
\begin{pmatrix}
X_{1,1} \\
X_{2,1} \\
X_{3,1}
\end{pmatrix}
\cdots
\begin{pmatrix}
X_{1,n_1} \\
X_{2,n_1} \\
X_{3,n_1}
\end{pmatrix}
\begin{pmatrix}
X_{1,n_1+1} \\
X_{2,n_1+1} \\
X_{1,n_1+n_2}
\end{pmatrix}
\cdots
\begin{pmatrix}
X_{1,n_1+n_2} \\
X_{1,n_1+n_2+1} \\
X_{1,n_1+n_2+n_3}
\end{pmatrix}
\]

We partition $\hat{\mu}$ in similar fashion into three subvectors, $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$. In this setting, we establish independence between $\hat{\mu}_1$ and $\{\hat{\mu}_2, \hat{\mu}_3\}$. Furthermore, we prove that when $\Sigma = I_d$, $d = p_1 + p_2 + p_3$, these subvectors are pairwise uncorrelated; therefore, although we have not established independence between $\hat{\mu}_2$ and $\hat{\mu}_3$, we have shown that they are uncorrelated for $\Sigma = I_d$. We establish the equivariance of $\hat{\mu}$ under a certain group of transformations and provide an extension of our alternative derivation for the distribution of $\hat{\mu}$ from two-step to three-step monotone incomplete data. Although we have not been able to find a joint stochastic representation for $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$, we believe that we have identified the six random variables, whose joint distribution is unknown, that form the basis of that representation.

Throughout the dissertation, we have also made an assumption on the process that generates the incomplete data. There are three main underlying processes that describe how observations are missing: *missing at random*, *missing completely at random*, and *not missing at random* [28]. Because of the independence structure we have assumed, we have also implicitly assumed our data is missing completely at random, that is, there is no reason or order as to why any one unit would be missing an observation as opposed
to another. Readers are referred to [23] and [30] for further discussion on the types of missingness.
Chapter 2

Preliminaries

2.1 Some matrix algebra

Let $X$ be a $p \times q$ matrix and let $\text{vec}(X)$ denote the $pq \times 1$ column vector formed by stacking the columns of $X$. Let $C > 0$ and $D > 0$ be $p \times p$ and $q \times q$ matrices, respectively, where “$>$ 0” denotes that the matrices are positive definite (and symmetric). We denote the inverse, trace, and determinant of $C$ by $C^{-1}$, $\text{tr}(C)$, and $|C|$, respectively. Also, we denote the Kronecker product of $C$ and $D$ by $C \otimes D$. Muirhead [25] provides a number of useful properties of these matrix operations and we collect together some of their properties in the following proposition.

Proposition 2.1.1. (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 the following equalities hold:

$$\text{vec}(BAX) = (X' \otimes B)\text{vec}(A)$$

$$\text{tr}(BAX) = (\text{vec}(B'))'(I \otimes A)\text{vec}(X)$$

$$\text{tr}(AX'CXB) = (\text{vec}(X'))'(BA \otimes C')\text{vec}(X).$$
(iv) Let $A$ be $p \times q$ and $B$ be $q \times p$ and $P = C + ADB$. The following expression for $P^{-1}$ is known as Woodbury’s formula:

$$P^{-1} = C^{-1} - C^{-1} AD (D + DBC^{-1} D)^{-1} DBC^{-1}.$$  

(v) Let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of $C$. Then there exists an orthogonal $p \times p$ matrix $H$ such that

$$H'CH = \text{diag}(\lambda_1, \ldots, \lambda_p).$$

Let the $(p + q) \times (p + q)$ matrix $M > 0$ be partitioned into $p$ and $q$ rows and columns, i.e.,

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where $M_{11}$ is $p \times p$, $M_{12} = M_{21}'$ is $p \times q$, and $M_{22}$ is $q \times q$. A large portion of our research involves the partitioning of data into blocks with similar patterns of missingness. We therefore make much use of the well-known Schur complement, that is $M_{22}^{-1} = M_{22} - M_{21} M_{11}^{-1} M_{12}$. There are a number of results involving Schur complements that we will need and so we list them in the following lemma, see Anderson [2] and Muirhead [25].

**Proposition 2.1.2.** Partition the positive definite matrix $M$ into $p$ and $q$ rows and columns, as above. Then

(i) The partial Iwasawa coordinates of $M$ are \{\(M_{11}^{-1} M_{12}^{-1} M_{22}\)\}, and

$$M = \begin{pmatrix} I_p & 0 \\ M_{21}' M_{11}^{-1} & I_q \end{pmatrix} \begin{pmatrix} M_{11} & 0 \\ 0 & M_{22}^{-1} \end{pmatrix} \begin{pmatrix} I_p & M_{11}^{-1} M_{12} \\ 0 & I_q \end{pmatrix}.$$  

(2.1.3)
Further,

$$M^{-1} = \begin{pmatrix} I_p & -M_{11}^{-1}M_{12} \\ 0 & I_q \end{pmatrix} \cdot \begin{pmatrix} M_{11}^{-1} & 0 \\ 0 & M_{22}^{-1} \end{pmatrix} \cdot \begin{pmatrix} I_p & 0 \\ -M_{21}M_{11}^{-1} & I_q \end{pmatrix}.$$ 

(ii) Let \( x = (x_1', x_2')' \), where \( x_1 \in \mathbb{R}^p, x_2 \in \mathbb{R}^q \). Then,

$$x'M^{-1}x = (x_1 - M_{12}M_{22}^{-1}x_2)'M_{11}^{-1}(x_1 - M_{12}M_{22}^{-1}x_2) + x_2'M_{22}^{-1}x_2. \quad (2.1.4)$$

(iii) Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

then,

$$Y|X \sim N_q(\mu_2 + \Sigma_{21} \Sigma_{11}^{-1}(X_j - \mu_1), \Sigma_{22}). \quad (2.1.5)$$

Finally, we define \( e_1 = (1, 0, \ldots, 0)' \) to be the vector whose first element is one followed by zeros for an arbitrary length that will be obvious when used in the context of the problem. Similarly, we define \( e_2 = (0, 1, 0, \ldots, 0)' \).

2.2 Some multivariate distributions

2.2.1 The matrix normal distribution

Let \( M \) be a \( p \times q \) matrix, \( C \) and \( D \) be \( p \times p \) and \( q \times q \) positive definite matrices, respectively. A \( p \times q \) random matrix \( W_{12} \) has a matrix normal distribution with mean \( M \) and covariance parameter \( C \otimes D \), denoted \( W_{12} \sim N(M, C \otimes D) \), if the probability
density function of $W_{12}$ is

$$(2\pi)^{-pq/2} |C|^{-q/2} |D|^{-p/2} \exp \left[ -\frac{1}{2} \text{tr} C^{-1} (W_{12} - M) D^{-1} (W_{12} - M)' \right], \quad (2.2.1)$$

$W_{12} \in \mathbb{R}^{p \times q}$. Then $W_{12} \sim N(M, C \otimes D)$ is equivalent to stating that $\text{vec}(W_{12}') \sim N(\text{vec}(M'), C \otimes D)$, the multivariate normal distribution discussed in (1.0.1); see Muirhead [25, p. 79].

### 2.2.2 The Wishart distribution

A $d \times d$ random matrix $W$ has a Wishart distribution with degrees of freedom $a > d - 1$ and covariance matrix $\Lambda > 0$, denoted $W \sim W_d(a, \Lambda)$, if its probability density function is

$$
\frac{1}{2^{ad/2} |\Lambda|^{a/2} \Gamma_d(a/2)} |W|^{\frac{1}{2}a - \frac{1}{2}(d+1)} \exp \left( -\frac{1}{2} \text{tr} \Lambda^{-1} W \right), \quad (2.2.2)
$$

where $W > 0$, $\text{Re}(a) > \frac{1}{2}(d - 1)$, and

$$
\Gamma_d(a) = \pi^{d(d-1)/4} \prod_{j=1}^d \Gamma(a - \frac{1}{2}(j - 1)), \quad (2.2.3)
$$

is the multivariate gamma function [2], [25]. The Wishart distribution is also defined when $W$ is singular, in which case, the density function does not exist. Let $\Sigma > 0$ and $T$ be $d \times d$ matrices and $Z \sim N(0, I_n \otimes \Sigma)$. If $W = Z'Z$, then $W \sim W_d(n, \Sigma)$ and has
characteristic function

\[ E[\exp(i \text{ tr } (TW))] = |I_d - 2iT\Sigma|^{-n/2}. \quad (2.2.4) \]

2.2.3 The multivariate beta distribution

A \( d \times d \) random matrix \( L \) has a multivariate beta distribution with degrees of freedom \((a, b)\), where \( a > d - 1 \) and \( b > d - 1 \), denoted \( \text{Beta}_d(a/2, b/2) \), if its probability density function is

\[
\frac{\Gamma_d((a + b)/2)}{\Gamma_d(a/2)\Gamma_d(b/2)} |L|^{(a-d-1)/2} |I - L|^{(b-d-1)/2},
\quad (2.2.5)
\]

\( L > 0, I - L > 0. \)

2.2.4 The noncentral \( \chi^2 \) distribution

Let \( Z \sim N(\mu, I_d) \), then \( v = Z'Z \sim \chi^2_d(\mu'\mu) \), has a noncentral chi-square distribution with noncentrality parameter \( \tau^2 = \mu'\mu \). It is well-known, \[25\], that the characteristic function of \( v \) is

\[
E \exp[itv] = (1 - 2it)^{-d/2} \exp \left[ \frac{it\tau^2}{1 - 2it} \right].
\quad (2.2.6)
\]

2.3 Some properties of these distributions

We begin by stating a result on the characteristic function of a quadratic form in multivariate normal variables. Results of this type have been stated in various forms
in the literature; notably, they can be deduced from a result of Khatri [18], p. 446, eq. (3.4).

**Lemma 2.3.1.** (Khatri [18]) Let $C$ be a real, symmetric $p \times p$ matrix, $t \in \mathbb{R}$, $v \in \mathbb{R}^p$, and $Z \sim N_p(0, \Sigma)$. Then

$$Ee^{it(Z'CZ + v'Z)} = |I_p - 2itC\Sigma|\|^{-1/2}\exp\left(-\frac{1}{2}t^2v'\Sigma(I_p - 2itC\Sigma)^{-1}v\right).$$

(2.3.1)

Moreover, (2.3.1) remains valid if $C$ is a complex symmetric matrix whose imaginary part is positive-definite and $v$ is a complex vector.

The following result extends Lemma 2.2 of Chang and Richards [6].

**Lemma 2.3.2.** Let $\Lambda$ be a $q \times q$ positive definite matrix and $U$ be a $p \times p$ positive semi-definite matrix. If $B_{12} \sim N(0, C \otimes D)$, then

$$E \exp(-\text{tr} UB_{12}D^{-1} \Lambda D^{-1} B_{12}') = |I_{pq} + 2C^{1/2}UC^{1/2} \otimes D^{-1/2} \Lambda D^{-1/2}|^{-1/2}.$$ (2.3.2)

This result remains valid if $U$ is a symmetric complex matrix with the real part of $U$ positive definite.

**Proof.** We attribute the following proof of (2.3.2) to an anonymous referee of Chang and Richards [6]. First, recall that if $X \sim N_d(0, I_d)$, then $XX' \sim W_d(1, I_d)$. Therefore, for any positive definite matrix $A$,

$$E \exp\left(-tX'AX\right) = E \exp\left(-t \text{ tr} AX X'\right) = |I + 2tA|^{-1/2},$$ (2.3.3)
for $t > 0$. Define $K = D^{-1/2}B_{12}'C^{-1/2}$, $\phi = D^{-1/2}\Lambda D^{-1/2}$, and $\psi = C^{-1/2}UC^{-1/2}$.

By Proposition 2.1.1(iii), $\text{vec}(K) = (C^{-1/2} \otimes D^{-1/2})\text{vec}(B_{12}')$. Because $\text{vec}(B_{12}') \sim N(0, C \otimes D)$, it follows that $\text{vec}(K) \sim N(0, (C^{-1/2} \otimes D^{-1/2})(C \otimes D)(C^{-1/2} \otimes D^{-1/2})')$.

By Proposition 2.1.1(ii), the covariance matrix of $\text{vec}(K)$ equals

$$
(C^{-1/2} \otimes D^{-1/2})(C \otimes D)(C^{-1/2} \otimes D^{-1/2}) = C^{-1/2}CC^{-1/2} \otimes DD^{-1/2}DD^{-1/2}
$$

$$
= I_p \otimes I_q = I_{pq};
$$

hence, $\text{vec}(K) \sim N_{pq}(0, I_{pq})$. Further, by Proposition 2.1.1(iii),

$$
(\text{vec } K)'(\psi' \otimes \phi)(\text{vec } K) \equiv (\text{vec } K)'\text{vec}(\phi K \psi) = \text{tr}(K'\phi K \psi) = \text{tr}(\psi K' \phi K),
$$

and from the definitions of $\psi$, $K$, and $\phi$, we have $\psi K' \phi K = UB_{12}D^{-1}\Lambda D^{-1}B_{12}'$.

Because $\text{vec}(K) \sim N_{pq}(0, I_{pq})$, the moment-generating function stated above, (2.3.3), with $t = 1$ and $A = \psi' \otimes \phi$, yields the desired result.

Chang and Richards [6] and Kanda and Fujikoshi [17] gathered together a collection of properties of the Wishart distribution that we will also need here, all of which are available from Anderson [2], Eaton [9], or Muirhead [25].

**Proposition 2.3.3.** Suppose that $W \sim W_{d}(a, \Lambda)$, and $W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}$ and $\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$ have been partitioned similarly to the matrix $M$ in Section 2.1. 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) For $k \leq d$, if $M$ is $k \times d$ of rank $k$, then $(MW^{-1} M')^{-1} \sim W_k(a-d+k, (MA^{-1} M')^{-1})$.

In particular, if $Y$ is a $d \times 1$ random vector which is independent of $W$ and satisfies $P(Y = 0) = 0$, then $Y$ is independent of $Y' \Lambda^{-1} Y / Y' Y^{-1} Y \sim \chi^2_{a-d+1}$.

(v) Let $G \sim W_d(a, \Sigma)$, $H \sim W_d(b, \Sigma)$, where $G$ and $H$ are independent. Then $L = (G + H)^{-1/2} G (G + H)^{-1/2} \sim \text{Beta}_d(a/2, b/2)$ and $L$ is independent of $G + H$.

Finally, in deriving any stochastic representation, we will use the standard notation, $\equiv$, for equal in distribution; $\succeq$, for stochastically greater; and $\preceq$, for stochastically smaller. That is to say, if $X$ and $Y$ are random entities, then $X \equiv Y$ signifies that $X$ and $Y$ have the same probability distribution; also if $X$ and $Y$ are scalar-valued random variables, then $X \succeq Y$ signifies that $P(X \geq t) \geq P(Y \geq t)$ for all $t \in \mathbb{R}$, and $X \preceq Y$ whenever $Y \succeq X$. 

Chapter 3

Two-step Monotone Incomplete
Multivariate Normal Data

This chapter begins with a thorough description of our notation and the maximum likelihood estimators for two-step monotone incomplete data. We then provide an alternative derivation of the exact distribution of \( \hat{\mu} \), the maximum likelihood estimator for \( \mu \), first derived by Chang and Richards, [6]. In this chapter we will also derive a stochastic representation for the exact distribution of a generalization of Hotelling’s \( T^2 \)-statistic. We will then derive upper and lower bounds for the exact distribution of the \( T^2 \)-statistic. As a consequence, we obtain exact ellipsoidal confidence regions for \( \mu \). We also apply the \( T^2 \)-statistic to derive simultaneous confidence intervals for linear functions of \( \mu \), and we apply these results to the Pennsylvania cholesterol data. We complete this chapter by studying prediction regions for complete observations.

3.1 Notation and maximum likelihood estimators

Let \( X \in \mathbb{R}^p \) and \( Y \in \mathbb{R}^q \). In the case of two-step monotone incomplete data, we suppose that the data are \( N \) mutually independent observations of the form,

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

(3.1.1)
where \( (X_j, Y_j), j = 1, \ldots, n \) 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.

One additional assumption necessary to guarantee that all means and variances are finite and that all integrals encountered later are absolutely convergent is that \( n > p + 2 \), [6]. Data of the form (3.1.1) have been widely studied; cf. Anderson [1], Bhargava [4, 5], Morrison [24], Eaton and Kariya [10], Fujisawa [12], Hao and Krishnamoorthy [14], Kanda and Fujikoshi [17], and most recently, Chang and Richards [6, 7].

Define the sample mean vectors

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

(3.1.2)

and the corresponding matrices of sums of squares and products

\[
A_{11,n} = \sum_{j=1}^{n} (X_j - \bar{X}_1)(X_j - \bar{X}_1)', \quad A_{12} = A_{21}' = \sum_{j=1}^{n} (X_j - \bar{X}_1)(Y_j - \bar{Y})', \\
A_{22} = \sum_{j=1}^{n} (Y_j - \bar{Y})(Y_j - \bar{Y})', \quad A_{11,N} = \sum_{j=1}^{N} (X_j - \bar{X})(X_j - \bar{X})'.
\]

(3.1.3)

In addition, we use the notation \( \tau = n/N \) for the proportion of data which are complete and denote \( 1 - \tau \) by \( \bar{\tau} \), so that \( \bar{\tau} = (N-n)/N \) is the proportion of incomplete observations.

By Anderson [1] (cf. Morrison [24], Anderson and Olkin [3], Jinadasa and Tracy [15]),
the maximum likelihood estimators of $\mu$ and $\Sigma$ are, respectively,

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \bar{Y} - \bar{\tau}A_{21}A^{-1}_{11,n}(\bar{X}_1 - \bar{X}_2) \end{pmatrix}, \quad (3.1.4)$$

and

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \frac{1}{N}A_{11,N} & \frac{1}{N}A_{11,N}A_{11,n}^{-1}A_{12} \\ \frac{1}{N}A_{21}A_{11,n}^{-1}A_{11,N} & \frac{1}{n}A_{22,1,n}^{-1}A_{11,N}A_{11,n}^{-1}A_{12} + \frac{1}{N}A_{21}A_{11,n}^{-1}A_{11,N}A_{11,n}^{-1}A_{12} \end{pmatrix}, \quad (3.1.5)$$

3.2 A new derivation of an exact stochastic representation for $\hat{\mu}$

Chang and Richards [6] derived an exact stochastic representation for $\hat{\mu}$ by means of a direct analysis of its characteristic function. In examining ways to extend their methods to three-step data, we discovered an alternative method to derive the exact distribution of $\hat{\mu}$ by means of the incomplete data given the complete data, namely $Y$ given $X_1, \ldots, X_N$.

Before we delve into the exact distribution, we show that $\hat{\mu}$ and $\hat{\Sigma}$ are equivariant. This can be derived from a general argument given by Davison [8], p. 185, however we have chosen to provide the explicit details here.
Proposition 3.2.1. Let $\Lambda_{11}$ and $\Lambda_{22}$ be $p \times p$ and $q \times q$ positive definite matrices, respectively, $\Lambda_{21}$ be $q \times p$, $\nu_1 \in \mathbb{R}^p$, $\nu_2 \in \mathbb{R}^q$, and

$$\Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix}, \quad C = \begin{pmatrix} I_p & 0 \\ \Lambda_{21} & I_q \end{pmatrix}, \quad \nu = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}.$$

Then the estimators $\hat{\mu}$ and $\hat{\Sigma}$ are equivariant under the transformation

$$\begin{pmatrix} X_j \\ Y_j \end{pmatrix} \mapsto \Lambda C \begin{pmatrix} X_j \\ Y_j \end{pmatrix} + \nu, \quad (3.2.1)$$

for $j = 1, \ldots, n$. For $j = n + 1, \ldots, N$, $X_j \mapsto \Lambda_{11} X_j + \nu_1$.

Proof. Let

$$\begin{pmatrix} X_j^* \\ Y_j^* \end{pmatrix} = \Lambda C \begin{pmatrix} X_j \\ Y_j \end{pmatrix} + \nu = \begin{pmatrix} \Lambda_{11} X_j + \nu_1 \\ \Lambda_{22} \Lambda_{21} X_j + \Lambda_{22} Y_j + \nu_2 \end{pmatrix}, \quad (3.2.2)$$

for $j = 1, \ldots, n$ and $X_j^* = \Lambda_{11} X_j + \nu_1$, $j = n + 1, \ldots, N$. Then

$$\begin{pmatrix} X_j^* \\ Y_j^* \end{pmatrix} \sim N_{p+q} \begin{pmatrix} \mu^* \\ \Sigma^* \end{pmatrix} = \begin{pmatrix} \mu_1^* \\ \mu_2^* \\ \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix}.$$
\( j = 1, \ldots, n \), and \( X^*_j \sim N_p(\mu^*_1, \Sigma^*_{11}) \), \( j = n + 1, \ldots, N \), where \( \mu^* = \Lambda C \mu + \nu \) and \( \Sigma^* = \Lambda C \Sigma C' \Lambda \). Define the sample mean vectors

\[
\bar{X}^*_1 = \frac{1}{n} \sum_{j=1}^{n} X^*_j, \quad \bar{X}^*_2 = \frac{1}{N-n} \sum_{j=n+1}^{N} X^*_j,
\]

\[
\bar{Y}^* = \frac{1}{n} \sum_{j=1}^{n} Y^*_j, \quad \bar{X}^* = \frac{1}{N} \sum_{j=1}^{N} X^*_j
\] (3.2.3)

and the corresponding matrices of sums of squares and products

\[
A^*_{11,N} = \sum_{j=1}^{N} (X^*_j - \bar{X}^*)(X^*_j - \bar{X}^*)', \quad A^*_{12} = A^*_{21} = \sum_{j=1}^{n} (X^*_j - \bar{X}^*_1)(Y^*_j - \bar{Y}^*)',
\]

\[
A^*_{11,n} = \sum_{j=1}^{n} (X^*_j - \bar{X}^*_1)(X^*_j - \bar{X}^*_1)', \quad A^*_{22} = \sum_{j=1}^{n} (Y^*_j - \bar{Y}^*_1)(Y^*_j - \bar{Y}^*_1)'.
\] (3.2.4)

Then, the maximum likelihood estimators of \( \mu^* \) and \( \Sigma^* \) are, respectively,

\[
\hat{\mu}^* = \begin{pmatrix} \bar{X}^* \\ \bar{Y}^* - \bar{X}^* \bar{A}^*_{21} A^*_{11,n}^{-1} (\bar{X}^* - \bar{X}^*_1) \end{pmatrix}
\] (3.2.5)

and

\[
\hat{\Sigma}^* = \begin{pmatrix} \frac{1}{N} A^*_{11,N} & \frac{1}{N} A^*_{11,N} A^*_{11,n}^{-1} A^*_{12} \\ \frac{1}{N} A^*_{21} A^*_{11,n}^{-1} A^*_{11,N} & \frac{1}{N} A^*_{21} A^*_{11,n}^{-1} A^*_{12} + \frac{1}{N} A^*_{22} A^*_{21,n}^{-1} A^*_{21} A^*_{11,n}^{-1} A^*_{12} \end{pmatrix}.
\] (3.2.6)
Our goal is to show that \( \hat{\mu} = \Lambda C \hat{\mu} + \nu \) and that \( \hat{\Sigma} = \Lambda C \hat{\Sigma} C' \Lambda' \). As a consequence of 3.2.3 - 3.2.4, we have the following relations:

\[
\tilde{X}_1^* = \Lambda_{11} \tilde{X}_1 + \nu_1, \quad \tilde{X}_2^* = \Lambda_{11} \tilde{X}_2 + \nu_1, \\
\bar{Y}^* = \Lambda_{22} (\bar{Y} + \Lambda_{21} \tilde{X}_1) + \nu_2, \quad \bar{Z}^* = \Lambda_{11} \bar{Z} + \nu_1,
\]

and

\[
A_{11,N}^* = \Lambda_{11} A_{11,N} \Lambda_{11},
\]

\[
A_{11,n}^* = \Lambda_{11} A_{11,n} \Lambda_{11},
\]

\[
A_{12}^* = A_{21}' = \sum_{j=1}^{n} \Lambda_{11} (X_j - \bar{X}_1) [\Lambda_{22} (\Lambda_{21} (X_j - \bar{X}_1) + Y_j - \bar{Y})]' = \Lambda_{11} (A_{12} + A_{11,n} A_{12}) \Lambda_{22},
\]

\[
A_{22}^* = \Lambda_{22} [\sum_{j=1}^{n} (\Lambda_{21} (X_j - \bar{X}_1) + Y_j - \bar{Y}) (\Lambda_{21} (X_j - \bar{X}_1) + Y_j - \bar{Y})] \Lambda_{22} = \Lambda_{22} [\Lambda_{21} A_{11,n} A_{12} + \Lambda_{21} A_{12} + A_{21} A_{12} + A_{22}] \Lambda_{22}.
\]

We may write both \( \hat{\mu}^* \) and \( \hat{\Sigma}^* \) in terms of the original means and matrices of sums of squares and products. It is straightforward that

\[
\hat{\mu}_1^* = \tau \tilde{X}_1^* + \bar{\tau} \tilde{X}_2^* = \tau \Lambda_{11} \tilde{X}_1 + \nu_1 + \bar{\tau} (\Lambda_{11} \tilde{X}_2 + \nu_1) = \Lambda_{11} \hat{\mu}_1 + \nu_1.
\]
The maximum likelihood estimator of $\mu^*_2$ is

$$
\hat{\mu}^*_2 = \Lambda_{22} \bar{Y} + \nu_2 + \Lambda_{22} \Lambda_{21} \bar{X}_1 - \bar{\tau} \Lambda_{22} (A_{21} + \Lambda_{21} A_{11,n}) A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2)
$$

$$
= \Lambda_{22} \bar{Y} + \nu_2 + \Lambda_{22} \Lambda_{21} (\bar{\tau} \bar{X}_1 + \bar{\tau} \bar{X}_2) - \bar{\tau} \Lambda_{22} A_{21} A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2) + \nu_2.
$$

(3.2.7)

Because $\tau \bar{X}_1 + \tau \bar{X}_2 = \bar{X}$, it follows that

$$
\hat{\mu}^*_2 = \Lambda_{22} (A_{21} \hat{\mu}_1 + \hat{\mu}_2) + \nu_2.
$$

(3.2.8)

Therefore

$$
\begin{pmatrix}
\hat{\mu}^*_1 \\
\hat{\mu}^*_2
\end{pmatrix} = \begin{pmatrix}
\Lambda_{11} \hat{\mu}_1 + \nu_1 \\
\Lambda_{22} (A_{21} \hat{\mu}_1 + \hat{\mu}_2) + \nu_2
\end{pmatrix} = \Lambda C \begin{pmatrix}
\hat{\mu}_1 \\
\hat{\mu}_2
\end{pmatrix} + \nu,
$$

(3.2.9)

so we have proved that $\hat{\mu}$ is equivariant. The maximum likelihood estimators of $\hat{\Sigma}^*_1$, $\hat{\Sigma}^*_2$, and $\hat{\Sigma}^*_2$ are, respectively,

$$
\hat{\Sigma}^*_1 = \frac{1}{N} \Lambda_{11} A_{11,N} \Lambda_{11},
$$

(3.2.10)

$$
\hat{\Sigma}^*_2 = \hat{\Sigma}^*_2 = \frac{1}{N} \Lambda_{11} A_{11,N} A_{11,n}^{-1} A_{12} + \frac{1}{N} \Lambda_{11} A_{11,N} \Lambda_{12} A_{22},
$$

(3.2.11)
and

\[ \hat{\Sigma}_{22}^* = \frac{1}{n} \Lambda_{22} \left[ \Lambda_{21} A_{11,n} \Lambda_{12} + \Lambda_{21} A_{12} + \Lambda_{21} A_{12} + A_{22} \right] \]

\[- \left( A_{21} + \Lambda_{21} A_{11,n} \right) A_{12}^{-1} \left( A_{12} + A_{11,n} A_{12} \right) \Lambda_{22} \]

\[ + \frac{1}{N} \Lambda_{22} \left( A_{21} + \Lambda_{21} A_{11,n} \right) A_{11,n}^{-1} A_{11,n}^{-1} \left( A_{12} + A_{11,n} A_{12} \right) \Lambda_{22} \]

\[ = \frac{1}{n} \Lambda_{22} A_{22,1,n} \Lambda_{22} + \frac{1}{N} \Lambda_{22} \left( A_{21} + \Lambda_{21} A_{11,n} A_{12} + \Lambda_{21} A_{11,n} A_{11,n} A_{12} \right) \Lambda_{22} \]

\[ + \Lambda_{21} A_{11,n} A_{11,n}^{-1} A_{12} + \Lambda_{21} A_{11,n} A_{11,n} A_{11,n} A_{12} \Lambda_{22}. \quad (3.2.12) \]

To establish the equivariance of \( \hat{\Sigma} \), let us evaluate \( \Lambda C \hat{\Sigma} C' \Lambda \). To that end,

\[ \Lambda C \hat{\Sigma} C' \Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ \Lambda_{22} \Lambda_{21} & \Lambda_{22} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{N} A_{11,N} & \frac{1}{N} A_{11,N} A_{11,n}^{-1} A_{12} \\ \frac{1}{N} A_{21} A_{11,n}^{-1} A_{11,N} & \frac{1}{N} A_{21} A_{11,n}^{-1} A_{11,n} A_{12} \end{pmatrix} \]

\[ \cdot \begin{pmatrix} \Lambda_{11} & \Lambda_{12} A_{22} \\ 0 & \Lambda_{22} \end{pmatrix} \]

\[ = \begin{pmatrix} \Lambda_{11} \hat{\Sigma}_{11} \Lambda_{11} & \Lambda_{11} \left( \hat{\Sigma}_{11} \Lambda_{12} + \hat{\Sigma}_{12} \right) \Lambda_{22} \\ \Lambda_{22} \left( \hat{\Sigma}_{21} + \Lambda_{21} \hat{\Sigma}_{11} \right) \Lambda_{11} & \Lambda_{22} \left( \Lambda_{21} \hat{\Sigma}_{11} A_{12} + \hat{\Sigma}_{21} A_{12} + \Lambda_{21} \hat{\Sigma}_{12} + \hat{\Sigma}_{22} \right) \Lambda_{22} \end{pmatrix}. \quad (3.2.13) \]

By straightforward matrix multiplication, it follows from \( (3.2.10) \), \( (3.2.11) \), and \( (3.2.12) \) that \( \Lambda C \hat{\Sigma} C' \Lambda = \hat{\Sigma}^* \). Therefore, \( \hat{\Sigma} \) is also equivariant under the transformation \( (3.2.2) \).
As a consequence of the equivariance of $\hat{\Sigma}$, we obtain the following result.

