ROBUST METHODS FOR MULTIVARIATE LINEAR MODELS WITH SPECTRAL MODELS FOR SCATTER MATRICES

by

Scott Kohler Hyde

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2004

APPROVAL

of a dissertation submitted by

Scott Kohler Hyde

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Robert J. Boik

Approved for the Department of Mathematical Sciences

Kenneth Bowers

Approved for the College of Graduate Studies

Bruce McLeod

STATEMENT OF PERMISSION TO USE

Dedicated to my wife Audrey,
and my children Leila, Naleo, Elena, and Emma (Meg).

ACKNOWLEDGEMENTS

There are many people whose influence has been felt in the writing of this dissertation, either indirectly or directly. First and foremost, I would like to acknowledge the support and devotion of my wife, Audrey. Without her, I would not have finished this dissertation. Second, I wish to thank my advisor, Robert Boik, for his enduring patience and encouragement. Through his help and advice I was able to progress on my dissertation during those times when I was discouraged. I am also grateful for all his difficult and satisfying classes.

I am grateful for the advice and constructive criticism of the members of my committee: John Borkowski, Steve Cherry, Jim Robison-Cox, and Thomas Blake.

I am grateful for the guidance of many educators, but can only mention a few: Marty Raebel, Scott Cooley, Bill Quimby, John Borkowski, Marty Hamilton, Jim Robison-Cox, and Robert Boik.

I would like to thank my family. To Mom, for wanting to show my master's paper to everyone who would listen. To Dad, for the wise counsel and encouragement you always provide me. To my brother and sister, for their love. To my children, Leila, Naleo, Elena, and Emma (Meg), who always wanted me to stay home and play (and I wanted to stay home and play!), I will always treasure your love for the simple things of life. Most of all, I thank my wife Audrey. I will always hold her in high regard for her belief that I could succeed. Her love, patience, advice, tolerance, endurance, affection, respect, devotion, and friendship will always be a source of strength to me.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The main goal of this dissertation is to present extensions to the robust estimation of multivariate location and scatter. These extensions include the estimation of structured scatter matrices embedded in a multivariate fixed effects model. Two different kinds of robust estimators are investigated. The first is based on Maronna's M-estimators of multivariate location and scatter (Maronna 1976). The second is a multivariate extension of the S-estimators introduced by Rousseeuw and Yohai (1984). In addition, asymptotic distributions of the estimators are derived from an estimating function approach. The sandwich estimator of the asymptotic variance is given, coupled with second order corrections to the bias of the estimators. Two different data sets are used to illustrate the techniques of the dissertation. A comparison of the new algorithms to Ruppert's SURREAL algorithm is made for each example. Models are fit to illustrate the flexibility of the new estimators. In addition, the sandwich estimate of the asymptotic variance is given for the examples. Simulations are performed evaluating the effectiveness of the estimators.

CHAPTER 1

INTRODUCTION

The main goal of this dissertation is to present extensions to the robust estimation of multivariate location and scatter. These extensions include the estimation of structured scatter matrices embedded in a multivariate fixed effects model. The scatter matrix is proportional to the covariance matrix. The proportionality constant depends on the underlying distribution. Two kinds of robust estimators are investigated. The first is based on Maronna's M-estimators of multivariate location and scatter (Maronna 1976). The second is a multivariate extension of the S-estimators introduced by Rousseeuw and Yohai (1984). In addition, asymptotic distributions of the estimators are given, coupled with second order corrections to the bias of the estimators.

The first chapter of the dissertation presents the linear model, describes some structures used for scatter matrices, reviews maximum likelihood theory, describes issues with maximum likelihood estimation, and provides a motivation for the robust estimation of multivariate regression parameters and the scatter matrix. Various robust estimators are introduced and their basic properties are described.

The second chapter introduces the parameterization of the multivariate regression model. The spectral model of Boik (2002a), along with the parameterization of the eigenvalue and eigenvector parameters is given.

Chapter 3 presents methods for estimating the parameters under both the M-estimating equations and the S-estimating equations. A modified Fisher-Scoring algorithm is presented for solving for the M-estimates using the likelihood from the multivariate-T distribution. First and second implicit derivatives that are necessary to solve for the S-estimators are given. A two-stage modified Newton-Raphson algorithm to solve for the S-estimates of the parameters is described.

In chapter 4, first order asymptotic distributions for estimators of the parameters are given. Second order expansions of the estimators, along with a correction for the bias of the estimators are described. The computation of the sandwich estimator of the asymptotic variance is described, as well as how to compute the estimated bias from the data. Previous asymptotic distributions (Lopuhaä 1989; Davies 1987) were derived from a influence function approach rather than an estimating function approach. This dissertation presents an estimating function approach.

In chapter 5, two data sets are used to illustrate the techniques of the dissertation. A comparison of the new algorithms to Ruppert's SURREAL algorithm is made for each example. Models are fit to illustrate the flexibility of the new estimators. In addition, the sandwich estimate of the asymptotic variance is given.

In chapter 6, simulations are given evaluating the effectiveness of the estimators.

Linear Model

Let $\boldsymbol{y}_k$ be a $p \times 1$ random vector from an unknown distribution with $\text{Var}(\boldsymbol{y}_k)$ proportional to $\boldsymbol{\Sigma}$. Generate a random sample of size n and arrange the $\boldsymbol{y}_k'$ vectors as rows of the $n \times p$ matrix

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1' \\ \boldsymbol{y}_2' \\ \vdots \\ \boldsymbol{y}_n' \end{bmatrix}.$$

Maronna (1976), Davies (1987), and Lopuhaä (1989) assume the underlying model for estimators of multivariate location and scatter is

$$\boldsymbol{Y} = \mathbf{1}_n \boldsymbol{\tau}' + \boldsymbol{E}, \tag{1.1}$$

where $\boldsymbol{\tau}$ is a $p \times 1$ vector of parameters, $\boldsymbol{E}$ is an $n \times p$ matrix of random errors, and $\boldsymbol{\tau}$ is described as the "location" with $\boldsymbol{\Sigma}$ as the "scatter" matrix. In this case, $E[\boldsymbol{y}_k] = \boldsymbol{\tau}$.

If covariates are available, it is natural to model the mean as $E[\boldsymbol{y}_k] = \boldsymbol{X}_k \boldsymbol{\beta}$. The specific structure of $\boldsymbol{X}_k$ is described below. The linear model to be employed in this dissertation is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{1.2}$$

where $\boldsymbol{X}$ is an $n \times d$ matrix of constants, $\boldsymbol{B}$ is a $d \times p$ matrix of parameters, and $\boldsymbol{E}$ is an $n \times p$ matrix of random errors. The focus of this dissertation is when the rows of $\boldsymbol{E}$ are independent and identically distributed from an elliptically contoured distribution. A definition of the family of elliptically contoured distributions is described in Chapter 3. Further,

$$E\left[\boldsymbol{Y}\right] = \boldsymbol{X}\boldsymbol{B}, \text{ and}$$

$$\operatorname{Disp}\left(\boldsymbol{Y}\right) \stackrel{\text{def}}{=} \operatorname{Var}\left(\operatorname{vec}\boldsymbol{Y}\right) = \boldsymbol{I}_n \otimes \alpha\boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is the "characteristic" matrix as defined by Chu (1973), α is a scalar constant that depends on the specific distribution, and vec is the operator which stacks the columns of a matrix into a column vector. For example, if $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_p\end{bmatrix}$, then

$$\operatorname{vec}\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_p\end{bmatrix}.$$

The robust estimators proposed in this dissertation are estimators of $\boldsymbol{B}$ and $\boldsymbol{\Sigma} = \kappa\boldsymbol{\Sigma}^*$, where $\boldsymbol{\Sigma}^*$ is the characteristic matrix of the actual underlying characteristic distribution, rather than the assumed underlying distribution. The scalar κ is a constant that depends on the actual distribution of $\boldsymbol{Y}$ as well as on the estimation method. As the constant κ is generally not know, the matrix $\boldsymbol{\Sigma}$ is known as the scatter matrix instead of either the covariance matrix or the characteristic matrix. If the underlying distribution is known, then $\kappa = 1$ and the scatter matrix and the characteristic matrix are the same. Further, when the actual underlying distribution is a multivariate normal distribution, then the characteristic matrix and the covariance matrix are the same (because $\alpha = 1$).

The rows of $\boldsymbol{X}$ are related to $\boldsymbol{X}_k = (\boldsymbol{I}_p \otimes \boldsymbol{x}_k')$ through the expectation of $\boldsymbol{y}_k$. Using the identity $\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A}) \operatorname{vec} \boldsymbol{B}$,

$$\begin{aligned} E[\boldsymbol{y}_k] &= E[\boldsymbol{Y}'\boldsymbol{e}_k] = (\boldsymbol{X}\boldsymbol{B})'\boldsymbol{e}_k \\ &= \boldsymbol{B}'\boldsymbol{X}'\boldsymbol{e}_k = \boldsymbol{B}'\boldsymbol{x}_k = \operatorname{vec}(\boldsymbol{B}'\boldsymbol{x}_k) \\ &= \operatorname{vec}(\boldsymbol{x}_k'\boldsymbol{B}) = \underbrace{(\boldsymbol{I}_p \otimes \boldsymbol{x}_k')}_{\boldsymbol{X}_k}\underbrace{\operatorname{vec}\boldsymbol{B}}_{\boldsymbol{\beta}} = \boldsymbol{X}_k\boldsymbol{\beta}, \end{aligned} \tag{1.3}$$

where $\boldsymbol{x}_k$ is the k^{th} column of $\boldsymbol{X}'$ and $\boldsymbol{e}_k$ is the k^{th} column of $\boldsymbol{I}_n$. Hence,

$$\boldsymbol{X}_k = (\boldsymbol{I}_p \otimes \boldsymbol{x}_k'), \tag{1.4}$$

$$\boldsymbol{\beta} = \operatorname{vec} \boldsymbol{B}, \text{ and} \tag{1.5}$$

$$\boldsymbol{B} = \operatorname{dvec}\big(\boldsymbol{\beta}, d, p\big),$$

where dvec is the inverse of the vec operator. In particular, for any matrix $\boldsymbol{A}$ with ab elements, $\operatorname{dvec}\big(\boldsymbol{A}, a, b\big)$ is the $a \times b$ matrix that satisfies $\operatorname{vec}(\boldsymbol{A}) = \operatorname{vec}\big\{\operatorname{dvec}\big(\boldsymbol{A}, a, b\big)\big\}$. Note that the linear model is equivalent to (1.1) when $\boldsymbol{X} = \boldsymbol{1}_n$. In this case, $\boldsymbol{X}_k = \boldsymbol{I}_p$, $\boldsymbol{B} = \boldsymbol{\tau}'$, and $\boldsymbol{\beta} = \boldsymbol{\tau}$.

In addition to estimating the parameter $\boldsymbol{\beta}$, estimation of the scatter matrix under a specified structure is sometimes desired. For example, one such structure on a 4×4 scatter matrix could be a variance components model (SAS Institute 1997),

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix},$$

or a Toeplitz structure:

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}.$$

The focus of this dissertation uses the principal components model for the scatter matrix.

The most common method for finding estimators of parameters is maximum likelihood estimation (Casella and Berger 1990, pg. 289). Using this method requires the knowledge of the distribution of the data. Suppose that $x_1, x_2, \cdots, x_n$ are a random sample from a population with the density function $f(x|\theta_1, \theta_2, \cdots, \theta_p)$. The likelihood function is defined as

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}).$$

The maximum likelihood estimator of $\boldsymbol{\theta}$ for a given set of data $\boldsymbol{x}$ is the value $\hat{\boldsymbol{\theta}}$ for which $\boldsymbol{L}(\boldsymbol{\theta}|\boldsymbol{x})$ attains its maximum. The maximum likelihood estimator has many optimal properties. First, the support of the maximum likelihood estimator is the same as the space of the parameters it is estimating. Hence, negative estimates of variances are impossible. The maximum likelihood estimator is consistent (Casella and Berger 1990, pg. 325), which means $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}$. For any function h, The maximum likelihood estimator is invariant in that if $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, then $h(\hat{\boldsymbol{\theta}})$ is the maximum likelihood estimator of $h(\boldsymbol{\theta})$ (Pawitan 2001, pg. 45). Another property it possesses is asymptotic efficiency, which means the asymptotic variance of the maximum likelihood estimator attains the Cramér-Rao Lower Bound as $n \to \infty$ (Casella and Berger 1990, pg. 325).

Although the maximum likelihood estimator possesses many nice properties, it is not robust to departures from model specification (Pawitan 2001, pg. 370). If the distribution is correctly specified, then the asymptotic distribution of the maximum likelihood estimator is

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\text{dist}} N(\mathbf{0}, \bar{\boldsymbol{I}}_{\theta,\infty}^{-1}), \tag{1.6}$$

where

$$\bar{\boldsymbol{I}}_{\theta,\infty} = \lim_{n\to\infty} \bar{\boldsymbol{I}}_{\theta,n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{I}_{\theta,k} \text{ and}$$

$\boldsymbol{I}_{\theta,k}$ is Fisher's information matrix for the k^{th} sample. If the distribution is wrongly specified, then the maximum likelihood estimator need not be asymptotically efficient (Pawitan 2001, pg. 370). Hence, the distribution specified in (1.6) is invalid.

Additionally, maximum likelihood estimators are not robust estimators. Small changes in the data can lead to large changes in the estimates (Casella and Berger 1990, pg. 297). A simple example of this property is the estimation of the average house sale prices. It is typical to use the median sale price as a measure of center instead of the mean sale price (the MLE under normality) because extreme values can influence the mean greatly. The influence of one house price can be sufficient to arbitrarily change the estimate of the mean house sale price. One point, however, is not sufficient to arbitrarily change the estimate of the median house price. The sample median is an example of a robust estimator. Robust estimators are used either when outliers are suspected in the data, or model specification departs from normality.

Several properties can be used to compare robust estimators. The breakdown point and the influence function were introduced by Hampel (1968) to measure the global and local sensitivity of an estimator. Donoho and Huber (1983) defined a finite sample of the breakdown point, which can be thought of as the minimum percentage of outliers in the data which can make the estimate arbitrarily large; the highest possible breakdown point is $\frac{1}{2}$. For example, the median has a breakdown point of $\frac{1}{2}$; e.g. half of the data points must be changed to drive the estimate to infinity. The influence function measures the effect on the estimator when adding one more data point. Those estimators that have bounded influence functions are typically considered robust. However, an estimator can have a bounded influence function, but a breakdown point that is not very high. Additionally, a robust estimator may be inefficient and/or biased for estimating the parameter of interest, especially when the data actually comes from a normal distribution. For example, suppose an estimator for the mean of a distribution for any data set always gives the estimate of 14. Adding a new data point to the data set does not change the estimate. Any percentage of outliers in the data will not arbitrarily change the estimate. Hence, this estimator has a high breakdown point and a bounded influence function. However, it is biased, and rarely gives useful information about the true parameter. Accordingly, an optimal robust estimator has good asymptotic efficiency under a Normal model and a reasonably high breakdown point.

Various types of robust estimators have been developed. One of the first was the Least Absolute Deviation (or L_1) estimator for regression. The LAD estimator of β was introduced by Edgeworth (1887) as a replacement for Least Sums of Squares and minimizes

$$LAD = \sum_{i=1}^{n} \left| y_i - \sum_{i=1}^{p} x_{ki}\beta_k \right| .$$

The LAD estimator has an asymptotic breakdown point of 0 since it does not protect against outlying x.

Huber's M-estimators of regression were introduced in 1964, and possess bounded influence functions. Maronna (1976) extended Huber's M-estimators to multivariate location and scatter. However, Maronna (1976) showed that the breakdown point of M-estimators is at most $\frac{1}{p+1}$, where p is the dimension of the data. Hence, in high dimension, M-estimators do not have good global robust properties. A description of M-estimators will be given in Chapter 3.

Other estimators were introduced in the 1980's and 1990's. In 1982, Siegel introduced the Repeated Median, which has an asymptotic breakdown point of $\frac{1}{2}$. However, the estimator is not affine equivariant, meaning that estimates change if the unit of measure of the data is changed.

Two high breakdown estimators used in the regression setting are the Least Trimmed Sum of Squares and the Least Median of Squares (Rousseeuw 1983, 1984).

The Least Trimmed Sum of Squares estimate is the minimizer of the sum of squares of the smallest 50% of the data. It has the same asymptotic efficiency as the the M-estimators and has a breakdown point of $\frac{1}{2}$. The Least Median of Squares estimate is obtained from minimizing the median of squared errors. It has a breakdown point of $\frac{1}{2}$. However, it converges to a non-normal asymptotic distribution (You 1999). Although the estimator is consistent, the estimator converges at a slow rate of $O_p(n^{-1/3})$ (You 1999). Because it converges at a rate slower than $O_p(n^{-1/2})$, it has an asymptotic relative efficiency of 0 when compared to the Gaussian model.

The S-estimator developed for the regression setting by Rousseeuw and Yohai (1984) represents a compromise between the Least Trimmed Sum of Squares and the Least Median of Squares. These estimators have high breakdown as well as high asymptotic efficiency when compared to the Normal model. A description of S-estimators will be given in Chapter 3.

Two of the first affine equivariant, high breakdown estimators were extensions of the Least Trimmed Sum of Squares and the Least Median of Squares estimators. The Minimum Volume Ellipsoid (MVE) estimator introduced by Rousseeuw (1985) is an extension of the Least Median of Squares estimator. The MVE location parameter estimator is the center of the smallest ellipsoid containing half of the data. The MVE scatter estimator is defined by the shape of that ellipsoid. Even though the MVE estimators have high breakdown points in high dimensions, they inherit the slow convergence rate of the Least Median of Squares (Davies 1992). Since it has a slow convergence rate, its asymptotic relative efficiency is 0 when compared to the Normal model estimators. Because other estimators exist that have good asymptotic properties along with high breakdown, the MVE estimators are not useful from a efficiency standpoint. The Minimum Scatter Determinant (MCD) estimator is an extension of the Least Trimmed Sum of Squares estimator (Rousseeuw 1985). Its computation requires finding the location vector and the covariance matrix which minimize the determinant of the covariance matrix over all sets of half of the data. It does not suffer from the slow convergence rate that the MVE does and has an asymptotic normal distribution. Both of these estimators have high breakdown in all dimensions, as compared to M-estimators.

Multivariate location and scatter versions of S-estimators were proposed by Davies (1987) and Lopuhaä (1989). These estimators have high breakdown point and relatively high asymptotic efficiency. With S-estimators, however, it is not possible to have high breakdown point and high efficiency at the same time. The smaller the breakdown point, the higher the asymptotic efficiency. Likewise, the larger the breakdown point, the smaller the asymptotic efficiency. A breakdown point of 50% is not realistic as the asymptotic relative efficiency under normality is 28.7%. A smaller breakdown point should be chosen which allows a relatively high breakdown point coupled with a good asymptotic efficiency. Table 1 lists the asymptotic relative efficiency under normality versus the asymptotic breakdown point. The Minimum

Volume Ellipsoid estimator can be shown to be a special case of a multivariate S-estimator. Under fairly general conditions, S-estimators, however, possess a $O_p(n^{-\frac{1}{2}})$ convergence rate and have an asymptotic normal distribution unlike the MVE estimator.