**Corollary 3.2.2.** The estimated covariance matrix of $\hat{\mu}$ is equivariant under the transformation (3.2.2).

**Proof.** Because $\hat{\Sigma}$ is equivariant under the transformation (3.2.2), it follows that

$$\hat{\text{Cov}}(\hat{\mu}^*) = \frac{1}{N} \hat{\Sigma}^* + \frac{(\gamma - 1)}{N} \begin{pmatrix} 0 & 0 \\ 0 & \hat{\Sigma}^*_{22-1} \end{pmatrix}$$

$$= \frac{1}{N} \Lambda C \hat{\Sigma} C' \Lambda + \frac{(\gamma - 1)}{N} \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_{22} \hat{\Sigma}_{22-1} \Lambda_{22} \end{pmatrix}$$

$$= \Lambda C \hat{\text{Cov}}(\hat{\mu}) C' \Lambda.$$ 

Therefore $\hat{\text{Cov}}(\hat{\mu})$ is equivariant. 

By taking

$$\Lambda_{11} = \Sigma_{11}^{-1/2}, \quad \Lambda_{22} = \Sigma_{22-1}^{-1/2}, \quad \Lambda_{21} = -\Sigma_{21} \Sigma_{11}^{-1}, \quad (3.2.14)$$

then under the transformation (3.2.2), we obtain

$$\text{Cov} \begin{pmatrix} X \\ Y \end{pmatrix} = \Lambda C \Sigma C' \Lambda' = \mathbf{I}_{p+q}. \quad (3.2.15)$$

Therefore, in analyzing the distribution of $\hat{\mu}$, we may assume, without loss of generality, that the population covariance matrix is $\mathbf{I}_{p+q}$. Furthermore, by choosing $\nu = -\Lambda C \mu$, we may also assume, without loss of generality, that $\mu = 0$. 
We will now provide an alternative proof for the exact stochastic representation of $\hat{\mu}$.

**Theorem 3.2.3.** (Chang and Richards [6]) Let $V_1 \sim N_{p+q}\left(0, \frac{1}{N} \Sigma + \frac{\tau}{n} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)$, $Q_1 \sim \chi^2_{n-p}$, $Q_2 \sim \chi^2_p$, $V_2 \sim N_q(0, I_q)$, where $V_1$, $V_2$, $Q_1$, and $Q_2$ are mutually independent. Then the distribution of $\hat{\mu}$ is given by the exact stochastic representation

$$\hat{\mu} \sim \mu + V_1 + \left(\frac{\bar{\tau}Q_2}{nQ_1}\right)^{1/2} \begin{pmatrix} 0 \\ \Sigma_{22}^{1/2} V_2 \end{pmatrix}, \tag{3.2.16}$$

**Proof.** Assume, without loss of generality, that $\mu = 0$ and $\Sigma = I_{p+q}$. Then, by (2.1.5), it follows that $Y_j|X_j \sim N_q(0, I_q)$ and therefore all the $X_j$ and $Y_j$ are mutually independent, normally distributed random vectors. Therefore $\bar{Y}|\{X_1, \ldots, X_N\} \sim N_q(0, \frac{1}{n} I_q)$. Conditional on the complete data, $X = \{X_1, \ldots, X_N\}$, the vector $\hat{\mu}_2$ is a linear combination of the incomplete data, $Y = (Y_1, \ldots, Y_n)$, hence $\hat{\mu}_2$ is normally distributed. Moreover, because $\hat{\mu}_1 = \bar{X}$, it then follows that $\hat{\mu}_1$ is fixed in the conditional distribution of $\hat{\mu}_2$ given $X$. We now need to find the conditional expected value and covariance matrix of $\hat{\mu}_2$. Let $c_j = (X_j - \bar{X})' A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2)$, $j = 1, \ldots, n$; noting that $\sum_{j=1}^n c_j = 0$, we may write $\hat{\mu}_2$ as

$$\hat{\mu}_2 = \bar{Y} - \bar{\tau} A_{21} A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2) = \frac{1}{n} \sum_{j=1}^n Y_j - \tau \sum_{j=1}^n c_j (Y_j - \bar{Y})$$

$$= \frac{1}{n} \sum_{j=1}^n Y_j - \tau \sum_{j=1}^n c_j Y_j.$$
Let \( \delta_{jk} \) be Kronecker’s delta, that is \( \delta_{jk} = 1, j = k \) and \( \delta_{jk} = 0, j \neq k \). Then it is of note that

\[
\text{Cov}(\sum_{j=1}^{n} Y_j, \sum_{j=1}^{n} c_j Y_j) = \sum_{j=1}^{n} \sum_{k=1}^{n} c_k E(Y_j Y'_k)
\]

\[
= \sum_{j=1}^{n} \sum_{k=1}^{n} c_k \delta_{jk} I_q
\]

\[
= (\sum_{j=1}^{n} c_j) I_q = 0.
\]

Because \( \sum_{j=1}^{n} Y_j \) and \( \sum_{j=1}^{n} c_j Y_j \) are jointly normally distributed and their covariance is zero, it then follows that \( \sum_{j=1}^{n} Y_j \) and \( \sum_{j=1}^{n} c_j Y_j \) are independent. Therefore, conditional on \( X \), \( \hat{\mu}_2 \) is a linear combination of independent normal vectors, hence \( \hat{\mu}_2 \) is normally distributed with mean

\[
E(\hat{\mu}_2 | X) = \sum_{j=1}^{n} E((\frac{1}{n} - \bar{\tau} c_j) Y_j | X)
\]

\[
= \sum_{j=1}^{n} (\frac{1}{n} - \bar{\tau} c_j) E(Y_j) = 0,
\]

(3.2.17)

and covariance matrix

\[
\text{Cov}(\hat{\mu}_2 | X) = \text{Cov}(\frac{1}{n} \sum_{j=1}^{n} Y_j) + \text{Cov}(\bar{\tau} \sum_{j=1}^{n} Y_j c_j)
\]

\[
= \frac{1}{n} I_q + \bar{\tau}^2 \sum_{j=1}^{n} \sum_{k=1}^{n} c_j c_k \delta_{jk} I_q
\]

\[
= \frac{1}{n} I_q + \bar{\tau}^2 (\sum_{j=1}^{n} c_j^2) I_q.
\]
Because

\[
\sum_{j=1}^{n} c_j^2 = (\bar{X}_1 - \bar{X}_2)' A_{11,n}^{-1} \sum_{j=1}^{n} (X_j - \bar{X}_1)(X_j - \bar{X}_1)' A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2)
\]

\[
= (\bar{X}_1 - \bar{X}_2)' A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2),
\]

it follows that

\[
\text{Cov}(\hat{\mu}_2|X) = \frac{1}{n}I_q + \bar{\tau}^2 (\bar{X}_1 - \bar{X}_2)' A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2) I_q.
\]

(3.2.18)

Therefore

\[
\hat{\mu}_2|X \sim N_q(0, \frac{1}{n}I_q + \bar{\tau}^2 (\bar{X}_1 - \bar{X}_2)' A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2) I_q).
\]

Observe that \(\hat{\mu}_2\) depends on \(X\) only through \(\bar{X}_1 - \bar{X}_2\) and \(A_{11,n}\); therefore \(\{\bar{X}_1 - \bar{X}_2, A_{11,n}\}\) is sufficient for \(\hat{\mu}_2\). By the independence of the sample mean and sample covariance matrix of a normal random sample, \(\bar{X}_1\) and \(A_{11,n}\) are independent and, consequently, \(\{\tau \bar{X}_1 + \bar{\tau} \bar{X}_2, \bar{X}_1 - \bar{X}_2\}\) and \(A_{11,n}\) are independent. Next, because

\[
\text{Cov}(\bar{X}_1 - \bar{X}_2, \tau \bar{X}_1 + \bar{\tau} \bar{X}_2) = \text{Cov}(\bar{X}_1, \tau \bar{X}_1) - \text{Cov}(\bar{X}_2, \tau \bar{X}_2) = N^{-1}I_p - N^{-1}I_p = 0
\]

and \((\bar{X}_1 - \bar{X}_2, \tau \bar{X}_1 + \bar{\tau} \bar{X}_2)\) has a joint multivariate normal distribution, it follows that

\[
\hat{\mu}_1 = \tau \bar{X}_1 + \bar{\tau} \bar{X}_2
\]

is independent of \(\bar{X}_1 - \bar{X}_2\). Therefore \(\bar{X}_1 - \bar{X}_2, A_{11,n}\), and \(\tau \bar{X}_1 + \bar{\tau} \bar{X}_2\), and consequently, \(\hat{\mu}_1\) and \(\hat{\mu}_2\), are mutually independent. By Proposition 2.3.3 (iv),

\[
Q_1 = \frac{(\bar{X}_1 - \bar{X}_2)' (\bar{X}_1 - \bar{X}_2)}{(\bar{X}_1 - \bar{X}_2)' A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2)} \sim \chi^2_{n-p},
\]
and $Q_1$ is independent of $\bar{X}_1 - \bar{X}_2$. Therefore,

$$\hat{\mu}_2|\{X, Q_1\} \sim N_q\left(0, \frac{1}{n^2} \left[1 + \tau^2/\bar{\tau}Q_1 \left(\bar{X}_1 - \bar{X}_2\right)\right] \right).$$

(3.2.19)

Because $\bar{X}_1 - \bar{X}_2 \sim N_p(0, (n^{-1} - (N-n)^{-1})I_p)$ and $n^{-1} - (N-n)^{-1} = 1/n\bar{\tau}$, it follows that $(\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2) \sim Q_2/n\bar{\tau}$, where $Q_2 \sim \chi^2_p$. We may now write the conditional distribution of $\hat{\mu}_2$ as

$$\hat{\mu}_2|\{Q_1, Q_2\} \sim N_q\left(0, \left(\frac{1}{n^2} + \frac{\tau Q_2}{n Q_1}\right) I_q\right).$$

By elementary properties of the normal distribution, it follows that $\hat{\mu}_2 \equiv \sqrt{\frac{1}{n}} V_{12} + \sqrt{\frac{\tau Q_2}{n Q_1}} V_2$, where $V_{12} \sim N_q(0, I_q)$, $V_2 \sim N_q(0, I_q)$, and $V_{12}, V_2, Q_1$, and $Q_2$ are mutually independent.

It is straightforward to see that $\hat{\mu}_1 = \bar{X} \sim N_p(0, \frac{1}{N} I_p)$ and therefore $\hat{\mu}_1 \equiv V_{11}$, where $V_{11} \sim N_p(0, \frac{1}{N} I_p)$. Therefore, a joint stochastic representation for $\hat{\mu}_1$ and $\hat{\mu}_2$ is

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \equiv V_1 + \sqrt{\frac{\tau Q_2}{n Q_1}} \begin{pmatrix} 0 \\ V_2 \end{pmatrix},$$

where $V_1 = \begin{pmatrix} V_{11} \\ V_{12} \end{pmatrix} \sim N_{p+q} \left(0, \begin{pmatrix} \frac{1}{N} I_p & 0 \\ 0 & \frac{1}{n} I_q \end{pmatrix} \right)$ and $V_2$ is as defined previously.
Our final step is to transform the data back to its original form for general $\mu$ and $\Sigma$. Recall that by (3.2.14), the transformation to $\mu = 0$ and $\Sigma = I$ is

$$
\begin{pmatrix}
\hat{\mu}_1^* \\
\hat{\mu}_2^*
\end{pmatrix} =
\begin{pmatrix}
\Sigma_{11}^{-1/2} & 0 \\
0 & \Sigma_{22}^{-1/2}
\end{pmatrix}
\begin{pmatrix}
I & 0 \\
-\Sigma_{21} \Sigma_{11}^{-1} & I
\end{pmatrix}
\begin{pmatrix}
\hat{\mu}_1 \\
\hat{\mu}_2
\end{pmatrix} - \mu.
$$

The inverse of the latter transformation is

$$
\begin{pmatrix}
\hat{\mu}_1 \\
\hat{\mu}_2
\end{pmatrix} =
\begin{pmatrix}
I & 0 \\
\Sigma_{21} \Sigma_{11}^{-1} & I
\end{pmatrix}
\begin{pmatrix}
\Sigma_{11}^{1/2} & 0 \\
0 & \Sigma_{22}^{1/2}
\end{pmatrix}
\begin{pmatrix}
\hat{\mu}_1^* \\
\hat{\mu}_2^*
\end{pmatrix} + \mu;
$$

therefore

$$
\hat{\mu} \overset{\text{L}}{=} 
\begin{pmatrix}
I & 0 \\
\Sigma_{21} \Sigma_{11}^{-1} & I
\end{pmatrix}
\begin{pmatrix}
\Sigma_{11}^{1/2} & 0 \\
0 & \Sigma_{22}^{1/2}
\end{pmatrix}
\begin{pmatrix}
V_1 + \sqrt{\frac{\bar{\tau} Q_2}{n Q_1}} \left( \begin{pmatrix} 0 \\ V_2 \end{pmatrix} \right)
\end{pmatrix} + \mu.
$$

Because $\Sigma_{22}^{1/2} V_2 \sim N_q(0, \Sigma_{22})$, and

$$
\begin{pmatrix}
I & 0 \\
\Sigma_{21} \Sigma_{11}^{-1} & I
\end{pmatrix}
\begin{pmatrix}
\Sigma_{11}^{1/2} & 0 \\
0 & \Sigma_{22}^{1/2}
\end{pmatrix}
V_1 \sim N_{p+q} \left( 0, \frac{1}{N} \Sigma + \frac{\bar{\tau}}{n} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right),
$$

we obtain (3.2.16).

One advantage of this proof is that it provides explicit formulas for $V_1$, $V_2$, $Q_1$, and $Q_2$ in terms of the data, whereas Chang and Richards showed only the existence of...
these random variable. Namely, those explicit formulas are:

$$Q_1 = \frac{(\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2)}{(\bar{X}_1 - \bar{X}_2)'A_{11,n}^{-1}(\bar{X}_1 - \bar{X}_2)},$$

$$Q_2 = n\bar{\tau}(\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2),$$

$$V_1 = \begin{pmatrix} V_{11} \\ V_{12} \end{pmatrix} = \begin{pmatrix} \bar{X} \\ Y \end{pmatrix},$$

$$V_2 = -\frac{\sum_{j=1}^{n} Y_j(X_j^* - \bar{X})'A_{11,n}^{-1}(\bar{X}_1 - \bar{X}_2)}{(\bar{X}_1 - \bar{X}_2)'A_{11,n}^{-1}(\bar{X}_1 - \bar{X}_2)}.$$

### 3.3 An exact stochastic representation for the $T^2$-statistic

Following Krishnamoorthy and Pannala [20] and Chang and Richards [6], we will study the pivotal quantity,

$$T^2 = (\hat{\mu} - \mu)'\widehat{\text{Cov}}(\hat{\mu})^{-1}(\hat{\mu} - \mu),\quad (3.3.1)$$

a generalization of Hotelling’s $T^2$-statistic in the setting of monotone incomplete data. An $F$-distribution approximation to the distribution of a statistic similar to (3.3.1) was given by Krishnamoorthy and Pannala [20]. Chang and Richards [6] obtained upper and lower bounds for its distribution, leading to conservative ellipsoidal confidence regions for $\mu$, and derived the asymptotic distribution of the $T^2$-statistic for the cases in which $n$, $N$, $p$, and $q$ satisfy $n > p+q$ for fixed $n$, or $n/N \to \delta \in (0, 1]$ as $n, N \to \infty$. Nevertheless, the exact finite-sample distribution of this statistic was unknown before the work in this dissertation.
Our primary motivation for deriving a stochastic representation for the exact distribution of the $T^2$ statistic (3.3.1) is that resulting ellipsoidal confidence regions for $\mathbf{\mu}$ will be less conservative than those previously derived.

Let

$$\gamma = 1 + \frac{(n-2)N\bar{r}}{n(n-p-2)}.$$  (3.3.2)

As shown by Chang and Richards [6], the maximum likelihood estimator of $\text{Cov}(\hat{\mathbf{\mu}})$ is

$$\hat{\text{Cov}}(\hat{\mathbf{\mu}}) = \frac{1}{N} \bar{\Sigma} + \frac{\gamma - 1}{N} \begin{pmatrix} 0 & 0 \\ 0 & \hat{\Sigma}_{22} \end{pmatrix},$$  (3.3.3)

where $\bar{\Sigma}$ is as defined in (3.1.5). Following Chang and Richards [7], we decompose $A_{11,N}$ as follows:

$$A_{11,N} = A_{11,n} + B_1 + B_2,$$  (3.3.4)

where

$$B_1 = \sum_{j=n+1}^N (X_j - \bar{X}_2)(X_j - \bar{X}_2)',$$  (3.3.5)

$$B_2 = \frac{n(N-n)}{N} (\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)',$$  (3.3.6)

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 [7].
Lemma 3.3.1. (Chang and Richards [6]) When $\Sigma_{12} = 0$, the random matrices and vectors $A_{22,1,n}, A_{21} A_{11,1,n}^{-1}, A_{11,1,n}, \bar{Y}, \bar{X}, B_1$, and $B_2$ are mutually independent.

In preparation for our derivation of an exact stochastic representation of the distribution of the $T^2$-statistic, we show that, without loss of generality, we may assume that $\mu = 0$ and $\Sigma = I_d$. Similar to Chang and Richards [6], we begin by writing

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

as a sum of two terms. Define

$$T^2_1 = n(\bar{Y} - A_{21} A_{11,1,n}^{-1} \bar{X}_1 - \mu_2 + A_{21} A_{11,1,n}^{-1} \mu_1)' A_{22,1,n}^{-1} (\bar{Y} - A_{21} A_{11,1,n}^{-1} \bar{X}_1 - \mu_2 + A_{21} A_{11,1,n}^{-1} \mu_1),
\quad (3.3.7)$$

and

$$T^2_2 = N(\bar{X} - \mu_1)' A_{11,n}^{-1} (\bar{X} - \mu_1)
= N(\bar{X} - \mu_1)' (A_{11,n} + B_1 + B_2)^{-1} (\bar{X} - \mu_1).
\quad (3.3.8)$$

Applying the quadratic identity (2.1.4) with $x \equiv \hat{\mu} - \mu$ and $\Lambda \equiv N \tilde{\text{Cov}}(\hat{\mu})$, we find that

$$\frac{1}{N} T^2 = (\hat{\mu} - \mu)' \left( N \tilde{\text{Cov}}(\hat{\mu}) \right)^{-1} (\hat{\mu} - \mu)
= (\hat{\mu}_2 - \mu_2 - A_{21} A_{11,1,n}^{-1} (\hat{\mu}_1 - \mu_1))' \left( N \tilde{\text{Cov}}(\hat{\mu})_{22,1} \right)^{-1} (\hat{\mu}_2 - \mu_2 - A_{21} A_{11,1,n}^{-1} (\hat{\mu}_1 - \mu_1))
+ (\hat{\mu}_1 - \mu_1)' \tilde{\Sigma}_{11}^{-1} (\hat{\mu}_1 - \mu_1).$$
By (3.1.4), \( \hat{\mu}_2 - A_{21} A_{11,n}^{-1} \hat{\mu}_1 = \bar{Y} - A_{21} A_{11,n}^{-1} \bar{X}_1 \); by (3.3.3), \( \text{NCov}(\hat{\mu})_{22,1} = \frac{2}{n} A_{22,1,n} \); and by (3.3.4), \( A_{11,N} = A_{11,n} + B_1 + B_2 \). Therefore

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

Krishnamoorthy and Pannala [20] also decomposed their \( T^2 \)-statistic into a corresponding sum \( \tilde{T}_1^2 + \tilde{T}_2^2 \) and showed that the marginal distribution of each \( \tilde{T}_j \) does not depend on \( (\mu, \Sigma) \). To deduce that the distribution of their \( \tilde{T}^2 \)-statistic depends neither on \( \mu \) or \( \Sigma \), it would need to be shown that the joint distribution of \( (\tilde{T}_1, \tilde{T}_2) \) also satisfies that property, a result which appears difficult to establish directly. We provide a proof that uses ideas of Yamada, et al. [32] to show that the distribution of the \( T^2 \)-statistic depends neither on \( \mu \) nor \( \Sigma \).

**Proposition 3.3.2.** The statistics \( T^2_1 \) in (3.3.7), and \( T^2_2 \) in (3.3.8) are algebraically invariant under the transformation (3.2.2). Consequently, the same holds for the \( T^2 \)-statistic in 3.3.1.

**Proof.** Let

\[
\begin{pmatrix}
X^* \\
Y^*
\end{pmatrix} = \Lambda C \begin{pmatrix} X \\
Y
\end{pmatrix} + \nu = \begin{pmatrix} \Lambda_{11} X + \nu_1 \\
\Lambda_{22} A_{21} X + \Lambda_{22} Y + \nu_2
\end{pmatrix}.
\]

By Proposition (3.2.1), \( \hat{\mu} \) and \( \hat{\Sigma} \) are equivariant under this transformation. Because

\[
(T^2_2)^* = (\hat{\mu}_1^* - \mu_1^*)' \hat{\Sigma}_{11}^{-1}(\hat{\mu}_1^* - \mu_1^*)
\]

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

\[
= (\hat{\mu}_1 - \mu_1)' \hat{\Sigma}_{11}^{-1}(\hat{\mu}_1 - \mu_1) \equiv T^2_2,
\]

Therefore, the distribution of \( T^2_2 \) is invariant under this transformation.
the statistic $T_2^2$ is invariant under the transformation (3.2.2). To show that the statistic $T_1^2$ is invariant under the transformation (3.2.2), let us analyze each term of

$$(T_1^2)^* \equiv (\hat{\mu}_2^* - \mu_2^* - \sum_{21}^* \sum_{11}^{-1} (\hat{\mu}_1^* - \mu_1^*))' \sum_{22-1}^* (\hat{\mu}_2^* - \mu_2^* - \sum_{21}^* \sum_{11}^{-1} (\hat{\mu}_1^* - \mu_1^*))$$

individually. The vector $\hat{\mu}_2^* - \mu_2^*$ transforms to

$$\Lambda_{22}(\Lambda_{21}\hat{\mu}_1 + \hat{\mu}_2) + \nu_2 - (\Lambda_{22}\mu_2 + \Lambda_{22}\Lambda_{21}\mu_1 + \nu_2)$$

$$= \Lambda_{22}(\hat{\mu}_2 - \mu_2) + \Lambda_{22}\Lambda_{21}(\hat{\mu}_1 - \mu_1).$$

The vector $\sum_{21}^* \sum_{11}^{-1} (\hat{\mu}_1^* - \mu_1^*)$ transforms to

$$\Lambda_{22}(\Lambda_{21}\hat{\Sigma}_{11} + \hat{\Sigma}_{21})\Lambda_{11}(\Lambda_{11}\hat{\Sigma}_{11}\Lambda_{11})^{-1}(\Lambda_{11}\hat{\mu}_1 + \nu_1 - \Lambda_{11}\mu_1 - \nu_1)$$

$$= \Lambda_{22}(\Lambda_{21} + \hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1})(\hat{\mu}_1 - \mu_1).$$

In addition,

$$\hat{\Sigma}_{22-1} = \Lambda_{22}(\Lambda_{21}\hat{\Sigma}_{11}A_{12} + \hat{\Sigma}_{21}\Lambda_{12} + \Lambda_{21}\hat{\Sigma}_{12} + \hat{\Sigma}_{22})\Lambda_{22}$$

$$- \Lambda_{22}(\Lambda_{21}\hat{\Sigma}_{11} + \hat{\Sigma}_{21})\Lambda_{11}(\Lambda_{11}\hat{\Sigma}_{11}\Lambda_{11})^{-1}A_{11}(\hat{\Sigma}_{11}A_{12} + \hat{\Sigma}_{12})\Lambda_{22}$$

$$= \Lambda_{22}(\hat{\Sigma}_{22} - \hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12})\Lambda_{22}$$

$$= \Lambda_{22}\hat{\Sigma}_{22-1}\Lambda_{22}. $$
Consequently, \((T^2_1)^*\) equals

\[
(\widehat{\mu}_2 - \mu_2 - \widehat{\Sigma}_{21} \widehat{\Sigma}_{11}^{-1} (\widehat{\mu}_1 - \mu_1))' \Lambda_{22} (\Lambda_{22} \widehat{\Sigma}_{22,1} \Lambda_{22})^{-1} \Lambda_{22} (\widehat{\mu}_2 - \mu_2 - \widehat{\Sigma}_{21} \widehat{\Sigma}_{11}^{-1} (\widehat{\mu}_1 - \mu_1))
\]

\[
= (\widehat{\mu}_2 - \mu_2 - \widehat{\Sigma}_{21} \widehat{\Sigma}_{11}^{-1} (\widehat{\mu}_1 - \mu_1))' \Sigma_{22,1}^{-1} (\widehat{\mu}_2 - \mu_2 - \widehat{\Sigma}_{21} \widehat{\Sigma}_{11}^{-1} (\widehat{\mu}_1 - \mu_1))
\]

\[
\equiv T^2_1.
\]

Therefore, \(T^2_1\) and \(T^2_2\) both are invariant and hence, by (3.3.9), \(T^2\) also is invariant. \(\square\)

By taking \(\Lambda_{11} = \Sigma_{11}^{-1/2}\), \(\Lambda_{22} = \Sigma_{22,1}^{-1/2}\), and \(\Lambda_{12} = -\Sigma_{21} \Sigma_{11}^{-1}\), the covariance matrix of \((X', Y')'\) under this transformation is \(\Lambda C \Sigma C' \Lambda' = I_{p+q}\). We may then assume that the population covariance matrix is \(I_{p+q}\). Furthermore, by choosing \(\nu = -\Lambda C \mu\) we may assume \(\mu = 0\). Hence, in deriving the distribution of the \(T^2\)-statistic, we assume, without loss of generality, that \(\mu = 0\) and \(\Sigma = I_{p+q}\).

We now derive a stochastic representation for the exact distribution of the \(T^2\)-statistic (3.3.1). The proof of this result is lengthy, relying on characteristic functions and repeated applications of the powerful method of orthogonal invariance. The resulting stochastic representation, however, involves only chi-square and Beta random variables, and a 2×2 Wishart matrix, all mutually independent. Thus, the stochastic representation is straightforward to simulate.