Asymptotic Breakdown Pt	Asymptotic Efficiency
50%	28.7%
45%	37.0%
40%	46.2%
35%	56.0%
30%	66.1%
25%	75.9%
20%	84.7%
15%	91.7%
10%	96.6%

Table 1. Asymptotic Relative Efficiency versus Breakdown Point for S-estimators.

Combinations of high-breakdown and high-efficiency estimators were proposed by Lopuhaä (1992). These estimators use a high breakdown estimator as the starting point of a highly efficient estimator. The combination of the two estimators retains the high breakdown characteristics of the starting point even if the highly efficient estimator does not have a high breakdown point.

Kent and Tyler (1996) proposed an extension of their Redescending M-estimators (Kent and Tyler 1991) coupled with S-estimators of location and scatter that have both a high breakdown point, as well as high asymptotic efficiency. Constrained M-estimators, or CM-estimators for short, use the objective function of the Redescending M-estimators with an inequality constraint like that of the S-estimators.

By choosing a tuning parameter appropriately, CM-estimators can have a high breakdown point and high efficiency. In addition, they have an asymptotic normal distribution and converge at a rate of $O_p(n^{-1/2})$.

Many of the robust estimators have an explicit objective function that is either solved or optimized. For example, the redescending M-estimators defined by Kent and Tyler (1991), have the objective function

$$L(\boldsymbol{\tau}, \boldsymbol{\Sigma}) = \sum_{k=1}^{n} \rho\left[(\boldsymbol{y}_k - \boldsymbol{\tau})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_k - \boldsymbol{\tau})\right] + \frac{1}{2} n \log |\boldsymbol{\Sigma}|,$$

where ρ is a function chosen by the investigator. The objective functions for the M and S-estimators of location and scatter are given in Chapter 3 of the dissertation. A list of objective functions of some the robust estimators is given in Table 2. Many of the objective functions are too complex to be given in a table. You (1999) and Yuan and Bentler (1998) contain the objective functions for many of the robust estimators. Others can be found in their defining papers.

A small list of estimators, along with their breakdown points is given in Table 3.

Estimator	Objective Function
M-Estimation (Maronna 1976)	See (3.2).
S-Estimation (Rousseeuw and Yohai 1984)	See (3.39).
GS-Estimation (Regression) (Croux et al. 1994)	Minimize s subject to $$\binom{n}{2}^{-1} \sum_{i<j} \rho\left(\frac{r_i - r_j}{s}\right) = k$$
Least Median of Squares (Regression) (Rousseeuw 1984)	$$\min_{\beta} \underset{i \leqslant n}{\text{median}} \left(y_i - \sum_{k=1}^{p} x_{ki}\beta_k\right)^2$$
Least Trimmed Sum of Squares (Regression) (Rousseeuw 1983, 1984)	$\min_{\beta} \sum_{i=1}^{J} u_{[i]}^2$, where $u_{[i]}^2$ are the ordered squared residuals and J is the largest integer less than or equal to $\frac{n}{2}+1$.
CM-Estimation (Kent and Tyler 1996)	$$L(\boldsymbol{\tau}, \boldsymbol{\Sigma}) = \sum_{k=1}^{n} \rho\left[(\boldsymbol{y}_k - \boldsymbol{\tau})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_k - \boldsymbol{\tau})\right] + \frac{1}{2} n \log \lvert\boldsymbol{\Sigma}\rvert$$ subject to $$\frac{1}{n}\sum_{k=1}^{n} \rho(dk) \leqslant \epsilon\rho(\infty),$$ where $0 < \epsilon < 1$.
Redescending M-Estimation (Kent and Tyler 1991)	$$L(\boldsymbol{\tau}, \boldsymbol{\Sigma}) = \sum_{k=1}^{n} \rho\left[(\boldsymbol{y}_k - \boldsymbol{\tau})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_k - \boldsymbol{\tau})\right] + \frac{1}{2} n \log \lvert\boldsymbol{\Sigma}\rvert$$

Table 2. Objective Functions for Robust Estimators.

Type	Asymptotic Breakdown pt	Other Properties
Normal Theory	0	
L_1-estimation (Edgeworth 1887)	0	Does not protect against outlying x.
M-Estimation (Huber 1964)	$\frac{1}{p+1}$	Has good asymptotic efficiency, but low breakdown point.
S-Estimation (Rousseeuw and Yohai 1984)	$\frac{1}{2}$	Has moderate asymptotic efficiency for high breakdown point, but high asymptotic efficiency for low breakdown point.
GS-Estimation (Croux et al. 1994)	$\frac{1}{2}$	Has high asymptotic efficiency in addition to high breakdown point.
Least Median of Squares (LMS) (Rousseeuw 1984)	$\frac{1}{2}$	Does not have good asymptotic properties (Converges at an slow rate).
Least Trimmed Sum of Squares (LTS) (Rousseeuw 1983, 1984)	$\frac{1}{2}$	Converges faster than LMS.
One-Step GM-Estimation (Coakley and Hettmansperger 1993)	$\frac{1}{2}$	By using a first step that has high breakdown point, this method maintains high breakdown point, but also retains good asymptotic efficiency.
CM-Estimation (Kent and Tyler 1996)	$\frac{1}{2}$	Has high breakdown point and high asymptotic efficiency.
Minimum Volume Ellipsoid (MVE) (Rousseeuw 1985)	$\frac{1}{2}$	Generalization of LMS for Multivariate Location and Scatter.
Minimum Covariance Determinant (MCD) (Rousseeuw 1985)	$\frac{1}{2}$	Generalization of LTS for Multivariate Location and Scatter.
Repeated Median (Siegel 1982)	$\frac{1}{2}$	Has high breakdown point, but is not affine equivariant (estimates change if changing units of measure).
MM-Estimation (Yohai 1987)	$\frac{1}{2}$	High efficiency under gaussian errors.

Table 3. Methods of Robust Estimation.

CHAPTER 2

MODEL AND PARAMETERIZATION

In this chapter, parameterizations of eigenvalues and eigenvectors of a scatter matrix are given. The scatter matrix is embedded within a multivariate fixed effects linear model. Together, the model for location and spectral parameters provide a framework within which inferences on Principal Components as well as on location parameters can be made.

Spectral Model

The parameterization for the scatter matrix was proposed by Boik (1998). A motivation for the parameterization follows. Modifications for the case of S-estimators is given in Chapter 3. Denote the eigenvalues of the $p \times p$ scatter matrix, $\mathbf{\Sigma}$, as $\lambda_1, \lambda_2, \ldots, \lambda_p$, some of which may be homogeneous. Further, denote the orthonormal eigenvectors of $\mathbf{\Sigma}$ by $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_p$. The spectral model of the scatter matrix is defined in terms of its spectral decomposition:

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}',$$

where $\mathbf{\Gamma} = \begin{pmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \cdots & \boldsymbol{\gamma}_p \end{pmatrix}$ and $\mathbf{\Lambda} = \text{Diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_p)$. One parameterization of the eigenvalues and eigenvectors could be

$$\mathbf{\Gamma} = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1p} \\ \theta_{21} & \theta_{22} & & \theta_{2p} \\ \vdots & & \ddots & \vdots \\ \theta_{p1} & \theta_{p2} & \cdots & \theta_{pp} \end{bmatrix} \text{ and } \mathbf{\Lambda} = \begin{bmatrix} \theta_{p^2+1} & 0 & \cdots & 0 \\ 0 & \theta_{p^2+2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \theta_{p^2+p} \end{bmatrix}.$$

However, this parameterization contains $p^2 + p$ parameters, which is larger than the number of entries in $\mathbf{\Sigma}$. The number of functionally independent parameters cannot be larger than the number of functionally independent entries in $\mathbf{\Sigma}$. Two sets of parameters are functionally dependent if one set of parameters is a function of the other set of parameters. If no such function exists, then the two sets of parameters are functionally independent. Because the number of functionally independent parameters is larger than the entries in $\mathbf{\Sigma}$, it follows that this model is over-parameterized. To obtain an identifiable parameterization, the number of parameters in $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ must be reduced. One option is through constraints.

For example, because $\mathbf{\Gamma}$ is an orthogonal matrix, then it satisfies the constraint $\mathbf{\Gamma}'\mathbf{\Gamma} = \mathbf{\Gamma}\mathbf{\Gamma}' = \boldsymbol{I}_p$. Hence, $\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_j = 0$ for all $i \neq j$ and $\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i = 1$ for each i. As a result, there are $p + \binom{p}{2}$ constraints on the p^2 entries in $\mathbf{\Gamma}$. This leaves at most $p^2 - p - \binom{p}{2} = \frac{p(p-1)}{2}$ parameters in $\mathbf{\Gamma}$. Organize these parameters as the entries in the upper half of the matrix $\mathbf{\Gamma}$; with the lower triangular part being occupied by the parameters which depend on the upper half. For example, in the case of $p = 4$, the matrices are

$$\boldsymbol{\Gamma} = \begin{bmatrix} \eta_1 & \mu_1 & \mu_2 & \mu_4 \\ \eta_2 & \eta_5 & \mu_3 & \mu_5 \\ \eta_3 & \eta_6 & \eta_8 & \mu_6 \\ \eta_4 & \eta_7 & \eta_9 & \eta_{10} \end{bmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}, \tag{2.1}$$

where $\{\mu_i\}$ is identified and $\{\eta_i\}$ is implicit.

Both the identified parameters and the implicit parameters can be organized into a vector. In this case,

$$\boldsymbol{\theta} = \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 & \mu_6 & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{pmatrix}' \text{ and}$$

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 & \eta_6 & \eta_7 & \eta_8 & \eta_9 & \eta_{10} \end{pmatrix}', \text{ where } \boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\mu}).$$

Without additional structure, $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\mu})$ need not exist. That is, given $\boldsymbol{\mu}$, there may be an infinite number of vectors $\boldsymbol{\eta}$ that satisfy $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_p$ or there may be no solution at all. A change in the parameterization of $\boldsymbol{\Gamma}$ must be made. To motivate this change, a simple example will be given.

Suppose $y_k \sim \text{iid } N(0, \sigma^2)$, for $k = 1, \ldots, n$. The likelihood function of $\boldsymbol{y}$ is

$$L(\sigma^2|\boldsymbol{y}) = \frac{1}{(\sigma^2)^{\frac{n}{2}}(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\frac{\boldsymbol{y}'\boldsymbol{y}}{\sigma^2}\right).$$

The mle is known to be $\hat{\sigma}^2 = \frac{\boldsymbol{y}'\boldsymbol{y}}{n}$. Define a new parameterization of σ^2, where $\sigma^2 = \sigma_0^2\mu$, where σ_0^2 is a known constant and μ is unknown. The log-likelihood becomes

$$\ell(\mu|\boldsymbol{y}) = -\frac{1}{2}\frac{\boldsymbol{y}'\boldsymbol{y}}{\sigma_0^2\mu} - \frac{n}{2}\log(\sigma_0^2\mu) - \frac{n}{2}\log(2\pi). \tag{2.2}$$

Take the derivative of (2.2) with respect to μ and set it equal to zero. The mle of μ can be computed as follows:

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu} &= \frac{1}{2}\frac{\boldsymbol{y}'\boldsymbol{y}}{\sigma_0^2\mu^2} - \frac{n\sigma_0^2}{2\sigma_0^2\mu} \overset{\text{set}}{=} 0 \\
\Longrightarrow \frac{\boldsymbol{y}'\boldsymbol{y}}{\sigma_0^2} &= n\mu \\
\Longrightarrow \hat{\mu} &= \frac{\boldsymbol{y}'\boldsymbol{y}}{n\sigma_0^2}.
\end{aligned}$$

Note that since $\sigma^2 = \sigma_0^2\mu$, then

$$\hat{\sigma}^2 = \sigma_0^2\hat{\mu} = \sigma_0^2\frac{\boldsymbol{y}'\boldsymbol{y}}{n\sigma_0^2} = \frac{\boldsymbol{y}'\boldsymbol{y}}{n}.$$

Accordingly, the mle of σ^2 is invariant with respect to the parameterization provided that $\sigma^2 > 0$.

In like manner, define

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_0\boldsymbol{G}(\boldsymbol{\mu}), \tag{2.3}$$

where $\boldsymbol{\Gamma}_0$ is treated as a known matrix, $\boldsymbol{\Gamma} \approx \boldsymbol{\Gamma}_0$; $\boldsymbol{G}(\boldsymbol{\mu})$ is in a neighborhood of the identity matrix; and $\boldsymbol{G}(\boldsymbol{\mu}) \in \mathcal{O}(p)$, where $\mathcal{O}(p)$ is the collection of all orthogonal matrices. Parameterize $\boldsymbol{G}$ in the same manner as $\boldsymbol{\Gamma}$ was parameterized in equation (2.1) and hold $\boldsymbol{\Gamma}_0$ fixed. Note that if $\boldsymbol{\mu} = \boldsymbol{0}$, then $\boldsymbol{G}'\boldsymbol{G} = \boldsymbol{I}_p$ implies that $\boldsymbol{G} = \boldsymbol{I}_p$. Hence, $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\mu})$ is well defined at $\boldsymbol{\mu} = \boldsymbol{0}$. It can be shown by the implicit function theorem (Fulks 1979, pg. 352) that $\boldsymbol{\eta}(\boldsymbol{\mu})$ is an implicit function of $\boldsymbol{\mu}$ in a neighborhood of $\boldsymbol{\mu} = \boldsymbol{0}$.

Constraints also can be placed on the eigenvalues to reduce the number of parameters. For example, if some eigenvalue multiplicities exceed one, then the number of eigenvalue and eigenvector parameters can be reduced. It is assumed that the eigenvalues are differentiable functions of eigenvalue parameters,

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\varphi}), \tag{2.4}$$

where $\boldsymbol{\varphi}$ is a vector of parameters.

Together with equations (2.3) and (2.4), the spectral model of $\boldsymbol{\Sigma}$ is specified as

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}) \boldsymbol{\Lambda}(\boldsymbol{\varphi}) \boldsymbol{G}(\boldsymbol{\mu})' \boldsymbol{\Gamma}_0', \tag{2.5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\varphi}$ are vectors of parameters.

Parameterization of the eigenvalues

Arrange the eigenvalues of the scatter matrix $\boldsymbol{\Sigma}$ in a $p \times 1$ vector $\boldsymbol{\lambda}$. The relationship between $\boldsymbol{\Lambda}$ and $\boldsymbol{\lambda}$ can be seen by using the Khatri-Rao column-wise product (Khatri and Rao 1968), which also will be used for derivatives in later chapters. Let $\boldsymbol{A}$ be an $n \times p$ matrix and $\boldsymbol{B}$ be an $m \times p$ matrix. Denote the columns of $\boldsymbol{A}$ and $\boldsymbol{B}$ as $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_p$, respectively. The Khatri-Rao column-wise product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is the $mn \times p$ matrix

$$\boldsymbol{A} * \boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_1 \otimes \boldsymbol{b}_1 & \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_p \otimes \boldsymbol{b}_p \end{bmatrix} = \sum_{k=1}^{p} (\boldsymbol{a}_k \otimes \boldsymbol{b}_k)\, \boldsymbol{e}_k', \tag{2.6}$$

where $\otimes$ is the Kronecker product, $\boldsymbol{e}_k$ is the k^{th} column of $\boldsymbol{I}_p$. Some properties of the product are given in Theorem 2.1.

Theorem 2.1 (Basic properties of the Khatri-Rao column-wise Product). *Suppose* $\boldsymbol{A}$ *is an* $n \times p$ *matrix,* $\boldsymbol{B}$ *is an* $m \times p$ *matrix,* $\boldsymbol{C}$ *is a* $q \times p$ *matrix,*

$\boldsymbol{D}$ is a $p \times p$ diagonal matrix, $\boldsymbol{E}$ is a $p \times s$ matrix, $\boldsymbol{F}$ is a $r \times n$ matrix, and $\boldsymbol{G}$ is a $t \times m$ matrix. Some useful properties of the Khatri-Rao product are

(a) $\boldsymbol{A} * (\boldsymbol{B} * \boldsymbol{C}) = (\boldsymbol{A} * \boldsymbol{B}) * \boldsymbol{C}$.

(b) $(\boldsymbol{A} * \boldsymbol{B}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \boldsymbol{L}_p$*, where* $\boldsymbol{L}_p \stackrel{\text{def}}{=} (\boldsymbol{I}_p * \boldsymbol{I}_p)$*. Note that* $\boldsymbol{L}_p$ *also can be written as* $\boldsymbol{L}_p = \sum_{k=1}^{p} (\boldsymbol{e}_k \otimes \boldsymbol{e}_k) \boldsymbol{e}_k'$ *and that* $\boldsymbol{L}_p$ *satisfies* $\boldsymbol{L}_p' \boldsymbol{L}_p = \boldsymbol{I}_p$.

(c) $(\boldsymbol{F} \otimes \boldsymbol{G}) (\boldsymbol{A} * \boldsymbol{B}) = (\boldsymbol{F}\boldsymbol{A} * \boldsymbol{G}\boldsymbol{B})$*, where* $\boldsymbol{F}\boldsymbol{A}$ *and* $\boldsymbol{G}\boldsymbol{B}$ *are conformable.*

(d) If $\boldsymbol{D}$ *is a* $p \times p$ *diagonal matrix, then* $\operatorname{vec}(\boldsymbol{C}\boldsymbol{D}\boldsymbol{E}) = (\boldsymbol{E}' \otimes \boldsymbol{C}) \boldsymbol{L}_p d(\boldsymbol{D}) = (\boldsymbol{E}' * \boldsymbol{C})\, d(\boldsymbol{D})$*, where* $d(\boldsymbol{D})$ *is a column vector whose components are the diagonal elements of* $\boldsymbol{D}$.

(e) $(\boldsymbol{A} * \boldsymbol{B}) = \boldsymbol{I}_{(m,n)} (\boldsymbol{B} * \boldsymbol{A})$*, where* $\boldsymbol{I}_{(m,n)}$ *is the commutation matrix (MacRae 1974).*

PROOF. Parts (a) and (c) are stated in Khatri and Rao (1968) without proof. Proofs are given below for completeness.