**Theorem 3.3.3.** Let \(\cos^2 \theta \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}(p-1)\right)\), \(Q_1 \sim \chi_p^2\), \(Q_2 \sim \chi_p^2\), \(Q_3 \sim \chi_{n-p-q}^2\), \(Q_4 \sim \chi_q^2\), \(W \sim W_2(N - p, I_2)\), and \(\beta \sim \text{Beta}\left(\frac{n-p-2}{2}, \frac{(N-n-1)}{2}\right)\) be mutually
independent. Then, 

\[ T^2 = \frac{NQ_4}{\gamma Q_3} \left( 1 + Q_1 \beta^{-1} e_1' W^{-1} e_1 \right) + \left( \begin{array}{c} \sqrt{n} Q_1^{1/2} + \sqrt{N-nQ_2^{1/2}} \cos \theta \\ \sqrt{N-nQ_2^{1/2}} \sin \theta \end{array} \right)' W^{-1} \left( \begin{array}{c} \sqrt{n} Q_1^{1/2} + \sqrt{N-nQ_2^{1/2}} \cos \theta \\ \sqrt{N-nQ_2^{1/2}} \sin \theta \end{array} \right) \]

\[ - \frac{\tau Q_1 + \tau Q_2 - 2(\tau Q_1 \tau Q_2)^{1/2} \cos \theta}{1 + (\tau Q_1 + \tau Q_2 - 2(\tau Q_1 \tau Q_2)^{1/2} \cos \theta) e_1' W^{-1} e_1} \cdot \left[ e_1' W^{-1} \left( \begin{array}{c} \sqrt{n} Q_1^{1/2} + \sqrt{N-nQ_2^{1/2}} \cos \theta \\ \sqrt{N-nQ_2^{1/2}} \sin \theta \end{array} \right) \right]^2 \] 

(3.3.10)

Proof. We assume, without loss of generality, that \( \mu = 0 \) and \( \Sigma = I_{p+q} \). Recall from (3.3.7) - (3.3.9) that

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

where

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

and

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

By elementary properties of the multivariate normal distribution,

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

(3.3.11)
and by Proposition 2.3.3(i), $\mathbf{A}_{22-1,n} \sim W_q(n-p-1, \mathbf{I}_q)$ and is independent of $\{\mathbf{A}_{12}, \mathbf{A}_{11,n}\}$.

Define

$$Q_3 = \frac{n(\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1)'(\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1)}{\mathbf{T}_1^2},$$

then, by Proposition 2.3.3(iv), $Q_3|\{\mathbf{A}_{12}, \mathbf{A}_{11,n}, \bar{\mathbf{X}}_1\} \sim \chi^2_{n-p-q}$, and $Q_3$ is independent of $\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1$. Because this distribution does not depend on $\{\mathbf{A}_{12}, \mathbf{A}_{11,n}, \bar{\mathbf{X}}_1\}$, then $Q_3$ is also independent of $\{\mathbf{A}_{12}, \mathbf{A}_{11,n}, \bar{\mathbf{X}}_1\}$. Therefore,

$$\mathbf{T}_1^2 = \frac{n(\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1)'(\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1)}{Q_3},$$

where $Q_3 \sim \chi^2_{n-p-q}$ and the numerator and denominator are mutually independent. By (3.3.11),

$$n(\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1)'(\bar{\mathbf{Y}} - \mathbf{A}_{21} \mathbf{A}_{11,n}^{-1} \bar{\mathbf{X}}_1) \sim \chi^2_q(n\bar{\mathbf{X}}_1'A_{11,n}^{-1}\mathbf{A}_{12}A_{21}A_{11,n}^{-1}\bar{\mathbf{X}}_1), \quad (3.3.12)$$

a noncentral chi-square distribution with $q$ degrees of freedom and noncentrality parameter $n\bar{\mathbf{X}}_1'A_{11,n}^{-1}\mathbf{A}_{12}A_{21}A_{11,n}^{-1}\bar{\mathbf{X}}_1$. 
Let \( t \in \mathbb{R} \). By Lemma 3.3.1, the characteristic function of \( T^2/N \) is

\[
E \exp[itN^{-1}T^2] = E \exp \left[ it \left( \frac{1}{\gamma Q_3} n \left( \bar{Y} - A_{21} A_{11,n}^{-1} \bar{X}_i \right)' \left( \bar{Y} - A_{21} A_{11,n}^{-1} \bar{X}_1 \right) + N \bar{X}' (A_{11,n} + B_1 + B_2)^{-1} \bar{X} \right) \right] 
\]

\[
= E_{Q_3} E_{X_1} E_{X_2} E_{B_1} E_{A_{21}, A_{11,n}} \exp \left[ itN \bar{X}' (A_{11,n} + B_1 + B_2)^{-1} \bar{X} \right] 
\cdot E_{Y | \{X_1, X_2, A_{21}, A_{11,n}, B_1\}} \exp \left[ it \frac{1}{\gamma Q_3} n \left( \bar{Y} - A_{21} A_{11,n}^{-1} \bar{X}_1 \right)' \left( \bar{Y} - A_{21} A_{11,n}^{-1} \bar{X}_1 \right) \right] 
\]  

(3.3.13)

Applying the formula (2.2.6) for the characteristic function of the noncentral \( \chi^2 \) distribution to (3.3.12) and inserting the result into (3.3.13) yields

\[
E \exp[itN^{-1}T^2] = E_{Q_3} E_{X_1} E_{X_2} E_{B_1} E_{A_{21}, A_{11,n}} \exp \left[ itN \bar{X}' \left( A_{11,n} + B_1 + B_2 \right)^{-1} \bar{X} \right] 
\cdot \left( 1 - \frac{2it}{\gamma Q_3} \right)^{-q/2} \exp \left[ \frac{itn \bar{X}' A_{11,n}^{-1} A_{12} A_{21} A_{11,n}^{-1} \bar{X}_1}{\gamma Q_3 - 2it} \right]. 
\]  

(3.3.14)

By Proposition 2.3.3 ii), \( A_{21} | A_{11,n} \sim N(0, I_q \otimes A_{11,n}) \); therefore (3.3.14) equals

\[
E_{Q_3} \left( 1 - \frac{2it}{\gamma Q_3} \right)^{-q/2} E_{X_1} E_{X_2} E_{B_1} E_{A_{11,n}} \exp \left[ itN \bar{X}' \left( A_{11,n} + B_1 + B_2 \right)^{-1} \bar{X} \right] 
\cdot E_{A_{21} | A_{11,n}} \exp \left[ \frac{itn \bar{X}' A_{11,n}^{-1} A_{12} A_{21} A_{11,n}^{-1} \bar{X}_1}{\gamma Q_3 - 2it} \right]. 
\]  

(3.3.15)
By Lemma 2.3.2 with $U = -itn/(\gamma Q_3 - 2it)I_q$, $B_{12} = A_{21}$, $D = A_{11,n}$, $C = I_q$, and $\Lambda = \bar{X}_1 \bar{X}_1'$, we have

$$E_{A_{21}|A_{11,n}} \exp \left[ \frac{itn \bar{X}_1' A_{11,n}^{-1} A_{12} A_{21} A_{11,n}^{-1} \bar{X}_1}{\gamma Q_3 - 2it} \right] = E_{A_{21}|A_{11,n}} \exp \left[ \frac{itn \text{ tr} \left( A_{21} A_{11,n}^{-1} \bar{X}_1 \bar{X}_1' A_{11,n}^{-1} A_{12} \right)}{\gamma Q_3 - 2it} \right]$$

$$= \left| I_{pq} - \frac{2itn}{\gamma Q_3 - 2it} A_{11,n}^{-1/2} \bar{X}_1 \bar{X}_1' A_{11,n}^{-1/2} \right|^{-1/2} \cdot \left( 1 - \frac{2itn}{\gamma Q_3 - 2it} \bar{X}_1 A_{11,n}^{-1} \bar{X}_1 \right)^{-q/2}.$$

Substituting this result into (3.3.15) yields,

$$E \exp[itN^{-1}T^2]$$

$$= E_{Q_3} \left( 1 - \frac{2it}{\gamma Q_3} \right)^{-q/2} E_{X_1} E_{X_2} E_{B_1} E_{A_{11,n}} \exp \left[ itN \bar{X}_1' \left( A_{11,n} + B_1 + B_2 \right)^{-1} \bar{X}_1 \right] \cdot \left( 1 - \frac{2itn}{\gamma Q_3 - 2it} \bar{X}_1 A_{11,n}^{-1} \bar{X}_1 \right)^{-q/2}.$$

Because

$$\left( 1 - \frac{2it}{\gamma Q_3} \right) \left( 1 - \frac{2itn}{\gamma Q_3} \bar{X}_1 A_{11,n}^{-1} \bar{X}_1 \right) = \frac{\gamma Q_3 - 2it - 2itn \bar{X}_1 A_{11,n}^{-1} \bar{X}_1}{\gamma Q_3},$$
it follows that
\[
E \exp [itN^{-1}T^2] = E_{Q_3} E_{\bar{X}_1} E_{\bar{X}_2} E_{B_1} E_{A_{11,n}} \exp \left[ itN \bar{X}' (A_{11,n} + B_1 + B_2)^{-1} \bar{X} \right]
\]
\[
\cdot \left( 1 - \frac{2it(1 + n\bar{X}' A_{11,n}^{-1} \bar{X})}{\gamma Q_3} \right) -q/2
\]
\[
= E_{Q_3} E_{\bar{X}_1} E_{\bar{X}_2} E_{B_1} E_{A_{11,n}} \exp \left[ itN \bar{X}' (A_{11,n} + B_1 + B_2)^{-1} \bar{X} \right]
\]
\[
\cdot E_{Q_4} \exp \left[ \frac{itQ_4}{\gamma Q_3} (1 + n\bar{X}' A_{11,n}^{-1} \bar{X}_1) \right],
\]
where \( Q_4 \sim \chi^2_q \) and \( Q_4 \) is independent of \( Q_3, \bar{X}_1, \bar{X}_2, B_1, A_{11,n} \). Hence,
\[
E \exp [itN^{-1}T^2]
\]
\[
= E_{Q_3} E_{Q_4} \exp \left[ \frac{itQ_4}{\gamma Q_3} \right] E_{\bar{X}_1} E_{\bar{X}_2} E_{B_1} E_{A_{11,n}} \exp \left[ itN \bar{X}' (A_{11,n} + B_1 + B_2)^{-1} \bar{X} \right]
\]
\[
\cdot \exp \left[ \frac{itnQ_4}{\gamma Q_3} \bar{X}' A_{11,n}^{-1} \bar{X}_1 \right].
\]
(3.3.16)

By applying the method of orthogonal invariance, we shall simplify the above characteristic function greatly. For fixed \( Q_3 \) and \( Q_4 \), define the function
\[
f(\bar{X}_1, \bar{X}_2) = E_{B_1} E_{A_{11,n}} \exp \left[ itN \bar{X}' (A_{11,n} + B_1 + B_2)^{-1} \bar{X} \right] \exp \left[ \frac{itnQ_4}{\gamma Q_3} \bar{X}' A_{11,n}^{-1} \bar{X}_1 \right].
\]

We first verify that \( f(\bar{X}_1, \bar{X}_2) \) is invariant under the transformation \( (\bar{X}_1, \bar{X}_2) \rightarrow (H \bar{X}_1, H \bar{X}_2) \), where \( H \in O(p) \), the set of orthogonal \( p \times p \) matrices. Recall that \( B_2 \) is a function of \( \bar{X}_1 \) and \( \bar{X}_2 \), (3.3.6). Suppose \( (\bar{X}_1, \bar{X}_2) \) is replaced by \( (H \bar{X}_1, H \bar{X}_2) \), then the last
two exponential terms in (3.3.16) have exponents

\[ itN(H(\tau \bar{X}_1 + \bar{X}_2)')(A_{11,n} + B_1 + n\bar{\tau}H(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)'H')^{-1}(H(\tau \bar{X}_1 + \bar{X}_2)) \]  \hspace{1cm} (3.3.17)

and

\[ \frac{itnQ_4}{\gamma Q_3} (H\bar{X}_1)'A_{11,n}^{-1}(H\bar{X}_1). \]  \hspace{1cm} (3.3.18)

Because \( A_{11,n} \sim W_p(n-1, I_p) \), then also \( H A_{11,n}H' \sim W_p(n-1, I_p) \) and a similar result holds for \( B_1 \). Then the random variable in (3.3.17) is equal in distribution to

\[
\bar{X}'H'(HA_{11,n}H' + HB_1H' + n\bar{\tau}H(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)'H')^{-1}H\bar{X}
= \bar{X}'H'(H(A_{11,n} + B_1 + n\bar{\tau}(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2))'H')^{-1}H\bar{X}
= \bar{X}'(A_{11,n} + B_1 + n\bar{\tau}(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2))^{-1}\bar{X},
\]

and, similarly, the random variable in (3.3.18) is equal in distribution to

\[
\frac{itnQ_4}{\gamma Q_3} (H\bar{X}_1)'A_{11,n}^{-1}(H\bar{X}_1) = \frac{itnQ_4}{\gamma Q_3} \bar{X}'A_{11,n}^{-1}\bar{X}.
\]

Therefore, \( f(\bar{X}_1, \bar{X}_2) = f(H\bar{X}_1, H\bar{X}_2) \).

Because \( B_2 = n\bar{\tau}(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)' \) is of rank 1, then by Proposition 2.1.1 (v), there exists \( \tilde{H} \in O(p) \) such that

\[
\tilde{H}(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)'\tilde{H}' = \left( \begin{array}{cc} \|\bar{X}_1 - \bar{X}_2\|^2 & 0 \\ 0 & 0 \end{array} \right) = \|\bar{X}_1 - \bar{X}_2\|^2 e_1 e_1'.
\]
where \( e_1 = (1, 0, \ldots, 0)' \in \mathbb{R}^p \). Therefore, we may replace 

\[(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}_2)' \] by \( \bar{H}'\|\bar{X}_1 - \bar{X}_2\|^2 e_1 e_1' \bar{H} \) in (3.3.16). In addition, by replacing \( (\bar{X}_1, \bar{X}_2) \) with \( (\tilde{H}'\bar{X}_1, \tilde{H}'\bar{X}_2) \) then, by orthogonal invariance, (3.3.16) becomes

\[
E_Q E_Q \exp \left[ \frac{itQ_4}{\gamma Q_3} \right] E_{\bar{X}_1} E_{\bar{X}_2} E_{B_1} \cdot E_{A_{11,n}} \exp \left[ \frac{itN}{\gamma Q_3} \tilde{X}' (A_{11,n} + B_1 + n\bar{\tau}\|\bar{X}_1 - \bar{X}_2\|^2 e_1 e_1')^{-1}\tilde{X} \right] 
\cdot \exp \left[ \frac{itnQ_4}{\gamma Q_3} \tilde{X}' A_{11,n}^{-1} \tilde{X} \right].
\]

(3.3.19)

We make one last orthogonal transformation. There exists an orthogonal matrix \( C \in O(p) \) with first row \( \bar{X}_1'/\|\bar{X}_1\| \); we may construct the remaining rows of \( C \) using the Gram-Schmidt orthogonalization process. We transform \( \bar{X}_1 \) to \( C\bar{X}_1 = \|\bar{X}_1\|e_1 \) and \( \bar{X}_2 \) to \( \alpha e_1 + \beta e_2 \), where \( e_1 \) is defined as before, \( e_2 = (0, 1, 0, \ldots, 0)' \in \mathbb{R}^p \), and \( \alpha \) and \( \beta \) are such that

\[
\alpha\|\bar{X}_1\| = \bar{X}_1'\bar{X}_2, \quad \alpha^2 + \beta^2 = \|\bar{X}_2\|^2.
\]

Let \( \theta \) be the angle between \( \bar{X}_1 \) and \( \bar{X}_2 \) and recall that \( \cos \theta = \bar{X}_1'\bar{X}_2/\|\bar{X}_1\|\|\bar{X}_2\| \). Then

\[
\alpha = \frac{\bar{X}_1'\bar{X}_2}{\|\bar{X}_1\|} = \|\bar{X}_2\| \cos \theta,
\]

and

\[
\beta = \left( \|\bar{X}_2\|^2 - \frac{(\bar{X}_1'\bar{X}_2)^2}{\|\bar{X}_1\|^2} \right)^{1/2} = \left( \frac{\|\bar{X}_1\|^2\|\bar{X}_2\|^2(1 - \cos^2 \theta)}{\|\bar{X}_1\|^2} \right)^{1/2} = \|\bar{X}_2\| \sin \theta.
\]

(3.3.20)
Therefore
\[ \bar{X} = \tau \bar{X}_1 + \bar{X}_2 = \tau \| \bar{X}_1 \| e_1 + \bar{X}_2 (e_1 \cos \theta + e_2 \sin \theta) \]

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

Because \( \sqrt{n} \bar{X}_1 \sim N_p(0, I_p) \) and \( \sqrt{N-n} \bar{X}_2 \sim N_p(0, I_p) \), then \( \bar{X}_1 \) and \( \bar{X}_2 \) are orthogonally invariant random vectors. Therefore \( \bar{X}_1 / \| \bar{X}_1 \| \) and \( \bar{X}_2 / \| \bar{X}_2 \| \) are mutually independent and uniformly distributed on \( S^{p-1} \), the unit sphere in \( \mathbb{R}^p \). Hence \( \cos \theta \overset{\mathcal{D}}{=} U'_1 U_2 \), where \( U_1 \) and \( U_2 \) are independent and uniformly distributed on \( S^{p-1} \). By Muirhead [25, p. 38], we then have \( \bar{X}_1, \bar{X}_2, \) and \( \theta \) are mutually independent, and \( \cos^2 \theta \sim \text{Beta}(1/2, (p-1)/2) \).

By (3.3.19),
\[
T^2 \overset{\mathcal{D}}{=} \frac{NQ_4}{\gamma Q_3} \left( 1 + n \| \bar{X}_1 \|^2 e_1' A_{11, n}^{-1} e_1 \right) \\
+ N^2 \left[ (\tau \| \bar{X}_1 \| + \bar{X}_2 \| \cos \theta) e_1 + \bar{X}_2 (e_1 \cos \theta + e_2 \sin \theta) \right]' \\
\cdot \left( A_{11, n} + B_1 + n \bar{X}_1 \| \bar{X}_2 \| \cos \theta \right) e_1 e_1' \left( \| \bar{X}_1 \|^2 + \| \bar{X}_2 \|^2 - 2 \| \bar{X}_1 \| \| \bar{X}_2 \| \cos \theta \right)^{-1} \\
\cdot \left[ (\tau \| \bar{X}_1 \| + \bar{X}_2 \| \cos \theta) e_1 + \bar{X}_2 (e_1 \cos \theta + e_2 \sin \theta) \right]. \tag{3.3.21}
\]

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

\begin{align*}
T^2 & \equiv \frac{NQ_4}{\gamma Q_3} \left( 1 + Q_1 e_1^t A_{11,n}^{-1} e_1 \right) \\
& \quad + N \left[ \left( \sqrt{\tau Q_1^{1/2}} + \sqrt{\tau Q_2^{1/2}} \cos \theta \right) e_1 + \sqrt{\tau Q_2^{1/2}} e_2 \sin \theta \right]' \\
& \quad \cdot \left( A_{11,n} + B_1 + \left( \tau Q_1 + \tau Q_2 - 2(\tau Q_1 \tau Q_2)^{1/2} \cos \theta \right) e_1 e_1^t \right)^{-1} \\
& \quad \cdot \left[ \left( \sqrt{\tau Q_1^{1/2}} + \sqrt{\tau Q_2^{1/2}} \cos \theta \right) e_1 + \sqrt{\tau Q_2^{1/2}} e_2 \sin \theta \right]. \quad (3.3.22)
\end{align*}

This representation involves \( p \times p \) Wishart matrices, so it would be nice to reduce the size of any such matrices appearing in the final result.

Next, we represent the distribution of \( T^2 \) in terms of a \( 2 \times 2 \) Wishart matrix. By Proposition 2.3.3 (v), \( L = (A_{11,n} + B_1)^{-1/2} A_{11,n} (A_{11,n} + B_1)^{1/2} \) is independent of \( P = A_{11,n} + B_1 \) and \( L \sim Beta_p((n - 1)/2, (N - n - 1)/2) \). Therefore, we may write (3.3.22) in terms of \( L \) and \( P \) as

\begin{align*}
T^2 & \equiv \frac{NQ_4}{\gamma Q_3} \left( 1 + Q_1 (P^{-1/2} e_1)' L^{-1} (P^{-1/2} e_1) \right) \\
& \quad + N \left[ \left( \sqrt{\tau Q_1^{1/2}} + \sqrt{\tau Q_2^{1/2}} \cos \theta \right) e_1 + \sqrt{\tau Q_2^{1/2}} e_2 \sin \theta \right]' \\
& \quad \cdot \left( P + (\tau Q_1 + \tau Q_2 - 2(\tau Q_1 \tau Q_2)^{1/2} \cos \theta) e_1 e_1^t \right)^{-1} \\
& \quad \cdot \left[ \left( \sqrt{\tau Q_1^{1/2}} + \sqrt{\tau Q_2^{1/2}} \cos \theta \right) e_1 + \sqrt{\tau Q_2^{1/2}} e_2 \sin \theta \right]. \quad (3.3.23)
\end{align*}

Because the distribution of \( L \) is invariant under orthogonal transformations, we may replace \( L \) by \( H L H' \), where \( H \in O(p) \). We now choose \( H \) to be the orthogonal matrix
with first row $P^{-1/2}e_1'/\|P^{-1/2}e_1\|$ and with all the remaining rows of $H$ constructed using the Gram-Schmidt orthogonalization process. Then $(P^{-1/2}e_1)'L^{-1}P^{-1/2}e_1 = e_1'e_1 L^{-1}e_1$. By Muirhead [25], $\beta^{-1} = e_1' L^{-1}e_1 \sim 1/Beta((n-p-2)/2,(N-n-1)/2)$.

In order to simplify the notation let us also define $u = (u_1, u_2)' = N^{1/2}((\sqrt{\tau Q_1^{1/2}} + \sqrt{\tau Q_2^{1/2}} \cos \theta), \sqrt{\tau Q_2^{1/2}} \sin \theta)'$, and $v = \tau Q_1 + \tau Q_2 - 2(\tau Q_1 \tau Q_2)^{1/2} \cos \theta$. Then the representation (3.3.23) becomes

$$T^2 = \frac{\tilde{c} N Q_4}{\gamma Q_3} (1 + Q_1 \beta^{-1} e_1' P^{-1} e_1) + (u_1 e_1 + u_2 e_2)' (P + v e_1 e_1')^{-1} (u_1 e_1 + u_2 e_2).$$

(3.3.24)

Our final step is to specify the distribution of the remaining terms that involve $P$. The first term only involves $e_1' P^{-1} e_1$ and the second term involving $P$ may be simplified by Woodbury’s formula, (2.1.1), as follows:

$$(u_1 e_1 + u_2 e_2)' (P + v e_1 e_1')^{-1} (u_1 e_1 + u_2 e_2)$$

$$= (u_1 e_1 + u_2 e_2)' \left( P^{-1} - \frac{v P^{-1} e_1 e_1' P^{-1}}{1 + v e_1' P^{-1} e_1} \right) (u_1 e_1 + u_2 e_2)$$

$$= (u_1 e_1 + u_2 e_2)' P^{-1} (u_1 e_1 + u_2 e_2) - \frac{v}{1 + v e_1' P^{-1} e_1} [(u_1 e_1 + u_2 e_2)' P^{-1} e_1]^2.$$
Therefore, we derive the joint distribution of $e_1^t P^{-1} e_1, e_1^t P^{-1} e_2$, and $e_2^t P^{-1} e_2$, when $P \sim W_p(N - 2, I_p)$. Let $M = (e_1, e_2)^t$; noting that $MM^t = I_2$, it follows by Proposition 2.3.3 (iv), that $(MP^{-1}M')^{-1} \sim W_2(N - p, I_2)$. Let

$$W^{-1} = MP^{-1}M' = \begin{pmatrix} e_1^t P^{-1} e_1 & e_1^t P^{-1} e_2 \\ e_2^t P^{-1} e_1 & e_2^t P^{-1} e_2 \end{pmatrix}.$$ 

Then the stochastic representation, (3.3.24), reduces to

$$T^2 \equiv \frac{\gamma Q_3}{NQ_4} \left(1 + Q_1 \beta^{-1} e_1^t W^{-1} e_1\right) + u^t W^{-1} u - \frac{v}{1 + ve_1^t W^{-1} e_1} (e_1^t W^{-1} u)^2,$$  

(3.3.25)

where $W \sim W_2(N - p, I_2)$. □

**Remark 3.3.4.** We may take this one step further and represent the entire stochastic representation in terms of scalar mutually independent random variables. Let $w_{11} = e_1^t W^{-1} e_1, w_{22} = e_2^t W^{-1} e_2, w_{12} = e_1^t W^{-1} e_2$, and $\rho = w_{12}/\sqrt{w_{11}w_{22}}$. We may rewrite Equation (3.3.25) in terms of scalar random variables:

$$T^2 \equiv t \left(1 + Q_1 \beta^{-1} w_{11}\right) + (u_1 e_1 + u_2 e_2)^t W^{-1} (u_1 e_1 + u_2 e_2)$$

$$- \frac{v(u_1 w_{11} + u_2 \rho \sqrt{w_{11}w_{22}})^2}{1 + vw_{11}}$$

$$= t \left(1 + Q_1 \beta^{-1} w_{11}\right) + u_1^2 w_{11} + u_2^2 w_{22} + 2u_1 u_2 \rho \sqrt{w_{11}w_{22}}$$

$$- \frac{v(u_1 w_{11} + u_2 \rho \sqrt{w_{11}w_{22}})^2}{1 + vw_{11}}.$$
By Anderson [2], \( W = TT' \), where \( T \) is lower-triangular, that is

\[
T = \begin{pmatrix}
t_{11} & 0 \\
t_{21} & t_{22}
\end{pmatrix},
\]

and the entries of \( T \) are mutually independent, \( t_{jj}^2 \sim \chi^2_{N-p} \), and \( t_{ij} \sim N(0, 1), i \neq j \). It follows that

\[
W^{-1} = (T^{-1})'T^{-1}
\]

\[
= \frac{1}{t_{11}t_{22}} \begin{pmatrix} t_{22} & -t_{21} \\ 0 & t_{11} \end{pmatrix} \begin{pmatrix} t_{22} & 0 \\ -t_{21} & t_{11} \end{pmatrix}
\]

\[
= \frac{1}{t_{11}^2 t_{22}^2} \begin{pmatrix} t_{22}^2 + t_{21}^2 & -t_{21} t_{11} \\ -t_{21} t_{11} & t_{11}^2 \end{pmatrix}.
\]

(3.3.26)

Therefore the joint distribution of \( \{w^{11}, w^{22}, \rho^2\} \) is equal to the joint distribution of

\[
\{(t_{22}^2 + t_{21}^2)/t_{11}^2 t_{22}^2, 1/t_{22}^2, t_{21}^2 / (t_{21}^2 + t_{22}^2)\} = \{(Q_5 + Q_7)/Q_5, Q_6, 1/Q_6, Q_7/(Q_6 + Q_7)\},
\]

where \( Q_5 \sim \chi^2_{N-p}, Q_6 \sim \chi^2_{N-p} \), and \( Q_7 \sim \chi^2_1 \).

**Remark 3.3.5.** We have provided in the stochastic representation the distribution of \( \cos^2 \theta \). Because the distribution of \( \cos \theta \) may also be desired, we provide the details here.

Recall from Theorem 3.3.10 that \( \cos^2 \theta \sim Beta\left(\frac{1}{2}, \frac{1}{2}(p-1)\right) \), where \( \cos \theta \in (-1, 1) \) and \( \cos \theta \leq - \cos \theta \). Let \( \alpha = \frac{1}{2}, \beta = \frac{1}{2}(p-1) \), \( X = \cos \theta \), and \( Y = \cos^2 \theta \sim Beta(\alpha, \beta) \). Let \( t \in (0, 1) \), then, because the distribution of \( \cos \theta \) is symmetric,

\[
P(X > t) = \frac{1}{2}(1 - P(-t < X < t)) = \frac{1}{2}(1 - P(|X| < t)) = \frac{1}{2}(1 - P(Y < t^2)).
\]
Therefore $P(X \leq t) = \frac{1}{2}(1 + P(Y < t^2))$. Because the probability density function, p.d.f, is the derivative of the cumulative distribution function, $f_X(t)$, the p.d.f of $X$, equals $tf_Y(t^2)$, $0 < t < 1$, where $f_Y(t)$ is the p.d.f of $Y$. Similarly, $f_X(t) = -tf_Y(t^2)$, $-1 < t \leq 0$ and therefore $f_X(t) = |t|f_Y(t^2)$, $-1 < t < 1$.

### 3.4 Probability inequalities for the $T^2$-statistic

Because the exact distribution of $T^2$ is complicated, it would be useful to find simpler upper and lower bounds on its distribution. Chang and Richards [6] found upper and lower bounds for the distribution function of the $T^2$-statistic [6] and Krishnamoorthy and Pannala [20] obtained an approximation to the distribution of their $T^2$-statistic by means of an $F$-distributed 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 in [6] and [20].

**Theorem 3.4.1.** Let $Q_1 \sim \chi^2_p$, $Q_3 \sim \chi^2_{n-p-q}$, $Q_4 \sim \chi^2_q$, $Q_8 \sim \chi^2_{N-p-1}$, and $\beta \sim \text{Beta}((n-p-2)/2,(N-n-1)/2)$ be mutually independent. For $t \geq 0$,

$$P(T^2 \leq t) \leq P\left(\frac{NQ_4}{\gamma Q_3} \left(1 + \frac{Q_1}{Q_8 \beta} \right) \leq t\right).$$

(3.4.1)

**Proof.** It is straightforward from (3.3.24) that

$$\hat{\mu}' \text{Cov}(\hat{\mu})^{-1} \hat{\mu} \geq \frac{NQ_4}{\gamma Q_3} \left(1 + Q_1 \beta^{-1} e_1' P^{-1} e_1 \right).$$

(3.4.2)
By Proposition 2.3.3(iv),

\[ Q_8 = \frac{e'_1 e_1}{e_1' P^{-1} e_1} \sim \chi^2_{N-p-1}. \]

The conclusion now follows from the definition of the stochastic bound, \( \geq \).

**Theorem 3.4.2.** Let \( Q_1 \sim \chi^2_p, Q_2 \sim \chi^2_p, Q_3 \sim \chi^2_{n-p-q}, Q_4 \sim \chi^2_q, Q_5 \sim \chi^2_{N-p}, Q_6 \sim \chi^2_{N-p}, Q_7 \sim \chi^2_1, \) and \( \beta \sim \text{Beta}((n-p-2)/2, (N-n-1)/2) \) be mutually independent.

For \( t \geq 0 \),

\[
\begin{align*}
\Pr(T^2 \leq t) & \geq \Pr \left[ \frac{NQ_4}{\gamma Q_3} \left( 1 + \frac{Q_1 (Q_6 + Q_7)}{Q_5 \tilde{Q}_6 \beta} \right) + \max \left( \frac{N \tilde{\tau} Q_2 (Q_6 + Q_7)}{Q_5 Q_6}, \frac{N \tilde{\tau} Q_2 Q_7}{Q_6 + Q_7} \right) + \frac{N (\tau Q_1 + \sqrt{\tau \tilde{\tau} Q_1 Q_2} (Q_6 + Q_7))}{Q_5 Q_6} + \frac{N (2 \sqrt{\tau \tilde{\tau} Q_1 Q_2} + \tau Q_2)}{Q_6} \leq t \right].
\end{align*}
\]

(3.4.3)

**Proof.** We refer to the stochastic representation for \( T^2 \) at the end of the proof, [3.3.25].

Because \( \cos \theta \leq 1 \), it follows that

\[
v = \tilde{\tau} Q_1 + \tau Q_2 - 2(\tilde{\tau} Q_1 \tau Q_2)^{1/2} \cos \theta
\]

\[
\geq \tilde{\tau} Q_1 + \tau Q_2 - 2(\tilde{\tau} Q_1 \tau Q_2)^{1/2}
\]

\[
= ((\tilde{\tau} Q_1)^{1/2} - (\tau Q_2)^{1/2})^2 \geq 0.
\]

Therefore,

\[
T^2 \geq \frac{NQ_4}{\gamma Q_3} \left( 1 + \frac{Q_1 e'_1 W^{-1} e_1}{\beta} \right) + (u_1 e_1 + u_2 e_2)' W^{-1} (u_1 e_1 + u_2 e_2)
\]

\[
= \frac{NQ_4}{\gamma Q_3} \left( 1 + \frac{Q_1 e'_1 W^{-1} e_1}{\beta} \right) + u_1^2 e_1 W^{-1} e_1 + u_2^2 e'_2 W^{-1} e_2 + 2u_1 u_2 e'_1 W^{-1} e_2.
\]
By Remark 3.3.4, the definition of \( u \), and using \( \cos \theta \sin \theta = \frac{1}{2} \sin 2\theta \), the lower bound becomes

\[
T^2 \leq \frac{NQ_1}{\gamma Q_3} \left( 1 + \frac{Q_1(Q_6 + Q_7)}{Q_5 Q_6 \theta} \right) + N(\tau Q_1 + \tau Q_2 \cos^2 \theta + \sqrt{\tau \tau Q_1 Q_2 \cos \theta}) \frac{Q_6 + Q_7}{Q_5 Q_6} + N\tau Q_2 \frac{Q_7}{Q_6 + Q_7} \sin^2 \theta
\]
\[
+ \frac{N(2\sqrt{\tau \tau Q_1}Q_1^{1/2}Q_2^{1/2} \sin \theta + \tau Q_2 \sin 2\theta)}{Q_6}.
\]  

(3.4.4)

Let \( a_1 = N\tau Q_2(Q_6 + Q_7)/Q_5 Q_6, a_2 = N\sqrt{\tau \tau Q_1 Q_2}(Q_6 + Q_7)/Q_5 Q_6, b_1 = N\tau Q_2 Q_7/(Q_6 + Q_7), b_2 = N2\sqrt{\tau \tau Q_1}Q_1^{1/2}Q_2^{1/2}/Q_6, \) and \( b_3 = N\tau Q_2/Q_6 \). To simplify this further, we now obtain an upper bound on all terms involving \( \theta \). The terms involving \( \theta \) may be written as

\[
f(\theta) = a_1 \cos^2 \theta + a_2 \cos \theta + b_1 \sin^2 \theta + b_2 \sin \theta + b_3 \sin 2\theta
\]
\[
= \frac{1}{2}(a_1 + b_1) + \left[ \frac{1}{2}(a_1 - b_1) \cos 2\theta + b_3 \sin 2\theta \right] + [a_2 \cos \theta + b_2 \sin \theta],
\]

where the second equality follows because \( \cos^2 \theta = 1 - \sin^2 \theta = (1 + \cos 2\theta)/2 \). Because \( a \cos \theta + b \sin \theta \leq (a^2 + b^2)^{1/2} \) for all \( \theta \), we obtain

\[
f(\theta) \leq \frac{1}{2}(a_1 + b_1) + \left( \frac{1}{4}(a_1 - b_1)^2 + b_3^2 \right)^{1/2} + (a_2^2 + b_2^2)^{1/2}
\]
\[
\leq \frac{1}{2}(a_1 + b_1) + \frac{1}{2}|a_1 - b_1| + b_3 + a_2 + b_2
\]
\[
= \max(a_1, b_1) + a_2 + b_2 + b_3,
\]  

(3.4.5)
where the second inequality holds because \( \max(a, b) = (a + b + |a - b|)/2 \). Substituting the terms for the lower bound on the cumulative distribution function of \( T^2 \), (3.4.5), yields

\[
T^2 \geq \frac{NQ_4}{\gamma Q_3} \left( 1 + \frac{Q_1(Q_6 + Q_7)}{Q_5Q_6\beta} \right) + \max \left( \frac{N\tau Q_2(Q_6 + Q_7)}{Q_5Q_6}, \frac{N\tau Q_2 Q_7}{Q_6 + Q_7} \right) + \frac{N(\tau Q_1 + \sqrt{\tau \tau Q_1 Q_2})(Q_6 + Q_7)}{Q_5Q_6} + \frac{N(2\sqrt{\tau \tau Q_1 Q_2} + \tau Q_2)}{Q_6}.
\]

(3.4.6)

By means of numerical simulations, it appears that our bounds are more precise than those previously available, leading to less conservative ellipsoidal confidence regions for \( \mu \). Using the statistical package R [26], we simulated our bounds, those derived by Chang and Richards [6], and the exact distribution 3,000 times for various values of \( p, q, n, \) and \( N \). We then plotted the empirical cumulative distribution function for 150 values of \( t \) from 0 to 100. It is clear from Figures 3.1 - 3.4 that the bounds in Theorem [3.4.1] and Theorem [3.4.2] perform the best overall; in all cases, the bounds either are closest to the exact distribution or are similar to the other bounds. Based on the simulations, our bounds appear to converge as \( t \to \infty \) to the exact distribution faster than the bounds obtained by Chang and Richards [6].

We also compared the bounds in Theorem [3.4.1] and Theorem [3.4.2] to the approximation of Krishnamoorthy and Pannala [20]. We have found their approximation to be very close to the exact distribution of the \( T^2 \)-statistic; see Figures 3.5 and 3.6. In order for their approximation to be valid, it is required that \( n > p + q + 4 \). Although their approximation is quite close to the exact distribution for those values of \( n, N, p, \) and \( q, \)
no error analysis is available, so there is no way to know when it will be above or below
the exact distribution. When it is above the exact distribution, it leads to unconservative
bounds. It is also of note that Krishnamoorthy and Pannala did not use the maximum
likelihood estimator for $\text{Cov}(\hat{\mu})$, so it is interesting that their approximation is close to
the exact distribution.

3.5 Applications of the $T^2$-statistic

We now provide two applications of the stochastic representation for the exact
finite-sample distribution of the $T^2$-statistic: the construction of simultaneous confidence
intervals for linear combinations of $\mu$ and prediction regions. At the end of this section,
we also apply these results to analyze the Pennsylvania cholesterol data.

3.5.1 Simultaneous confidence intervals for linear functions of $\mu$

The $T^2$-statistic has been used in the complete data setting, ([2], Section 5.3.3),
to construct ellipsoidal confidence regions for linear combinations, $v' \mu$, of $\mu$. Similar
results have been obtained in the monotone incomplete data setting for the likelihood
ratio test (cf. Little [22], Srivastava [31]).

As in the classical case, for $v \in \mathbb{R}^{p+q}$, we apply the generalized Cauchy-Schwarz
inequality

$$[v' (\hat{\mu} - \mu)]^2 \leq v' \text{Cov}(\hat{\mu}) v \cdot (\hat{\mu} - \mu)' \text{Cov}(\hat{\mu})^{-1} (\hat{\mu} - \mu) = v' \text{Cov}(\hat{\mu}) v \cdot T^2$$
to obtain
\[
\sqrt{T^2} \geq \frac{|v'\hat{\mu} - v'\mu|}{\sqrt{v'\text{Cov}(\hat{\mu})v}}.
\]
Solving this inequality for \(v'\mu\), it follows that if \(T^2_{\alpha}\) is a 100(1 - \(\alpha\))% percentage point for the \(T^2\)-statistic. Then, with confidence at least 1 - \(\alpha\), \(\mu\) satisfies for all \(v\) the simultaneous inequalities
\[
v'\mu \in \left( v'\hat{\mu} \pm \sqrt{T^2_{\alpha} \cdot \sqrt{v'\text{Cov}(\hat{\mu})v}} \right),
\]
and this result is applicable using the bounds for \(T^2_{\alpha}\) that were discussed in Section (3.4).

3.5.2 Ellipsoidal prediction regions for future observations

Consider the problem of predicting \(Z\), a future, independent, complete observation from \(N_d(\mu, \Sigma)\). In the complete sample case, the problem of constructing a prediction ellipsoid for \(Z\) is well-known (Johnson and Wichern [16]). We explore a possible path to a solution of this problem in the case of monotone samples using our methods. We hit a barrier in this part of our research, nevertheless, we present our ideas up to that point.
As before, let \( \hat{\mu} \) and \( \hat{\Sigma} \) be constructed as usual from the monotone incomplete sample (3.1.1). Because \( Z \sim N_d(\mu, \Sigma) \) and \( Z \) is independent of \( \hat{\mu} \) then, by (3.3.3),

\[
\text{Cov}(Z - \hat{\mu}) = \text{Cov}(Z) + \text{Cov}(\hat{\mu})
\]

\[
= \Sigma + \frac{1}{N} \Sigma + \frac{(n - 2)\bar{\tau}}{n(n - p - 2)} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{22} \end{pmatrix}
\]

\[
= (1 + \frac{1}{N})\Sigma + \frac{(n - 2)\bar{\tau}}{n(n - p - 2)} \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{22} \end{pmatrix}
\];

therefore, the maximum likelihood estimator of \( \text{Cov}(Z - \hat{\mu}) \) is

\[
\hat{\text{Cov}}(Z - \hat{\mu}) = (1 + \frac{1}{N})\hat{\Sigma} + \frac{(n - 2)\bar{\tau}}{n(n - p - 2)} \begin{pmatrix} 0 & 0 \\ 0 & \hat{\Sigma}_{22} \end{pmatrix}.
\quad (3.5.1)
\]

Because the form of this estimated covariance matrix is similar to that of \( \hat{\text{Cov}}(\hat{\mu}) \), it is natural to utilize the statistic \( T_2^2 = (Z - \hat{\mu})'(\hat{\text{Cov}}(Z - \hat{\mu}))^{-1}(Z - \hat{\mu}) \) in constructing a prediction region for \( Z \).

Because \( Z, \hat{\mu}, \) and \( \hat{\Sigma} \) are equivariant under the transformation (3.2.2), it follows that the \( T_2^2 \)-statistic is also invariant under that transformation. Therefore, we assume, without loss of generality, that \( \mu = 0 \) and \( \Sigma = I_{p+q} \); see (3.2.14) for the specific transformation. Following our approach to the exact distribution of the \( T_2^2 \)-statistic, we first apply the quadratic identity (2.1.4) with \( x \equiv Z - \hat{\mu} \) and \( \Lambda \equiv N\hat{\text{Cov}}(Z - \hat{\mu}) \),
obtaining

\[
\frac{1}{N} T_{Z}^{2} = (Z - \hat{\mu})' \left( N\tilde{\text{Cov}}(Z - \hat{\mu}) \right)^{-1} (Z - \hat{\mu})
\]
\[
= \begin{pmatrix} Z_1 - \hat{\mu}_1 \\ Z_2 - \hat{\mu}_2 \end{pmatrix}' \left[ (N + 1) \tilde{\Sigma} + (\gamma - 1) \begin{pmatrix} 0 & 0 \\ 0 & \tilde{\Sigma}_{22} \end{pmatrix} \right]^{-1} \begin{pmatrix} Z_1 - \hat{\mu}_1 \\ Z_2 - \hat{\mu}_2 \end{pmatrix}
\]
\[
= (Z_1 - \hat{\mu}_1)' \left( (N + 1) \tilde{\Sigma}_{11}^{-1} \right) (Z_1 - \hat{\mu}_1)
\]
\[
+ (Z_2 - \hat{\mu}_2 - \tilde{\Sigma}_{21} \tilde{\Sigma}_{11}^{-1} (Z_1 - \hat{\mu}_1))' \left( N\tilde{\text{Cov}}(Z - \hat{\mu})_{22,1} \right)^{-1} \cdot (Z_2 - \hat{\mu}_2 - \tilde{\Sigma}_{21} \tilde{\Sigma}_{11}^{-1} (Z_1 - \hat{\mu}_1))
\]

By (3.1.4) and (3.1.5),

\[
Z_2 - \hat{\mu}_2 - \tilde{\Sigma}_{21} \tilde{\Sigma}_{11}^{-1} (Z_1 - \hat{\mu}_1) = Z_2 - \bar{Y} - A_{21} A_{11,n}^{-1} (Z_1 - \bar{X}_1);
\]

by (3.5.1), \( N\tilde{\text{Cov}}(Z - \hat{\mu})_{22,1} = n^{-1}(N + \gamma) A_{22,1,n} \); and by (3.3.4), \( A_{11,N} = A_{11,n} + B_1 + B_2 \). Therefore,

\[
\frac{1}{N} T_{Z}^{2} = (N + \gamma)^{-1} T_{Z1}^{2} + T_{Z2}^{2},
\]

where

\[
T_{Z1}^{2} = n(Z_2 - \bar{Y} - A_{21} A_{11,n}^{-1} (Z_1 - \bar{X}_1))' A_{22,1,n}^{-1} (Z_2 - \bar{Y} - A_{21} A_{11,n}^{-1} (Z_1 - \bar{X}_1)); \quad (3.5.2)
\]

and

\[
T_{Z2}^{2} = N(N + 1)(Z_1 - \bar{X})' (A_{11,n} + B_1 + B_2)^{-1} (Z_1 - \bar{X}). \quad (3.5.3)
\]
By elementary properties of the multivariate normal distribution, the conditional distribution of \( \sqrt{n}(Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1)) \) is

\[
\sqrt{n}(Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1)) \sim N_q((-\sqrt{n}(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1)), I_q), (3.5.4)
\]

and, by Proposition 2.3.3(i), \( A_{22,1,n} \sim W_q(n - p - 1, I_q) \) independently of \( \{A_{12}, A_{11,n}\} \).

Also, by Proposition 2.3.3(iv), conditional on \( \{A_{12}, A_{11,n}, \bar{X}_1, Z_1\} \),

\[
Q_3 \equiv \frac{n(Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))' (Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))}{T^2_{Z1}} \sim \chi^2_{n-p-q};
\]

hence, \( Q_3 \) is independent of \( Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1) \). Because this distribution does not depend on \( \{A_{12}, A_{11,n}, \bar{X}_1, Z_1\} \), it follows that \( Q_3 \) is also independent of \( \{A_{12}, A_{11,n}, \bar{X}_1, Z_1\} \). Therefore,

\[
T^2_{Z1} = \frac{n(Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))' (Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))}{Q_3},
\]

where \( Q_3 \sim \chi^2_{n-p-q} \) and, by (3.5.4),

\[
n(Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))' (Z_2 - \bar{Y} - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1)) \sim \chi^2_q(n(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))' (Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1)), (3.5.5)
\]

a noncentral chi-square distribution with \( q \) degrees of freedom and noncentrality parameter \( n(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1))' (Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}_1)) \).
By Lemma 3.3.1 and the independence of $Z$ from the remaining data, we may express the characteristic function of $N^{-1}T^2_Z$ as follows:

$$E \exp[itN^{-1}T^2_Z] = E \exp \left[ it \left( \frac{1}{N + \gamma} Q_3 - n(Z_2 - \bar{Y} - A_{21}A^{-1}_{11,n}(Z_1 - \bar{X}_1))' \right) \cdot (Z_2 - \bar{Y} - A_{21}A^{-1}_{11,n}(Z_1 - \bar{X}_1))' \right]
+ N(N + 1)(Z_1 - \bar{X})'(A_{11,n} + B_1 + B_2)^{-1}(Z_1 - \bar{X})
$$

$$= E_Z E_{Q_3} E_{X_1} E_{X_2} E_{B_1} E_{A_{21},A_{11,n}} \exp \left[ it(N + 1)(Z_1 - \bar{X})' \cdot (A_{11,n} + B_1 + B_2)^{-1}(Z_1 - \bar{X}) \right]
\cdot E_{Y} \exp \left[ \frac{it}{(N + \gamma)Q_3} n(Z_2 - \bar{Y} - A_{21}A^{-1}_{11,n}(Z_1 - \bar{X}_1))' \cdot (Z_2 - \bar{Y} - A_{21}A^{-1}_{11,n}(Z_1 - \bar{X}_1)) \right]. \quad (3.5.6)$$

Applying the formula (2.2.6) for the characteristic function of a noncentral chi-square random variable to (3.5.5) and inserting the result into (3.5.6) yields

$$E_Z E_{Q_3} E_{X_1} E_{X_2} E_{B_1} \cdot E_{A_{21},A_{11,n}} \exp \left[ itN(N + 1)(Z_1 - \bar{X})' (A_{11,n} + B_1 + B_2)^{-1}(Z_1 - \bar{X}) \right]$$

$$\cdot \left( 1 - \frac{2it}{(N + \gamma)Q_3} \right)^{-\frac{q}{2}}$$

$$\cdot \exp \left[ itn(Z_2 - A_{21}A^{-1}_{11,n}(Z_1 - \bar{X}_1))' (Z_2 - A_{21}A^{-1}_{11,n}(Z_1 - \bar{X}_1)) \right] \left( \frac{N + \gamma}Q_3 - 2it \right).$$
By Proposition 2.3.3(ii), \( A_{21}|A_{11,n} \sim N(0, I_q \otimes A_{11,n}) \); therefore,

\[
E \exp[itN^{-1} T^2_Z] = E_Q^3 \left(1 - \frac{2it}{(N + \gamma)Q_3}\right)^{-q/2}
\]

\[
\cdot E_Z E_{\tilde{X}_1} E_{\tilde{X}_2} E_{B_1} E_{A_{11,n}} \exp \left[itN(N + 1)(Z_1 - \bar{X})'(A_{11,n} + B_1 + B_2)^{-1}(Z_1 - \bar{X})\right]
\]

\[
\cdot E_{A_{21}} \cdot A_{11,n} \exp \left[\frac{itn(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}))'(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}))}{(N + \gamma)Q_3 - 2it}\right].
\]

(3.5.7)

We may rewrite the last term in (3.5.7) as

\[
E_{A_{21}}|A_{11,n} \exp \left[\frac{itn(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}))'(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}))}{(N + \gamma)Q_3 - 2it}\right.
\]

\[
\left. + \frac{itn}{(N + \gamma)Q_3 - 2it} \right. \cdot \left. \exp \left[\frac{itn}{(N + \gamma)Q_3 - 2it} tr \left(A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}')(Z_1 - \bar{X}')(A_{11,n}^{-1}A_{12})\right.\right.\right.
\]

\[
\left. + 2(Z_2 - A_{21}A_{11,n}^{-1}(Z_1 - \bar{X}'))(A_{11,n}^{-1}A_{12})\right] \right]
\]

(3.5.8)

By Proposition 2.1.1 (iii), the last term in (3.5.8) becomes

\[
E_{A_{21}}|A_{11,n} \exp \left[\frac{itn}{(N + \gamma)Q_3 - 2it} \right.
\]

\[
\cdot \left. \left(\text{vec}(A_{12})'(I_q \otimes A_{11,n}^{-1}(Z_1 - \bar{X}')(Z_1 - \bar{X}'))A_{11,n}^{-1}\right)\text{vec}(A_{12})\right.\right.
\]

\[
\left. + 2(\text{vec}(I_q))'(I_q \otimes Z_2(Z_1 - \bar{X}')(A_{11,n}^{-1})\text{vec}(A_{12}))\right]\]

(3.5.9)
By Lemma 2.3.1 with

\[ C = \frac{n}{(N + \gamma)Q_3 - 2it} (I_q \otimes A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n}), \]

\[ Z = \text{vec}(A_{12}), \] and

\[ v' = -\frac{2n}{(N + \gamma)Q_3 - 2it} (\text{vec}(I_q))'(I_q \otimes Z_2(Z_1 - \bar{X}_1)'A^{-1}_{11,n}), \]

we see that (3.5.9) equals

\[ \left| I_{pq} - \frac{2itn}{(N + \gamma)Q_3 - 2it} (I_q \otimes A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n})(I_q \otimes A_{11,n}) \right|^{-1/2} \]

\[ \cdot \exp \left[ -\frac{2t^2n^2}{[(N + \gamma)Q_3 - 2it]^2} (\text{vec}(I_q))'(I_q \otimes Z_2(Z_1 - \bar{X}_1)'A^{-1}_{11,n})(I_q \otimes A_{11,n}) \right] \]

\[ \cdot (I_{pq} - \frac{2itn}{(N + \gamma)Q_3 - 2it} (I_q \otimes A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n})(I_q \otimes A_{11,n})^{-1} \]

\[ \cdot (I_q \otimes Z_2(Z_1 - \bar{X}_1)'A^{-1}_{11,n})'(\text{vec}(I_q)) \].

Next, by Proposition 2.1.1 (i),

\[ \left| I_{pq} - \frac{2itn}{(N + \gamma)Q_3 - 2it} (I_q \otimes A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n})(I_q \otimes A_{11,n}) \right|^{-1/2} \]

\[ = \left| I_{pq} - \frac{2itn}{(N + \gamma)Q_3 - 2it} (I_q \otimes A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)') \right|^{-1/2} \]

\[ = \left| I_{p} - \frac{2itn}{(N + \gamma)Q_3 - 2it} A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'^{-q/2} \right|^{-1/2} \]

\[ = \left( 1 - \frac{2itn}{(N + \gamma)Q_3 - 2it} (Z_1 - \bar{X}_1)'A^{-1}_{11,n} (Z_1 - \bar{X}_1) \right)^{-q/2}, \quad (3.5.10) \]
and, similarly,

\[
(I_{pq} - \frac{2itn}{(N + \gamma)Q_3 - 2it}(I_q \otimes A^{-1}_{11,n}(Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(I_q \otimes A_{11,n}))^{-1}

= (I_{pq} - \frac{2itn}{(N + \gamma)Q_3 - 2it}(I_q \otimes A^{-1}_{11,n}(Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(I_q \otimes A_{11,n}))^{-1}

= I_q \otimes (I_p - \frac{2itn}{(N + \gamma)Q_3 - 2it}A^{-1}_{11,n}(Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(I_q \otimes A_{11,n}))^{-1}.
\] (3.5.11)

In addition, by Proposition 2.1.1 (iii),

\[
(I_q \otimes Z_2(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(vec(I_q)) = (I_q \otimes A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z'_2)(vec(I_q))

= vec(A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z'_2),
\] (3.5.12)

and

\[
(vec(I_q))'(I_q \otimes Z_2(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(I_q \otimes A_{11,n})

= vec(A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z'_2)'(I_q \otimes A_{11,n})

= vec((Z_1 - \bar{X}_1)Z'_2)').
\] (3.5.13)
By Proposition 2.1.1 (iii), the product of the last three relations, (3.5.11), (3.5.12), and (3.5.13), equals

\[
(\text{vec}(Z_1 - \bar{X}_1)Z'_2)' \cdot (I_q \otimes (I_p - \frac{2itn}{(N + \gamma)Q_3} - 2it A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'^{-1}))
\]

\[
\cdot \text{vec}(A^{-1}_{11,n} (Z_1 - \bar{X}_1)Z'_2)
\]

\[
= (\text{vec}((Z_1 - \bar{X}_1)Z'_2)')
\]

\[
\cdot \text{vec}((I_p - \frac{2itn}{(N + \gamma)Q_3} - 2it A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'^{-1})A^{-1}_{11,n} (Z_1 - \bar{X}_1)Z'_2).
\]

(3.5.14)

Substituting (3.5.10) and (3.5.14) into (3.5.7) yields

\[
E \exp[itN^{-1}T^2_Z]
\]

\[
= E_{Q_3} E_Z E_{X_1} E_{X_2} E_B E A_{11,n} \left[ \left( 1 - \frac{2it}{(N + \gamma)Q_3} \right) \right.
\]

\[
\cdot \left( 1 - 2 \frac{itn}{(N + \gamma)Q_3} - 2it A^{-1}_{11,n} (Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1) \right)^{-q/2}
\]

\[
\cdot \exp \left[ itN(N + 1)(Z_1 - \bar{X})' (A_{11,n} + B_1 + B_2)^{-1} (Z_1 - \bar{X}) \right]
\]

\[
\cdot \exp \left[ \frac{itn}{(N + \gamma)Q_3} - 2it Z'_2 Z_2 \right]
\]

\[
\cdot \exp \left[ -\frac{2it^2 n^2}{[(N + \gamma)Q_3 - 2it]^2} (\text{vec}((Z_1 - \bar{X}_1)Z'_2))' \right]
\]

\[
\cdot \text{vec}((I_p - \frac{2itn}{(N + \gamma)Q_3} - 2it A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)'^{-1})A^{-1}_{11,n} (Z_1 - \bar{X}_1)Z'_2)
\].

(3.5.15)
Because

\[
(1 - \frac{2it}{(N + \gamma)Q_3}) \left( 1 - \frac{itn}{(N + \gamma)Q_3 - 2it} (Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1) \right)
\]

\[
= \frac{(N + \gamma)Q_3 - 2it - 2itn(Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1)}{(N + \gamma)Q_3},
\]

it follows that

\[
E \exp[itN^{-1}T_z^2] = E_{Q_3} E_{X_1} E_{X_2} E_{B_1} E_{A_{11,n}} \left( 1 - \frac{2it(1 + n(Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1))}{(N + \gamma)Q_3} \right)^{-q/2}
\]

\[
\cdot \exp \left[ itN(N + 1)(Z_1 - \bar{X})'(A_{11,n} + B_1 + B_2)^{-1} (Z_1 - \bar{X}) \right]
\]

\[
\cdot \exp \left[ \frac{itn}{(N + \gamma)Q_3 - 2it} Z_2' Z_2 \right]
\]

\[
\cdot \exp \left[ \frac{-2t^2 n^2}{(N + \gamma)Q_3 - 2it)^2} \text{vec}((Z_1 - \bar{X}_1)Z_2')' \right]
\]

\[
\cdot \text{vec} \left( I_p - \frac{2itn}{(N + \gamma)Q_3 - 2it} A^{-1}_{11,n} (Z_1 - \bar{X}_1)(Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1)Z_2' \right)
\].