(a) By equation (2.6),

$$\begin{aligned}
\boldsymbol{A} * (\boldsymbol{B} * \boldsymbol{C}) &= \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_p \end{pmatrix} * \begin{pmatrix} \boldsymbol{b}_1 \otimes \boldsymbol{c}_1 & \boldsymbol{b}_2 \otimes \boldsymbol{c}_2 & \cdots & \boldsymbol{b}_p \otimes \boldsymbol{c}_p \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{a}_1 \otimes \boldsymbol{b}_1 \otimes \boldsymbol{c}_1 & \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 \otimes \boldsymbol{c}_2 & \cdots & \boldsymbol{a}_p \otimes \boldsymbol{b}_p \otimes \boldsymbol{c}_p \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{a}_1 \otimes \boldsymbol{b}_1 & \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_p \otimes \boldsymbol{b}_p \end{pmatrix} * \begin{pmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \cdots & \boldsymbol{c}_p \end{pmatrix} \\
&= (\boldsymbol{A} * \boldsymbol{B}) * \boldsymbol{C}.
\end{aligned}$$

(b) Using the definition of $\boldsymbol{L}_p$,

$$\begin{aligned}(\boldsymbol{A}\otimes\boldsymbol{B})\,\boldsymbol{L}_p &= (\boldsymbol{A}\otimes\boldsymbol{B})\,(\boldsymbol{I}_p * \boldsymbol{I}_p) = (\boldsymbol{A}\otimes\boldsymbol{B})\sum_{k=1}^{p}(\boldsymbol{e}_k\otimes\boldsymbol{e}_k)\,\boldsymbol{e}_k' \\ &= \sum_{k=1}^{p}(\boldsymbol{A}\boldsymbol{e}_k\otimes\boldsymbol{B}\boldsymbol{e}_k)\,\boldsymbol{e}_k' = \sum_{k=1}^{p}(\boldsymbol{a}_k\otimes\boldsymbol{b}_k)\,\boldsymbol{e}_k' \\ &= (\boldsymbol{A} * \boldsymbol{B})\,.\end{aligned}$$

(c) Using result (b),

$$\begin{aligned}(\boldsymbol{F}\otimes\boldsymbol{G})\,(\boldsymbol{A} * \boldsymbol{B}) &= (\boldsymbol{F}\otimes\boldsymbol{G})\,(\boldsymbol{A}\otimes\boldsymbol{B})\,\boldsymbol{L}_p = (\boldsymbol{F}\boldsymbol{A}\otimes\boldsymbol{G}\boldsymbol{B})\,\boldsymbol{L}_p \\ &= (\boldsymbol{F}\boldsymbol{A} * \boldsymbol{G}\boldsymbol{B})\,.\end{aligned}$$

(d) Note that

$$\begin{aligned}\operatorname{vec}\boldsymbol{D} &= \sum_{k=1}^{p}\operatorname{vec}(\boldsymbol{e}_k d_{kk}\boldsymbol{e}_k') = \sum_{k=1}^{p}(\boldsymbol{e}_k\otimes\boldsymbol{e}_k)\,d_{kk} \\ &= \sum_{k=1}^{p}(\boldsymbol{e}_k\otimes\boldsymbol{e}_k)\,\boldsymbol{e}_k' d(\boldsymbol{D}) = \boldsymbol{L}_p d(\boldsymbol{D}).\end{aligned}$$

It follows that

$$\operatorname{vec}(\boldsymbol{C}\boldsymbol{D}\boldsymbol{E}) = (\boldsymbol{E}'\otimes\boldsymbol{C})\operatorname{vec}\boldsymbol{D} = (\boldsymbol{E}'\otimes\boldsymbol{C})\,\boldsymbol{L}_p d(\boldsymbol{D}) = (\boldsymbol{E}' * \boldsymbol{C})\,d(\boldsymbol{D}).$$

(e) The commutation matrix, $\boldsymbol{I}_{(n,p)}$ is the $np\times np$ matrix that transforms the vec of the $n\times p$ matrix $\boldsymbol{A}$ to the vec of its transpose. In particular,

$$\operatorname{vec}(\boldsymbol{A}) = \boldsymbol{I}_{(n,p)}\operatorname{vec}(\boldsymbol{A}')\,. \tag{2.7}$$

It is apparent from the above definition that

$$\boldsymbol{I}_{(n,1)} = \boldsymbol{I}_{(1,n)} = \boldsymbol{I}_n,$$

$$\boldsymbol{I}_{(n,p)} = \boldsymbol{I}'_{(p,n)}, \text{ and}$$

$$\boldsymbol{I}_{(n,p)}\boldsymbol{I}_{(p,n)} = \boldsymbol{I}_{np}.$$

The commutation matrix can also be expressed as

$$\boldsymbol{I}_{(n,p)} = \sum_{k=1}^{n}\left(\boldsymbol{e}_k^{n'} \otimes \boldsymbol{I}_p \otimes \boldsymbol{e}_k^{n}\right) = \sum_{k=1}^{p}\left(\boldsymbol{e}_k^{p} \otimes \boldsymbol{I}_n \otimes \boldsymbol{e}_k^{p'}\right), \tag{2.8}$$

where $\boldsymbol{e}_k^n$ is the k^{th} column from $\boldsymbol{I}_n$ (Magnus and Neudecker 1979). The commutation matrix can be used to reverse the order of a Kronecker product (MacRae 1974). Suppose that $\boldsymbol{F}$ is an $r \times n$ matrix and that $\boldsymbol{G}$ is a $t \times m$ matrix. Then

$$(\boldsymbol{F} \otimes \boldsymbol{G}) = \boldsymbol{I}_{(t,r)}\left(\boldsymbol{G} \otimes \boldsymbol{F}\right)\boldsymbol{I}_{(n,m)} \tag{2.9}$$

The proof for (e) proceeds as follows. Because

$$\boldsymbol{I}_{(p,p)}\boldsymbol{L}_p = \sum_{k=1}^{p}\boldsymbol{I}_{(p,p)}\left(\boldsymbol{e}_k \otimes \boldsymbol{e}_k\right)\boldsymbol{e}'_k = \sum_{k=1}^{p}\left(\boldsymbol{e}_k \otimes \boldsymbol{e}_k\right)\boldsymbol{e}'_k = \boldsymbol{L}_p,$$

it follows that

$$\begin{aligned}(\boldsymbol{A} * \boldsymbol{B}) = (\boldsymbol{A} \otimes \boldsymbol{B})\,\boldsymbol{L}_p &= \boldsymbol{I}_{(m,n)}\left(\boldsymbol{B} \otimes \boldsymbol{A}\right)\boldsymbol{I}_{(p,p)}\boldsymbol{L}_p \\ &= \boldsymbol{I}_{(m,n)}\left(\boldsymbol{B} \otimes \boldsymbol{A}\right)\boldsymbol{L}_p = \boldsymbol{I}_{(m,n)}\left(\boldsymbol{B} * \boldsymbol{A}\right).\end{aligned}$$

□

An interesting property of commutation matrices is that they always commute. The result appears to be new, so it will be described. The proof can be found in Theorem 2.3. However, a lemma is first needed.

LEMMA 2.2 (BALESTRA (1976)). *Suppose that m, r, and s are positive integers. Then the following properties hold:*

(a) $\boldsymbol{I}_{(m,rs)}\boldsymbol{I}_{(r,ms)} = \boldsymbol{I}_{(r,ms)}\boldsymbol{I}_{(m,rs)} = \boldsymbol{I}_{(mr,s)}$,

(b) $\boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(rs,m)} = \boldsymbol{I}_{(rs,m)}\boldsymbol{I}_{(ms,r)} = \boldsymbol{I}_{(s,mr)}$, *and*

(c) $\boldsymbol{I}_{(s,mr)}\boldsymbol{I}_{(ms,r)} = \boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(s,mr)}$.

PROOF. First, the proof for (a) will be given. According to Balestra (1976),

$$\boldsymbol{I}_{(m,rs)} = \left(\boldsymbol{I}_r \otimes \boldsymbol{I}_{(m,s)}\right)\left(\boldsymbol{I}_{(m,r)} \otimes \boldsymbol{I}_s\right) = \left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(m,r)}\right)\left(\boldsymbol{I}_{(m,s)} \otimes \boldsymbol{I}_r\right). \tag{2.10}$$

Switching r and m in equation (2.10) yields

$$\boldsymbol{I}_{(r,ms)} = \left(\boldsymbol{I}_m \otimes \boldsymbol{I}_{(r,s)}\right)\left(\boldsymbol{I}_{(r,m)} \otimes \boldsymbol{I}_s\right) = \left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(r,m)}\right)\left(\boldsymbol{I}_{(r,s)} \otimes \boldsymbol{I}_m\right).$$

Premultiplying $\boldsymbol{I}_{(r,ms)}$ by $\boldsymbol{I}_{(m,rs)}$ yields

$$\boldsymbol{I}_{(m,rs)}\boldsymbol{I}_{(r,ms)} = \left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(m,r)}\right)\left(\boldsymbol{I}_{(m,s)} \otimes \boldsymbol{I}_r\right)\left(\boldsymbol{I}_m \otimes \boldsymbol{I}_{(r,s)}\right)\left(\boldsymbol{I}_{(r,m)} \otimes \boldsymbol{I}_s\right). \tag{2.11}$$

Note that (2.11) can be simplified using the properties $\boldsymbol{I}_{(mr,s)} = \left(\boldsymbol{I}_{(m,s)} \otimes \boldsymbol{I}_r\right)\left(\boldsymbol{I}_m \otimes \boldsymbol{I}_{(r,s)}\right)$ and $\left(\boldsymbol{I}_{(r,m)} \otimes \boldsymbol{I}_s\right) = \boldsymbol{I}_{(s,mr)}\left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(r,m)}\right)\boldsymbol{I}_{(mr,s)}$. Hence,

$$\begin{aligned}\boldsymbol{I}_{(m,rs)}\boldsymbol{I}_{(r,ms)} &= \left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(m,r)}\right)\boldsymbol{I}_{(mr,s)}\boldsymbol{I}_{(s,mr)}\left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(r,m)}\right)\boldsymbol{I}_{(mr,s)} \\ &= \left(\boldsymbol{I}_s \otimes \boldsymbol{I}_{(m,r)}\boldsymbol{I}_{(r,m)}\right)\boldsymbol{I}_{(mr,s)} \\ &= \boldsymbol{I}_{(mr,s)}. \end{aligned} \tag{2.12}$$

Because mr can also be written as rm, (2.12) can be written as

$$\boldsymbol{I}_{(m,rs)}\boldsymbol{I}_{(r,ms)} = \boldsymbol{I}_{(r,ms)}\boldsymbol{I}_{(m,rs)} = \boldsymbol{I}_{(mr,s)}. \tag{2.13}$$

Transposing all three sides of (2.13) gives the result for (b):

$$\boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(rs,m)} = \boldsymbol{I}_{(rs,m)}\boldsymbol{I}_{(ms,r)} = \boldsymbol{I}_{(s,mr)}.$$

Finally, the result in (c) can be obtained as follows:

$$\begin{aligned} & \boldsymbol{I}_{(rs,m)}\boldsymbol{I}_{(ms,r)} = \boldsymbol{I}_{(s,mr)} \\ \Longrightarrow\ & \boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(rs,m)}\boldsymbol{I}_{(ms,r)} = \boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(s,mr)} \\ \Longrightarrow\ & \boldsymbol{I}_{(s,mr)}\boldsymbol{I}_{(ms,r)} = \boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(s,mr)}. \end{aligned}$$

□

THEOREM 2.3. *Suppose that* $\boldsymbol{I}_{(a,b)}$ *and* $\boldsymbol{I}_{(c,d)}$ *are conformable for multiplication. Then*

$$\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} = \boldsymbol{I}_{(c,d)}\boldsymbol{I}_{(a,b)}.$$

PROOF. Because the matrix product exists, then $ab = cd \Longrightarrow a = \frac{cd}{b}$. In order for a to be an integer, b must divide evenly into the product cd. The number b is either prime or composite. First suppose b is a prime number. Then b divides c or b divides d.

First, suppose b divides c. Then there exists a nonzero integer f such that $c = fb$. The product $\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)}$ can be simplified as follows:

$$\begin{aligned}
\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} &= \boldsymbol{I}_{(\frac{cd}{b},b)}\boldsymbol{I}_{(c,d)} \\
&= \boldsymbol{I}_{(fd,b)}\boldsymbol{I}_{(fb,d)} \\
&= \boldsymbol{I}_{(f,db)} && \text{by Lemma 2.2, part (b)} \\
&= \boldsymbol{I}_{(fb,d)}\boldsymbol{I}_{(fd,b)} && \text{by Lemma 2.2, part (b)} \\
\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} &= \boldsymbol{I}_{(c,d)}\boldsymbol{I}_{(a,b)}.
\end{aligned}$$

Second, suppose b divides d. Then there exists a nonzero integer f such that $d = fb$. The product $\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)}$ can be simplified as follows:

$$\begin{aligned}
\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} &= \boldsymbol{I}_{(\frac{cd}{b},b)}\boldsymbol{I}_{(c,d)} \\
&= \boldsymbol{I}_{(fc,b)}\boldsymbol{I}_{(c,fb)} \\
&= \boldsymbol{I}_{(c,fb)}\boldsymbol{I}_{(fc,b)} && \text{by Lemma 2.2, part (c)} \\
\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} &= \boldsymbol{I}_{(c,d)}\boldsymbol{I}_{(a,b)}.
\end{aligned}$$

Suppose b is a composite number. Then b can be split up into two parts, b_1 and b_2, where b_1 divides c and b_2 divides d. Hence,

$$a = \frac{c}{b_1}\frac{d}{b_2} = fg, \text{ where } c = fb_1, \; d = gb_2, \text{ and } b_1b_2 = b.$$

The product $\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)}$ can be simplified as follows:

$$\begin{aligned}
\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} &= \boldsymbol{I}_{(fg,b_1b_2)}\boldsymbol{I}_{(fb_1,gb_2)} \\
&= \boldsymbol{I}_{(s,mr)}\boldsymbol{I}_{(fb_1,gb_2)}, \text{ where } m = b_1, r = b_2, s = fg \\
&= \boldsymbol{I}_{(ms,r)}\boldsymbol{I}_{(rs,m)}\boldsymbol{I}_{(fb_1,gb_2)}, \text{ by Lemma 2.2, part (b)}
\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{I}_{(fb_1g,b_2)}\boldsymbol{I}_{(fgb_2,b_1)}\boldsymbol{I}_{(fb_1,gb_2)} \\
&= \boldsymbol{I}_{(mr,s)}\boldsymbol{I}_{(fgb_2,b_1)}\boldsymbol{I}_{(fb_1,gb_2)}, \text{ where } m = fb_1, r = g, s = b_2. \\
&= \boldsymbol{I}_{(m,rs)}\boldsymbol{I}_{(r,ms)}\boldsymbol{I}_{(fgb_2,b_1)}\boldsymbol{I}_{(fb_1,gb_2)}, \text{ by Lemma 2.2, part (a)} \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(g,fb_1b_2)}\boldsymbol{I}_{(fgb_2,b_1)}\boldsymbol{I}_{(fb_1,gb_2)} \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(g,fb_1b_2)}\boldsymbol{I}_{(sm,r)}\boldsymbol{I}_{(sr,m)}, \text{ where } m = gb_2, r = b_1, s = f. \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(g,fb_1b_2)}\boldsymbol{I}_{(s,mr)}, \text{ by Lemma 2.2, part (b)} \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(g,fb_1b_2)}\boldsymbol{I}_{(f,gb_1b_2)} \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(m,rs)}\boldsymbol{I}_{(r,ms)}, \text{ where } m = g, r = f, s = b_1b_2. \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(mr,s)}, \text{ by Lemma 2.2, part (a)} \\
&= \boldsymbol{I}_{(fb_1,gb_2)}\boldsymbol{I}_{(fg,b_1b_2)} \\
\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} &= \boldsymbol{I}_{(c,d)}\boldsymbol{I}_{(a,b)}
\end{aligned}
$$

Therefore, if $ab = cd$, then $\boldsymbol{I}_{(a,b)}\boldsymbol{I}_{(c,d)} = \boldsymbol{I}_{(c,d)}\boldsymbol{I}_{(a,b)}$. □

The relationship between $\boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}$ is made clear by using Theorem 2.1. Using part (d) of the theorem,

$$
\begin{aligned}
\operatorname{vec}\boldsymbol{\Lambda} &= \boldsymbol{L}_p\boldsymbol{\lambda}, \\
\boldsymbol{\lambda} &= d(\boldsymbol{\Lambda}), \text{ and} \\
\boldsymbol{\lambda} &= \boldsymbol{L}_p' \operatorname{vec}\boldsymbol{\Lambda}.
\end{aligned}
$$

Several models may be employed in parameterizing the eigenvalues. Three models proposed by Boik (2002a) are the following:

$$\boldsymbol{\lambda} = \boldsymbol{T}_3\boldsymbol{\varphi}, \tag{2.14}$$

$$\boldsymbol{\lambda} = \boldsymbol{T}_1 \exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}, \text{ and} \tag{2.15}$$

$$\boldsymbol{\lambda} = \exp\left[\odot\boldsymbol{T}_1 \exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}\right]; \tag{2.16}$$

where $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, and $\boldsymbol{T}_3$ are eigenvalue design matrices with sizes $p\times q_1$, $q_1\times\nu_3$, and $p\times\nu_3$ respectively; and $\odot$ indicates that the operation is to be performed element-wise. For example, if $\boldsymbol{w} = \begin{pmatrix} w_1 & w_2 & \cdots & w_p \end{pmatrix}'$, then

$$\exp\{\odot\boldsymbol{w}\} = \begin{bmatrix} e^{w_1} & e^{w_2} & \cdots & e^{w_p} \end{bmatrix}'$$

and

$$\boldsymbol{w}^{\odot z} = \begin{bmatrix} w_1^z & w_2^z & \cdots & w_p^z \end{bmatrix}'.$$

Model (2.14) illustrates a basic linear relationship between the eigenvalues and the eigenvalue parameters. It can be used for modeling a simple repeated eigenvalue structure, or any linear structure of the eigenvalues. While each of the models has an appropriate use, only model (2.15) is used in this dissertation. The other models, however, can be employed easily because the derivatives in Chapter 3 and 4 are specified in terms of the derivatives of $\boldsymbol{\lambda}$.

Model (2.15) uses the design matrix $\boldsymbol{T}_1$ to model linear relationships between the eigenvalues and uses $\boldsymbol{T}_2$ to model exponential relationships between the eigenvalues.

Many different models can be employed by different choices for $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$. The following are examples of different models.

Arrange the multiplicities of the eigenvalues of $\boldsymbol{\lambda}$ in the vector $\boldsymbol{m}$. For example, if $\boldsymbol{\lambda} = \begin{pmatrix}7 & 7 & 7 & 4 & 3 & 3 & 2\end{pmatrix}'$, then $\boldsymbol{m} = \begin{pmatrix}3 & 1 & 2 & 1\end{pmatrix}'$. Often the number of distinct eigenvalues is the same as the size of the vector $\boldsymbol{\varphi}$ in (2.15). For other cases, however, the dimension of $\boldsymbol{\varphi}$ is smaller than the number of distinct eigenvalues.

Suppose that $p = 5$, all eigenvalues are distinct, and the eigenvalues are ordered, but otherwise unstructured. Then $\boldsymbol{m} = \mathbf{1}_5$, and $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be chosen as

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{T}_2 = \boldsymbol{I}_5$$

to ensure that the eigenvalues are positive and are ordered from largest to smallest.

Suppose that $p = 6$, $\boldsymbol{m} = \begin{pmatrix}2 & 1 & 3\end{pmatrix}'$ and the three distinct eigenvalues are ordered but otherwise unstructured. Then $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be chosen as

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{T}_2 = \boldsymbol{I}_3$$

to ensure that the eigenvalues are positive and ordered from largest to smallest.

Suppose that $p = 5$, $\boldsymbol{m} = \begin{pmatrix}2 & 1 & 2\end{pmatrix}'$, the first eigenvalue is the largest but the eigenvalues are otherwise unstructured. Then $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be chosen as

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \text{ and } \boldsymbol{T}_2 = \boldsymbol{I}_3$$

to force the first eigenvalue to be the largest.

Suppose that $p = 6$, $\boldsymbol{m} = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 \end{bmatrix}'$, the first three distinct eigenvalues are ordered but otherwise unstructured, and the last three are modeled with a linear relationship. Then $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be chosen as

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 1 & 1 & 3 & 1 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \text{ and } \boldsymbol{T}_2 = \boldsymbol{I}_4$$

to order the first three eigenvalues from largest to smallest and model the last three with a linear relationship.

Suppose $p = 6$, $\boldsymbol{m} = \boldsymbol{1}_6$, and the six distinct eigenvalues follow the model $\lambda_k = \exp\{\alpha + \beta(6 - k)\}, k = 1, 2, \ldots, 6$, where α and β are constants. Then $\boldsymbol{\lambda} = \boldsymbol{T}_1 \exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}$, where

$$\boldsymbol{T}_1 = \boldsymbol{I}_6, \quad \boldsymbol{T}_2 = \begin{bmatrix} 1 & 5 \\ 1 & 4 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \boldsymbol{\varphi} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Derivatives of model (2.15) are needed when estimating parameters and constructing confidence intervals. First and second derivatives are used to estimate parameters; and third-order derivatives are used to obtain the bias and skewness of the estimators. The derivatives are given in Theorem 2.4.

THEOREM 2.4 (BOIK (2002B)). *The first three derivatives of* $\boldsymbol{\lambda}$ *in* (2.15) *with respect to* $\boldsymbol{\varphi}$ *can be expressed as*

$$\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)} &\stackrel{\text{def}}{=} \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}'} = (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' * \boldsymbol{T}_1)\, \boldsymbol{T}_2, \\
\boldsymbol{D}_{\boldsymbol{\lambda}}^{(2)} &\stackrel{\text{def}}{=} \frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' * \boldsymbol{T}_1)\, (\boldsymbol{T}_2' * \boldsymbol{T}_2')', \\
\boldsymbol{D}_{\boldsymbol{\lambda}}^{(3)} &\stackrel{\text{def}}{=} \frac{\partial^3 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' * \boldsymbol{T}_1)\, (\boldsymbol{T}_2' * \boldsymbol{T}_2' * \boldsymbol{T}_2')',
\end{aligned}$$

where $*$ *indicates the Khatri-Rao column-wise product.*

PROOF. Let $\boldsymbol{e}_k^{q_1}$ be the k^{th} column of $\boldsymbol{I}_{q_1}$. By the definition of the $\odot$ symbol,

$$\boldsymbol{\lambda} = \boldsymbol{T}_1 \exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\} = \boldsymbol{T}_1 \sum_{k=1}^{q_1} \boldsymbol{e}_k^{q_1} \boldsymbol{e}_k^{q_1\prime} \exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\} = \boldsymbol{T}_1 \sum_{k=1}^{q_1} \boldsymbol{e}_k^{q_1} \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\}.$$

Accordingly, the first derivative of $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\varphi}'$ is

$$\begin{aligned}
\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}'} &= \boldsymbol{T}_1 \sum_{k=1}^{q_1} \boldsymbol{e}_k^{q_1} \frac{\partial \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\}}{\partial \boldsymbol{\varphi}'} \\
&= \boldsymbol{T}_1 \sum_{k=1}^{q_1} \boldsymbol{e}_k^{q_1} \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\} \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \qquad (2.17) \\
&= \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \boldsymbol{e}_k^{q_1\prime} \exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\} \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \\
&= \sum_{k=1}^{q_1} (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \boldsymbol{e}_k^{q_1} \otimes \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1})\, \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \\
&= (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \otimes \boldsymbol{T}_1) \sum_{k=1}^{q_1} (\boldsymbol{e}_k^{q_1} \otimes \boldsymbol{e}_k^{q_1})\, \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \\
&= (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \otimes \boldsymbol{T}_1)\, (\boldsymbol{I}_{q_1} * \boldsymbol{I}_{q_1})\, \boldsymbol{T}_2. \qquad (2.18)
\end{aligned}$$

By Theorem 2.1(d), equation (2.18) can be written as

$$\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}'} = (\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' * \boldsymbol{T}_1)\, \boldsymbol{T}_2.$$

Continuing from (2.17), the second derivative of $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\varphi}'$ and $\boldsymbol{\varphi}'$ is

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} &= \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \frac{\partial \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\}}{\partial \boldsymbol{\varphi}'} \left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&= \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\} \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \qquad (2.19) \\
&= \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \boldsymbol{e}_k^{q_1\prime} \exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\} \left(\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&= \sum_{k=1}^{q_1} \left(\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \boldsymbol{e}_k^{q_1} \otimes \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \right) \left(\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&= \left(\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \otimes \boldsymbol{T}_1 \right) \sum_{k=1}^{q_1} \left(\boldsymbol{e}_k^{q_1} \otimes \boldsymbol{e}_k^{q_1} \right) \left(\boldsymbol{e}_k^{q_1} \otimes \boldsymbol{e}_k^{q_1} \right)' \left(\boldsymbol{T}_2 \otimes \boldsymbol{T}_2 \right) \\
&= \left(\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \otimes \boldsymbol{T}_1 \right) \left(\boldsymbol{I}_{q_1} * \boldsymbol{I}_{q_1} \right) \left(\boldsymbol{I}_{q_1} * \boldsymbol{I}_{q_1} \right)' \left(\boldsymbol{T}_2 \otimes \boldsymbol{T}_2 \right). \qquad (2.20)
\end{aligned}$$

By Theorem 2.1(d), equation (2.20) can be written as

$$\frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \left(\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' * \boldsymbol{T}_1 \right) \left(\boldsymbol{T}_2' * \boldsymbol{T}_2' \right)'.$$

Differentiating (2.19) with respect to $\boldsymbol{\varphi}'$ yields

$$\begin{aligned}
&\frac{\partial^3 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \\
&\quad = \frac{\partial}{\partial \boldsymbol{\varphi}'} \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\} \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&\quad = \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \frac{\partial \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\}}{\partial \boldsymbol{\varphi}'} \left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&\quad = \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \exp\{\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \boldsymbol{\varphi}\} \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&\quad = \sum_{k=1}^{q_1} \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \boldsymbol{e}_k^{q_1\prime} \exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\} \left(\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right) \\
&\quad = \sum_{k=1}^{q_1} \left(\exp\{\odot \boldsymbol{T}_2 \boldsymbol{\varphi}\}' \boldsymbol{e}_k^{q_1} \otimes \boldsymbol{T}_1 \boldsymbol{e}_k^{q_1} \right) \left(\boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \otimes \boldsymbol{e}_k^{q_1\prime} \boldsymbol{T}_2 \right)
\end{aligned}$$

$$= \left(\exp\{\odot \boldsymbol{T}_2\boldsymbol{\varphi}\}' \otimes \boldsymbol{T}_1\right) \sum_{k=1}^{q_1} \left(\boldsymbol{e}_k^{q_1} \otimes \boldsymbol{e}_k^{q_1}\right) \left(\boldsymbol{e}_k^{q_1} \otimes \boldsymbol{e}_k^{q_1} \otimes \boldsymbol{e}_k^{q_1}\right)' \left(\boldsymbol{T}_2 \otimes \boldsymbol{T}_2 \otimes \boldsymbol{T}_2\right)$$

$$= \left(\exp\{\odot \boldsymbol{T}_2\boldsymbol{\varphi}\}' \otimes \boldsymbol{T}_1\right) \left(\boldsymbol{I}_{q_1} * \boldsymbol{I}_{q_1}\right) \left(\boldsymbol{I}_{q_1} * \boldsymbol{I}_{q_1} * \boldsymbol{I}_{q_1}\right)' \left(\boldsymbol{T}_2 \otimes \boldsymbol{T}_2 \otimes \boldsymbol{T}_2\right). \qquad (2.21)$$

By Theorem 2.1, parts (b) and (d) simplify (2.21) to

$$\frac{\partial^3 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \left(\exp\{\odot \boldsymbol{T}_2\boldsymbol{\varphi}\}' * \boldsymbol{T}_1\right) \left(\boldsymbol{T}_2' * \boldsymbol{T}_2' * \boldsymbol{T}_2'\right)'.$$

□

Parameterization of the eigenvectors

The form of the $\boldsymbol{G}$ matrix changes when some eigenvalue multiplicities exceed one. To illustrate, suppose that there are $h < p$ distinct eigenvalues of $\boldsymbol{\Sigma}$. Specify the multiplicities of the eigenvalues in the vector $\boldsymbol{m}$. Denote the distinct eigenvalues as $\lambda_1^*, \lambda_2^*, \ldots, \lambda_h^*$. Then $\boldsymbol{\Lambda}$ can be written as

$$\begin{aligned}\boldsymbol{\Lambda} &= \operatorname{Diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_m\right) \\ &= \operatorname{Diag}\Big(\underbrace{\lambda_1^*, \ldots, \lambda_1^*}_{m_1 \text{ of these}}, \underbrace{\lambda_2^*, \ldots, \lambda_2^*}_{m_2 \text{ of these}}, \ldots, \underbrace{\lambda_h^*, \ldots, \lambda_h^*}_{m_h \text{ of these}}\Big) \\ &= \begin{bmatrix} \boldsymbol{I}_{m_1}\lambda_1^* & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{m_2}\lambda_2^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{I}_{m_h}\lambda_h^* \end{bmatrix}.\end{aligned}$$

Partition $\boldsymbol{\Gamma}$ according to the multiplicities of the eigenvalues. It follows that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \underset{p\times m_1}{\boldsymbol{\Gamma}_1} & \underset{p\times m_2}{\boldsymbol{\Gamma}_2} & \cdots & \underset{p\times m_h}{\boldsymbol{\Gamma}_h} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_{m_1}\lambda_1^* & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{m_2}\lambda_2^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{I}_{m_h}\lambda_h^* \end{bmatrix} \begin{bmatrix} \underset{m_1\times p}{\boldsymbol{\Gamma}_1'} \\ \underset{m_2\times p}{\boldsymbol{\Gamma}_2'} \\ \vdots \\ \underset{m_h\times p}{\boldsymbol{\Gamma}_h'} \end{bmatrix}$$

$$= \sum_{k=1}^{h} \boldsymbol{\Gamma}_k \lambda_k^* \boldsymbol{I}_{m_k} \boldsymbol{\Gamma}_k' = \sum_{k=1}^{h} \lambda_k^* \boldsymbol{\Gamma}_k \boldsymbol{\Gamma}_k'.$$

Further, write $\boldsymbol{\Gamma}$ as $\boldsymbol{\Gamma}_0 \boldsymbol{G}$ and partition $\boldsymbol{G}$ conformably with $\boldsymbol{m}$:

$$\begin{aligned} \boldsymbol{\Gamma} &= \boldsymbol{\Gamma}_0 \begin{bmatrix} \underset{p \times m_1}{\boldsymbol{G}_1} & \underset{p \times m_2}{\boldsymbol{G}_2} & \cdots & \underset{p \times m_h}{\boldsymbol{G}_h} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Gamma}_0 \boldsymbol{G}_1 & \boldsymbol{\Gamma}_0 \boldsymbol{G}_2 & \cdots & \boldsymbol{\Gamma}_0 \boldsymbol{G}_h \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_h \end{bmatrix}. \end{aligned}$$

Accordingly, the spectral model (2.5) simplifies to

$$\boldsymbol{\Sigma} = \sum_{k=1}^{h} \lambda_k^* \boldsymbol{\Gamma}_k \boldsymbol{\Gamma}_k' = \sum_{k=1}^{h} \lambda_k^* \boldsymbol{\Gamma}_0 \boldsymbol{G}_k \boldsymbol{G}_k' \boldsymbol{\Gamma}_0' = \boldsymbol{\Gamma}_0 \left[\sum_{k=1}^{h} \lambda_k^* \boldsymbol{G}_k \boldsymbol{G}_k' \right] \boldsymbol{\Gamma}_0'.$$

Note that $\boldsymbol{G}_k \boldsymbol{G}_k' = \boldsymbol{G}_k \boldsymbol{Q}_k \boldsymbol{Q}_k' \boldsymbol{G}_k'$, where $\boldsymbol{Q}_k$ is any orthogonal matrix. The orthogonal matrix can be chosen to introduce zeros into the parameterization for $\boldsymbol{\mu}$ when an eigenvalue is repeated. In general, each partition $\boldsymbol{G}_k$, can be further partitioned into three parts,

$$\boldsymbol{G}_k = \begin{bmatrix} \underset{c \times m_k}{\boldsymbol{G}_{k1}} \\ \underset{m_k \times m_k}{\boldsymbol{G}_{k2}} \\ \underset{d \times m_k}{\boldsymbol{G}_{k3}} \end{bmatrix}, \text{ where } c = \sum_{j=1}^{k-1} m_j, \; d = p - \sum_{j=1}^{k} m_j.$$

When $m_k > 1$, a suitable $\boldsymbol{Q}_k$ can be found through the QR decomposition of $\boldsymbol{G}_{k2}' = \boldsymbol{Q}_k \boldsymbol{R}_k$, where $\boldsymbol{R}_k$ is a full rank upper triangular matrix. Consequently,

$$\boldsymbol{G}_k \boldsymbol{Q}_k = \begin{bmatrix} \boldsymbol{G}_{k1} \\ \boldsymbol{G}_{k2} \\ \boldsymbol{G}_{k3} \end{bmatrix} \boldsymbol{Q}_k = \begin{bmatrix} \boldsymbol{G}_{k1} \boldsymbol{Q}_k \\ \boldsymbol{G}_{k2} \boldsymbol{Q}_k \\ \boldsymbol{G}_{k3} \boldsymbol{Q}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_{k1} \boldsymbol{Q}_k \\ \boldsymbol{R}_k' \\ \boldsymbol{G}_{k3} \boldsymbol{Q}_k \end{bmatrix}.$$

The $m_k \times m_k$ upper triangular matrix $\boldsymbol{R}_k$ contains $\frac{m_k(m_k-1)}{2}$ zeros.

During the estimation process, the QR factorization is not explicitly used, but the parameters are solved for assuming the structure of the zeros in the matrices

$\boldsymbol{G}_k\boldsymbol{Q}_k$ and calling the matrix $\boldsymbol{G}$. To illustrate, suppose that the multiplicity vector is $\boldsymbol{m} = \begin{pmatrix}2 & 3 & 1 & 1\end{pmatrix}'$. It follows that the structure of the matrix $\boldsymbol{G}$ is

$$\boldsymbol{G} = \left[\begin{array}{cc|ccc|c|c} \eta_1 & 0 & \mu_1 & \mu_3 & \mu_5 & \mu_7 & \mu_{12} \\ \eta_2 & \eta_8 & \mu_2 & \mu_4 & \mu_6 & \mu_8 & \mu_{13} \\ \eta_3 & \eta_9 & \eta_{14} & 0 & 0 & \mu_9 & \mu_{14} \\ \eta_4 & \eta_{10} & \eta_{15} & \eta_{19} & 0 & \mu_{10} & \mu_{15} \\ \eta_5 & \eta_{11} & \eta_{16} & \eta_{20} & \eta_{23} & \mu_{11} & \mu_{16} \\ \eta_6 & \eta_{12} & \eta_{17} & \eta_{21} & \eta_{24} & \eta_{26} & \mu_{17} \\ \eta_7 & \eta_{13} & \eta_{18} & \eta_{22} & \eta_{25} & \eta_{27} & \eta_{28} \end{array}\right]. \tag{2.22}$$

The number of parameters in the $\boldsymbol{G}$ matrix varies for different multiplicity vectors $\boldsymbol{m}$. When the eigenvalues are all distinct and unstructured ($\boldsymbol{m} = \boldsymbol{1}_p$), then

$$\operatorname{Dim}(\boldsymbol{\mu}) = p^2 - \frac{p(p+1)}{2} = \frac{p(p-1)}{2}.$$

If $m_k \geqslant 1$, then $\binom{m_k}{2}$ zeros are introduced into $\boldsymbol{G}_k$, where $\binom{n}{m}$ is extended to $n \in \mathbb{R}$, as defined by Abramowitz and Stegun (1992, pg. 822). In general, for any multiplicity vector $\boldsymbol{m}$,

$$\begin{aligned} \operatorname{Dim}(\boldsymbol{\mu}) &= \frac{p(p-1)}{2} - \sum_{k=1}^{h} \binom{m_k}{2} \\ &= \frac{p(p-1)}{2} - \sum_{k=1}^{h} \frac{m_k(m_k-1)}{2} \\ &= \frac{p^2}{2} - \frac{p}{2} - \frac{1}{2}\sum_{k=1}^{h} m_k^2 + \frac{1}{2}\sum_{k=1}^{h} m_k \\ &= \frac{p^2}{2} - \frac{p}{2} - \frac{1}{2}\sum_{k=1}^{h} m_k^2 + \frac{1}{2}p \\ &= \frac{p^2 - \boldsymbol{m}'\boldsymbol{m}}{2}. \end{aligned}$$

Derivatives of $\boldsymbol{\Sigma}$ with respect to the eigenvalue and eigenvector parameters are needed when estimating parameters and constructing confidence intervals. Prior to obtaining an expression for the first derivative of $\operatorname{vec}\boldsymbol{\Sigma}$, an expression for the first derivative of $\operatorname{vec}\boldsymbol{G}$ must be found. The first derivative of $\operatorname{vec}\boldsymbol{G}$ and $\operatorname{vec}\boldsymbol{\Sigma}$ are given in Theorem 2.5 and 2.6 respectively.