Because

\[
\left( 1 - \frac{2it(1 + n(Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1))}{(N + \gamma)Q_3} \right)^{-q/2}
\]

\[
= E_{Q_4} \exp \left[ \frac{it(1 + n(Z_1 - \bar{X}_1)' A^{-1}_{11,n} (Z_1 - \bar{X}_1))}{(N + \gamma)Q_3} Q_4 \right],
\]
where \( Q_4 \sim \chi^2_q \), then

\[
\begin{align*}
E \exp[itN^{-1}T^2_Z] &= E_{Q_3} E_Z E_{\bar{X}_1} E_{\bar{X}_2} E_{B_1} E_{A_{11,n}} E_{Q_4} \exp \left[ \frac{itQ_4(1 + n(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(Z_1 - \bar{X}_1))}{(N + \gamma)Q_3} \right] \\
& \quad \cdot \exp \left[ \frac{itn}{(N + \gamma)Q_3 - 2it} Z_2'Z_2 \right] \\
& \quad \cdot \exp \left[ \frac{-2it^2n^2}{[(N + \gamma)Q_3 - 2it]^2} \right] \text{vec}((Z_1 - \bar{X}_1)Z_2')' \\
& \quad \cdot \text{vec}((I_p - \frac{2itn}{(N + \gamma)Q_3 - 2it} A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z_2')' - 1 A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z_2')',
\end{align*}
\]

and \( Q_4 \) is independent of \( Q_3, Z, \bar{X}_1, \bar{X}_2, B_1, A_{11,n} \). Hence,

\[
\begin{align*}
E \exp[itN^{-1}T^2_Z] &= E_{Q_3} E_{Q_4} \exp \left[ \frac{it}{(N + \gamma)Q_3} \right] \\
& \quad \cdot E_Z E_{\bar{X}_1} E_{\bar{X}_2} E_{B_1} E_{A_{11,n}} \exp \left[ \frac{itnQ_4(Z_1 - \bar{X}_1)'A^{-1}_{11,n}(Z_1 - \bar{X}_1)}{(N + \gamma)Q_3} \right] \\
& \quad \cdot \exp \left[ \frac{itn}{(N + \gamma)Q_3 - 2it} Z_2'Z_2 \right] \\
& \quad \cdot \exp \left[ \frac{-2it^2n^2}{[(N + \gamma)Q_3 - 2it]^2} \right] \text{vec}((Z_1 - \bar{X}_1)Z_2')' \\
& \quad \cdot \text{vec}((I_p - \frac{2itn}{(N + \gamma)Q_3 - 2it} A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z_2')' - 1 A^{-1}_{11,n}(Z_1 - \bar{X}_1)Z_2')'.
\end{align*}
\]
Therefore
\[ E(\exp(itN^{-1}T^2_{Z})) = E_{Q_3,Q_4} \exp \left[ \frac{itQ_4}{(N + \gamma)Q_3} \right] E_{\bar{X}_1,\bar{X}_2} \psi(\bar{X}_1, \bar{X}_2), \]

where \( \psi(x_1, x_2) \) equals
\[
E_{Z,A_{11,n},B_1} \exp \left[ \frac{itnQ_4(Z_1 - x_1)'A^{-1}_{11,n}(Z_1 - x_1)}{(N + \gamma)Q_3} \right]
\cdot \exp \left[ itN(N + 1)(Z_1 - \tau x_1 - \bar{\tau}x_2)'(A_{11,n} + B_1 + B_2)^{-1}(Z_1 - \tau x_1 - \bar{\tau}x_2) \right]
\cdot \exp \left[ \frac{itn}{(N + \gamma)Q_3 - 2it} Z_2'Z_2 \right]
\cdot \exp \left[ \frac{-2t^2n^2}{[(N + \gamma)Q_3 - 2it]^2} (\text{vec}(Z_1 - x_1)Z_2')' \right]
\cdot \text{vec}(I_p - \frac{2itn}{(N + \gamma)Q_3 - 2it} A^{-1}_{11,n}(Z_1 - x_1)(Z_1 - x_1)'^{-1} A^{-1}_{11,n}(Z_1 - x_1)Z_2').
\]

It can be shown that \( \psi \) is orthogonally invariant, that is \( \psi(Hx_1, Hx_2) = \psi(x_1, x_2) \) for all \( H \in O(p) \). Therefore \( \psi(x_1, x_2) = \tilde{\psi}(\|x_1\|, \|x_2\|, x_1'x_2) \). Even when we use this transformation, we have been unable to simplify this expectation. Similar to Krishnamoorthy and Pannala [20], we can approximate the distribution of the \( T^2_{Z} \)-statistic with an \( F \)-distribution and we reserve this for future work.

3.5.3 Analysis of the Pennsylvania cholesterol data

Consider the cholesterol data provided in Table 1.1 of Chapter 1. This data set being two-step monotone incomplete, we wish to apply our results from this chapter. We first assess the assumption of multivariate normality for this data set. In Figure 3.7, we provide marginal quantile-quantile plots for each day and scatterplots for each pair
of days. By examining these graphs, we infer that no outliers appear to be present in
the data set and that the marginal distribution of each variable is reasonably normally
distributed. Because little is known about testing for multivariate normality with incom-
plete data, we also calculated Mardia’s test statistics for normality of the complete data;
\( b_{1,3} = 5.08; \ p\text{-value} = .886 \) and \( b_{2,3} = -1.47; \ p\text{-value} = .143 \). Because neither test is
significant at the 5% level, the assumption of multivariate normality is not unreasonable.

We simulated 1,000 values of the \( T^2\)-statistic and our upper and lower bounds to estimate the 97.5\( th \) percentile of each distribution. We then calculated 95% confidence intervals for various linear combinations of \( \mu \), namely the mean cholesterol levels for Day 2, for Day 4, for Day 14, and for the difference in mean cholesterol levels between Days 14 and 2. Because the data are repeated measurements, the proper way to calculate an interval for the difference in Days 14 and 2 would be, first, to find the individual differences and then, second, to obtain a confidence interval for the mean difference. Because this analysis is simply an illustration of our distributional results, we present the linear combination of Day 14 - Day 2.

Table 3.1 95% Confidence Interval for Mean Cholesterol Levels

<table>
<thead>
<tr>
<th></th>
<th>Lower Bound</th>
<th>Exact</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 2</td>
<td>(205.56, 302.29)</td>
<td>(214.41, 293.45)</td>
<td>(224.12, 283.73)</td>
</tr>
<tr>
<td>Day 4</td>
<td>(184.09, 277.19)</td>
<td>(190.68, 270.61)</td>
<td>(204.27, 257.02)</td>
</tr>
<tr>
<td>Day 14</td>
<td>(183.12, 263.05)</td>
<td>(188.78, 257.39)</td>
<td>(200.44, 245.73)</td>
</tr>
<tr>
<td>Day 14 - Day 2</td>
<td>(-82.83, 21.14)</td>
<td>(-75.48, 13.79)</td>
<td>(-60.30, -1.39)</td>
</tr>
</tbody>
</table>
As expected, in Table 3.1, the lower bound provides a wider, more conservative interval, and the upper bound provides a narrower, anti-conservative interval.
Fig. 3.1 Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 2; q = 1; n = 19; N = 28$)
Dashed Red = Chang & Richards; Dotted Blue = Theorem 3.4.1 and Theorem 3.4.2; Solid Black = Exact

Fig. 3.2 Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 2; q = 2; n = 10; N = 15$)
Dashed Red = Chang & Richards; Dotted Blue = Theorem 3.4.1 and Theorem 3.4.2; Solid Black = Exact
Fig. 3.3 Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 3; q = 3; n = 15; N = 20$)
Dashed Red = Chang & Richards; Dotted Blue = Theorem 3.4.1 and Theorem 3.4.2; Solid Black = Exact

Fig. 3.4 Simulated Cumulative Distribution Function of the $T^2$-statistic: Comparison of Bounds ($p = 4; q = 4; n = 15; N = 20$)
Dashed Red = Chang & Richards; Dotted Blue = Theorem 3.4.1 and Theorem 3.4.2; Solid Black = Exact
Fig. 3.5 Simulated Cumulative Distribution Function of the $T^2$-statistic: Approximation ($p = 2; q = 2; n = 10; N = 15$)  
Dotted Blue = Theorem 3.4.1 and Theorem 3.4.2  
Solid Black = Exact; Dashed Green = Krishnamoorthy & Pannala

Fig. 3.6 Simulated Cumulative Distribution Function of the $T^2$-statistic: Approximation ($p = 4; q = 4; n = 15; N = 20$)  
Dotted Blue = Theorem 3.4.1 and Theorem 3.4.2  
Solid Black = Exact; Dashed Green = Krishnamoorthy & Pannala
Fig. 3.7 The Pennsylvania Cholesterol Data: Assessment of Multivariate Normality
Chapter 4

Three-Step Monotone Incomplete Multivariate Normal Data

In this chapter we consider the problem of inference for $\mu$ and $\Sigma$ with three-step monotone incomplete data from $N_d(\mu, \Sigma)$, where $\mu$ and $\Sigma$ are partitioned,

$$
\mu = \begin{pmatrix} 
\mu_1 \\
\mu_2 \\
\mu_3 
\end{pmatrix}, \\
\Sigma = \begin{pmatrix} 
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\
\Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\
\Sigma_{31} & \Sigma_{32} & \Sigma_{33} 
\end{pmatrix},
$$

with $\mu_l$ of dimension $p_l \times 1$ and $\Sigma_{ij}$ of dimension $p_l \times p_j$, $l, j = 1, 2, 3$. For the case in which $\Sigma_{ij} = 0$ for all $l \neq j$, we establish independence between the maximum likelihood estimator of $\mu_1$, denoted $\hat{\mu}_1$, and the maximum likelihood estimators of $\mu_2$ and $\mu_3$, denoted $\hat{\mu}_2$ and $\hat{\mu}_3$, respectively, and also show that $\hat{\mu}_2$ is uncorrelated with $\hat{\mu}_3$. The last section of this chapter establishes the equivariance of the maximum likelihood estimator of $\mu$, denoted $\hat{\mu}$, and examines its full distribution.

4.1 Notation and maximum likelihood estimators

Let $X_1 \in \mathbb{R}^{p_1}$, $X_2 \in \mathbb{R}^{p_2}$, and $X_3 \in \mathbb{R}^{p_3}$. Suppose $(X_1', X_2', X_3')' \sim N_{p_1+p_2+p_3}(\mu, \Sigma)$.

In the three-step setting, we observe $n_1$ mutually independent observations on $(X_1', X_2', X_3')'$. 

an additional \( n_2 \) independent observations on \((X_1', X_2')'\), and an additional \( n_3 \) observations on \( X_1 \) only. Therefore, the data are mutually independent vectors of the form

\[
\begin{pmatrix}
X_{1,1} \\
X_{2,1} \\
X_{3,1}
\end{pmatrix}
\cdots
\begin{pmatrix}
X_{1,n_1} \\
X_{2,n_1} \\
X_{3,n_1}
\end{pmatrix}
\begin{pmatrix}
X_{1,n_1+1} \\
X_{2,n_1+1} \\
X_{3,n_1+1}
\end{pmatrix}
\cdots
\begin{pmatrix}
X_{1,n_1+n_2} \\
X_{2,n_1+n_2} \\
X_{3,n_1+n_2}
\end{pmatrix}
\begin{pmatrix}
X_{1,n_1+n_2+1} \\
X_{2,n_1+n_2+1} \\
X_{3,n_1+n_2+1}
\end{pmatrix}
\cdots
\begin{pmatrix}
X_{1,n_1+n_2+n_3} \\
X_{2,n_1+n_2+n_3} \\
X_{3,n_1+n_2+n_3}
\end{pmatrix}
\]

where \((X_{1,j}', X_{2,j}', X_{3,j}')', j = 1, \ldots, n_1\) are mutually independent observations from \(N_{p_1+p_2+p_3} (\mu, \Sigma)\). The observations with incomplete data \((X_{1,j}', X_{2,j}')', j = n_1+1, \ldots, N_2\), and \( X_{1,j}, j = N_2+1, \ldots, N_3 \) are mutually independent observations on the first \( p_1 \) and \( p_1 \) characteristics from the same population, respectively. Define \( N_1 = n_1, N_2 = n_1 + n_2, N_3 = n_1 + n_2 + n_3 \); consequently \( n_2 = N_2 - N_1 \) and \( n_3 = N_3 - N_2 \).

We define the data matrices

\[X_1 = \begin{pmatrix}
X_{1,1} & \cdots & X_{1,N_1} & X_{1,N_1+1} & \cdots & X_{1,N_2} & X_{1,N_2+1} & \cdots & X_{1,N_3}
\end{pmatrix}, \tag{4.1.1}
\]

\[X_2 = \begin{pmatrix}
X_{1,1} & \cdots & X_{1,N_1} & X_{1,N_1+1} & \cdots & X_{1,N_2} \\
X_{2,1} & \cdots & X_{2,N_1} & X_{2,N_1+1} & \cdots & X_{2,N_2}
\end{pmatrix}, \tag{4.1.2}
\]

and

\[X_3 = \begin{pmatrix}
X_{1,1} & \cdots & X_{1,N_1} \\
X_{2,1} & \cdots & X_{2,N_1} \\
X_{3,1} & \cdots & X_{3,N_1}
\end{pmatrix}. \tag{4.1.3}
\]
Similar to Kanda and Fujikoshi [17], we decompose the data into \( p_j \times n_j \) segments

\[
\begin{pmatrix}
X_{1,1} & \cdots & X_{1, N_1} \\
X_{1, N_1+1} & \cdots & X_{1, N_2} \\
X_{1, N_2+1} & \cdots & X_{1, N_3}
\end{pmatrix}
\begin{pmatrix}
X_{2,1} & \cdots & X_{2, N_1} \\
X_{2, N_1+1} & \cdots & X_{2, N_2} \\
X_{2, N_2+1} & \cdots & X_{2, N_3}
\end{pmatrix}
\begin{pmatrix}
X_{3,1} & \cdots & X_{3, N_1} \\
X_{3, N_1+1} & \cdots & X_{3, N_2} \\
X_{3, N_2+1} & \cdots & X_{3, N_3}
\end{pmatrix}
\]

so that, in the case of block-diagonal \( \Sigma \), the segments are mutually independent. Similar to the two-step case, we define \( \tau_{ij} = n_i/N_j, i, j = 1, 2, 3 \) to represent the proportion of data within each segment, and we define \( \bar{\tau}_{ij} = 1 - \tau_{ij} \).

We define the corresponding segment means

\[
\bar{X}_{1,1} = \frac{1}{n_1} \sum_{j=1}^{N_1} X_{1,j}, \quad \bar{X}_{1,2} = \frac{1}{n_2} \sum_{j=N_1+1}^{N_2} X_{1,j}, \quad \bar{X}_{1,3} = \frac{1}{n_3} \sum_{j=N_2+1}^{N_3} X_{1,j},
\]

\[
\bar{X}_{2,1} = \frac{1}{n_1} \sum_{j=1}^{N_1} X_{2,j}, \quad \bar{X}_{2,2} = \frac{1}{n_2} \sum_{j=N_1+1}^{N_2} X_{2,j},
\]

\[
\bar{X}_{3,1} = \frac{1}{n_1} \sum_{j=1}^{N_1} X_{3,j}.
\]

For \( l = 1, 2, 3 \), let \( 1_{N_4-l} = (1, \ldots, 1)' \) be the \( N_{4-l} \times 1 \) vector of 1’s. Because of the iterative nature of the maximum likelihood estimators, we will also need notation for
\[
\bar{X}_l = \frac{1}{N_{4-l}} X_l 1'_{N_{4-l}}, \ l = 1, 2, 3; \text{ that is,}
\]

\[
\bar{X}_1 = \frac{1}{N_3} \sum_{j=1}^{N_3} X_{1,j} = \tau_{13} \bar{X}_{1,1} + \tau_{23} \bar{X}_{1,2} + \tau_{33} \bar{X}_{1,3},
\]

\[
\bar{X}_2 = \frac{1}{N_2} \sum_{j=1}^{N_2} \begin{pmatrix} X_{1,j} \\ X_{2,j} \end{pmatrix} = \begin{pmatrix} \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} \\ \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2} \end{pmatrix},
\]

and

\[
\bar{X}_3 = \frac{1}{N_1} \sum_{j=1}^{N_1} \begin{pmatrix} X_{1,j} \\ X_{2,j} \\ X_{3,j} \end{pmatrix} = \begin{pmatrix} \bar{X}_{1,1} \\ \bar{X}_{2,1} \\ \bar{X}_{3,1} \end{pmatrix}.
\]

Further, we define the \textit{stepwise} sums of squares and cross-products

\[
A_1 = \sum_{j=1}^{N_3} \left( X_{1,j} - \bar{X}_1 \right) \left( X_{1,j} - \bar{X}_1 \right)',
\]

\[
A_2 = \sum_{j=1}^{N_3} X_{1,j} X_{1,j}' - N_3 \bar{X}_1 \bar{X}_1', \quad (4.1.4)
\]

\[
A_2 = \sum_{j=1}^{N_2} \begin{pmatrix} \left( X_{1,j} - \bar{X}_1 \right) \\ \left( X_{2,j} - \bar{X}_2 \right) \end{pmatrix} \begin{pmatrix} \left( X_{1,j} - \bar{X}_1 \right) \\ \left( X_{2,j} - \bar{X}_2 \right) \end{pmatrix}' = \begin{pmatrix} A_{2,11} & A_{2,12} \\ A_{2,21} & A_{2,22} \end{pmatrix}, \quad (4.1.5)
\]

where \(A_{2,ij}\) is a \(p_i \times p_j\) matrix; and

\[
A_3 = \sum_{j=1}^{N_1} \begin{pmatrix} \left( X_{1,j} - \bar{X}_3 \right) \\ \left( X_{2,j} - \bar{X}_3 \right) \\ \left( X_{3,j} - \bar{X}_3 \right) \end{pmatrix} \begin{pmatrix} \left( X_{1,j} - \bar{X}_3 \right) \\ \left( X_{2,j} - \bar{X}_3 \right) \\ \left( X_{3,j} - \bar{X}_3 \right) \end{pmatrix}' = \begin{pmatrix} A_{3,11} & A_{3,12} & A_{3,13} \\ A_{3,21} & A_{3,22} & A_{3,23} \\ A_{3,31} & A_{3,32} & A_{3,33} \end{pmatrix}, \quad (4.1.6)
\]
where $A_{3,lj}$ also is a $p_l \times p_j$ matrix.

The maximum likelihood estimator of $\mu$, $\hat{\mu}$, is well-known [15], [19], [12]:

\[
\hat{\mu}_1 = \bar{X}_1, \tag{4.1.7}
\]

\[
\hat{\mu}_2 = \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2} - B_{21} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \hat{\mu}_1 \right), \tag{4.1.8}
\]

\[
\hat{\mu}_3 = \bar{X}_{3,1} - B_{31} \left( \bar{X}_{1,1} - \hat{\mu}_1 \right) - B_{32} \left( \bar{X}_{2,1} - \hat{\mu}_2 \right), \tag{4.1.9}
\]

where

\[
B_{21} = B'_{12} = A^{-1}_{2,21} \tag{4.1.10}
\]

and

\[
(B_{31}, B_{32}) = (B'_{13}, B'_{23})' = \left( A_{3,31}, A_{3,32} \right) \left( \begin{array}{cc} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{array} \right)^{-1}. \tag{4.1.11}
\]

We may write the maximum likelihood estimators in terms of our segment means.

Noting that

\[
\tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 = \frac{n_1}{N_2} \bar{X}_{1,1} + \frac{n_2}{N_2} \bar{X}_{1,2} - \left( \frac{n_1}{N_3} \bar{X}_{1,1} + \frac{n_2}{N_3} \bar{X}_{1,2} + \frac{n_3}{N_3} \bar{X}_{1,3} \right)
\]

\[
= n_1 \left( \frac{1}{N_2} - \frac{1}{N_3} \right) \bar{X}_{1,1} + n_2 \left( \frac{1}{N_2} - \frac{1}{N_3} \right) \bar{X}_{1,2} - \frac{n_3}{N_3} \bar{X}_{1,3}
\]

\[
= \frac{n_1}{N_2 N_3} \bar{X}_{1,1} + \frac{n_2}{N_2 N_3} \bar{X}_{1,2} - \frac{n_3}{N_3} \bar{X}_{1,3}
\]

\[
= \tau_{33} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 \right);
\]
it follows that

\[
\begin{align*}
\hat{\mu}_1 &= \tau_{13} \bar{X}_{1,1} + \tau_{23} \bar{X}_{1,2} + \tau_{33} \bar{X}_{1,3}, \\
\hat{\mu}_2 &= \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2} - B_{21} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 \right) \\
&= \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2} - \tau_{33} B_{21} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 \right), \quad (4.1.12)
\end{align*}
\]

and

\[
\begin{align*}
\hat{\mu}_3 &= \bar{X}_{3,1} - B_{31} \left( \bar{X}_{1,1} - \bar{X}_1 \right) \\
&- B_{32} \left[ \bar{X}_{2,1} - \tau_{12} \bar{X}_{2,1} - \tau_{22} \bar{X}_{2,2} + \tau_{33} B_{21} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 \right) \right] \\
&= \bar{X}_{3,1} - B_{31} \left( \tau_{13} \bar{X}_{1,1} - \tau_{23} \bar{X}_{1,2} - \tau_{33} \bar{X}_{1,3} \right) \\
&- B_{32} \left[ \tau_{22} (\bar{X}_{2,1} - \bar{X}_{2,2}) + \tau_{33} B_{21} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 \right) \right]. \quad (4.1.13)
\end{align*}
\]

4.2 Correlation properties of \( \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\mu}_3 \)

Chang and Richards [6] derived an exact, finite-sample stochastic representation for \( \hat{\mu} \) in the two-step monotone incomplete setting. The distribution of \( \hat{\mu} \) for \( k > 2 \) is still unknown today. If we consider only the first \( p_1 + p_2 \) rows of data in the three step setting, we have a two-step monotone incomplete data set. We have verified that the three-step formulas for \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) reduce to the two-step formulas when considered in this way. This is a very useful result because we may now apply all that is known from the two-step setting for \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) to the three-step setting. Because the joint distribution of \( \hat{\mu}_1 \) and
\( \hat{\mu}_2 \) is known from the two-step setting, the conditional distribution of \( \hat{\mu}_3 \) given \( \{\hat{\mu}_1, \hat{\mu}_2\} \) would yield the joint distribution of all three estimators.

In order to study the exact distribution for \( k = 3 \), we first examine the independence properties of \( \hat{\mu}_1 \), \( \hat{\mu}_2 \), and \( \hat{\mu}_3 \). In the two-step setting, Chang and Richards\[6\] found that \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) are independent if and only if \( \Sigma \) is block-diagonal, \( \Sigma_{12} = 0 \); therefore mutual independence in the three-step setting is plausible in this case also.

We have one critical piece of the proof remaining to conclude mutual independence for block-diagonal \( \Sigma \), that is that \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) are independent. Although we were unable to establish this independence, we have proved that if \( \Sigma \) is block-diagonal, then the cross-covariance between \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) is 0. In each of the following results, we assume without loss of generality that \( \mu = 0 \); this can be achieved by replacing each \( X_{lj} \) by \( X_{lj} - \mu_l \), \( l = 1, 2, 3 \).

If we consider the first \( p_1 + p_2 \) rows of data as two-step, monotone incomplete data, the formulas for \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) in the three-step setting reduce to \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) in the two-step setting. Therefore, the distribution and covariance matrix of \( (\hat{\mu}_1, \hat{\mu}_2) \) in the three-step setting, follow directly from the results of Chang and Richards\[6\]. We will first verify that the formulas for \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) are the same and therefore attribute our theorem of independence to Chang and Richards\[6\], however we still provide our new derivation that does not use the joint distribution of \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \).

**Lemma 4.2.1.** Consider the first \( p_1 + p_2 \) rows of data as two-step, monotone incomplete data. Then the maximum likelihood estimators for \( \mu_1 \) and \( \mu_2 \) from the two-step case,
namely,

\[ \hat{\mu}_1 = \bar{X}, \quad \hat{\mu}_2 = \bar{Y} - \tau A_{21} A_{11,n}^{-1} (\bar{X}_1 - \bar{X}_2), \]

are the same as the maximum likelihood estimators for \( \mu_1 \) and \( \mu_2 \) in the three-step case, namely

\[ \hat{\mu}_1 = \bar{X}_1, \quad \hat{\mu}_2 = \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2} - \tau_{33} B_{21} \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_{1,3} \right). \]

**Proof.** The maximum likelihood estimator for \( \mu_1 \) in both sets of notation is the overall mean of the first \( p_1 \) rows, therefore \( \hat{\mu}_1 \) is the same. In the three-step setting, the first \( N_2 \) columns of the \( p_2 \) rows of data comprise the \( Y \)-data in the two-step setting. The first term in the three-step maximum likelihood estimator, \( \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2} \), is exactly \( \bar{Y} \), the mean of the \( Y \)-data. The proportion of incomplete data is represented by \( \bar{\tau} \) in the two-step setting and \( \tau_{33} \) in the three-step setting. In the two-step setting, \( A_{21} \) is the covariance matrix between the first \( N_2 \) columns of the \( p_1 \) rows of data and the first \( N_2 \) columns of the \( p_2 \) rows of data, defined as \( A_{2,21} \) in the three-step setting. Similarly, \( A_{11,n} \) is the variance-covariance matrix of the first \( N_2 \) columns of the \( p_1 \) rows of data, defined as \( A_{2,11} \) in the three-step setting. Therefore \( B_{21} \) in the three-step setting is \( A_{21} A_{11,n}^{-1} \) in the two-step setting. The linear combination, \( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} \), is exactly the mean of the first \( N_2 \) columns of the \( p_1 \) rows of complete data, defined as \( \bar{X}_1 \) in the two-step setting, and \( \bar{X}_{1,3} \) is the mean of the incomplete data, defined as \( \bar{X}_2 \) in the two-step setting. \( \square \)
**Theorem 4.2.2.** If $\Sigma$ is block-diagonal, then $\hat{\mu}_1$ is independent of $\{\hat{\mu}_2, \hat{\mu}_3\}$.

We will prove this theorem by means of the following results.

**Proof.** We begin with an alternative proof for the independence of $\hat{\mu}_1$ and $\hat{\mu}_2$ when $\Sigma$ is block-diagonal.

**Proposition 4.2.3.** (Chang and Richards [2]) If $\Sigma$ is block-diagonal, then $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent.

**Proof.** By the well-known independence of the sample mean and sample covariance matrix of a sample from a multivariate normal population, it follows that $A_2$ is independent of $\{\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2}, \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}\}$. In addition, because the individual vector observations are mutually independent, we find that $A_2, \{\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2}, \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}\}$, and $\bar{X}_{1,3}$ are mutually independent.

Because $B_{21} = A_{2,21}A_{2,11}^{-1}$ is a function of $A_2$, it follows that $B_{21}, \{\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2}, \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}\}$, and $\bar{X}_{1,3}$ are mutually independent. It then follows that $B_{21}$ and $\{\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}, \bar{X}_{1,1}, \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}\}$ are mutually independent.

Because $\Sigma$ is block-diagonal, then $B_{21}, \{\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}, \bar{X}_{1,1}\}$, and $\{\tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}\}$ are mutually independent.

Because $\bar{X}_{1}$ and $\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}$ are linear functions of multivariate normal data, we need only to examine their cross-covariance to determine whether they
are independent. Then

\[
E \bar{X}_1 \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_1 \right)'
\]

\[
= E \bar{X}_1 \left( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} \right)' - E \bar{X}_1 \bar{X}'_1
\]

\[
= E \sum_{j=1}^{2} \left( \tau_{13} \bar{X}_{1,1} + \tau_{23} \bar{X}_{1,2} + \tau_{33} \bar{X}_{1,3} \right) \frac{n_j}{N_2} \bar{X}'_{1,j} - \frac{1}{N_3} \Sigma_{11}.
\]

Because \( E \bar{X}_{1,j} \bar{X}'_{1,j} = \frac{1}{n_j} \Sigma_{11} \), this cross-covariance becomes

\[
\frac{\tau_{13}}{N_2} \Sigma_{11} + \frac{\tau_{23}}{N_2} \Sigma_{11} - \frac{1}{N_3} \Sigma_{11} = \frac{\tau_{12} + \tau_{22} - 1}{N_3} \Sigma_{11} = 0.
\]

Therefore \( \bar{X}_1, \ B_{21}, \ \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2}, \) and \( \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_{1,3} \) are mutually

independent; consequently, \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) are independent.

In order to examine the cross-covariance between \( \hat{\mu}_1 \) and \( \hat{\mu}_3 \), we need to rewrite

\( A_2 \) in terms of mutually independent matrices similar to the way in which Chang and

Richards express \( A_{22,\nu} \) in the two-step case [7].

**Proposition 4.2.4.** The matrix \( A_2 \) may be expressed as

\[
A_2 = J A_3 J' + C + D,
\]
where $A_3$ is defined in (4.1.6),

$$J = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_q & 0 \end{pmatrix}, \quad (4.2.1)$$

$$C = \sum_{j=n_1+1}^{N_2} \begin{pmatrix} X_{1j} - \bar{X}_{1,2} \\ X_{2j} - \bar{X}_{2,2} \end{pmatrix} \begin{pmatrix} X_{1j} - \bar{X}_{1,2} \\ X_{2j} - \bar{X}_{2,2} \end{pmatrix}', \quad (4.2.2)$$

and

$$D = \frac{n_1 n_2}{n_1 + n_2} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix}'. \quad (4.2.3)$$

Further, the matrices $A_3, C, \text{ and } D$ are mutually independent.

Proof. By adding and subtracting means for the first two blocks of data, we obtain

$$A_2 = \sum_{j=1}^{n_1} \begin{pmatrix} X_{1j} - \bar{X}_{1,1} + \bar{X}_{1,1} - (\tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2}) \\ X_{2j} - \bar{X}_{2,1} + \bar{X}_{2,1} - (\tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2}) \end{pmatrix} + \sum_{j=n_1+1}^{n_1+n_2} \begin{pmatrix} X_{1j} - \bar{X}_{1,2} + \bar{X}_{1,2} - (\tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2}) \\ X_{2j} - \bar{X}_{2,2} + \bar{X}_{2,2} - (\tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2}) \end{pmatrix}'. $$
Because the cross-products each simplify to 0, we obtain

\[
A_2 = \sum_{j=1}^{n_1} \begin{pmatrix} X_{1,j} - \bar{X}_{1,1} \\ X_{2,j} - \bar{X}_{2,1} \end{pmatrix} \begin{pmatrix} X_{1,j} - \bar{X}_{1,1} \\ X_{2,j} - \bar{X}_{2,1} \end{pmatrix}' + \sum_{j=n_1+1}^{n_1+n_2} \begin{pmatrix} X_{1,j} - \bar{X}_{1,2} \\ X_{2,j} - \bar{X}_{2,2} \end{pmatrix} \begin{pmatrix} X_{1,j} - \bar{X}_{1,2} \\ X_{2,j} - \bar{X}_{2,2} \end{pmatrix}'
\]

\[
+ \sum_{j=1}^{n_1} \tau_{22} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix}'
\]

\[
+ \sum_{j=n_1+1}^{n_1+n_2} \tau_{12} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix}'
\]

\[
= \begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix} + C + D,
\]

and it is elementary that

\[
JA_3J' = \begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix}.
\]

By the independence of the sample mean and sample covariance matrix, and the independence of the individual observations, it follows that \(A_3, C, \text{ and } D\) are mutually independent.

\[\square\]

**Proposition 4.2.5.** If \(\Sigma\) is block-diagonal, then \(\hat{\mu}_1\) and \(\hat{\mu}_3\) are independent.

**Proof.** Because \(B_{21}, B_{31}, \text{ and } B_{32}\) are functions of \(A_2 = JA_3J' + C + D\) and \(A_3\), we examine the independence properties of the random variables appearing in \(\hat{\mu}_1\) and \(\hat{\mu}_3\):

\[
\bar{X}_1, A_3, C, D, \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}), \tau_{13} \bar{X}_{1,1} - \tau_{23} \bar{X}_{1,2} - \tau_{33} \bar{X}_{1,3}, \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_{1,3}.
\]
By the independence of the sample mean and sample covariance matrix, we see that 
\( \{A_3, C\} \) and \( \{D, \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}), \bar{\tau}_{13}\bar{X}_{1,1} - \tau_{23}\bar{X}_{1,2} - \tau_{33}\bar{X}_{1,3}, \tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}, \bar{X}_1 \} \) are independent. Our final step is to show that \( \hat{\mu}_1 = \tau_{13}\bar{X}_{1,1} + \tau_{23}\bar{X}_{1,2} + \tau_{33}\bar{X}_{1,3} \) is independent of the set of random variables:

\[ \{D, \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}), \bar{\tau}_{13}\bar{X}_{1,1} - \tau_{23}\bar{X}_{1,2} - \tau_{33}\bar{X}_{1,3}, \tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3} \}. \]

Because all the variables in this set are linear combinations of multivariate normal vectors, we need only show that the cross-covariance between \( \hat{\mu}_1 \) and this set of variables is 0. To that end,

\[
E \left( \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \\ \bar{\tau}_{13}\bar{X}_{1,1} - \tau_{23}\bar{X}_{1,2} - \tau_{33}\bar{X}_{1,3} \\ \tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3} \end{pmatrix} ' \right) \left( \begin{pmatrix} \tau_{13}\bar{X}_{1,1} + \tau_{23}\bar{X}_{1,2} + \tau_{33}\bar{X}_{1,3} \end{pmatrix} \right)'
\]

\[
= \left( \begin{pmatrix} \frac{1}{N_3} \Sigma_{11} - \frac{1}{N_3} \Sigma_{11} \\ 0 \\ \left( \frac{n_2 + n_3}{N_2^2} - \frac{\tau_{23}^2}{n_2} - \frac{\tau_{33}^2}{n_3} \right) \Sigma_{11} \\ \left( \frac{\tau_{13}^2}{N_2^2} + \frac{\tau_{23}^2}{N_2^2} - \frac{1}{N_3} \right) \Sigma_{11} \end{pmatrix} \right) \left( \begin{pmatrix} 0 \\ 0 \\ \left( \frac{n_2 + n_3}{N_2^2} - \frac{n_2}{N_2^2} - \frac{n_3}{N_3^2} \right) \Sigma_{11} \end{pmatrix} \right) = 0.
\]

This completes the proof.

Proof of Theorem 4.2.2: By including the vector \( \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2} \) with

\[ \{D, \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}), \bar{\tau}_{13}\bar{X}_{1,1} - \tau_{23}\bar{X}_{1,2} - \tau_{33}\bar{X}_{1,3}, \tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}, \bar{X}_1 \} , \]


we would have all of the variables that arise in \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \). Because the cross-covariance between \( \bar{X}_1 \) and \( \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2} \) is 0, we could have included \( \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2} \) with the linear combinations of normal random variables within \( \hat{\mu}_3 \) from above. This proves independence between \( \bar{X}_1 \) and \( \{A_3, C, D, \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}), (\tau_{13}\bar{X}_{1,1} - \tau_{23}\bar{X}_{1,2} - \tau_{33}\bar{X}_{1,3}), (\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}), \tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}\} \); therefore \( \hat{\mu}_1 \) is independent of \( \{\hat{\mu}_2, \hat{\mu}_3\} \).

If \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) also are independent for the case in which \( \Sigma \) is block-diagonal, then we would deduce that \( \hat{\mu}_1 \), \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) are mutually independent. Unfortunately, it appears that methods similar to those in the proof of Theorem 4.2.2 cannot be used to show independence between \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \). Because the second term in \( \hat{\mu}_3 \) includes 
\[
\bar{X}_{2,1} - \hat{\mu}_2 = \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}) - \tau_{33}A_{2,21}A_{2,11}^{-1}(\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3})
\]
and because \( \tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3} \) is not independent of itself, we cannot decompose \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) into mutually independent terms using the same method. It is therefore unclear at this time whether or not \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) are independent. There is hope, however, because we have proved that if \( \Sigma = I \), then the covariance between \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) is 0.

**Theorem 4.2.6.** If \( \Sigma = I_d \), then \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \) are uncorrelated.

**Proof.** Because we assume that \( \mu = 0 \), without loss of generality, and we assume that \( \Sigma = I \), then the \( X_{l,j} \) are mutually independent \( \mathcal{N}(0, I) \) random vectors, \( l = 1, 2, 3 \), \( j = 1, \ldots, N_3 \). If we replace each \( X_{3,j} \) by \( -X_{3,j} \), \( j = 1, \ldots, N_1 \), and leave \( X_{1,j} \) and \( X_{2,j} \) unchanged, then \( \hat{\mu}_1 \) and \( \hat{\mu}_2 \) are unchanged, whereas \( \hat{\mu}_3 \) changes to \( -\hat{\mu}_3 \). However,
−X_{3,j} has the same distribution as X_{3,j}; hence (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3) \overset{d}{=} (\mu_1, \mu_2, -\mu_3). Therefore,

\[ E(\hat{\mu}_2\hat{\mu}_3') = E(\hat{\mu}_2(-\hat{\mu}_3')) = -E(\hat{\mu}_2\hat{\mu}_3'), \]

so we obtain \[ E(\hat{\mu}_2\hat{\mu}_3') = 0. \]

By Theorem (4.2.2) and Theorem (4.2.6), it follows that if \[ \Sigma = I_d, \]

it is block-diagonal.

4.3 The distribution of \( \hat{\mu}_3 \)

We approached the joint distribution of \( \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\mu}_3 \) from many directions. Our most promising idea was to condition on the first \( p_1 + p_2 \) rows of data to find the conditional distribution of \( \hat{\mu}_3 \) given \( X_1 \) and \( X_2 \), \[ \text{(4.1.1) and (4.1.2)}. \] Similar to the two-step proof, we first establish that \( \hat{\mu} \) is equivariant, and assume, without loss of generality, that \( \mu = 0 \) and \( \Sigma = I_{p_1 + p_2 + p_3} \). We then derive the conditional distribution of \( \hat{\mu}_3 \) given \( X_1 \) and \( X_2 \) and extend this distribution as far as we can in our attempt to find the joint distribution of \( \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\mu}_3 \).

**Proposition 4.3.1.** Let \( \Lambda_{jj} \) be a \( p_j \times p_j \) positive definite matrix, \( j = 1, 2, 3 \), \( \Lambda_{21} \) be \( p_2 \times p_1 \), \( \Lambda_{31} \) be \( p_3 \times p_1 \), \( \Lambda_{32} \) be \( p_3 \times p_2 \), \( \nu_1 \in \mathbb{R}^{p_1}, \nu_2 \in \mathbb{R}^{p_2}, \nu_3 \in \mathbb{R}^{p_3} \), and

\[
\Lambda = \begin{pmatrix} \Lambda_{11} & 0 & 0 \\ 0 & \Lambda_{22} & 0 \\ 0 & 0 & \Lambda_{33} \end{pmatrix}, \quad C = \begin{pmatrix} I_{p_1} & 0 & 0 \\ \Lambda_{21} & I_{p_2} & 0 \\ \Lambda_{31} & \Lambda_{32} & I_{p_3} \end{pmatrix}, \quad \nu = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}.
\]
Then the estimator \( \hat{\mu} \) is equivariant under the transformation

\[
\begin{pmatrix}
X_{1,j} \\
X_{2,j} \\
X_{3,j}
\end{pmatrix}
\mapsto \Lambda C
\begin{pmatrix}
X_{1,j} \\
X_{2,j} \\
X_{3,j}
\end{pmatrix}
+ \nu,
\quad (4.3.1)
\]

for \( j = 1, \ldots, N_3 \).

**Proof.** Let

\[
\begin{pmatrix}
X_{1,j}^* \\
X_{2,j}^* \\
X_{3,j}^*
\end{pmatrix}
= \Lambda C
\begin{pmatrix}
X_{1,j} \\
X_{2,j} \\
X_{3,j}
\end{pmatrix}
+ \nu
\]

\[
= \begin{pmatrix}
\Lambda_{11} X_{1,j} + \nu_1 \\
\Lambda_{22} \Lambda_{21} X_{1,j} + \Lambda_{22} X_{2,j} + \nu_2 \\
\Lambda_{33} \Lambda_{31} X_{1,j} + \Lambda_{33} \Lambda_{32} X_{2,j} + \Lambda_{33} X_{3,j} + \nu_3
\end{pmatrix},
\quad (4.3.2)
\]

for \( j = 1, \ldots, N_3 \), then

\[
\begin{pmatrix}
X_{1,j}^* \\
X_{2,j}^* \\
X_{3,j}^*
\end{pmatrix}
\sim N
\begin{pmatrix}
\mu^*_1 \\
\mu^*_2 \\
\mu^*_3
\end{pmatrix},
\quad \Sigma^* = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\
\Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\
\Sigma_{31} & \Sigma_{32} & \Sigma_{33}
\end{pmatrix}
\]
Define the sample mean vectors

\[ \bar{X}^*_{1,1} = \frac{1}{n_1} \sum_{j=1}^{n_1} X^*_{1,j}, \quad \bar{X}^*_{1,2} = \frac{1}{n_2} \sum_{j=1}^{n_2} X^*_{1,j}, \quad \bar{X}^*_{1,3} = \frac{1}{n_3} \sum_{j=1}^{n_3} X^*_{1,j}, \]

\[ \bar{X}^*_{2,1} = \frac{1}{n_1} \sum_{j=1}^{n_1} X^*_{2,j}, \quad \bar{X}^*_{2,2} = \frac{1}{n_2} \sum_{j=1}^{n_2} X^*_{2,j}, \quad \bar{X}^*_{2,3} = \frac{1}{n_3} \sum_{j=1}^{n_3} X^*_{2,j}, \]

\[ \bar{X}^*_{1} = \frac{1}{N_3} \sum_{j=1}^{N_3} X^*_{1,j}, \quad \bar{X}^*_{2} = \frac{1}{N_2} \sum_{j=1}^{N_2} \left( \begin{array}{c} X^*_{1,j} \\ X^*_{2,j} \end{array} \right), \]

and the corresponding matrices of sums of squares and products

\[ A_1^* = \sum_{j=1}^{N_3} (X^*_{1,j} - \bar{X}^*_{1})(X^*_{1,j} - \bar{X}^*_{1})', \]

\[ A_2^* = \sum_{j=1}^{N_2} \left( \begin{array}{c} X^*_{1,j} \\ X^*_{2,j} \end{array} \right) \left( \begin{array}{c} X^*_{1,j} \\ X^*_{2,j} \end{array} \right)' - \bar{X}^*_{2}, \]

\[ A_3^* = \sum_{j=1}^{N_1} \left( \begin{array}{c} X^*_{1,j} \\ X^*_{2,j} \\ X^*_{3,j} \end{array} \right) \left( \begin{array}{c} X^*_{1,j} \\ X^*_{2,j} \\ X^*_{3,j} \end{array} \right)' - \bar{X}^*_{1,1}, \]

Let \( J \) be defined as in 4.2.1. We also define

\[ B^*_{21} = A^*_{2,21} A^*_{2,11}^{-1}, \quad (B^*_{31}, B^*_{32}) = (A^*_{3,31}, A^*_{3,32})(J A^* J')^{-1}. \]
The maximum likelihood estimators of $\mu_1^*$, $\mu_2^*$, and $\mu_3^*$ are, respectively,

$$\hat{\mu}_1^* = \bar{X}_{1,1}^*$$

(4.3.3)

$$\hat{\mu}_2^* = \tau_{12} \bar{X}_{2,1}^* + \tau_{22} \bar{X}_{2,2}^* - \tau_{33} B_{11}^* \left( \tau_{12} \bar{X}_{1,1}^* + \tau_{22} \bar{X}_{1,2}^* - \bar{X}_{1,3}^* \right)$$

(4.3.4)

and

$$\hat{\mu}_3^* = \bar{X}_{3,1}^* - B_{31}^* \left( \tau_{13} \bar{X}_{1,1}^* - \tau_{23} \bar{X}_{1,2}^* - \tau_{33} \bar{X}_{1,3}^* \right)$$

$$- B_{32}^* \left( \tau_{22} \bar{X}_{2,1}^* - \bar{X}_{2,2}^* \right) + \tau_{33} B_{21}^* \left( \tau_{12} \bar{X}_{1,1}^* + \tau_{22} \bar{X}_{1,2}^* - \bar{X}_{1,3}^* \right)$$

(4.3.5)

Our goal is to show that $\hat{\mu}^* = (\hat{\mu}_1^*, \hat{\mu}_2^*, \hat{\mu}_3^*)' = \Lambda C\hat{\mu} + \nu$. As a consequence of the definitions of the sample mean vectors and matrices of sums of squares and products, we have the following relations

$$\bar{X}_{1,1}^* = \Lambda_{11} \bar{X}_{1,1} + \nu_1, \quad \bar{X}_{1,2}^* = \Lambda_{11} \bar{X}_{1,2} + \nu_1, \quad \bar{X}_{1,3}^* = \Lambda_{11} \bar{X}_{1,3} + \nu_1,$$

and

$$\bar{X}_{2,1}^* = \Lambda_{22} (\Lambda_{21} \bar{X}_{1,1} + \bar{X}_{2,1}) + \nu_2, \quad \bar{X}_{2,2}^* = \Lambda_{22} (\Lambda_{21} \bar{X}_{1,2} + \bar{X}_{2,2}) + \nu_2,$$

$$\bar{X}_{3,1}^* = \Lambda_{33} (\Lambda_{31} \bar{X}_{1,1} + \Lambda_{32} \bar{X}_{2,1} + \bar{X}_{3,1}) + \nu_3, \quad \bar{X}_{1}^* = \Lambda_{11} \bar{X}_{1} + \nu_1,$$

and

$$\bar{X}_2^* = \begin{pmatrix} \Lambda_{11} (\tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2}) + \nu_1 \\ \Lambda_{22} (\Lambda_{21} (\tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2}) + \tau_{12} \bar{X}_{2,1} + \tau_{22} \bar{X}_{2,2}) + \nu_2 \end{pmatrix}$$
The corresponding sums of squares and products are therefore

\[ A_1^* = \Lambda_{11} A_1 A_{11}, \]

\[ A_2^* = \begin{pmatrix} \Lambda_{11} & 0 \\ A_{22} A_{21} & \Lambda_{22} \end{pmatrix} A_2 \begin{pmatrix} \Lambda_{11} & 0 \\ A_{22} A_{21} & \Lambda_{22} \end{pmatrix}', \]

and

\[ A_3^* = \sum_{j=1}^{N_1} \left( \Lambda_{11} \begin{pmatrix} X_{1,j} \\ X_{2,j} \\ X_{3,j} \end{pmatrix} + \nu - \Lambda C \begin{pmatrix} \bar{X}_{1,1} \\ \bar{X}_{2,1} \\ \bar{X}_{3,1} \end{pmatrix} - \nu \right) \begin{pmatrix} \Lambda_{11} \begin{pmatrix} X_{1,j} \\ X_{2,j} \\ X_{3,j} \end{pmatrix} + \nu - \Lambda C \begin{pmatrix} \bar{X}_{1,1} \\ \bar{X}_{2,1} \\ \bar{X}_{3,1} \end{pmatrix} - \nu \end{pmatrix}'. \]

\[ = \Lambda C A_3 C' \Lambda. \]

It also follows from our definition of the sample means and sums of squares and products that

\[ B_{21}^* = \Lambda_{22} (A_{21} A_{21,11} + A_{2,21}) A_{11} (\Lambda_{11} A_{2,11} A_{11})^{-1} = \Lambda_{22} (A_{21} + B_{21}) A_{11}^{-1}, \]
and

\[
(B^*_31, B^*_32) = (\Lambda_{33} A_{31}, \Lambda_{33} A_{32}, \Lambda_{33}) \begin{pmatrix}
A_{3,11} & A_{3,12} \\
A_{3,21} & A_{3,22} \\
A_{3,31} & A_{3,32}
\end{pmatrix}
\begin{pmatrix}
\Lambda_{11} & \Lambda_{12} \\
0 & \Lambda_{22}
\end{pmatrix}^{-1}.
\]

Because

\[
\begin{pmatrix}
\Lambda_{11} & 0 \\
\Lambda_{22} & \Lambda_{22}
\end{pmatrix}^{-1} = \begin{pmatrix} I_{p_1} & 0 \\
0 & I_{p_2}
\end{pmatrix}^{-1}
\begin{pmatrix} \Lambda_{11} & 0 \\
0 & \Lambda_{22}
\end{pmatrix}^{-1} = \begin{pmatrix} \Lambda_{11}^{-1} & 0 \\
0 & \Lambda_{22}^{-1}
\end{pmatrix},
\]

it follows that

\[
(B^*_31, B^*_32) = (\Lambda_{33} (A_{31} + A_{33} B_{31}, A_{33} A_{32} + \Lambda_{33} B_{33})) \begin{pmatrix}
\Lambda_{11}^{-1} & 0 \\
-\Lambda_{21} & \Lambda_{22}^{-1}
\end{pmatrix},
\]

\[
= (\Lambda_{33} (A_{31} + B_{31} - A_{32} A_{21} - B_{32} A_{21}) \Lambda_{11}^{-1}, \Lambda_{33} (A_{32} + B_{32}) \Lambda_{22}^{-1}).
\]

We may write \(\hat{\mu}^*\), \ref{4.3.3} in terms of the original means and matrices of sums of squares and products. It is straightforward that

\[
\hat{\mu}_1^* = \Lambda_{11} \bar{X}_1 + \nu_1 = \Lambda_{11} \hat{\mu}_1 + \nu_1.
\]
The maximum likelihood estimator of $\mu_2^*$, (4.3.4), is

$$
\hat{\mu}_2^* = \tau_{12}(A_{22}(A_{21}X_{1,1} + \bar{X}_{2,1}) + \nu_2) + \tau_{22}(A_{22}(A_{21}X_{1,2} + \bar{X}_{2,2}) + \nu_2)
$$

$$
- \tau_{33}A_{22}(A_{21} + B_{21})A^{-1}_{11}(\tau_{12}(A_{11}X_{1,1} + \nu_1) + \tau_{22}(A_{11}X_{1,2} + \nu_1) - A_{11}\bar{X}_{1,3} - \nu_1)
$$

$$
= A_{22}A_{21}(\tau_{12}X_{1,1} + \tau_{22}X_{2,2}) - \tau_{33}A_{22}A_{21}(\tau_{12}X_{1,1} + \tau_{22}X_{1,2} - \bar{X}_{1,3})
$$

$$
+ A_{22}(\tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2}) + \nu_2 - \tau_{33}A_{22}B_{21}(\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}).
$$

Because $\tau_{12}(1 - \tau_{33}) = \tau_{12}N_2/N_3 = \tau_{13}$ and $\tau_{22}(1 - \tau_{33}) = \tau_{22}N_2/N_3 = \tau_{23}$, it follows that

$$
\hat{\mu}_2^* = A_{22}A_{21}\bar{X}_{1} + A_{22}[\tau_{12}\bar{X}_{2,1} + \tau_{22}\bar{X}_{2,2} - \tau_{33}B_{21}(\tau_{12}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3})] + \nu_2
$$

$$
= A_{22}(A_{21}\bar{\mu}_1 + \bar{\mu}_2) + \nu_2. \quad (4.3.7)
$$

The maximum likelihood estimator of $\mu_3^*$, (4.3.5), is

$$
A_{33}A_{31}\bar{X}_{1,1} + A_{33}A_{32}\bar{X}_{2,1} + A_{33}\bar{X}_{3,1} + \nu_3
$$

$$
- \tau_{33}A_{33}(A_{31} + B_{31} - A_{32}A_{21} - B_{32}A_{21})A^{-1}_{11}
$$

$$
\cdot (\tau_{13}(A_{11}X_{1,1} + \nu_1) - \tau_{23}(A_{11}X_{1,2} + \nu_1) - \tau_{33}(A_{11}X_{1,3} + \nu_1))
$$

$$
- \tau_{33}A_{32}B_{32}A_{22}(\tau_{22}(A_{22}A_{21}\bar{X}_{1,1} + A_{22}\bar{X}_{2,1} + \nu_2 - A_{22}A_{21}\bar{X}_{1,2} - A_{22}\bar{X}_{2,2} - \nu_2)
$$

$$
+ \tau_{33}A_{22}(A_{21} + B_{21})A^{-1}_{11}
$$

$$
\cdot (\tau_{12}(A_{11}\bar{X}_{1,1} + \nu_1) + \tau_{22}(A_{11}\bar{X}_{1,2} + \nu_1) - A_{11}\bar{X}_{1,3} - \nu_1). \quad (4.3.8)
$$
After combining terms, $\hat{\mu}_3^*$ equals

\[
\Lambda_{33}[(\Lambda_{31}\tau_{13} + \Lambda_{32}\Lambda_2[(-\tau_{23} + \tau_{22} - \tau_{33}\tau_{12})]X_{1,1} \\
+ (\Lambda_{31}\tau_{23} + \Lambda_{32}\Lambda_2[-\tau_{33}\tau_{22}]X_{1,2} + \Lambda_{31}\tau_{33}X_{1,3}] + \nu_3 \\
+ \Lambda_{33}\Lambda_3\tau_{12}X_{2,1} + \tau_{22}X_{2,2} - \tau_{33}B_{21}(\tau_{12}X_{1,1} + \tau_{22}X_{1,2} - X_{1,3})] \\
+ \Lambda_{33}\Lambda_3\tau_{33}X_{3,1} - B_{31}(\tau_{33}\tau_{13}X_{1,1} - \tau_{33}\tau_{23}X_{1,2} - \tau_{33}X_{1,3}) \\
- B_{32}(\tau_{22}(X_{2,1} - X_{2,2}) + \tau_{33}B_{21}(\tau_{12}X_{1,1} + \tau_{22}X_{1,2} - X_{1,3})) \\
+ \Lambda_{33}\Lambda_3\Lambda_2[(-\tau_{23} + \tau_{22} - \tau_{33}\tau_{22})X_{1,2} + (-\tau_{33} + \tau_{33})X_{1,3}].
\]

Because

\[
\tau_{33}\tau_{12} - \tau_{13} = \frac{n_3n_1 - N_3N_2 + n_1N_2}{N_2N_3} = \frac{n_1N_3 - N_2N_3}{N_2N_3} = -\tau_{22},
\]

and

\[
\tau_{33}\tau_{22} + \tau_{23} = \frac{n_3n_2 + n_2N_2}{N_2N_3} = \tau_{22},
\]

it follows that

\[
\hat{\mu}_3^* = \Lambda_{33}\Lambda_{31}X_1 + \nu_3 + \Lambda_{33}\Lambda_{32}[\tau_{12}X_{2,1} + \tau_{22}X_{2,2} - \tau_{33}B_{21}(\tau_{12}X_{1,1} + \tau_{22}X_{1,2} - X_{1,3})] \\
+ \Lambda_{33}\Lambda_3[\tau_{33}(X_{3,1} - B_{31}(\tau_{33}\tau_{13}X_{1,1} - \tau_{33}\tau_{23}X_{1,2} - \tau_{33}X_{1,3}) \\
- B_{32}(\tau_{22}(X_{2,1} - X_{2,2}) + \tau_{33}B_{21}(\tau_{12}X_{1,1} + \tau_{22}X_{1,2} - X_{1,3}))]] \\
= \Lambda_{33}[\Lambda_{31}\hat{\mu}_1 + \Lambda_{32}\hat{\mu}_2 + \hat{\mu}_3] + \nu_3.
\]
Combining equations (4.3.6), (4.3.7), and (4.3.10), yields

\[
\begin{pmatrix}
\hat{\mu}_1
\hat{\mu}_2
\hat{\mu}_3
\end{pmatrix} =
\begin{pmatrix}
\Lambda_{11} \hat{\mu}_1 + \nu_1 \\
\Lambda_{22} (\Lambda_{21} \hat{\mu}_1 + \hat{\mu}_2) + \nu_2 \\
\Lambda_{33} (\Lambda_{31} \hat{\mu}_1 + \Lambda_{32} \hat{\mu}_2 + \hat{\mu}_3) + \nu_3
\end{pmatrix} = \Lambda C \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{pmatrix} + \nu;
\]

(4.3.11)

therefore we have proved that \( \hat{\mu} \) is equivariant.

In order to find the transformation that yields \( \Sigma = I \), we must first find the partial Iwasawa coordinates in the three-step setting.