Theorem 2.5 (First derivative of $\operatorname{vec}\boldsymbol{G}$ (Boik 2002a)). *Denote by* $\boldsymbol{A}_1$ *the* $p^2 \times \nu_2$ *indicator matrix that satisfies the property* $\boldsymbol{A}_1' \operatorname{vec}\boldsymbol{G} = \boldsymbol{\mu}$. *Similarly, denote by* $\boldsymbol{A}_2$ *the* $p^2 \times \frac{p(p+1)}{2}$ *indicator matrix that satisfies* $\boldsymbol{A}_2' \operatorname{vec}\boldsymbol{G} = \boldsymbol{\eta}$. *Note that* $\boldsymbol{A}_1'\boldsymbol{A}_2 = \boldsymbol{0}$, *and* $\operatorname{vec}\boldsymbol{G} = \boldsymbol{A}_1\boldsymbol{\mu} + \boldsymbol{A}_2\boldsymbol{\eta}$. *Then, the first derivative of* $\operatorname{vec}\boldsymbol{G}$ *with respect to* $\boldsymbol{\mu}'$ *evaluated at* $\boldsymbol{\mu} = \boldsymbol{0}$ *is*

$$\boldsymbol{D}_{\boldsymbol{G}}^{(1)} = \left.\frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = \boldsymbol{A}_1 - \boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{A}_1 = \left[\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right]\boldsymbol{A}_1.$$

Proof. The proof uses the implicit function theorem (Fulks 1979, pg. 352). Define

$$\boldsymbol{F}(\boldsymbol{\mu}) \stackrel{\text{def}}{=} \operatorname{vech}(\boldsymbol{G}\boldsymbol{G}') - \operatorname{vech}(\boldsymbol{I}_p) = \boldsymbol{H}\operatorname{vec}(\boldsymbol{G}\boldsymbol{G}') - \boldsymbol{H}\operatorname{vec}\boldsymbol{I}_p,$$

where $\boldsymbol{H}$ is any generalized inverse of the $p^2 \times \frac{p(p+1)}{2}$ duplication matrix $\boldsymbol{D}_p$ (Magnus 1988). The duplication matrix relates the vec and vech operators for a symmetric matrix. For example, if $\boldsymbol{A}$ is a symmetric $p \times p$ matrix, then

$$\operatorname{vec}(\boldsymbol{A}) = \boldsymbol{D}_p \operatorname{vech}(\boldsymbol{A}).$$

The parameter vector η is defined implicitly by $\boldsymbol{F}(\boldsymbol{\mu}) = \boldsymbol{0}\ \forall \mu$. If the determinant of $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}'}$ evaluated at $\boldsymbol{\mu} = \boldsymbol{0}$ is not zero, then $\boldsymbol{\eta}$ is a function of $\boldsymbol{\mu}$ (e.g. $\boldsymbol{G}$ is solely a function of $\boldsymbol{\mu}$) in a neighborhood of $\boldsymbol{\mu} = \mathbf{0}$. The derivative of $\boldsymbol{F}$ with respect to $\boldsymbol{\eta}'$ is

$$\begin{aligned}\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}'} &= \frac{\partial}{\partial \boldsymbol{\eta}'} \left[\operatorname{vech}(\boldsymbol{G}\boldsymbol{G}') - \operatorname{vech}(\boldsymbol{I}_p)\right] \\ &= \frac{\partial}{\partial \boldsymbol{\eta}'} \left[\boldsymbol{H} \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}') - \boldsymbol{H} \operatorname{vec} \boldsymbol{I}_p\right] \\ &= \frac{\partial \boldsymbol{H} \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\eta}'} \\ &= \boldsymbol{H} \frac{\partial \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\eta}'}. \end{aligned} \tag{2.23}$$

The derivative of $\operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')$ in (2.23) can be obtained by using the total derivative. Let $\boldsymbol{V} = \boldsymbol{W}(\boldsymbol{Y}, \boldsymbol{Z})$ be a matrix function of the matrices $\boldsymbol{Y}$ and $\boldsymbol{Z}$, which are functions of the matrix $\boldsymbol{X}$. The total derivative (called the Decomposition Rule by MacRae (1974)) is

$$\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{X}'} = \left.\frac{\partial \boldsymbol{W}(\boldsymbol{Y}, \boldsymbol{Z})}{\partial \boldsymbol{X}'}\right|_{\boldsymbol{Y} \text{ fixed}} + \left.\frac{\partial \boldsymbol{W}(\boldsymbol{Y}, \boldsymbol{Z})}{\partial \boldsymbol{X}'}\right|_{\boldsymbol{Z} \text{ fixed}}.$$

First, note that

$$\operatorname{vec}(\boldsymbol{G}\boldsymbol{G}') = (\boldsymbol{G} \otimes \boldsymbol{I}_p) \operatorname{vec} \boldsymbol{G} \tag{2.24}$$

$$\begin{aligned} &= (\boldsymbol{I}_p \otimes \boldsymbol{G}) \operatorname{vec}(\boldsymbol{G}') = \boldsymbol{I}_{(p,p)} (\boldsymbol{G} \otimes \boldsymbol{I}_p) \boldsymbol{I}_{(p,p)} \boldsymbol{I}_{(p,p)} \operatorname{vec} \boldsymbol{G} \\ &= \boldsymbol{I}_{(p,p)} (\boldsymbol{G} \otimes \boldsymbol{I}_p) \operatorname{vec} \boldsymbol{G}. \end{aligned} \tag{2.25}$$

Combine (2.24) and (2.25) to obtain

$$\begin{aligned}\frac{\partial \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\eta}'} &= \left.\frac{\partial \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\eta}'}\right|_{\boldsymbol{\eta} \text{ in first } \boldsymbol{G} \text{ constant}} + \left.\frac{\partial \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\eta}'}\right|_{\boldsymbol{\eta} \text{ in second } \boldsymbol{G} \text{ constant}} \\ &= \left[\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right] (\boldsymbol{G} \otimes \boldsymbol{I}_p) \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\eta}'} \\ &= 2\boldsymbol{N}_p (\boldsymbol{G} \otimes \boldsymbol{I}_p) \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\eta}'} = 2\boldsymbol{N}_p (\boldsymbol{G} \otimes \boldsymbol{I}_p) \boldsymbol{A}_2, \end{aligned} \tag{2.26}$$

where $\boldsymbol{N}_p = \frac{1}{2}\left[\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right]$. Using $\boldsymbol{N}_p = \boldsymbol{D}_p\left(\boldsymbol{D}_p'\boldsymbol{D}_p\right)^{-1}\boldsymbol{D}_p'$, $\boldsymbol{H}\boldsymbol{D}_p = \boldsymbol{I}_{\frac{p(p+1)}{2}}$, and (2.26), it follows that (2.23) can be simplified as

$$\begin{aligned}\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}'} &= 2\boldsymbol{H}\boldsymbol{D}_p\left(\boldsymbol{D}_p'\boldsymbol{D}_p\right)^{-1}\boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\boldsymbol{A}_2\\ &= 2\left(\boldsymbol{D}_p'\boldsymbol{D}_p\right)^{-1}\boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\boldsymbol{A}_2. \end{aligned}\tag{2.27}$$

If $\boldsymbol{\mu} = \boldsymbol{0}$, then $\boldsymbol{G} = \boldsymbol{I}_p$ and (2.27) becomes

$$\left.\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\eta}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = 2\left(\boldsymbol{D}_p'\boldsymbol{D}_p\right)^{-1}\boldsymbol{D}_p'\boldsymbol{A}_2.$$

Since $\boldsymbol{D}_p'\boldsymbol{D}_p$ is a diagonal matrix with ones and twos on the diagonal, it is invertible. Further, $\boldsymbol{D}_p'\boldsymbol{A}_2$ is the identity matrix. Therefore, the conditions of the implicit function theorem are satisfied and $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\mu})$. The derivative of $\operatorname{vec}\boldsymbol{G}$ with respect to the explicit parameters, $\boldsymbol{\mu}$, is obtained as follows. Write $\operatorname{vec}\boldsymbol{G}$ as $\operatorname{vec}\boldsymbol{G} = \boldsymbol{A}_1\boldsymbol{\mu}+\boldsymbol{A}_2\boldsymbol{\eta}$. Then

$$\begin{aligned}\frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}'} &= \left.\frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\eta}\text{ fixed}} + \frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\eta}'}\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}\\ &= \boldsymbol{A}_1 + \boldsymbol{A}_2\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}. \end{aligned}\tag{2.28}$$

The implicit function theorem guarantees that $\boldsymbol{\eta}$ is an implicit function of $\boldsymbol{\mu}$ in a neighborhood of $\boldsymbol{\mu} = \boldsymbol{0}$. Accordingly, (2.28) evaluated at $\boldsymbol{\mu} = \boldsymbol{0}$ becomes

$$\boldsymbol{D}_{\boldsymbol{G}}^{(1)} = \boldsymbol{A}_1 + \boldsymbol{A}_2\left.\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}}.$$

The implicit derivative $\left.\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}}$ is found by using the constraint

$$\boldsymbol{F}(\boldsymbol{\mu}) = \operatorname{vech}(\boldsymbol{G}\boldsymbol{G}') - \operatorname{vech}\boldsymbol{I}_p = \boldsymbol{0}.$$

Specifically,

$$
\begin{aligned}
&\operatorname{vech}(\boldsymbol{G}\boldsymbol{G}') - \operatorname{vech}\boldsymbol{I}_p = \mathbf{0} \;\; \forall \boldsymbol{\mu} \\
\Longrightarrow\; & \frac{\partial \operatorname{vech}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\mu}'} = \mathbf{0} \;\; \forall \boldsymbol{\mu} \\
\Longrightarrow\; & \boldsymbol{H}\frac{\partial \operatorname{vec}(\boldsymbol{G}\boldsymbol{G}')}{\partial \boldsymbol{\mu}'} = \mathbf{0} \;\; \forall \boldsymbol{\mu} \\
\Longrightarrow\; & 2\boldsymbol{H}\boldsymbol{D}_p\left(\boldsymbol{D}_p'\boldsymbol{D}_p\right)^{-1}\boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}'} = \mathbf{0} \;\; \forall \boldsymbol{\mu} \\
\Longrightarrow\; & \boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{A}_1 + \boldsymbol{A}_2\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}\right) = \mathbf{0} \;\; \forall \boldsymbol{\mu} \text{ because } \boldsymbol{H}\boldsymbol{D}_p = \boldsymbol{I}_{\frac{p(p+1)}{2}} \\
\Longrightarrow\; & \boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\boldsymbol{A}_2\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\boldsymbol{A}_1 \;\; \forall \boldsymbol{\mu} \\
\Longrightarrow\; & \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'} = -\left[\boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\boldsymbol{A}_2\right]^{-1}\boldsymbol{D}_p'\left(\boldsymbol{G}\otimes\boldsymbol{I}_p\right)\boldsymbol{A}_1 \;\; \forall \boldsymbol{\mu} \\
\Longrightarrow\; & \left.\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = -\left(\boldsymbol{D}_p'\boldsymbol{A}_2\right)^{-1}\boldsymbol{D}_p'\boldsymbol{A}_1 \\
\Longrightarrow\; & \left.\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = -\boldsymbol{D}_p'\boldsymbol{A}_1, \text{ because } \boldsymbol{D}_p'\boldsymbol{A}_2 = \boldsymbol{I}_{\frac{p(p+1)}{2}}.
\end{aligned}
$$

It is not too difficult to show that $\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{A}_1 = \boldsymbol{I}_{(p,p)}\boldsymbol{A}_1$. Accordingly, the derivative of $\operatorname{vec}\boldsymbol{G}$ with respect to $\boldsymbol{\mu}'$ is

$$
\boldsymbol{D}_{\boldsymbol{G}}^{(1)} = \left.\frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \boldsymbol{A}_1 - \boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{A}_1 = \left[\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right]\boldsymbol{A}_1.
$$

□

Theorem 2.6 (First derivatives of $\operatorname{vec}\boldsymbol{\Sigma}$ (Boik 2002b)). *The first derivatives of* $\operatorname{vec}\boldsymbol{\Sigma}$ *with respect to the transpose of the parameters* $\boldsymbol{\beta}$, $\boldsymbol{\mu}$ *and* $\boldsymbol{\varphi}$ *are as follows:*

$$
\boldsymbol{F}_{\boldsymbol{\beta}}^{(1)} = \left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \mathbf{0} \text{ (by construction)},
$$

$$\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)} = \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = 2\boldsymbol{N}_p \left(\boldsymbol{\Gamma}_0 \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0\right) \boldsymbol{D}_{\boldsymbol{G}}^{(1)},$$
$$\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} = \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \left(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0\right) \boldsymbol{L}_p \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)} = \left(\boldsymbol{\Gamma}_0 * \boldsymbol{\Gamma}_0\right) \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)},$$

where $\boldsymbol{L}_p$ is as defined in Theorem 2.1(c)*, and $\boldsymbol{N}_p$ and $\boldsymbol{D}_{\boldsymbol{G}}^{(1)}$ are given in Theorem 2.5.*

Proof. The total derivative is used to find the derivative of $\operatorname{vec} \boldsymbol{\Sigma}$ with respect to $\boldsymbol{\mu}'$. The scatter matrix is parameterized as $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}) \boldsymbol{\Lambda} \boldsymbol{G}'(\boldsymbol{\mu}) \boldsymbol{\Gamma}_0'$. The derivative of $\operatorname{vec}(\boldsymbol{\Sigma})$ with respect the vector $\boldsymbol{\mu}'$ is

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'} = \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu} \text{ in first } \boldsymbol{G} \text{ constant}} + \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu} \text{ in second } \boldsymbol{G} \text{ constant}}.$$

For convenience, specify $\boldsymbol{\mu}$ as $\boldsymbol{\mu}_i$ in the i^{th} $\boldsymbol{G}$ matrix. Using the identity

$$\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{B}) \operatorname{vec} \boldsymbol{A},$$

the vec of the scatter matrix is

$$\begin{aligned}
\operatorname{vec} \boldsymbol{\Sigma} &= \operatorname{vec}(\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_1) \boldsymbol{\Lambda} \boldsymbol{G}(\boldsymbol{\mu}_2)' \boldsymbol{\Gamma}_0') \\
&= (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_2) \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}_1)) \\
&= (\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_1) \boldsymbol{\Lambda}) \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}_2)').
\end{aligned}$$

Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are each $p \times p$ matrices. Using the identities that $\boldsymbol{I}_{(p,p)} \operatorname{vec} \boldsymbol{A} = \operatorname{vec}(\boldsymbol{A}')$, $\boldsymbol{I}_{(p,p)} (\boldsymbol{A} \otimes \boldsymbol{B}) \boldsymbol{I}_{(p,p)} = (\boldsymbol{B} \otimes \boldsymbol{A})$, and $\boldsymbol{I}_{(p,p)}^2 = \boldsymbol{I}_{p^2}$; $\operatorname{vec} \boldsymbol{\Sigma}$ becomes

$$\begin{aligned}
\operatorname{vec} \boldsymbol{\Sigma} &= \operatorname{vec}(\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_1) \boldsymbol{\Lambda} \boldsymbol{G}(\boldsymbol{\mu}_2)' \boldsymbol{\Gamma}_0') \\
&= (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_2) \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}_1)) \qquad (2.29) \\
&= \boldsymbol{I}_{(p,p)} (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_1) \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}_2)). \qquad (2.30)
\end{aligned}$$

Using (2.29) and (2.30), the derivative of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\mu}'$ is

$$\begin{aligned}\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'} &= \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}_2 \text{ fixed}} + \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}_1 \text{ fixed}} \\ &= (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_2)\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \frac{\partial \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}_1))}{\partial \boldsymbol{\mu}_1'} + \boldsymbol{I}_{(p,p)} (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu}_1)\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \frac{\partial \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}_2))}{\partial \boldsymbol{\mu}_2'} \\ &= (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu})\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \frac{\partial \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}))}{\partial \boldsymbol{\mu}'} + \boldsymbol{I}_{(p,p)} (\boldsymbol{\Gamma}_0 \boldsymbol{G}(\boldsymbol{\mu})\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \frac{\partial \operatorname{vec}(\boldsymbol{G}(\boldsymbol{\mu}))}{\partial \boldsymbol{\mu}'} \\ &= 2\boldsymbol{N}_p (\boldsymbol{\Gamma}_0 \boldsymbol{G}\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}.\end{aligned}$$

Accordingly, the derivative of $\operatorname{vec} \boldsymbol{\Sigma}$ with respect to $\boldsymbol{\mu}'$ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ is

$$\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)} = \left.\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = 2\boldsymbol{N}_p (\boldsymbol{\Gamma}_0 \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}_0) \boldsymbol{D}_{\boldsymbol{G}}^{(1)}.$$

□

The remaining derivatives of $\boldsymbol{G}$ and $\operatorname{vec} \boldsymbol{\Sigma}$ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\varphi}$ can be obtained in a similar manner to those obtained in Theorems 2.5 and 2.6. To obtain expressions for the second and third derivatives of $\boldsymbol{G}$, second and third derivatives of the constraint $\operatorname{vech}(\boldsymbol{G}\boldsymbol{G}' - \boldsymbol{I}_p) = \mathbf{0}$ are used. This allows solving for second and third implicit derivatives of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\mu}$. These derivatives will be used for successive derivatives of $\operatorname{vec} \boldsymbol{\Sigma}$. A complete list of these derivatives can be found in Appendix A.

Solving for the Implicit Parameter $\boldsymbol{\eta}$

According to the implicit function theorem, $\boldsymbol{\eta}$ is a function of $\boldsymbol{\mu}$ in a neighborhood of $\boldsymbol{\mu} = \mathbf{0}$. Because $\boldsymbol{G} \approx \boldsymbol{I}_p$, it follows that the diagonal entries of $\boldsymbol{G}$ are positive and near one. Because $\boldsymbol{G}$ is an orthogonal matrix, the columns of $\boldsymbol{G}$ are orthogonal to each other and each has unit norm. An algorithm can be derived to solve for $\boldsymbol{\eta}$. The algorithm can be illustrated using $\boldsymbol{G}$ from equation (2.22). Partition $\boldsymbol{G}$ according to its columns:

$$\boldsymbol{G} = \begin{bmatrix} \eta_1 & 0 & \mu_1 & \mu_3 & \mu_5 & \mu_7 & \mu_{12} \\ \eta_2 & \eta_8 & \mu_2 & \mu_4 & \mu_6 & \mu_8 & \mu_{13} \\ \eta_3 & \eta_9 & \eta_{14} & 0 & 0 & \mu_9 & \mu_{14} \\ \eta_4 & \eta_{10} & \eta_{15} & \eta_{19} & 0 & \mu_{10} & \mu_{15} \\ \eta_5 & \eta_{11} & \eta_{16} & \eta_{20} & \eta_{23} & \mu_{11} & \mu_{16} \\ \eta_6 & \eta_{12} & \eta_{17} & \eta_{21} & \eta_{24} & \eta_{26} & \mu_{17} \\ \eta_7 & \eta_{13} & \eta_{18} & \eta_{22} & \eta_{25} & \eta_{27} & \eta_{28} \end{bmatrix} = \left[\boldsymbol{g}_1 \quad \boldsymbol{g}_2 \quad \boldsymbol{g}_3 \quad \boldsymbol{g}_4 \quad \boldsymbol{g}_5 \quad \boldsymbol{g}_6 \quad \boldsymbol{g}_7\right].$$

The algorithm proceeds as follows: The last column of $\boldsymbol{G}$, $\boldsymbol{g}_7$ has unit norm. Therefore

$$\eta_{28} = \pm\sqrt{1 - \mu_{12}^2 - \mu_{13}^2 - \mu_{14}^2 - \mu_{15}^2 - \mu_{16}^2 - \mu_{17}^2}.$$

However, since η_{28} lies on the diagonal of $\boldsymbol{G}$, it must be positive and near one. Hence,

$$\eta_{28} = \sqrt{1 - \mu_{12}^2 - \mu_{13}^2 - \mu_{14}^2 - \mu_{15}^2 - \mu_{16}^2 - \mu_{17}^2}.$$

Second, the dot product between $\boldsymbol{g}_6$, and $\boldsymbol{g}_7$ is zero and $\boldsymbol{g}_6$ has unit norm. Using these two conditions, it is possible to solve for η_{26} and η_{27}. In general, the implicit parameters in the unit vector $\boldsymbol{g}_i$ can be solved for by recognizing it is orthogonal to $\boldsymbol{g}_{i+1}, \cdots, \boldsymbol{g}_p$. A system of equations results which is easy to solve.

CHAPTER 3

ESTIMATING PARAMETERS

There are various methods of deriving estimators for parameters. Maximum likelihood estimation is most frequently used. Other methods of deriving estimators involve the method of moments, Bayes estimators, or invariant estimators (Casella and Berger 1990, Chapter 7). Another method of estimation is minimizing a loss function. Maximum likelihood is a special case of minimizing a loss function, where the loss function is the negative log-likelihood function. Another example of a loss function is the least squares criterion. The least squares estimator of $\boldsymbol{\Sigma}$ is the minimizer of

$$\ell(\boldsymbol{\Sigma}, \hat{\boldsymbol{\Sigma}}) = \|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}\|^2,$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{Y}'(\boldsymbol{I}_p - \boldsymbol{H})\boldsymbol{Y}$ and $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Maximum likelihood estimators for the eigenvalues and eigenvectors of the spectral model (2.5) under a multivariate normal distribution have been thoroughly explored by Boik (2002a, 2003). Particular choices for loss functions are the basis for finding robust estimators.

Another method for finding robust estimators is through an estimating function. Yuan and Jennrich (1998) define an estimating function as a function that has the form

$$G_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_i(\boldsymbol{\theta}), \tag{3.1}$$

where $\boldsymbol{g}_i(\boldsymbol{\theta})$ are $p \times 1$ independent random vectors. Further, they give conditions for which the zeros of an estimating function have a normal asymptotic distribution. Their results are given in Theorem 4.1.

This dissertation proposes an extension of Maronna's M-estimators of location and scatter (Maronna 1976), and an extension of Rousseeuw and Yohai's S-estimators (Rousseeuw and Yohai 1984). Both robust estimators are generalized by allowing multivariate regression along with modeling of the eigenvalues. For example, the model could allow some of the eigenvalues to have multiplicities larger than one.

This chapter discusses extensions of M-estimators and S-estimators, along with methods of solving the corresponding estimating equations. Multivariate regression and modeling of eigenvalues are allowed in the extension.

M-estimators

M-estimators for multivariate location and scatter were first defined by Maronna (1976). Huber (1981, pg. 212) extended Maronna's definition of M-estimators for multivariate location and scatter to be the solution, $(\tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\Sigma}})$, to the system

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n} v_1(d_k)(\boldsymbol{y}_k - \boldsymbol{\tau}) &= \mathbf{0}, \\ \frac{1}{n}\sum_{k=1}^{n} \left[v_2(d_k^2)(\boldsymbol{y}_k - \boldsymbol{\tau})(\boldsymbol{y}_k - \boldsymbol{\tau})' - v_3(d_k)\boldsymbol{\Sigma}\right] &= \mathbf{0}, \end{aligned} \tag{3.2}$$

where $\boldsymbol{\Sigma}$ is an unstructured scatter matrix, $\boldsymbol{\tau}$ is an unstructured vector of location parameters, $d_k = d_k(\boldsymbol{y}_k; \boldsymbol{\tau}, \boldsymbol{\Sigma}) = \left\{(\boldsymbol{y}_k - \boldsymbol{\tau})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_k - \boldsymbol{\tau})\right\}^{\frac{1}{2}}$, and v_1, v_2, and v_3 are nonnegative scalar valued functions. The existence and uniqueness of solutions to

(3.2) were shown for the case of $v_3 = 1$ by Maronna (1976) and Huber (1981, pg. 226-7). In addition, in this case Maronna (1976) showed that the estimators are consistent and asymptotically normal. In this dissertation, v_3 will be taken to be 1.

I propose to generalize the structure of the location parameters to that of multivariate regression; e.g. $\boldsymbol{\tau}_k = \boldsymbol{X}_k\boldsymbol{\beta}$, where $\boldsymbol{X}_k$ and $\boldsymbol{\beta}$ are defined in (1.3). Further, the spectral model will allow eigenvalue modeling of the M-estimators.

Estimating Equations

I will motivate the derivation of the estimating equations through the specific derivation of the maximum likelihood estimator for the family of elliptical distributions (Fang et al. 1990). The probability density function of an elliptically contoured distribution has the form

$$f(\boldsymbol{y}_k) = \frac{g(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}, \text{ where} \tag{3.3}$$

$$d_k = \left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})' \, \boldsymbol{\Sigma}^{-1} \, (\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}) \right]^{\frac{1}{2}},$$

$\boldsymbol{\Sigma}$ is the characteristic matrix, and g is a non-negative function. The density function for $\boldsymbol{Y}$ is

$$f(\boldsymbol{Y}) = \prod_{k=1}^{n} \frac{g(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} = \frac{\prod_{k=1}^{n} g(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}}.$$

If the observed data actually follow the elliptical distribution in (3.3), then the estimation method to be derived will provide estimation of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$. If, however, the observed data follow some other elliptical distribution, then the proposed method provides estimation of $\boldsymbol{B}$ and $\boldsymbol{\Sigma} = \kappa\boldsymbol{\Sigma}^*$, where $\boldsymbol{\Sigma}^*$ is the characteristic matrix of the underlying distribution.

For organization purposes, it is convenient to place the parameters for both the linear model and the spectral model in a vector,

$$\underset{\nu\times 1}{\boldsymbol{\theta}} = \begin{bmatrix} \underset{\nu_1\times 1}{\boldsymbol{\beta}} \\ \underset{\nu_2\times 1}{\boldsymbol{\mu}} \\ \underset{\nu_3\times 1}{\boldsymbol{\varphi}} \end{bmatrix}, \tag{3.4}$$

where $\nu_1 + \nu_2 + \nu_3 = \nu$.

The log-likelihood function for an elliptically contoured density is

$$\ell(\boldsymbol{\theta}) = \log f(\boldsymbol{Y}) = -\frac{n}{2}\log|\boldsymbol{\Sigma}| + \sum_{k=1}^{n} \log\left[g(d_k^2)\right]$$

where $\boldsymbol{\theta}$ is defined in (3.4). For convenience, combine the $\boldsymbol{\mu}$ and $\boldsymbol{\varphi}$ parts into the vector

$$\boldsymbol{\zeta} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\varphi} \end{pmatrix}.$$

THEOREM 3.1 (FIRST DERIVATIVES OF THE FAMILY OF ELLIPTICALLY CONTOURED DISTRIBUTIONS). *The first derivatives of the log likelihood function of the family of elliptically contoured distributions with respect to the parameters* $\boldsymbol{\beta}$ *and* $\boldsymbol{\zeta} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\varphi} \end{pmatrix}$ *are*

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = \sum_{k=1}^{n} \frac{-2g^{(1)}(d_k^2)}{g(d_k^2)} \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_k \text{ and}$$

$$\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\left\{\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)} \boldsymbol{z}_k \boldsymbol{z}_k' - \boldsymbol{\Sigma}\right\}\right]' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)},$$

where $g^{(1)}(d_k^2) = \frac{\partial g(d_k^2)}{\partial d_k^2}$, $\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}$, $\boldsymbol{X}_k$ *is defined in* (1.4), $\boldsymbol{F}^{(1)} = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)} \\ \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \end{bmatrix}$, *and* $\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}$ *and* $\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}$ *are defined in Theorem 2.6. Note that the equations above are equivalent to*

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = \operatorname{vec}\big(\boldsymbol{X}'\boldsymbol{T}_g\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\big)' \textit{ and}$$

$$\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \frac{n}{2}\operatorname{vec}\left[\frac{1}{n}\boldsymbol{Z}'\boldsymbol{T}_g\boldsymbol{Z} - \boldsymbol{\Sigma}\right]\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)},$$

where $\boldsymbol{T}_g = \operatorname{Diag}\left(\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\right)$.

Proof. The derivative of the log likelihood function with respect to the location parameters, $\boldsymbol{\beta}$, is

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = -\frac{n}{2}\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta}'} + \sum_{k=1}^{n}\frac{\partial \log\left[g(d_k^2)\right]}{\partial \boldsymbol{\beta}'}. \tag{3.5}$$

The first part of (3.5) is zero because $\boldsymbol{\Sigma}$ does not depend on $\boldsymbol{\beta}$. Accordingly, (3.5) becomes

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}} &= \sum_{k=1}^{n}\frac{\partial \log\left[g(d_k^2)\right]}{\partial \boldsymbol{\beta}'} \\
&= \sum_{k=1}^{n}\frac{\partial \log\left[g(d_k^2)\right]}{\partial g(d_k^2)}\underbrace{\frac{\partial g(d_k^2)}{\partial d_k^2}}_{g^{(1)}(d_k^2)}\frac{\partial d_k^2}{\partial \boldsymbol{\beta}'} \\
&= \sum_{k=1}^{n}\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\frac{\partial d_k^2}{\partial \boldsymbol{\beta}'}.
\end{aligned} \tag{3.6}$$

The derivative of d_k^2 with respect to $\boldsymbol{\beta}'$ can be written in terms of $\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}$. Because

$$d_k^2 = \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k,$$

it follows that

$$\frac{\partial d_k^2}{\partial \boldsymbol{\beta}'} = 2\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{z}_k}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k. \tag{3.7}$$

Using (3.7), (3.6) simplifies to

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = \sum_{k=1}^{n} -2\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k. \tag{3.8}$$

The derivative of the log likelihood function with respect to the scale parameters, $\boldsymbol{\zeta}$, is

$$\frac{\partial \ell}{\partial \boldsymbol{\zeta}'} = -\frac{n}{2}\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\zeta}'} + \sum_{k=1}^{n}\frac{\partial \log\left[g(d_k^2)\right]}{\partial \boldsymbol{\zeta}'}. \tag{3.9}$$

Searle (1982) showed that the derivative of the log determinant of a matrix with respect to a scalar can be written as

$$\begin{aligned}\frac{\partial \log|\boldsymbol{X}|}{\partial y} &= \operatorname{tr}\left(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial y}\right) = \operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\operatorname{vec}\left(\frac{\partial \boldsymbol{X}}{\partial y}\right)\\ &= \operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y}.\end{aligned} \tag{3.10}$$

The derivative of the log determinant of a $p \times p$ matrix with respect to a row vector can be found by using (3.10). It is

$$\begin{aligned}\frac{\partial \log|\boldsymbol{X}|}{\partial \boldsymbol{y}'} &= \begin{pmatrix}\frac{\partial \log|\boldsymbol{X}|}{\partial y_1} & \frac{\partial \log|\boldsymbol{X}|}{\partial y_2} & \cdots & \frac{\partial \log|\boldsymbol{X}|}{\partial y_p}\end{pmatrix}\\ &= \begin{pmatrix}\operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y_1} & \operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y_2} & \cdots & \operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y_p}\end{pmatrix}\\ &= \operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\begin{pmatrix}\frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y_1} & \frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y_2} & \cdots & \frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial y_p}\end{pmatrix}\\ &= \operatorname{vec}\left[(\boldsymbol{X}')^{-1}\right]'\frac{\partial \operatorname{vec}\boldsymbol{X}}{\partial \boldsymbol{y}'}.\end{aligned} \tag{3.11}$$

Therefore, the first part of (3.9) is

$$\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\zeta}'} = \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big)' \, \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}. \tag{3.12}$$

The right-hand side of (3.9) is

$$\begin{aligned} \frac{\partial \log[g(d_k^2)]}{\partial \boldsymbol{\zeta}'} &= \frac{\partial \log[g(d_k^2)]}{\partial d_k^2} \frac{\partial d_k^2}{\partial \boldsymbol{\zeta}'} \\ &= \frac{g^{(1)}(d_k^2)}{g(d_k^2)} \frac{\partial d_k^2}{\partial \boldsymbol{\zeta}'}. \end{aligned} \tag{3.13}$$

It follows from $d_k^2 = \operatorname{tr}(d_k^2) = \operatorname{tr}(\boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k) = \operatorname{tr}(\boldsymbol{z}_k \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1}) = \operatorname{vec}(\boldsymbol{z}_k \boldsymbol{z}_k')' \operatorname{vec}(\boldsymbol{\Sigma}^{-1})$, that

$$\frac{\partial d_k^2}{\partial \boldsymbol{\zeta}'} = \operatorname{vec}(\boldsymbol{z}_k \boldsymbol{z}_k')' \, \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\zeta}'}. \tag{3.14}$$

To compute the derivative on the right-hand side of (3.14), the total derivative is used in conjunction with $\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A}) \operatorname{vec} \boldsymbol{B}$. Because $\operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}) = \operatorname{vec} \boldsymbol{I}_p = (\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}^{-1}) \operatorname{vec} \boldsymbol{\Sigma} = (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_p) \operatorname{vec}(\boldsymbol{\Sigma}^{-1})$, it follows that

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma})}{\partial \boldsymbol{\zeta}'} = 0 = \big(\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}^{-1}\big) \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'} + (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_p) \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\zeta}'}. \tag{3.15}$$

Solving (3.15) for $\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\zeta}'}$ yields

$$\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\zeta}'} = -\big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\big) \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}. \tag{3.16}$$

Hence, (3.14) becomes

$$\frac{\partial d_k^2}{\partial \boldsymbol{\zeta}'} = -\operatorname{vec}(\boldsymbol{z}_k \boldsymbol{z}_k')' \big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\big) \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}. \tag{3.17}$$

By combining (3.12), (3.13), and (3.17), equation (3.9) simplifies to

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\zeta}'} &= -\frac{n}{2}(\operatorname{vec}\boldsymbol{\Sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'} \\
&\quad + \sum_{k=1}^{n}\frac{-g^{(1)}(d_k^2)}{g(d_k^2)}\operatorname{vec}(\boldsymbol{z}_k\boldsymbol{z}_k')'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}, \\
&= \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k\boldsymbol{z}_k' - n\boldsymbol{\Sigma}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}, \\
&= \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\left\{\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right\}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}.
\end{aligned} \tag{3.18}$$

Lastly, evaluating (3.18) at $\boldsymbol{\mu} = \mathbf{0}$ yields

$$\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\left\{\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right\}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}, \tag{3.19}$$

where $\boldsymbol{F}^{(1)} = \left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \begin{bmatrix}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \\ \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\end{bmatrix}$, and $\boldsymbol{F}^{(1)}_{\cdot}$ are defined in Appendix A.

To show the matrix equivalent of (3.8), the relation in equation (1.4) is used. That is,

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\beta}'} &= \sum_{k=1}^{n} -2\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{I}_p\otimes\boldsymbol{x}_k'\right) = \sum_{k=1}^{n} -2\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\left(\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{x}_k'\right) \\
&= \sum_{k=1}^{n} -2\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{x}_k\right)' = \sum_{k=1}^{n} -2\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\operatorname{vec}\big(\boldsymbol{x}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\big)' \\
&= \sum_{k=1}^{n}\operatorname{vec}\left(\boldsymbol{x}_k\left[\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\right]\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)' \\
&= \operatorname{vec}\left(\sum_{k=1}^{n}\boldsymbol{x}_k\left[\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\right]\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)' \\
&= \operatorname{vec}\big(\boldsymbol{X}'\boldsymbol{T}_g\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\big)'.
\end{aligned}$$

Similarly, it can be shown that the matrix equivalent of equation (3.19) is

$$\frac{\partial \ell}{\partial \boldsymbol{\zeta}'} = \frac{n}{2}\operatorname{vec}\left[\frac{1}{n}\boldsymbol{Z}'\boldsymbol{T}_g\boldsymbol{Z} - \boldsymbol{\Sigma}\right]\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}.$$

□

Setting the derivative, $\frac{\partial \ell}{\partial \boldsymbol{\beta}}$, equal to zero yields the first generalized Maronna M-estimating equation:

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{X}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \underbrace{\left(-2\frac{g^{(1)}(d_k^2)}{g(d_k^2)}\right)}_{v_1(d_k)} = 0 \Longrightarrow \frac{1}{n}\sum_{k=1}^{n} v_1(d_k)\boldsymbol{X}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k = 0. \quad (3.20)$$

The first Maronna equation (3.2) is a special case of (3.20). If $\boldsymbol{\beta} = \boldsymbol{\tau}$ and $\boldsymbol{X} = \mathbf{1}_n$ (no explanatory variables), then $\boldsymbol{X}_k = \boldsymbol{I}_p$. In this case, equations (3.2) and (3.20) are identical.

Setting the derivative, $\dfrac{\partial \ell}{\partial \boldsymbol{\zeta}}$, equal to zero yields the second generalized Maronna M-estimating equation:

$$\begin{aligned}
\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}}\right|_{\boldsymbol{\mu}=\mathbf{0}} &= \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\left\{\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right\}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}, \\
&\Longrightarrow \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\left\{\underbrace{\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}}_{v_2(d_k^2)}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right\}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}, \\
&\Longrightarrow \frac{1}{n}\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left[\sum_{k=1}^{n}\left\{v_2(d_k^2)\boldsymbol{z}_k\boldsymbol{z}_k' - v_3(d_k)\boldsymbol{\Sigma}\right\}\right] = 0, \quad (3.21)
\end{aligned}$$

where $\frac{1}{n}\sum_{k=1}^{n} v_3(d_k) = 1$.

The second Maronna equation (3.2) is a special case of (3.21). If the scatter matrix is unstructured, as in (3.2), then $\boldsymbol{F}^{(1)} = \boldsymbol{D}_p$. The perpendicular projection operator of $\boldsymbol{D}_p$ is $\boldsymbol{N}_p = \boldsymbol{D}_p(\boldsymbol{D}_p'\boldsymbol{D}_p)^{-1}\boldsymbol{D}_p$, which projects onto the space of all symmetric matrices ($\boldsymbol{N}_p \operatorname{vec}(\boldsymbol{A}) = \operatorname{vec}(\boldsymbol{A})$, if $\boldsymbol{A}$ is symmetric). For convenience, define $\diamondsuit =$

$\sum_{k=1}^{n} \left\{ v_2(d_k^2) \boldsymbol{z}_k \boldsymbol{z}_k' - v_3(d_k) \boldsymbol{\Sigma} \right\}$. If $\boldsymbol{F}^{(1)} = \boldsymbol{D}_p$, then (3.21) becomes

$$\begin{aligned}
& \frac{1}{n} \boldsymbol{F}^{(1)\prime} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \operatorname{vec} [\diamondsuit] = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \boldsymbol{D}_p' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \operatorname{vec} [\diamondsuit] = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \boldsymbol{D}_p' \operatorname{vec} \left[\boldsymbol{\Sigma}^{-1} \diamondsuit \boldsymbol{\Sigma}^{-1} \right] = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \underbrace{\boldsymbol{D}_p (\boldsymbol{D}_p' \boldsymbol{D}_p)^{-1} \boldsymbol{D}_p'}_{\boldsymbol{N}_p} \operatorname{vec} \left[\boldsymbol{\Sigma}^{-1} \diamondsuit \boldsymbol{\Sigma}^{-1} \right] = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \operatorname{vec} \left[\boldsymbol{\Sigma}^{-1} \diamondsuit \boldsymbol{\Sigma}^{-1} \right] = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \boldsymbol{\Sigma}^{-1} \diamondsuit \boldsymbol{\Sigma}^{-1} = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \diamondsuit = \mathbf{0} \\
\Longrightarrow\ & \frac{1}{n} \sum_{k=1}^{n} \left\{ v_2(d_k^2) \boldsymbol{z}_k \boldsymbol{z}_k' - v_3(d_k) \boldsymbol{\Sigma} \right\} = \mathbf{0}.
\end{aligned}$$

I propose the solution to the two equations (3.20) and (3.21) as a generalization of Maronna's M-estimators. They are the solution $\boldsymbol{\theta}$ to the system

$$\begin{gathered}
\frac{1}{n} \sum_{k=1}^{n} v_1(d_k) \boldsymbol{X}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k = \mathbf{0} \\
\frac{1}{n} \boldsymbol{F}^{(1)\prime} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right) \operatorname{vec} \left[\sum_{k=1}^{n} \left\{ v_2(d_k^2) \boldsymbol{z}_k \boldsymbol{z}_k' - v_3(d_k) \boldsymbol{\Sigma} \right\} \right] = \mathbf{0},
\end{gathered} \tag{3.22}$$

where $d_k = d_k(\boldsymbol{y}_k; \boldsymbol{\theta}) = \{ \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \}^{\frac{1}{2}}$, $\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta}$, $\boldsymbol{F}^{(1)} = \left. \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\zeta}'} \right|_{\boldsymbol{\mu}=\mathbf{0}}$, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} & \boldsymbol{\zeta} \end{pmatrix}'$, and v_1, v_2, and v_3 are nonnegative scalar valued functions.