**Proposition 4.3.2.** Let the \((p_1 + p_2 + p_3) \times (p_1 + p_2 + p_3)\) positive definite matrix \( M \) be partitioned into \( p_1, p_2, \) and \( p_3 \) rows and columns, i.e.,

\[
M = \begin{pmatrix}
M_{11} & M_{12} & M_{13} \\
M_{21} & M_{22} & M_{23} \\
M_{31} & M_{32} & M_{33}
\end{pmatrix},
\]

where \( M_{11} \) is \( p_1 \times p_1 \), \( M_{12} = M_{21}' \) is \( p_1 \times p_2 \), and so on. Define \( M_{33:12} = M_{33:1} - M_{32:1} M_{22:1}^{-1} M_{23:1} \). Then, the partial Iwasawa coordinates of \( M \) are \( \{M_{11}, M_{22:1}, M_{33:12}, M_{11}^{-1} M_{12}, M_{11}^{-1} M_{13}, M_{22:1}^{-1} M_{23:1}\} \) and \( M \) equals

\[
\begin{pmatrix}
I_{p_1} & 0 & 0 \\
M_{21} M_{11}^{-1} & I_{p_2} & 0 \\
M_{31} M_{11}^{-1} & M_{32:1} M_{22:1}^{-1} & I_{p_3}
\end{pmatrix}
\begin{pmatrix}
M_{11} & 0 & 0 \\
0 & M_{22:1} & 0 \\
0 & 0 & M_{33:12}
\end{pmatrix}
\begin{pmatrix}
I_{p_1} & M_{11}^{-1} M_{12} & M_{11}^{-1} M_{13} \\
0 & I_{p_2} & M_{22:1}^{-1} M_{23:1}
\end{pmatrix}.
\]

(4.3.12)
Further $M^{-1}$ equals

$$
\begin{pmatrix}
I_{p_1} & -M^{-1}_{11}M_{12} & -M^{-1}_{11}[M_{13} - M_{12}M^{-1}_{22}M^{-1}_{22}]M_{22}^{-1}M_{23}^{-1} \\
0 & I_{p_2} & -M^{-1}_{22}M_{23}^{-1} \\
0 & 0 & I_{p_3}
\end{pmatrix}
\cdot
\begin{pmatrix}
M^{-1}_{11} & 0 & 0 \\
0 & M^{-1}_{22} & 0 \\
0 & 0 & M^{-1}_{33}
\end{pmatrix}
\cdot
\begin{pmatrix}
I_{p_1} & 0 & 0 \\
-M_{21}M^{-1}_{11} & I_{p_2} & 0 \\
-[M_{31} - M_{32}M^{-1}_{22}M_{21}]M^{-1}_{11} & -M_{32}M^{-1}_{22} & I_{p_3}
\end{pmatrix}
$$

Proof. The form of the partial Iwasawa coordinates is such that

$$
M = \begin{pmatrix}
M_{11} & M_{12} & M_{13} \\
M_{21} & M_{22} & M_{23} \\
M_{31} & M_{32} & M_{33}
\end{pmatrix}
= \begin{pmatrix}
A_{11} & 0 & 0 \\
A_{21} & A_{22} & 0 \\
A_{31} & A_{32} & A_{33}
\end{pmatrix}
\cdot
\begin{pmatrix}
A_{11} & A_{12} & A_{13} \\
0 & A_{22} & A_{23} \\
0 & 0 & A_{33}
\end{pmatrix}
= \begin{pmatrix}
A^{2}_{11} & A_{11}A_{12} & A_{11}A_{13} \\
A_{21}A_{11} & A_{21}A_{12} + A^{2}_{22} & A_{21}A_{13} + A_{22}A_{23} \\
A_{31}A_{11} & A_{31}A_{12} + A_{32}A_{22} & A_{31}A_{13} + A_{32}A_{23} + A^{2}_{33}
\end{pmatrix}.
$$

Solving for each term sequentially yields

$$
A_{11} = M_{11}^{1/2}, \quad A_{21} = M_{21}M_{11}^{-1/2}, \quad A_{31} = M_{31}M_{11}^{-1/2},
$$

$$
A_{22} = M_{22}^{1/2}, \quad A_{32} = M_{32}M_{22}^{-1/2}, \quad A_{33} = M_{33}^{1/2}.
$$
Because
\[
\begin{pmatrix}
A_{11} & 0 & 0 \\
A_{21} & A_{22} & 0 \\
A_{31} & A_{32} & A_{33}
\end{pmatrix} =
\begin{pmatrix}
I_{p_1} & 0 & 0 \\
A_{21}A_{11}^{-1} & I_{p_2} & 0 \\
A_{31}A_{11}^{-1} & A_{32}A_{22}^{-1} & I_{p_3}
\end{pmatrix}
\begin{pmatrix}
A_{11} & 0 & 0 \\
0 & A_{22} & 0 \\
0 & 0 & A_{33}
\end{pmatrix},
\]
it follows that \( M \) equals
\[
\begin{pmatrix}
I_{p_1} & 0 & 0 \\
M_{21}M_{11}^{-1} & I_{p_2} & 0 \\
M_{31}M_{11}^{-1} & M_{32}M_{22}^{-1} & I_{p_3}
\end{pmatrix}
\begin{pmatrix}
M_{11} & 0 & 0 \\
0 & M_{22} & 0 \\
0 & 0 & M_{33}
\end{pmatrix}
\begin{pmatrix}
I_{p_1} & M_{11}^{-1}M_{12} & M_{11}^{-1}M_{13} \\
0 & I_{p_2} & M_{22}^{-1}M_{23} \\
0 & 0 & I_{p_3}
\end{pmatrix}.
\]

By taking
\[
\Lambda_{11} = \Sigma_{11}^{-1/2}, \quad \Lambda_{21} = -\Sigma_{21} \Sigma_{11}^{-1}
\]
\[
\Lambda_{22} = \Sigma_{22}^{-1/2}, \quad \Lambda_{31} = -(\Sigma_{31} - \Sigma_{32} \Sigma_{22}^{-1} \Sigma_{21}) \Sigma_{11}^{-1}
\]
\[
\Lambda_{33} = \Sigma_{33}^{-1/2}, \quad \Lambda_{32} = -\Sigma_{32} \Sigma_{22}^{-1} \Sigma_{21}^{-1}
\]
then under the transformation (4.3.2), we obtain
\[
\text{Cov}
\begin{pmatrix}
X_1 \\
X_2 \\
X_3
\end{pmatrix}
= \Lambda \Sigma \Sigma' \Lambda = I_{p_1 + p_2 + p_3}.
\]
Therefore, in analyzing the distribution of $\hat{\mu}$, we may assume, without loss of generality, that $\Sigma = I_{p_1+p_2+p_3}$. Furthermore, by choosing $\nu = -\Lambda C \mu$ we may also assume, without loss of generality, that $\mu = 0$.

The following theorem provides the conditional distribution of $\hat{\mu}_3$ given the data $X_1$ and $X_2$ defined in (4.1.1) and (4.1.2), respectively. Following this theorem, we extend this distribution as far as we can in our attempt to find the joint distribution of $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$.

**Theorem 4.3.3.** Let $v' = (v'_1, v'_2) = (v'_1, (v_{21} + \tau_{33} B_{21} v_{22})')$, where $v_1 = \bar{\tau}_{13} \bar{X}_{1,1} - \tau_{23} \bar{X}_{1,2} - \tau_{33} \bar{X}_{1,3}$, $v_{21} = \tau_{22} (\bar{X}_{2,1} - \bar{X}_{2,2})$, and $v_{22} = \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_{1,3}$. Then

$$\hat{\mu}_3 | \{X_1, X_2\} \sim N\left(0, \left(\frac{1}{n_1} + v'(JA_3J')^{-1}v\right)I_{p_3}\right), \quad (4.3.15)$$

where $J$ and $A_3$ are defined in (4.2.1) and (4.1.6).

**Proof.** Assume, without loss of generality, that $\mu = 0$ and $\Sigma = I_{p_1+p_2+p_3}$; then, for $j = 1, \ldots, n_1$, $X_{3,j} | \{X_{1,j}, X_{2,j}\} \sim N_{p_3}(0, I_{p_3})$. Therefore $\bar{X}_{3,1} | \{X_1, X_2\} \sim N_{p_3}(0, \frac{1}{n} I_{p_3})$. Conditional on the first $p_1 + p_2$ rows of data, $X_1$ and $X_2$, $\hat{\mu}$ is a linear function of the last $p_3$ rows of data, $X_3$, (4.1.3), and therefore is normal. We now find the conditional expected value and covariance matrix of $\hat{\mu}_3$. By Proposition (2.3.3)(ii),

$$\begin{pmatrix} A_{3,31} \\ A_{3,32} \end{pmatrix} \begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix} \sim N\left(0, I_{p_3} \otimes \begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix}\right).$$
Because $EX_{3,j} = E\bar{X}_{3,1} = 0$,

$$E((B_{31}, B_{32}) \mid \{X_1, X_2\}) = E((A_{3,31}, A_{3,32}) \mid \{X_1, X_2\}) \left( \begin{array}{cc} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{array} \right)^{-1}$$

$$= (\sum_{j=1}^{n_1} E[(X_{3,j} - \bar{X}_{3,1}) \mid \{X_1, X_2\}] (X_{1,j} - \bar{X}_{1,1}')), \quad \sum_{j=1}^{n_1} E[(X_{3,j} - \bar{X}_{3,1}) \mid \{X_1, X_2\}] (X_{2,j} - \bar{X}_{2,1}')) (JA_3J')^{-1} = 0.$$

Therefore,

$$E(\hat{\mu}_3 \mid \{X_1, X_2\})$$

$$= E\left(\left[\bar{X}_{3,1} - B_{31}(\bar{X}_{1,1} - \tau_{33} \bar{X}_{1,2} - \tau_{33} \bar{X}_{1,3})ight.\right.$$  

$$\left. - B_{32}(\tau_{33} B_{21}(\bar{X}_{1,2,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_{1,3})) \right] \mid \{X_1, X_2\})$$

$$= 0. \quad (4.3.16)$$

Let $v$ be as defined in Theorem 4.3.3; then $E(\hat{\mu}_3 \hat{\mu}_3' \mid \{X_1, X_2\})$, the conditional covariance matrix of $\hat{\mu}_3$ given $\{X_1, X_2\}$, equals

$$E\left[\bar{X}_{3,1} \bar{X}_{3,1}' - \bar{X}_{3,1} v_1 B_{13} - B_{31} v_1 \bar{X}_{3,1}' - \bar{X}_{3,1} [v_21 + \tau_{33} B_{21} v_{22}]' B_{23}ight.$$

$$- B_{32} [v_21 + \tau_{33} B_{21} v_{22}] \bar{X}_{3,1}' + B_{31} v_1 v_1' B_{13} + B_{31} v_1 [v_21 + \tau_{33} B_{21} v_{22}]' B_{23}$$

$$+ B_{32} [v_21 + \tau_{33} B_{21} v_{22}] v_1' B_{13}$$

$$+ B_{32} [v_21 + \tau_{33} B_{21} v_{22}] [v_21 + \tau_{33} B_{21} v_{22}]' B_{23} \mid \{X_1, X_2\}] \right]. \quad (4.3.17)$$
We examine each expected value separately. Because $X_{3,1}$ is independent of $\{X_1, X_2\}$ and $X_{3,1} \sim N(0, \frac{1}{n_1} I_{p_3})$, then $E(X_{3,1}X_3') = \frac{1}{n_1} I_{p_3}$. Define

$$A_{3,ij-k} = A_{3,ij} - A_{3,ik}A_{3,kk}^{-1}A_{3,kj}$$

$$= \sum_{s=1}^{n_1} (X_{i,s} - \bar{X}_{i,1})(X_{j,s} - \bar{X}_{j,1}) - A_{3,jk}A_{3,kk}^{-1}(X_{k,s} - \bar{X}_{k,1})$$

for $i = 1, 2, 3$, and $j, k = 1, 2$. By decomposing $\begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix}^{-1}$ into its partial Iwasawa coordinates, (2.1.3), we see that

$$\begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{3,11}^{-1} & -A_{3,11}^{-1}A_{3,12}A_{3,22}^{-1} \\ -A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} & A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} + A_{3,22}^{-1} \\ -A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} & A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} + A_{3,22}^{-1} \\ -A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} & A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} + A_{3,22}^{-1} \end{pmatrix}. $$

Therefore, it follows that

$$B_{31} = A_{3,31}A_{3,11}^{-1} - A_{3,32}A_{3,22}^{-1}A_{3,21}A_{3,11}^{-1} = A_{3,31}A_{3,11}^{-1}$$

(4.3.18)

and, similarly, $B_{32} = A_{3,32}A_{3,22}^{-1}$. We will use these definitions for $B_{31}$ and $B_{32}$ to calculate the other expected values found in (4.3.17).

Next, $E[B_{31}v_1\tilde{X}'_{3,1} | \{X_1, X_2\}]$ equals

$$E\left[ \sum_{s=1}^{n_1} (X_{3,s} - \bar{X}_{3,1})(X_{1,s} - \bar{X}_{1,1}) - A_{3,21}A_{3,22}^{-1}(X_{2,s} - \bar{X}_{2,1})'A_{3,11}^{-1}v_1 \tilde{X}_{3,1}' | \{X_1, X_2\} \right]$$

$$= \sum_{s=1}^{n_1} \left[ (X_{1,s} - \bar{X}_{1,1}) - A_{3,12}A_{3,22}^{-1}(X_{2,s} - \bar{X}_{2,1})'A_{3,11}^{-1}v_1 \right]$$

$$\cdot E[(X_{3,s} - \bar{X}_{3,1})\tilde{X}_{3,1}' | \{X_1, X_2\}].$$
Because $E[(X_{3,s} - \bar{X}_{3,1})' \bar{X}_{3,1} | \{X_1, X_2\}]$ does not depend on $s$, it equals its average over $s$ and therefore

$$E[(X_{3,s} - \bar{X}_{3,1})' \bar{X}_{3,1} | \{X_1, X_2\}] = \frac{1}{n_1} \sum_{j=1}^{n_1} E[(X_{3,j} - \bar{X}_{3,1})' \bar{X}_{3,1} | \{X_1, X_2\}]$$

$$= \frac{1}{n_1} E \left[ \sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})' \bar{X}_{3,1} | \{X_1, X_2\} \right] = 0,$$

(4.3.19)

because $\sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1}) = 0$. Therefore $E[B_{31} v_1 \bar{X}_{3,1} | \{X_1, X_2\}] = 0$. Similarly, by (4.3.19), $E[B_{32} (v_{21} + \tau_{33} B_{21} v_{22}) \bar{X}_{3,1} | \{X_1, X_2\}]$ equals

$$E \left[ \sum_{s=1}^{n_1} (X_{3,s} - \bar{X}_{3,1})(X_{2,s} - \bar{X}_{2,1}) - A_{3,21} A_{3,11}^{-1} (X_{1,s} - \bar{X}_{1,1})' A_{3,221}^{-1} v_2 \bar{X}_{3,1} | \{X_1, X_2\} \right]$$

$$= \sum_{s=1}^{n_1} \left[ (X_{2,s} - \bar{X}_{2,1}) - A_{3,21} A_{3,11}^{-1} (X_{1,s} - \bar{X}_{1,1})' A_{3,221}^{-1} v_2 \right]$$

$$\cdot E[(X_{3,s} - \bar{X}_{3,1})' \bar{X}_{3,1} | \{X_1, X_2\}] = 0.$$

(4.3.20)

We may write $B_{31} v_1 v_1' B_{13}$ as $A_{3,312}^{-1} A_{3,112}^{-1} v_1 v_1' A_{3,132}$, which equals

$$\sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})(X_{1,j} - \bar{X}_{1,1}) - A_{3,12} A_{3,22}^{-1} (X_{2,j} - \bar{X}_{2,1})' A_{3,112}^{-1} v_1 v_1' A_{3,112}^{-1}$$

$$\cdot \sum_{k=1}^{n_1} \left[ (X_{1,k} - \bar{X}_{1,1}) - A_{3,12} A_{3,22}^{-1} (X_{2,k} - \bar{X}_{2,1}) \right] (X_{3,k} - \bar{X}_{3,1})'$$

$$= \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} c_{jk} (X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})',$$

(4.3.21)
where the constant scalar

\[ c_{jk} = [X_{1,j} - \bar{X}_{1,1} - A_{3,12} A_{3,22}^{-1} (X_{2,j} - \bar{X}_{2,1})]'A_{3,11-2}^{-1} \nu_1 \]

\[ \cdot v_1' A_{3,11-2}^{-1} [X_{1,k} - \bar{X}_{1,1} - A_{3,12} A_{3,22}^{-1} (X_{2,k} - \bar{X}_{2,1})]. \]  

(4.3.22)

By interchanging summation and expectation, we need only calculate

\[ E[(X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})'] \]  

(\{X_1, X_2\}). Because \( \Sigma = I_d \), so that the \( \{X_{3,j}\} \) are independent of \( \{X_{1,j}, X_{2,j}\} \), we may drop the condition on \( \{X_1, X_2\} \). Because \( \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})' = 0 \), then

\[ \sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})(X_{3,j} - \bar{X}_{3,1})' = -\sum_{j=1}^{n_1} \sum_{k \neq j} (X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})'. \]  

(4.3.23)

Because we may permute the \( n_1 \) vectors within \( X_3 \), defined in (4.1.3), without changing the distribution of \( X_{3,j} - \bar{X}_{3,1} \), then \( E(X_{3,j} - \bar{X}_{3,1})(X_{3,j} - \bar{X}_{3,1})' \) does not depend on \( j \) and \( E(X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})' \), \( j \neq k \), does not depend on \( j \) or \( k \). Therefore each equals its average. That is, \( E \sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})(X_{3,j} - \bar{X}_{3,1})' = (n_1 - 1)I_{p_3} \), and therefore

\[ E[(X_{3,j} - \bar{X}_{3,1})(X_{3,j} - \bar{X}_{3,1})'] = \frac{n_1 - 1}{n_1} I_{p_3}, \]

and by (4.3.23),

\[ E[(X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})'] = -\frac{n_1 - 1}{n_1^2 - n_1} I_{p_3} = -\frac{1}{n_1} I_{p_3}, \quad j \neq k. \]
Therefore, for any $j, k$,

$$E[(X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})] = \delta_{jk} I_{p_3} - \frac{1}{n_1} I_{p_3} = \left(\delta_{jk} - \frac{1}{n_1}\right) I_{p_3}, \quad (4.3.24)$$

where $\delta_{jk}$ is Kronecker’s delta. Also, because

$$v' A^{-1}_{3,11} \sum_{j=1}^{n_1} [X_{1,j} - \bar{X}_{1,1} - A_{3,12} A^{-1}_{3,22} (X_{2,j} - \bar{X}_{2,1})] = 0, \quad (4.3.25)$$

then $\sum_{j=1}^{n_1} \sum_{k=1}^{n_1} c_{jk} = 0, \quad (4.3.22)$. Therefore,

$$E\left[\sum_{j=1}^{n_1} \sum_{k=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})' c_{jk} | \{X_1, X_2\}\right] = \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \delta_{jk} c_{jk} I_{p_3}.$$ 

However, $\sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \delta_{jk} c_{jk} = \sum_{j=1}^{n_1} c_{jj}$, which, by (4.3.22) and (4.3.25), equals

$$v' A^{-1}_{3,11} \sum_{j=1}^{n_1} [X_{1,j} - \bar{X}_{1,1} - A_{3,12} A^{-1}_{3,22} (X_{2,j} - \bar{X}_{2,1})]$$

$$\cdot [X_{1,j} - \bar{X}_{1,1} - A_{3,12} A^{-1}_{3,22} (X_{2,j} - \bar{X}_{2,1})]' A^{-1}_{3,11} v_1$$

$$= v' A^{-1}_{3,11} \left( A_{3,11} - A_{3,12} A^{-1}_{3,22} A_{3,21} - A_{3,12} A^{-1}_{3,22} A_{3,21} + A_{3,12} A^{-1}_{3,22} A_{3,22} A^{-1}_{3,22} A_{3,21} \right)$$

$$\cdot A^{-1}_{3,11} v_1$$

$$= v' A^{-1}_{3,11} v_1.$$ 

Therefore, $E[B_{31} v_1 v' B_{13} | \{X_1, X_2\}] = v' A^{-1}_{3,11} v_1 I_{p_3}.$
Similarly, we may rewrite \( B_{32}[v_{21} + \tau_{33}B_{21}v_{22}][v_{21} + \tau_{33}B_{21}v_{22}]'B_{23} \) as

\[
A_{3,32-1} A_{3,32}^{-1} v_2 v_2' A_{3,32-1}^{-1} A_{3,32-1},
\]

which equals

\[
\sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1}) (X_{2,j} - \bar{X}_{2,1} - A_{3,21} A_{3,11}^{-1} (X_{1,j} - \bar{X}_{1,1}))' A_{3,32-1}^{-1} v_2 v_2' A_{3,32-1}^{-1}
\]

\[
\cdot \sum_{k=1}^{n_1} [X_{2,k} - \bar{X}_{2,1} - A_{3,21} A_{3,11}^{-1} (X_{1,k} - \bar{X}_{1,1})] (X_{3,k} - \bar{X}_{3,1})'
\]

\[
= \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} c_{jk} (X_{3,j} - \bar{X}_{3,1}) (X_{3,k} - \bar{X}_{3,1})',
\]

(4.3.26)

where the constant scalar

\[
c_{jk} = [X_{2,j} - \bar{X}_{2,1} - A_{3,21} A_{3,11}^{-1} (X_{1,j} - \bar{X}_{1,1})]' A_{3,32-1}^{-1} v_2
\]

\[
v_2' A_{3,32-1}^{-1} [X_{2,k} - \bar{X}_{2,1} - A_{3,21} A_{3,11}^{-1} (X_{1,k} - \bar{X}_{1,1})].
\]

(4.3.27)

By (4.3.24),

\[
E[B_{32} v_2 v_2' B_{23} | \{X_1, X_2\}] = \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} (\delta_{jk} - \frac{1}{n_1}) c_{jk} I_p.
\]

(4.3.28)

Again, \( \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} c_{jk} = 0 \), (4.3.27), because

\[
v_2' A_{3,32-1}^{-1} \sum_{k=1}^{n_1} [X_{2,k} - \bar{X}_{2,1} - A_{3,21} A_{3,11}^{-1} (X_{1,k} - \bar{X}_{1,1})] = 0.
\]

(4.3.29)
However, \( \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \delta_{jk} c_{jk} = \sum_{j=1}^{n_1} c_{jj} \), which, by (4.3.27), equals

\[
v'_2 A^{-1}_{3,22,1} \sum_{j=1}^{n_1} [X_{2,j} - \bar{X}_{2,1} - A_{3,21} A^{-1}_{3,11} (X_{1,j} - \bar{X}_{1,1})] \\
\quad \cdot [X_{2,j} - \bar{X}_{2,1} - A_{3,21} A^{-1}_{3,11} (X_{1,j} - \bar{X}_{1,1})]' A^{-1}_{3,22,1} v_2 \\
= v'_2 A^{-1}_{3,22,1} (A_{3,22} - A_{3,21} A^{-1}_{3,11} A_{3,12} - A_{3,21} A^{-1}_{3,11} A_{3,12} + A_{3,21} A^{-1}_{3,11} A_{3,11} A^{-1}_{3,11} A_{3,12}) \\
\quad \cdot A^{-1}_{3,22,1} v_2 \\
= v'_2 A^{-1}_{3,22,1} v_2.
\]

Therefore, \( E[B_{32} v_2 v'_2 B_{23}] \{X_1, X_2\} = v'_2 A^{-1}_{3,22,1} v_2 I_{p_3} \).

The final term in (4.3.17) is the cross-product \( B_{32} v_2 v'_1 B_{13} = A_{3,32,1} A^{-1}_{3,22,1} v_2 \cdot v'_1 A^{-1}_{3,11,2} A_{3,13,2} \), which equals

\[
\sum_{j=1}^{n_1} (X_{3,j} - \bar{X}_{3,1}) [X_{2,j} - \bar{X}_{2,1} - A_{3,21} A^{-1}_{3,11} (X_{1,j} - \bar{X}_{1,1})]' A^{-1}_{3,22,1} v_2 v'_1 A^{-1}_{3,11,2} \\
\quad \cdot \sum_{k=1}^{n_1} [X_{1,k} - \bar{X}_{1,1} - A_{3,12} A^{-1}_{3,22} (X_{2,k} - \bar{X}_{2,1})]' (X_{3,k} - \bar{X}_{3,1})' \\
= \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} (X_{3,j} - \bar{X}_{3,1})(X_{3,k} - \bar{X}_{3,1})' c_{jk}, \quad (4.3.30)
\]

where

\[
c_{jk} = [X_{2,j} - \bar{X}_{2,1} - A_{3,21} A^{-1}_{3,11} (X_{1,j} - \bar{X}_{1,1})]' A^{-1}_{3,22,1} v_2 \\
v'_1 A^{-1}_{3,11,2} [X_{1,k} - \bar{X}_{1,1} - A_{3,12} A^{-1}_{3,22} (X_{2,k} - \bar{X}_{2,1})]. \quad (4.3.31)
\]
By (4.3.24),

$$E[B_{32}v_2v_1^TB_{13}|\{X_1, X_2\}] = \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \left( \delta_{jk} - \frac{1}{n_1} \right) c_{kj} I_{p_3}.$$  \hspace{1cm} (4.3.32)

By (4.3.25) and (4.3.29), \(\sum_{j=1}^{n_1} \sum_{k=1}^{n_1} c_{jk} = 0\), (4.3.31). As before, \(\sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \delta_{jk} c_{jk} = \sum_{j=1}^{n_1} c_{jj}\), which, by (4.3.31), equals

$$v_2^{-1} A_{3,22}^{-1} \frac{n_1}{2} \left[ X_{2,j} - \bar{X}_{2,1} - A_{3,21} A_{3,11}^{-1} (X_{1,j} - \bar{X}_{1,1}) \right]$$

$$\cdot \left[ X_{1,j} - \bar{X}_{1,1} - A_{3,12} A_{3,22}^{-1} (X_{2,j} - \bar{X}_{2,1}) \right] A_{3,11}^{-1} v_1$$

$$= v_2^{-1} A_{3,21} A_{3,11}^{-1} \left( A_{3,21} - A_{3,22} A_{3,22}^{-1} A_{3,21} - A_{3,21} A_{3,11}^{-1} A_{3,11} + A_{3,21} A_{3,11}^{-1} A_{3,12} A_{3,22} A_{3,21} \right)$$

$$\cdot A_{3,11}^{-1} v_1$$

$$= v_2^{-1} A_{3,21} A_{3,11}^{-1} \left( -A_{3,11} + A_{3,12} A_{3,22} A_{3,21} \right) A_{3,11}^{-1} v_1$$

$$= -v_2^{-1} A_{3,21} A_{3,11}^{-1} v_1.$$  \hspace{1cm}

Therefore \(E[B_{32}v_2v_1^TB_{13}|\{X_1, X_2\}] = -v_2^{-1} A_{3,21} A_{3,11}^{-1} v_1 I_{p_3}\).

Substituting each expected value into (4.3.17), we obtain

$$E(\hat{\mu}_3\hat{\mu}_3'|\{X_1, X_2\}) = \left( \frac{1}{n_1} + v_1 A_{3,11}^{-1} v \right) I_{p_3}$$

$$= \left( \frac{1}{n_1} + v(JA_3J')^{-1}v \right) I_{p_3},$$  \hspace{1cm} (4.3.33)
because

\[
\begin{align*}
\begin{pmatrix} v_1 & A_{3,11}^{-1} v_{12} & -2v_2 & A_{3,21}^{-1} v_1 + v_2 A_{3,22}^{-1} v_2 \\
\end{pmatrix} & = \\
\begin{pmatrix} v_1 \\
v_2 \\
\end{pmatrix} & = \\
\begin{pmatrix} A_{3,11}^{-1} & -A_{3,12} A_{3,21}^{-1} A_{3,11} \\
-A_{3,21} A_{3,21}^{-1} & A_{3,22}^{-1} A_{3,21}^{-1} A_{3,22} \\
\end{pmatrix} & \begin{pmatrix} v_1 \\
v_2 \\
\end{pmatrix}.
\end{align*}
\]

Therefore,

\[
\hat{\mu}_3 | \{X_1, X_2\} \sim N_{p_3} \left( 0, \left( \frac{1}{n_1} + v'(JA_3J')^{-1}v \right) I_{p_3} \right).
\]

We may not apply Proposition 2.3.3 (iv) to \(v'(JA_3J')^{-1}v\) because \(v\) and \(A_3\) are not independent, nor do we know their joint distribution. This is essentially the problem we have not been able to overcome in order to find the joint distribution of \(\hat{\mu}_1, \hat{\mu}_2,\) and \(\hat{\mu}_3\). We have, however, been able to simplify this difficulty a little further by identifying a Beta random variable. Because \(v\) and \(A_3\) are sufficient statistics, the conditional characteristic function of \(\hat{\mu}_3\) given \(\{X_1, X_2\}\) is, for \(t \in \mathbb{R}^{p_3}\),

\[
E(e^{it'\hat{\mu}_3}|\{X_1, X_2\}) = E(e^{it'\hat{\mu}_3}|\{v, A_3\})
\]

\[
= \exp\left[-\frac{1}{2n_1} t't\right] \exp\left[-\frac{1}{2} v'(JA_3J')^{-1} v't\right].
\]
Therefore, the unconditional characteristic function of \( \hat{\mu}_3 \) is

\[
E(e^{it\hat{\mu}_3}) = E_{v,A_3}E_{\{\check{x}_{31},B_{31},B_{32}\} \mid \{v,A_3\}}(e^{it\hat{\mu}_3}) \\
= \exp[-\frac{1}{2n_1}t't]E_{v,A_3}(\exp[-\frac{1}{2}v'(J A_3 J')^{-1}v't]) \tag{4.3.34}
\]

Let \( A = JA_3 J' \) and \( B = C + D \), where \( C \) and \( D \) are as defined in \([4.2.2]\) and \([4.2.3]\).

By Proposition \([2.3.3]\) (v), it follows that \( L = (A + B)^{-1/2}A(A + B)^{-1/2} \) is independent of \( P = A + B \equiv A_2 \), and \( L \sim Beta_p((n_1 - 1)/2, n_2/2) \). Therefore,

\[
E(e^{it\hat{\mu}_3}) = \exp[-\frac{1}{2n_1}t't]E_{v,P,L}(\exp[-\frac{1}{2}v'(P^{1/2}LP^{1/2})^{-1}v't]) \\
= \exp[-\frac{1}{2n_1}t't]E_{v,P,L}(\exp[-\frac{1}{2}\text{tr}(L^{-1}P^{-1/2}vtv'P^{-1/2})]) \tag{4.3.35}
\]

Let \( M = P^{-1/2}vtv'P^{-1/2} = P^{-1/2}vv'P^{-1/2}t' \). Because \( vv' \) is of rank one, then \( M \) also is of rank one. By Proposition \([2.1.1]\) (v), there exists an orthogonal matrix \( H \in O(p_1 + p_2) \) such that

\[
HP^{-1/2}vv'P^{-1/2}H' = HMH' = \begin{pmatrix} \lambda_1 & 0 \\ 0 & 0 \end{pmatrix}.
\]
By taking traces of both sides, we see that $\lambda_1 = \text{tr}(P^{-1/2}vv'P^{-1/2}) = v'P^{-1}v$. Because $L$ is orthogonally invariant, then \((4.3.35)\) becomes

$$
E_{v,P}E_{L\{v,P\}} \exp\left[-\frac{1}{2}\text{tr}(L^{-1}M)\right] = E \exp\left[-\frac{1}{2}\text{tr}\left((H'LH)^{-1}M\right)\right]
$$

$$
= E \exp\left[-\frac{1}{2}\text{tr}\left(L^{-1}HMH'\right)\right]
$$

$$
= E \exp\left[-\frac{1}{2}\text{tr}\left(L^{-1}\begin{pmatrix} v'P^{-1}v & 0 \\ 0 & 0 \end{pmatrix}\right)\right]
$$

$$
= E \exp\left[-\frac{1}{2}L_{11:2}^{-1}v'P^{-1}vt\right], \quad (4.3.36)
$$

where $L_{11:2}^{-1}$ is formed from $L$ partitioned as the $(1, 1)$ element and the $(p_1 + p_2 - 1) \times (p_1 + p_2 - 1)$ sub-matrix. Our next idea was to apply Anderson’s identity, \((2.1.4)\), to $v'P^{-1}v$ as follows

$$
v'P^{-1}v = v'_1P_{11}^{-1}v_1 + (v_2 - P_{21}P_{11}^{-1}v_1)'P_{22:1}^{-1}(v_2 - P_{21}P_{11}^{-1}v_1), \quad (4.3.37)
$$

where $v_2 - P_{21}P_{11}^{-1}v_1$ equals

$$
\tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2})
$$

$$
+ P_{21}P_{11}^{-1}\left(\tau_{33}(\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2} - \bar{X}_{1,3}) - \tau_{13}\bar{X}_{1,1} + \tau_{23}\bar{X}_{1,2} + \tau_{33}\bar{X}_{1,3}\right).
$$
By 4.3.8, \( \tau_{33}\tau_{12} - \tau_{13} = -\tau_{22} \), and by 4.3.9, \( \tau_{33}\tau_{22} + \tau_{23} = \tau_{22} \). Therefore,

\[
v_2 - P_{21}P_{11}^{-1}v_1 = \tau_{22}(\bar{X}_{2,1} - \bar{X}_{2,2}) + P_{21}P_{11}^{-1}(-\tau_{22}\bar{X}_{1,1} + \tau_{22}\bar{X}_{1,2})
\]

\[
= \tau_{22}[\bar{X}_{2,1} - \bar{X}_{2,2} - P_{21}P_{11}^{-1}(\bar{X}_{1,1} - \bar{X}_{1,2})]
\]

\[
= \tau_{22} \begin{pmatrix} -P_{21}P_{11}^{-1}, & I_{p_2} \end{pmatrix} \begin{pmatrix} \bar{X}_{1,1} - \bar{X}_{1,2} \\ \bar{X}_{2,1} - \bar{X}_{2,2} \end{pmatrix}.
\]

Inserting this into (4.3.37) yields

\[
v'P^{-1}v = v'_1P_{11}^{-1}v_1 + \frac{N_2}{n_1n_2} \tau_{22} \text{tr} \left[ P_{22,1}^{-1}\left(-P_{21}P_{11}^{-1}, I_{p_2}\right)D\left(-P_{21}P_{11}^{-1}, I_{p_2}\right)' \right]
\]

\[
= v'_1P_{11}^{-1}v_1 + \frac{\tau_{22}}{n_1} \text{tr} \left[ P_{22,1}^{-1}\left(-P_{21}P_{11}^{-1}, I_{p_2}\right)D\left(-P_{21}P_{11}^{-1}, I_{p_2}\right)' \right].
\]

(4.3.38)

At this point, in order to derive a stochastic representation for \( \hat{\mu}_3 \), we would need to derive the joint distribution of \( L, P, \) and \( v \). Even with Anderson’s identity, this distribution remains beyond our reach; hence, we have not been able to derive the desired stochastic representation. Although, we were unable to find the joint distribution of \( \hat{\mu}_1, \hat{\mu}_2, \) and \( \hat{\mu}_3 \), we hope that our new methods will eventually lead to the exact finite-sample distribution of \( \hat{\mu} \) in the three-step setting and give us insight into the general case of \( k \)-step monotone incomplete data.

4.4 The covariance matrix of \( \hat{\mu}_3 \)

Although we were unable to derive the distribution of \( \hat{\mu}_3 \), we have derived the covariance matrix of \( \hat{\mu}_3 \). By Lemma 4.2.1 and Theorems 4.2.2 and 4.2.6, the covariance
matrix of \( \hat{\mu}_3 \) yields the entire covariance matrix of \( \hat{\mu} \) under the assumption that \( \Sigma = I \).

Kanda and Fujikoshi \cite{17} have also studied the covariance matrix of \( \hat{\mu} \) in the three-step monotone incomplete setting, however our final result differs from theirs. Kanda and Fujikoshi stated (without proof) that \( B_{21} \) and \((B_{31}, B_{32})\) are independent, which we believe is not valid.

**Theorem 4.4.1.** When \( \Sigma = I_d \), the covariance matrix of \( \hat{\mu}_3 \) is

\[
\left( \frac{1}{n_1} + \frac{p_1(n_2 + n_3)}{n_1N_3(N_2 - p_1 - p_2 - 2)} \right) + \frac{n_2p_2}{n_1N_2(N_2 - p_1 - p_2 - 2)} + \frac{N_3(p_2 \cdot \text{Etr} A_{2,11}^{-3})}{N_2n_3(N_2 - p_1 - p_2 - 2)} \right) I_{p_3}
\]

(4.4.1)

**Proof.** Because we assume \( \mu = 0 \) and \( \Sigma = I_d \), the vectors \( X_{l,j} \), \( 1 \leq j \leq N_l, 1 \leq l \leq 3 \), are mutually independent with \( X_{l,j} \sim \mathcal{N}(0, I_{p_l}) \). Because the calculation is lengthy, we present the computation in separate parts. As in Theorem 4.3.3, let \( v' = (v'_1, v'_2) = (v_1', (v_{21} + \tau_{33} B_{21} v_{22})') \), where \( v_1 = \bar{\tau}_{13} \bar{X}_{1,1} - \tau_{23} \bar{X}_{1,2} - \tau_{33} \bar{X}_{1,3}, v_{21} = \tau_{22} (\bar{X}_{2,1} - \bar{X}_{2,2}) \), and \( v_{22} = \tau_{12} \bar{X}_{1,1} + \tau_{22} \bar{X}_{1,2} - \bar{X}_{1,3} \). By the formula for \( \hat{\mu}_3 \), (4.1.13), \( E\hat{\mu}_3 \hat{\mu}_3' \) equals

\[
E[\bar{X}_{3,1} - B_{31} v_1 - B_{32}(v_{21} - \tau_{33} B_{21} v_{22})](\bar{X}_{3,1} - B_{31} v_1 - B_{32}(v_{21} - \tau_{33} B_{21} v_{22}))'
\]

(4.4.2)

We calculate the right-hand side, term-by-term, as follows. First,

\[
E\hat{\mu}_3 \bar{X}_{3,1}' = E[\bar{X}_{3,1} - B_{31} v_1 - B_{32}(v_{21} - \tau_{33} B_{21} v_{22})]\bar{X}_{3,1}'
\]

\[
= \frac{1}{n_1} I_{p_3} - E[B_{31} v_1 - B_{32}(v_{21} - \tau_{33} B_{21} v_{22})] \bar{X}_{3,1}'.
\]
Because $\Sigma = I_d$ and because the sample mean is independent of the sample covariance matrix, $\bar{X}_{3,1}$ is independent of $\{B_{31}, v_1, B_{32}, v_{21}, B_{21}, v_{22}\}$. Because $E\bar{X}_{3,1} = 0$, it follows that every cross-term involving $\bar{X}_{3,1}$ is 0. The only nonzero term involving $\bar{X}_{3,1}$ is $E\bar{X}_{3,1}^{'} = I_{p3}/n_1$. The remaining terms in (4.4.2) are $EB_{31}v_1v_{12}B_{21} + EB_{32}v_2v_{21}B_{13} + EB_{32}v_2v_{22}B_{23}$. By the independence of the sample mean and sample covariance matrix and our assumption that $\Sigma = I_d$, the cross-term becomes

$$EB_{31}v_1v_{12}B_{23} = EB_{31}v_1v_{12}B_{23} + \tau_{33}EB_{31}v_1v_{22}B_{23}$$

$$= E_{B_{31}, B_{32}}B_{31}[E_{v_1, v_{21}}v_{21}]B_{23} + \tau_{33}EB_{31}v_1v_{22}B_{23}$$

$$= 0 + \tau_{33}EB_{31}v_1v_{22}B_{23} = \tau_{33}EB_{31}v_1v_{22}B_{23}.$$

If we replace each $X_{1,j}$ by $-X_{1,j}$, $j = 1, \ldots, N_3$, and leave $X_{2,j}$ and $X_{3,j}$ unchanged, then $B_{23}$ does not change, whereas $B_{31}$ changes to $-B_{31}$, $v_1$ changes to $-v_1$, and $v_{22}$ changes to $-v_{22}$. However, $-X_{1,j}$ has the same distribution as $X_{1,j}$; hence, $EB_{31}v_1v_{22}B_{23}$

$$\leq -B_{31}v_1v_{22}B_{23}.$$

Therefore, $EB_{31}v_1v_{22}B_{23} = -EB_{31}v_1v_{22}B_{23}$, and consequently $EB_{31}v_1v_{22}B_{23} = 0$. As a result, $EB_{31}v_1v_{22}B_{23} = 0$, and (4.4.2) becomes

$$E\hat{\mu}_3\hat{\mu}_3' = \frac{1}{n_1}I_{p3} + EB_{31}v_1v_{12}B_{13} + EB_{32}v_2v_{22}B_{23}$$

(4.4.3)

Note that $v_1 \sim N(0, (\tau_{13}^2/n_1 + \tau_{23}^2/n_2 + \tau_{33}^2/n_3)I_{p1})$, that is, $v_1 \sim N(0, (n_2 + n_3)/n_1 N_3 I_{p1})$.

By the independence of the sample mean and sample covariance matrix, $v_1$ and $B_{31}$ are
independent. Therefore

\[
EB_{31}v_1v'_1B_{13} = E[B_{31}E[v_1v'_1|B_{13}]] \\
= \frac{n_2 + n_2}{n_1N_3} EB_{31}B_{13},
\]

(4.4.4)

Because

\[
JA_3J' \equiv \begin{pmatrix} A_{3,11} & A_{3,12} \\ A_{3,21} & A_{3,22} \end{pmatrix},
\]

then by Proposition 2.3.3 (ii),

\[
(A_{3,31}, A_{3,32})|JA_3J' \sim N(0, I_{p_3} \otimes JA_3J').
\]

Therefore, the conditional distribution of \((B_{31}, B_{32}) = (A_{3,31}, A_{3,32})(JA_3J')^{-1}\) given \(JA_3J'\) is

\[
(B_{31}, B_{32})|JA_3J' \sim N(0, I_{p_3} \otimes (JA_3J')^{-1}).
\]

Noting that

\[
B_{31} = (B_{31}, B_{32}) \begin{pmatrix} I_{p_1} \\ 0_{p_2 \times p_1} \end{pmatrix},
\]

we obtain

\[
B_{31}|JA_3J' \sim N\left(0, I_{p_3} \otimes (I_{p_1}, 0)(JA_3J')^{-1}\begin{pmatrix} I_{p_1} \\ 0 \end{pmatrix}\right).
\]
By Proposition 2.1.2 (i), the $p_1 \times p_1$ submatrix of $(JA_3J')^{-1}$ is $A_{3,11,2}^{-1}$; therefore

$$B_{31}JA_3J' \sim N(0, I_{p_3} \otimes A_{3,11,2}^{-1})$$

By the definition of the matrix normal distribution, (2.2.1), $\text{vec} (B_{13}) \mid JA_3J' \sim N(0, I_{p_3} \otimes A_{3,11,2}^{-1})$; therefore $E(\text{vec} (B_{13}) \mid JA_3J') = I_{p_3} \otimes A_{3,11,2}^{-1}$. Let $A = (A_{ij})$ and $W = (W_{ij})$ be symmetric $p_3 \times p_3$ matrices. Let $\partial \partial A = (1/2(1+\delta_{ij})\partial \partial A_{ij})$ denote the matrix of partial derivatives with respect to the entries of $A$. Then

$$\frac{\partial}{\partial A} E \text{tr} AW = \frac{\partial}{\partial A} E \left[ \sum_i \sum_j A_{ij}W_{ij} \right] = \frac{\partial}{\partial A} \sum_i \sum_j A_{ij} EW_{ij} = EW. \quad (4.4.5)$$

Apply this relation with $W = B_{31}B_{13}$ to obtain

$$E(B_{31}B_{13} \mid JA_3J') = \frac{\partial}{\partial A} E[\text{tr} AB_{31}B_{13} \mid JA_3J']. \quad (4.4.6)$$

By Proposition 2.1.1 (iii),

$$\text{tr} AB_{31}B_{13} = \text{tr} \text{vec} (B_{13})' (A \otimes I) \text{vec} (B_{13}) = \text{tr} (A \otimes I) \text{vec} (B_{13}) \text{vec} (B_{13})'.$$

By interchanging trace and expected value, it follows that

$$EB_{31}B_{13} \mid JA_3J' = \frac{\partial}{\partial A} \text{tr} (A \otimes I)(I \otimes A_{3,11,2}^{-1})$$

$$= \frac{\partial}{\partial A} \text{tr} (A \otimes A_{3,11,2}^{-1})$$

$$= [\text{tr} A_{3,11,2}^{-1}] I_{p_3}. $$
By Proposition 2.3.3 (i), $A_{3,11:2} \sim W_{p_1}(N_2 - p_2 - 1, I_{p_1})$, and consequently,

$$EA_{3,11:2}^{-1} = \frac{1}{N_2 - p_1 - p_2 - 2} I_{p_1}.$$ 

Therefore,

$$EB_{31}B_{13} = E[\text{tr } A_{3,11:2}^{-1}] I_{p_3} = \text{tr} E[A_{3,11:2}^{-1}] I_{p_3} = \frac{p_1}{N_2 - p_1 - p_2 - 2} I_{p_3}.$$  

Substituting this result into (4.4.4) yields

$$EB_{31} v_1 v_1' B_{13} = \frac{p_1(n_2 + n_3)}{n_1 N_3 (N_2 - p_1 - p_2 - 2)} I_{p_3}. \quad (4.4.7)$$

It now remains to calculate

$$EB_{32} v_2 v_2' B_{23} = EB_{32} v_2 v_2' B_{23} + EB_{32} v_2 B_{12} v_2' B_{23} + EB_{32} v_2 B_{21} v_2' B_{12} B_{23}.$$  

By the independence of the sample mean and sample covariance matrix, it follows that $v_2$ and $\{B_{32}, B_{21}\}$ are independent. By the assumption that $\Sigma = I$, it follows that $v_2$ and $\{B_{32}, B_{21}, v_2\}$ are independent. Because $Ev_{22} = 0$, each of the cross-terms are 0. Note that $\text{Cov}(v_2) = \tau^2 (n_1^{-1} + n_2^{-1}) I_{p_2} = n_2 (n_1 N_2^{-1}) I_{p_2}$ and therefore $v_{21} \sim N(0, n_2(n_1 N_2^{-1}) I_{p_2})$. Also by the independence of the sample mean and sample
covariance matrix, $v_{21}$ and $B_{32}$ are independent and therefore

$$EB_{32}v_{21}v'_{21}B_{23} = EB_{32}E[v_{21}v'_{21}]B_{23} = \frac{n_2}{n_1 n_2} EB_{32}B_{23}.$$  \hfill (4.4.9)

Noting that

$$B_{32} = (B_{31}, B_{32}) \begin{pmatrix} 0_{p_1 \times p_2} \\ I_{p_2} \end{pmatrix},$$

it follows by the calculation of $EB_{31}B_{13}$, \hfill (4.4.5), and the expectation of the inverse of a Wishart matrix \hfill [25] p. 97, that

$$EB_{32}B_{23} = E[\text{tr} A_{3,22,1}^{-1}] I_{p_3} = \text{tr} E[A_{3,22,1}^{-1}] I_{p_3} = \frac{p_2}{N_2 - p_1 - p_2 - 2} I_{p_3}.$$  

Inserting this result into \hfill (4.4.9) yields

$$EB_{32}v_{21}v'_{21}B_{23} = \frac{n_2 p_2}{n_1 N_2(N_2 - p_1 - p_2 - 2)} I_{p_3}.$$  \hfill (4.4.10)

It remains only to calculate $EB_{32}B_{21}v_{22}v'_{22}B_{12}B_{23}$. Note that Cov($v_{22}$) = \left(\tau_{22}^{-1} + n_2^{-1} + \tau_{22}^{-1} + n_3^{-1}\right) I_{p_1} = N_3(N_2 n_3)^{-1} I_{p_1}$, and therefore $v_{22} \sim N(0, N_3(N_2 n_3)^{-1} I_{p_1})$. Because $v_{22}$ and $\{B_{32}, B_{21}\}$ are independent, it follows that

$$EB_{32}B_{21}v_{22}v'_{22}B_{12}B_{23} = EB_{32}B_{21}E[v_{22}v'_{22}]B_{12}B_{23} = \frac{N_3}{N_2 n_3} EB_{32}B_{21}B_{12}B_{23}.$$  \hfill (4.4.11)
Given \( \{JA_3J', C, D\} \), in equations (4.2.1), (4.2.2), (4.2.3), \( B_{21} \) is a constant and \( B_{32} \) is independent of \( \{C, D\} \). Recall that \( B_{32}|JA_3J' \sim N(0, I_{p_3} \otimes A_{3,22,1}^{-1}) \). Let \( A \) be a symmetric \( p_3 \times p_3 \) matrix. By Proposition 2.1.1 (iii),

\[
\text{tr } AB_{32}B_{21}B_{12}B_{23} = \text{tr } \text{vec}(B_{23})'(A \otimes B_{21}B_{12})\text{vec}(B_{23}) \\
= \text{tr } (A \otimes B_{21}B_{12})\text{vec}(B_{23})\text{vec}(B_{23})'.
\]

By interchanging trace and expected value, it follows that

\[
EB_{32}B_{21}B_{12}B_{23}|JA_3J' = \frac{\partial}{\partial A}\text{tr}(A \otimes B_{21}B_{12})(I \otimes A_{3,22,1}^{-1}) \\
= \frac{\partial}{\partial A}\text{tr}(A \otimes B_{21}B_{12}A_{3,22,1}^{-1}) \\
= [\text{tr } B_{21}B_{12}A_{3,22,1}^{-1}]I_{p_3}.
\]

Now,

\[
B_{21}B_{12}A_{3,22,1}^{-1} = A_{2,21}A_{2,11}^{-1}A_{2,11}^{-1}A_{2,12}A_{3,22,1}^{-1} \\
= (A_{3,21} + C_{21} + D_{21})(A_{3,11} + C_{11} + D_{11})^{-2}(A_{3,12} + C_{12} + D_{12})A_{3,22,1}^{-1}.
\]

By Proposition 2.3.3 (i), \( A_{3,22,1} \) is independent of \( \{A_{3,11}, A_{3,21}\} \), and therefore \( B_{21} \) is independent of \( A_{3,22,1} \). Also, by Proposition 2.3.3 (i), \( A_{3,22,1} \sim W_{p_2}(N_2 - p_1 - 1, I_{p_2}) \).

By Muirhead [25], p. 97,

\[
EA_{3,22,1}^{-1} = \frac{1}{N_2 - p_1 - p_2 - 2}I_{p_2}.
\]
Therefore,

\[ EB_{32} B_{21} B_{12} B_{23} | B_{21} = E[\text{tr} B_{21} B_{12} A_{321}^{-1}] I_{p_3} \]

\[ = \text{tr} B_{21} B_{12} E[A_{321}^{-1}] I_{p_3} = \frac{\text{tr} B_{21} B_{12}}{N_2 - p_1 - p_2 - 2} I_{p_3}. \]

By Proposition 2.3.3 (ii), \( A_{2,21} | A_{2,11} \sim N(0, I_{p_2} \otimes A_{2,11}) \), therefore

\[ B_{21} | A_{2,11} = A_{2,21} A_{2,11}^{-1} | A_{2,11} \sim N(0, I_{p_2} \otimes A_{2,11}^{-1}). \]

Let \( A \) be a symmetric \( p_2 \times p_2 \) matrix. By Proposition 2.1.1 (iii),

\[ \text{tr} A A_{2,21} A_{2,11}^{-2} A_{2,12} = \text{tr} \text{vec} (A_{2,12}) (A \otimes A_{2,11}^{-2}) \text{vec} (A_{2,12}) \]

\[ = \text{tr} (A \otimes A_{2,11}^{-2}) \text{vec} (A_{2,12}) \text{vec} (A_{2,12})'. \]

By interchanging trace and expected value, it follows that

\[ EA_{2,21} A_{2,11}^{-2} A_{2,12} | A_{2,11} = \frac{\partial}{\partial A} \text{tr} (A \otimes A_{2,11}^{-2}) (I \otimes A_{2,11}^{-1}) \]

\[ = \frac{\partial}{\partial A} \text{tr} (A \otimes A_{2,11}^{-3}) \]

\[ = [\text{tr} A_{2,11}^{-3}] I_{p_2}. \]

Now, \( A_{2,11} \sim W_{p_1} (N_1 - 1, I_{p_1}) \), however \( E \text{tr} A_{2,11}^{-3} \) is quite complicated, so we will leave the covariance matrix in terms of this expected value. Therefore \( E \text{tr} B_{21} B_{12} = \)
\[ p_2 \cdot Etr A_{2,11}^{-3} \] and \((4.4.11)\) becomes

\[ EB_{32}B_{21}v_2 v'_2 B_{12}B_{23} = \frac{N_3(p_2 \cdot Etr A_{2,11}^{-3})}{N_2n_3(N_2 - p_1 - p_2 - 2)} I_{p_3}. \]

and \((4.4.8)\) becomes

\[ EB_{32}v_2 v'_2 B_{23} = \left( \frac{n_2p_2}{n_1N_2(N_2 - p_1 - p_2 - 2)} + \frac{N_3(p_2 \cdot Etr A_{2,11}^{-3})}{N_2n_3(N_2 - p_1 - p_2 - 2)} \right) I_{p_3} \quad (4.4.12) \]

Substituting \((4.4.7)\) and \((4.4.12)\) into \((4.4.3)\) yields that \(E\hat{\mu}_3\hat{\mu}'_3\) equals

\[ \left( \frac{1}{n_1} + \frac{p_1(n_2 + n_3)}{n_1N_3(N_2 - p_1 - p_2 - 2)} + \frac{n_2p_2}{n_1N_2(N_2 - p_1 - p_2 - 2)} + \frac{N_3(p_2 \cdot Etr A_{2,11}^{-3})}{N_2n_3(N_2 - p_1 - p_2 - 2)} \right) I_{p_3}. \]

\[ (4.4.13) \]

**Remark 4.4.2.** (i) In our derivation of \(\text{Cov}(\hat{\mu}_3)\), we showed that \(B_{32} | J A_3 J' \sim N(0, I_{p_3} \otimes A_{3,22}^{-1})\). Therefore \(A_{3,22}^{-1}\) is sufficient. By Proposition 2.3.3 (i), \(A_{3,22}^{-1}\) is independent of \(\{A_{3,11}, A_{3,21}\}\), and therefore \(B_{32}\) is independent of \(B_{21}\). However, \(B_{31} | J A_3 J' \sim N(0, I_{p_3} \otimes A_{3,11,2}^{-1})\), and \(A_{3,11,2}\) is not independent of \(B_{21}\).

(ii) An exact value for \(Etr A_{2,11}^{-3}\) may be obtained by means of the theory of zonal polynomials (Muirhead [25], Chapter 7).
Chapter 5

Concluding Remarks

In summary, we have made a number of theoretical advances in the statistical analysis of monotone incomplete multivariate normal data. In the case of two-step, monotone incomplete data, we proved that $\hat{\mu}$ and $\hat{\Sigma}$ are equivariant and obtained a new derivation of Chang and Richards’ \cite{6} stochastic representation for $\hat{\mu}$. With our new derivation, we identified explicitly in terms of the data the independent random variables that arise in that stochastic representation. We also provided a stochastic representation for a generalization of Hotelling’s $T^2$-statistic and shown that it is invariant under certain transformations. We provided applications of this result, including simultaneous confidence intervals for linear combinations of $\mu$. With our stochastic representation, we evaluated the performance of previous approximations and bounds for the $T^2$-statistic. We proposed a pivotal quantity, $T^2_{z^t}$, to create prediction intervals, made progress on a stochastic representation for its exact distribution, and we plan to develop approximations in future work.

We then extended our derivation of a stochastic representation for $\hat{\mu}$ in the two-step setting to the three-step setting. We hope that our methods will eventually lead to the distribution of $\hat{\mu}$ in the three-step, and also in the general $k$-step, setting. Certainly, establishing the equivariance of $\hat{\mu}$ was a very useful result we relied upon. By Davison \cite{8}, the maximum likelihood estimator is always equivariant and therefore this result holds for
general $k$. Although we did not completely derive the desired stochastic representation, we proved that $\hat{\mu}_1$ is independent of $\{\hat{\mu}_2, \hat{\mu}_3\}$, and that, under certain assumptions, the subvectors of $\hat{\mu}$ are uncorrelated.

We conclude this dissertation with a number of open problems.

1. Exact stochastic representations for $\hat{\mu}$, $\hat{\Sigma}$, and the $T^2$-statistic in the general $k$-step case remain unknown. As open problems, we propose the derivation of stochastic representations for each of the aforementioned statistics when $k > 2$.

2. Richards and Yamada \cite{27} have derived recently a class of shrinkage results for $\hat{\mu}$. Bearing in mind that this dissertation, together with the results of Chang and Richards \cite{6, 7}, Yamada, et al \cite{32}, and \cite{27}, now provide a wealth of explicit formulas for the analysis of monotone incomplete data, it would be useful to extend to monotone incomplete multivariate normal data the empirical Bayes results of Efron and Morris \cite{11} and Lindley and Smith \cite{21}.

3. In the analysis of monotone incomplete data, the subject of imputation-based procedures are still largely unexplored. In this area, we offer as open problems the development of results, in a form as explicit as the results in this dissertation, imputation-based analyses that apply methods such as hot deck imputation, mean imputation, regression imputation, nearest neighbor imputation, and multiple imputation, all of which are treated in the classical case by Little and Rubin \cite{23}. These results have the potential for profound impact on statistical analyses of data arising from national panel surveys, social and behavioral research, toxicology research, and any area in which incomplete data commonly appear.
Bibliography


[27] D. St. P. Richards and T. Yamada. The stein phenomenon for monotone incomplete multivariate normal data. *submitted for publication*.


Vita
Megan M. Romer

Education

The Pennsylvania State University  State College, Pennsylvania  2006–Present
Ph.D. in Statistics, expected in August 2009
Area of Specialization: Monotone Incomplete Multivariate Normal Data

M.S. in Statistics

State University of New York at Oswego  Oswego, NY  1996–2000
B.S. in Mathematics and Applied Mathematical Economics, Statistics Minor,
summa cum laude with distinction

Research Experience

Doctoral Research  The Pennsylvania State University  2007–Present
Thesis Advisor: Prof. Donald St. P. Richards
I researched the statistical analysis of monotone incomplete multivariate normal data.
More specifically, I found the exact finite sample distribution of a generalization of
Hotelling’s $T^2$ statistic for two-step monotone incomplete data, provided applications,
and researched maximum likelihood estimators for two- and three-step data.

Graduate Research  The Pennsylvania State University  2001–2002
Research Advisor: Prof. Joseph L. Schafer
As work for my Master’s in statistics, I researched the area of nonresponse in cluster sampling.

Teaching Experience

Instructor  The Pennsylvania State University  2007–2008
I taught four sections of STAT 100, Statistical Concepts & Reasoning. The course
covers descriptive statistics, probability, the normal distribution, statistical inference,
simple linear regression, and contingency tables. I also taught STAT 200, Elementary
Statistics. The course covers descriptive statistics, frequency distributions, probability,
binomial and normal distributions, statistical inference, simple linear regression, and
analysis of variance.

I taught computer labs, held office hours, and graded for STAT 100, STAT 200, STAT
511, Regression Models, STAT 512, Design and Analysis of Experiments, STAT 415,

Work Experience

Statistical Analyst  Clinical Trials Coordination Center  2004–2006
I provided technical support on research projects including data analysis and interpre-
tation. I spent most of my time programming in SAS.

Senior Research Support Associate  Health Evaluation Sciences, HMC  2003-2004
I provided technical support on research projects including data analysis and interpre-
tation. I programmed in SAS and GIS. I also co-taught SAS programming labs in
introductory statistics classes.