Newton-Raphson and Fisher-Scoring

To find the solution for (3.22), a numerical iterative procedure must be used. For a general choice of v_1, v_2, and v_3, a modified Newton-Raphson method can be used. To use the Newton-Raphson method, derivatives of both of the equations in (3.22) with respect to the parameters are needed in order to construct the Hessian matrix. When the functions v_1, v_2, and v_3 correspond to a specific distribution, then a modified Fisher-Scoring iteration scheme can be used instead of Newton-Raphson. Generally, Fisher-Scoring is more stable than the Newton-Raphson method. Since the Fisher-Scoring method and the Newton-Raphson method are similar, a description of the Newton-Raphson method will be given.

Denote $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$ as the log-likelihood function of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$. Also denote $\hat{\boldsymbol{\theta}}_i$ as the guess for $\boldsymbol{\theta}$ after the i^{th} iteration. Expand $\ell(\boldsymbol{\theta})$ in a Taylor series about the i^{th} iteration guess for $\boldsymbol{\theta}$. That is,

$$\ell(\boldsymbol{\theta}|\boldsymbol{Y}) \approx \ell(\hat{\boldsymbol{\theta}}_i|\boldsymbol{Y}) + \left.\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)' \left[\left.\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right] (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i).$$

Let

$$\boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i} = \left.\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \quad \text{and} \quad \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i} = \left.\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}.$$

Then

$$\ell(\boldsymbol{\theta}|\boldsymbol{Y}) \approx \ell(\hat{\boldsymbol{\theta}}_i|\boldsymbol{Y}) + \boldsymbol{g}'_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)'\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i). \tag{3.23}$$

To find the vector which approximately maximizes $\ell(\boldsymbol{\theta}|\boldsymbol{Y})$, take the derivative of (3.23) with respect to $\boldsymbol{\theta}$ and set equal to zero. Solving for $\boldsymbol{\theta}$ results in

$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_i - \boldsymbol{H}^{-1}_{\hat{\boldsymbol{\theta}}_i}\boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{3.24}$$

The right-hand side of (3.24) becomes the new guess $\hat{\boldsymbol{\theta}}_{i+1}$, and the procedure is repeated. The Newton Raphson procedure can be summarized as

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i - \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i}.$$

The Fisher-Scoring algorithm is almost the same as the Newton Raphson algorithm. Rather than using the Hessian matrix ($\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i}$), the Fisher-Scoring algorithm (named for R.A. Fisher) uses the expected value of the Hessian matrix. The negative of the expected value of the Hessian matrix is called Fisher's information matrix and is denoted as $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i} = -E\left[\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i}\right]$. The Fisher-Scoring algorithm in terms of Fisher's Information matrix is therefore

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i + \boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{3.25}$$

Both the Fisher-Scoring method and the Newton-Raphson method described in this section are for traditional problems. These methods need to be modified to work with the implicit parameterization used in this dissertation.

Solving the M-Equations

For a general choice of v_1, v_2, and v_3, a modified Newton-Raphson method can be employed. When the functions v_1, v_2, and v_3 correspond to a particular distribution, then a modified Fisher-Scoring method can be used. To illustrate the modified Fisher-Scoring technique, the functions v_1, v_2, and v_3 corresponding to a multivariate-T distribution will be used.

Suppose the random p-vector, $\mathring{\boldsymbol{n}}_k$, has a multivariate Normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_p$. Further, suppose that $\mathring{\boldsymbol{c}}_k$ has a χ^2 distribution with ξ degrees of freedom. Then the p-dimensional random variable

$$\boldsymbol{t} = \frac{\boldsymbol{\Sigma}^{\frac{1}{2}}\mathring{\boldsymbol{n}}_k}{\sqrt{\frac{\mathring{\boldsymbol{c}}_k}{\xi}}}$$

has a multivariate-T distribution with mean vector $\boldsymbol{\tau}$, characteristic matrix $\boldsymbol{\Sigma}$, and covariance matrix $\alpha\boldsymbol{\Sigma}$, where $\alpha = \frac{\xi}{\xi+2}$. In notational terms, $\boldsymbol{t} \sim \boldsymbol{T}(\boldsymbol{\tau}, \frac{\xi}{\xi+2}\boldsymbol{\Sigma})$. The density of $\boldsymbol{t}$ with ξ degrees of freedom (Press 1982) is

$$f(\boldsymbol{t}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \frac{\boldsymbol{\Gamma}\left(\frac{\xi+p}{2}\right)}{\boldsymbol{\Gamma}\left(\frac{\xi}{2}\right)\pi^{\frac{p}{2}}} \frac{\xi^{\frac{\xi}{2}}}{\left[\xi + (\boldsymbol{t}-\boldsymbol{\tau})'\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{t}-\boldsymbol{\tau})\right]^{\frac{\xi+p}{2}}},$$

where $-\infty < t_j < \infty,\ \xi > 0.$

In terms of the linear model (1.2) and the spectral model (2.5), the distribution of $\boldsymbol{y}_k$ is

$$f(\boldsymbol{y}_k) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \frac{\boldsymbol{\Gamma}\left(\frac{\xi+p}{2}\right)}{\boldsymbol{\Gamma}\left(\frac{\xi}{2}\right)\pi^{\frac{p}{2}}} \frac{\xi^{\frac{\xi}{2}}}{\left[\xi + d_k^2\right]^{\frac{\xi+p}{2}}},$$

where $d_k = \left\{(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})\right\}^{\frac{1}{2}}$.

The v functions for the multivariate-T distribution are

$$v_1(d_k) = \frac{\xi+p}{\xi+d_k^2}, \quad v_2(d_k^2) = \frac{\xi+p}{\xi+d_k^2}, \text{ and } v_3(d_k) = 1. \tag{3.26}$$

Theorem 3.2 (First derivatives of multivariate-T). *The first derivatives of the log-likelihood of the multivariate-T distribution with respect to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\zeta} = \begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\varphi}\end{pmatrix}$ are*

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = (\xi+p)\sum_{k=1}^{n} \frac{z_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k}{\xi+d_k^2} = \operatorname{vec}\big(\boldsymbol{X}'\boldsymbol{T}_g\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\big)',$$

$$\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \frac{n}{2}\operatorname{vec}\left[\frac{1}{n}\sum_{k=1}^{n}\frac{\xi+p}{\xi+d_k^2}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}$$

$$= \frac{n}{2}\operatorname{vec}\left[\frac{1}{n}\boldsymbol{Z}'\boldsymbol{T}_g\boldsymbol{Z} - \boldsymbol{\Sigma}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)},$$

where $\boldsymbol{T}_g = \operatorname{Diag}\left(\frac{\xi+p}{\xi+d_k^2}\right)$ *and* ξ *is known.*

Proof. The multivariate-T distribution belongs to the family of elliptically contoured distributions. The function $g(d_k^2)$ is equal to

$$g(d_k^2) = \frac{\Gamma\left(\frac{\xi+p}{2}\right)\xi^{\frac{\xi}{2}}}{\Gamma\left(\frac{\xi}{2}\right)\pi^{\frac{p}{2}}}(\xi+d_k^2)^{-\frac{\xi+p}{2}},$$

and its derivative is equal to

$$g^{(1)}(d_k^2) = \frac{\partial g(d_k^2)}{\partial d_k^2} = -\left(\frac{\xi+p}{2}\right)\frac{\Gamma\left(\frac{\xi+p}{2}\right)\xi^{\frac{\xi}{2}}}{\Gamma\left(\frac{\xi}{2}\right)\pi^{\frac{p}{2}}}\frac{1}{(\xi+d_k^2)^{\frac{\xi+p}{2}+1}}.$$

Therefore, the ratio of the derivative of g and g is

$$\frac{g^{(1)}(d_k^2)}{g(d_k^2)} = \frac{-\left(\frac{\xi+p}{2}\right)\frac{\Gamma\left(\frac{\xi+p}{2}\right)\xi^{\frac{\xi}{2}}}{\Gamma\left(\frac{\xi}{2}\right)\pi^{\frac{p}{2}}}\frac{1}{(\xi+d_k^2)^{\frac{\xi+p}{2}+1}}}{\frac{\Gamma\left(\frac{\xi+p}{2}\right)\xi^{\frac{\xi}{2}}}{\Gamma\left(\frac{\xi}{2}\right)\pi^{\frac{p}{2}}}\frac{1}{(\xi+d_k^2)^{\frac{\xi+p}{2}}}} = -\frac{1}{2}\left(\frac{\xi+p}{\xi+d_k^2}\right).$$

Using Theorem 3.1, the derivative of the log-likelihood function for the multivariate-T distribution with respect to $\boldsymbol{\beta}'$ is

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = \sum_{k=1}^{n}\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k = \sum_{k=1}^{n}\left(\frac{\xi+p}{\xi+d_k^2}\right)\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k$$

$$= (\xi+p)\sum_{k=1}^{n}\frac{z_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k}{\xi+d_k^2}.$$

Similarly,

$$\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \frac{1}{2}\operatorname{vec}\left[\sum_{k=1}^{n}\left\{\frac{\xi+p}{\xi+d_k^2}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right\}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}.$$

The equivalent matrix expressions for the multivariate-T distribution are

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = \operatorname{vec}\big(\boldsymbol{X}'\boldsymbol{T}_g\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\big)' \text{ and}$$
$$\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \frac{n}{2}\operatorname{vec}\left[\frac{1}{n}\boldsymbol{Z}'\boldsymbol{T}_g\boldsymbol{Z} - \boldsymbol{\Sigma}\right]\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)},$$

where $\boldsymbol{T}_g = \operatorname{Diag}\left(\frac{\xi+p}{\xi+d_k^2}\right)$. □

In order to use Fisher-Scoring for the multivariate-T Distribution, Fisher's Information matrix is needed. First, some lemmas are needed.

LEMMA 3.3 (CASELLA AND BERGER (1990)). *Suppose X_1 and X_2 are independent χ^2 random variables with ν_1 and ν_2 degrees of freedom, respectively. The distribution of $\frac{X_1}{X_1+X_2}$ is*

$$\frac{X_1}{X_1+X_2} \sim \operatorname{Beta}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right).$$

PROOF. The result is given as an exercise in Casella and Berger (1990, pg. 193). The proof is given here for completeness.

The joint distribution of X_1 and X_2 is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}} x_1^{\frac{\nu_1}{2}-1} x_2^{\frac{\nu_2}{2}-1} e^{-\frac{x_1+x_2}{2}}.$$

To find the distribution of $\frac{X_1}{X_1+X_2}$, define the bivariate transformation

$$h(X_1,X_2) = \left(\frac{X_1}{X_1+X_2}, X_2\right) = (Y,Z).$$

The inverse function of $h(X_1,X_2)$ is

$$h^{-1}(Y,Z) = \left(\frac{YZ}{1-Y}, Z\right).$$

Therefore, the joint distribution of Y and Z is

$$
\begin{aligned}
f_{Y,Z}(y,z) &= f_{X_1,X_2}\left(h^{-1}(y,z)\right)|\boldsymbol{J}| \\
&= \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}}\left(\frac{yz}{1-y}\right)^{\frac{\nu_1}{2}-1} z^{\frac{\nu_2}{2}-1}e^{-\frac{yz}{2(1-y)}+\frac{z}{2}}\left[\frac{z}{(1-y)^2}\right] \\
&= \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}}\frac{y^{\frac{\nu_1}{2}-1}}{(1-y)^{\frac{\nu_1}{2}+1}} z^{\frac{\nu_1+\nu_2}{2}-1}e^{-\frac{z}{2(1-y)}}. \qquad (3.27)
\end{aligned}
$$

Integrating z out of (3.27) yields

$$
\begin{aligned}
f_Y(y) &= \int f_{Y,Z}(y,z)\ dz \\
&= \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}}\frac{y^{\frac{\nu_1}{2}-1}}{(1-y)^{\frac{\nu_1}{2}+1}}\int_0^\infty z^{\frac{\nu_1+\nu_2}{2}-1}e^{-\frac{z}{2(1-y)}}\ dz \\
&= \frac{y^{\frac{\nu_1}{2}-1}(1-y)^{\frac{\nu_2}{2}-2}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}}\int_0^\infty \left(\frac{z}{1-y}\right)^{\frac{\nu_1+\nu_2}{2}-1} e^{-\frac{z}{2(1-y)}}\ dz.
\end{aligned}
$$

Making the substitution $u = \frac{z}{1-y}$ yields

$$
\begin{aligned}
f_Y(y) &= \frac{y^{\frac{\nu_1}{2}-1}(1-y)^{\frac{\nu_2}{2}-1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}}\int_0^\infty u^{\frac{\nu_1+\nu_2}{2}-1}e^{-\frac{u}{2}}\ dz \\
&= \frac{y^{\frac{\nu_1}{2}-1}(1-y)^{\frac{\nu_2}{2}-1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}}}\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)2^{\frac{\nu_1+\nu_2}{2}} \\
&= \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}y^{\frac{\nu_1}{2}-1}(1-y)^{\frac{\nu_2}{2}-1}.
\end{aligned}
$$

It follows that

$$
\frac{X_1}{X_1+X_2} \sim \text{Beta}\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right).
$$

□

Corollary 3.4. *Suppose X_1 and X_2 are independent χ^2 random variables with ν_1 and ν_2 degrees of freedom, respectively. The distribution of $\frac{X_1}{X_1+X_2}$ is*

$$
\frac{X_1}{X_1+X_2} \sim \text{Beta}\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right).
$$

Then,

$$\text{(a)}\quad E\left[\left(\frac{X_1}{X_1+X_2}\right)^r\right] = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\Gamma\left(\frac{\nu_1}{2}+r\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_1+\nu_2}{2}+r\right)},$$

$$\text{(b)}\quad E\left[\left(\frac{X_2}{X_1}\right)^k\left(\frac{X_1}{X_1+X_2}\right)^2\right] = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\Gamma\left(\frac{\nu_1}{2}+2-k\right)\Gamma\left(\frac{\nu_2}{2}+k\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\Gamma\left(\frac{\nu_1+\nu_2}{2}+2\right)},$$

where $r > -\frac{\nu_1}{2}$, $-\frac{\nu_2}{2} < k < \frac{\nu_1}{2}+2$.

Proof. The distribution of $\frac{X_1}{X_1+X_2}$ follows from Lemma 3.3. Part (a) is a well known result. A proof is given in Casella and Berger (1990, pg. 108). Note that if $b = \frac{X_1}{X_1+X_2}$, then $\frac{1-b}{b} = \frac{X_2}{X_1}$. So

$$\begin{aligned} E\left[\left(\frac{X_2}{X_1}\right)^k\left(\frac{X_1}{X_1+X_2}\right)^2\right] &= E\left[\left(\frac{1-b}{b}\right)^k b^2\right] = E\left[b^{2-k}(1-b)^k\right] \\ &= \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\int_0^1 b^{\frac{\nu_1}{2}+2-k-1}(1-b)^{\frac{\nu_2}{2}+k-1}\,db \\ &= \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\frac{\Gamma\left(\frac{\nu_1}{2}+2-k\right)\Gamma\left(\frac{\nu_2}{2}+k\right)}{\Gamma\left(\frac{\nu_1+\nu_2}{2}+2\right)} \\ &= \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\Gamma\left(\frac{\nu_1}{2}+2-k\right)\Gamma\left(\frac{\nu_2}{2}+k\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\Gamma\left(\frac{\nu_1+\nu_2}{2}+2\right)}. \end{aligned}$$

□

Lemma 3.5 (Lopuhaä (1989)). *Suppose $\boldsymbol{Y}$ is a random p-vector following an elliptically contoured distribution with mean vector $\mathbf{0}$ and covariance matrix $\alpha\boldsymbol{I}_p$. Then it follows that*

$$\frac{\boldsymbol{Y}}{\sqrt{\boldsymbol{Y}'\boldsymbol{Y}}} \perp\!\!\!\perp \boldsymbol{Y}'\boldsymbol{Y}.$$

Proof. Define the transformation

$$h(\boldsymbol{Y}) = \left(\frac{Y_1}{\sqrt{\boldsymbol{Y}'\boldsymbol{Y}}}, \frac{Y_2}{\sqrt{\boldsymbol{Y}'\boldsymbol{Y}}}, \cdots, \frac{Y_{p-1}}{\sqrt{\boldsymbol{Y}'\boldsymbol{Y}}}, \sqrt{\boldsymbol{Y}'\boldsymbol{Y}}\right) = (Z_1, Z_2, \cdots, Z_p).$$

Further, define

$$\boldsymbol{Z}_{(p)} = (Z_1 \;\; Z_2 \;\; \cdots \;\; Z_{p-1})'.$$

Note that if i equals any number between 1 and $p-1$, then $Y_i = Z_p Z_i$. If $i = p$, then

$Y'Y = Z_p^2$,

$$Y_p^2 = Z_p^2 - \sum_{i=1}^{p-1} Y_i^2 = Z_p^2 \left(1 - \sum_{i=1}^{p-1} Z_i^2\right) = Z_p^2 \left(1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}\right), \text{ and}$$
$$Y_p = \pm Z_p \sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}.$$

Hence, the inverse transformation of h is

$$h^{-1}(\boldsymbol{Z}) = \left(Z_p Z_1, Z_p Z_2, \cdots, Z_p Z_{p-1}, \pm Z_p \sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}\right). \tag{3.28}$$

The Jacobian of the transformation is

$$|\boldsymbol{J}| = \begin{vmatrix} Z_p & 0 & \cdots & 0 & Z_1 \\ 0 & Z_p & \ddots & \vdots & Z_2 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & Z_p & Z_{p-1} \\ c^* Z_1 & c^* Z_2 & \cdots & c^* Z_{p-1} & \pm\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}} \end{vmatrix}, \tag{3.29}$$

where $c^* = \mp \frac{Z_p}{\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}}$. Equation (3.29) can be written in matrix form as

$$|\boldsymbol{J}| = \begin{vmatrix} Z_p \boldsymbol{I}_{p-1} & \boldsymbol{Z}_{(p)} \\ c^* \boldsymbol{Z}'_{(p)} & \pm\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}} \end{vmatrix}. \tag{3.30}$$

Using results from the determinant of a partitioned matrix, equation (3.30) can be simplified as

$$|\boldsymbol{J}| = |Z_p \boldsymbol{I}_{p-1}| \left| \pm\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}} - \left(\mp \frac{Z_p}{\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}} \frac{\boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}{Z_p}\right) \right|$$

$$= Z_p^{p-1} \left| \pm \left(\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}} + \frac{\boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}{\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}} \right) \right|$$
$$= \frac{Z_p^{p-1}}{\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}}.$$

To find the joint distribution of $Z_1,\ Z_2,\ \cdots,\ Z_p$, standard transformation techniques are used. Because the transformation in (3.28) is not monotone, the domain of h must be split into two parts, over which h is monotone in each part. Notate h in each particular domain as h_i. Define

$$h_1^{-1}(\boldsymbol{Z}) = \left(Z_pZ_1, Z_pZ_2, \cdots, Z_pZ_{p-1}, Z_p\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}} \right) \text{ and}$$
$$h_2^{-1}(\boldsymbol{Z}) = \left(Z_pZ_1, Z_pZ_2, \cdots, Z_pZ_{p-1}, -Z_p\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}} \right).$$

Note that $\left[h_i^{-1}(\boldsymbol{Z})\right]' h_i^{-1}(\boldsymbol{Z}) = Z_p^2$ for $i = 1$ or $i = 2$. So the distribution of $\boldsymbol{Z}$ is

$$f_{\boldsymbol{Z}}(\boldsymbol{Z}) = f_{\boldsymbol{Y}}\left[h_1^{-1}(\boldsymbol{Z})\right]|\boldsymbol{J}| + f_{\boldsymbol{Y}}\left[h_2^{-1}(\boldsymbol{Z})\right]|\boldsymbol{J}|$$
$$= 2Z_p^{p-1}g(Z_p^2)\frac{1}{\sqrt{1 - \boldsymbol{Z}'_{(p)}\boldsymbol{Z}_{(p)}}}.$$

Because the support for Z_p is $(0, \infty)$ and the support for $\boldsymbol{Z}_{(p)}$ is the interior of a $p-1$ dimensional ball, the joint distribution can be factored into two parts. It follows that

$$Z_p \perp\!\!\!\perp Z_1,\ Z_2,\ \cdots,\ Z_{p-1}. \tag{3.31}$$

Lastly, it follows from (3.31) that

$$\frac{\boldsymbol{Y}}{\sqrt{\boldsymbol{Y}'\boldsymbol{Y}}} \quad \perp\!\!\!\perp \quad \boldsymbol{Y}'\boldsymbol{Y}.$$

□

The next Lemma is a precursor to the proof for the derivation of Fisher's Information matrix (Theorem 3.7).

Lemma 3.6. *Suppose* $\boldsymbol{y}_k$ *and* $\boldsymbol{z}_k$ *are the* k^{th} *rows of the matrices* $\boldsymbol{Y}$ *and* $\boldsymbol{Z}$ *respectively, where* $\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}$*, and the marginal distributions of* $\boldsymbol{y}_k$ *and* $\boldsymbol{z}_k$ *are*

$$\begin{aligned}\boldsymbol{y}_k &\sim \boldsymbol{T}(\boldsymbol{X}_k\boldsymbol{\beta}, \frac{\xi}{\xi-2}\boldsymbol{\Sigma})\\ \boldsymbol{z}_k &\sim \boldsymbol{T}(\boldsymbol{0}, \frac{\xi}{\xi-2}\boldsymbol{\Sigma}).\end{aligned}$$

Note that the distribution of $\boldsymbol{z}_k$ *also can be written as*

$$\boldsymbol{z}_k \sim \frac{\boldsymbol{\Sigma}^{\frac{1}{2}}\mathring{\boldsymbol{n}}_k}{\sqrt{\frac{\mathring{\boldsymbol{c}}_k}{\xi}}}, \tag{3.32}$$

where $\mathring{\boldsymbol{n}}_k \sim N(\boldsymbol{0}, \boldsymbol{I}_p)$ *and* $\mathring{\boldsymbol{c}}_k \sim \chi^2_\xi$*. Then the following are true:*

(a) $E\left[\dfrac{\boldsymbol{z}_k'}{\xi + \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right] = \boldsymbol{0}$,

(b) $E\left[\dfrac{\boldsymbol{z}_k\boldsymbol{z}_k'}{(\xi + \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)^2}\right] = \dfrac{\boldsymbol{\Sigma}}{(p+\xi)(p+\xi+2)}$,

(c) $E\left[\dfrac{\boldsymbol{z}_i\boldsymbol{z}_k'}{(\xi + \boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i)(\xi + \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)}\right] = \boldsymbol{0}$.

Proof. A proof of (a) will first be given.

Because the expectation of the score function is zero, it follows that

$$\begin{aligned}\boldsymbol{0} = E\left[\frac{\partial \ell}{\partial \boldsymbol{\beta}'}\right] &= E\left[(\xi+p)\sum_{k=1}^{n}\frac{\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k}{\xi + \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right]\\ &= (\xi+p)\sum_{k=1}^{n}E\left[\frac{\boldsymbol{z}_k'}{\xi + \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k.\end{aligned} \tag{3.33}$$

Because each $\boldsymbol{z}_k$ is identically distributed, and the sum of the expectation in (3.33) is $\mathbf{0}$; then it follows that $E\left[\frac{\boldsymbol{z}_k'}{\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k = 0$. Finally, because $\boldsymbol{X}_k$ has full row rank, it follows that $E\left[\frac{\boldsymbol{z}_k'}{\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right] = \mathbf{0}$.

It follows from part (a) that

$$E\left[\frac{\boldsymbol{z}_i\boldsymbol{z}_k'}{(\xi+\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i)(\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)}\right] = E\left[\frac{\boldsymbol{z}_i}{\xi+\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i}\right]E\left[\frac{\boldsymbol{z}_k}{\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right] = \mathbf{0}.$$

Using the relationship in (3.32), (b) simplifies to

$$\begin{aligned}
E\left[\frac{\boldsymbol{z}_k\boldsymbol{z}_k'}{(\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)^2}\right] &= E\left[\frac{\frac{\xi}{\mathring{\boldsymbol{c}}_k}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathring{\boldsymbol{n}}_k\mathring{\boldsymbol{n}}_k'\boldsymbol{\Sigma}^{\frac{1}{2}}}{(\xi+\frac{\xi}{\mathring{\boldsymbol{c}}_k}\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k)^2}\right] \\
&= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}}E\left[\frac{\mathring{\boldsymbol{c}}_k\mathring{\boldsymbol{n}}_k\mathring{\boldsymbol{n}}_k'}{(\mathring{\boldsymbol{c}}_k+\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k)^2}\right]\boldsymbol{\Sigma}^{\frac{1}{2}} \\
&= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}}E\left[\frac{\mathring{\boldsymbol{c}}_k\mathring{\boldsymbol{n}}_k\mathring{\boldsymbol{n}}_k'}{(\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k)^2}\frac{(\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k)^2}{(\mathring{\boldsymbol{c}}_k+\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k)^2}\right]\boldsymbol{\Sigma}^{\frac{1}{2}} \\
&= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}}E\left[\frac{\mathring{\boldsymbol{n}}_k}{\sqrt{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}}\frac{\mathring{\boldsymbol{n}}_k'}{\sqrt{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}}\frac{\mathring{\boldsymbol{c}}_k}{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\left(\frac{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k+\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\right)^2\right]\boldsymbol{\Sigma}^{\frac{1}{2}}. \quad (3.34)
\end{aligned}$$

Define $\mathring{\boldsymbol{\ell}}_k \overset{\text{def}}{=} \frac{\mathring{\boldsymbol{n}}_k}{\sqrt{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}}$. Because $\mathring{\boldsymbol{\ell}}_k \perp\!\!\!\perp \mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k$ by Lemma 3.5 and $\mathring{\boldsymbol{n}}_k \perp\!\!\!\perp \mathring{\boldsymbol{c}}_k$, it follows that (3.34) simplifies to

$$\begin{aligned}
E\left[\frac{\boldsymbol{z}_k\boldsymbol{z}_k'}{(\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)^2}\right] &= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}}E\left[\frac{\mathring{\boldsymbol{n}}_k}{\sqrt{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}}\frac{\mathring{\boldsymbol{n}}_k'}{\sqrt{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}}\frac{\mathring{\boldsymbol{c}}_k}{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\left(\frac{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k+\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\right)^2\right]\boldsymbol{\Sigma}^{\frac{1}{2}} \\
&= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}}E\left[\mathring{\boldsymbol{\ell}}_k\mathring{\boldsymbol{\ell}}_k'\right]E\left[\frac{\mathring{\boldsymbol{c}}_k}{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\left(\frac{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k+\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\right)^2\right]\boldsymbol{\Sigma}^{\frac{1}{2}}. \quad (3.35)
\end{aligned}$$

Each of the expectations in (3.35) can easily be dealt with. Because $\mathring{\boldsymbol{\ell}}_k \perp\!\!\!\perp \mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k$, it follows that

$$E\left[\mathring{\boldsymbol{\ell}}_k\mathring{\boldsymbol{\ell}}_k'\right]E\left[\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k\right] = E\left[\mathring{\boldsymbol{\ell}}_k\mathring{\boldsymbol{\ell}}_k'\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k\right] = E\left[\frac{\mathring{\boldsymbol{n}}_k\mathring{\boldsymbol{n}}_k'}{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k\right] = E\left[\mathring{\boldsymbol{n}}_k\mathring{\boldsymbol{n}}_k'\right].$$

Because $\mathring{\boldsymbol{n}}_k \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k \sim \chi^2_p$, then

$$E\left[\mathring{\boldsymbol{\ell}}_k\mathring{\boldsymbol{\ell}}_k'\right] = \frac{E\left[\mathring{\boldsymbol{n}}_k\mathring{\boldsymbol{n}}_k'\right]}{E\left[\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k\right]} = \frac{\operatorname{Var}\left(\mathring{\boldsymbol{n}}_k\right)}{p} = \frac{1}{p}\boldsymbol{I}_p. \tag{3.36}$$

Applying Corollary 3.4 simplifies the second expectation in (3.35) to

$$\begin{aligned} E\left[\frac{\mathring{\boldsymbol{c}}_k}{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\left(\frac{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k + \mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\right)^2\right] &= \frac{\Gamma\left(\frac{p+\xi}{2}\right)\Gamma\left(\frac{p}{2}+1\right)\Gamma\left(\frac{\xi}{2}+1\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{\xi}{2}\right)\Gamma\left(\frac{p+\xi}{2}+2\right)} \\ &= \frac{p\xi}{(p+\xi)(p+\xi+2)}. \end{aligned} \tag{3.37}$$

Together, (3.36) and (3.37) can be used to simplify (3.35). The result is

$$\begin{aligned} E\left[\frac{\boldsymbol{z}_k\boldsymbol{z}_k'}{(\xi + \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)^2}\right] &= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}} E\left[\mathring{\boldsymbol{\ell}}_k\mathring{\boldsymbol{\ell}}_k'\right] E\left[\frac{\mathring{\boldsymbol{c}}_k}{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\left(\frac{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k + \mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\right)^2\right]\boldsymbol{\Sigma}^{\frac{1}{2}} \\ &= \frac{1}{\xi}\boldsymbol{\Sigma}^{\frac{1}{2}}\left[\frac{1}{p}\boldsymbol{I}_p\right]\left[\frac{p\xi}{(p+\xi)(p+\xi+2)}\right]\boldsymbol{\Sigma}^{\frac{1}{2}} = \frac{\boldsymbol{\Sigma}}{(p+\xi)(p+\xi+2)}. \end{aligned}$$

□

THEOREM 3.7. *Suppose the p-dimensional random vector, T, has a multivariate-T distribution with ξ degrees of freedom. Fisher's information matrix for T evaluated at $\boldsymbol{\mu} = \mathbf{0}$ is*

$$E\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = \left[\begin{matrix} E\left(\frac{\partial \ell}{\partial \boldsymbol{\beta}}\frac{\partial \ell}{\partial \boldsymbol{\beta}'}\right) & E\left(\frac{\partial \ell}{\partial \boldsymbol{\beta}}\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right) \\ E\left(\frac{\partial \ell}{\partial \boldsymbol{\zeta}}\frac{\partial \ell}{\partial \boldsymbol{\beta}'}\right) & E\left(\frac{\partial \ell}{\partial \boldsymbol{\zeta}}\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right) \end{matrix}\right]\Bigg|_{\boldsymbol{\mu}=\mathbf{0}} = \tfrac{\xi+p}{\xi+p+2}\left[\begin{matrix} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{X} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{F}^{(1)\prime}\boldsymbol{T}_{\boldsymbol{\Sigma}}\boldsymbol{F}^{(1)} \end{matrix}\right],$$

where $\boldsymbol{F}^{(1)}$ is given in Theorem 3.1, the derivatives are given in Theorem 3.2, and

$$\boldsymbol{T}_{\boldsymbol{\Sigma}} = \frac{n}{2}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \frac{1}{\xi+p}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)'\right].$$

Proof. The expectation of $\frac{\partial \ell}{\partial \boldsymbol{\beta}} \frac{\partial \ell}{\partial \boldsymbol{\beta}'}$ proceeds as follows:

$$E\left[\frac{\partial \ell}{\partial \boldsymbol{\beta}} \frac{\partial \ell}{\partial \boldsymbol{\beta}'}\right] = E\left\{\left[(\xi+p)\sum_{i=1}^{n}\frac{\boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i}{\xi+\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i}\right]\left[(\xi+p)\sum_{k=1}^{n}\frac{\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k}{\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}\right]\right\}$$
$$= (\xi+p)^2\sum_{i=1}^{n}\sum_{k=1}^{n}\boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}E\left[\frac{\boldsymbol{z}_i\boldsymbol{z}_k'}{(\xi+\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i)(\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k. \quad (3.38)$$

By using Lemma 3.6, equation (3.38) becomes

$$E\left[\frac{\partial \ell}{\partial \boldsymbol{\beta}} \frac{\partial \ell}{\partial \boldsymbol{\beta}'}\right] = (\xi+p)^2\sum_{k=1}^{n}\boldsymbol{X}_k'\boldsymbol{\Sigma}^{-1}E\left[\frac{\boldsymbol{z}_k\boldsymbol{z}_k'}{(\xi+\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)^2}\right]\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k$$
$$= (\xi+p)^2\sum_{k=1}^{n}\frac{\boldsymbol{X}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k}{(p+\xi)(p+\xi+2)}$$
$$= \frac{p+\xi}{p+\xi+2}\sum_{k=1}^{n}\boldsymbol{X}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k = \frac{p+\xi}{p+\xi+2}\sum_{k=1}^{n}(\boldsymbol{I}_p\otimes\boldsymbol{x}_k)\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_p\otimes\boldsymbol{x}_k')$$
$$= \frac{p+\xi}{p+\xi+2}\sum_{k=1}^{n}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{x}_k\boldsymbol{x}_k') = \frac{p+\xi}{p+\xi+2}\left(\boldsymbol{\Sigma}^{-1}\otimes\sum_{k=1}^{n}\boldsymbol{x}_k\boldsymbol{x}_k'\right)$$
$$= \frac{p+\xi}{p+\xi+2}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{X}'\boldsymbol{X}\right).$$

The computation for the remaining parts of the information matrix are similar to that derived above. The proofs are not included. □

To solve the M-estimating equations in (3.22) for a given gradient vector and Information matrix, a modified algorithm using (3.25) was developed based on the algorithm in Boik (2002a). The modified Fisher-Scoring algorithm has the following steps:

Step 1. Compute initial guesses for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}$. Denote these guesses by $\hat{\boldsymbol{\Gamma}}_{0,0}$ and $\hat{\boldsymbol{\Lambda}}_0$.

Step 2. Denote the maximum likelihood estimate of $\boldsymbol{\Sigma}$ after the i^{th} iteration by

$$\hat{\boldsymbol{\Sigma}}_i = \hat{\boldsymbol{\Gamma}}_{0,i}\hat{\boldsymbol{\Lambda}}_i\hat{\boldsymbol{\Gamma}}_{0,i}'.$$

Step 3. Set $\boldsymbol{\mu}_i = \mathbf{0}$, then $\hat{\boldsymbol{\theta}}_i = (\begin{array}{ccc} \hat{\boldsymbol{\beta}}_i' & \mathbf{0}' & \hat{\boldsymbol{\varphi}}_i' \end{array})'$. Update $\hat{\boldsymbol{\theta}}_i$ as

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i + \alpha \hat{\boldsymbol{I}}_{\hat{\boldsymbol{\theta}}_i} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i},$$

where $\hat{\ell}_i(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}; \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\Gamma}}_{0,i}, \hat{\boldsymbol{\Lambda}}_i)$, $\boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i} = \left.\frac{\partial \hat{\ell}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}$, $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i} = -E\left[\left.\frac{\partial^2 \hat{\ell}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}\right]$, and $\alpha \in (0, 1]$.

Step 4. Compute $\boldsymbol{G}(\hat{\boldsymbol{\mu}}_{i+1})$ following the algorithm from the last section of Chapter 2 and set $\hat{\boldsymbol{\Lambda}}_{i+1} = \boldsymbol{\Lambda}(\hat{\boldsymbol{\varphi}}_{i+1})$.

Step 5. Set $\hat{\boldsymbol{\Gamma}}_{0,i+1} = \hat{\boldsymbol{\Gamma}}_{0,i} \boldsymbol{G}(\hat{\boldsymbol{\mu}}_{i+1})$.

Step 6. Iterate steps 2–5 until convergence.

S-estimators

Another type of robust estimator is Rousseeuw and Yohai's S-estimators, which were introduced in the setting of robust regression (Rousseeuw and Yohai 1984). A desirable property of S-estimators is that they have high breakdown point. The largest asymptotic breakdown point of S-estimators is 50% (Rousseeuw and Yohai 1984). If the maximum breakdown point is used however, then the asymptotic efficiency of the estimators is low. A generalization of Rousseeuw and Yohai's S-estimator in the setting of multivariate location and scatter is defined by Davies (1987) and Lopuhaä (1989). They defined S-estimators of multivariate location and scatter, $\theta_n = (\boldsymbol{\tau}_n, \boldsymbol{\Sigma}_n)$, as the minimizers of $\det(\mathbf{\Sigma})$ among all $(\boldsymbol{\tau}, \mathbf{\Sigma})$ in the parameter space subject to

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left[\left\{(\boldsymbol{y}_i-\boldsymbol{\tau})'\mathbf{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{\tau})\right\}^{\frac{1}{2}}\right]=b_0, \tag{3.39}$$

where b_0 is a constant satisfying $0 < b_0 < \sup\rho$, and ρ is a function chosen to down-weight extreme values. The choice of $\rho(t) = t^2$ leads to least squares estimates. In order to maintain asymptotic normality yet have robust properties, Rousseeuw and Yohai (1984) required the function ρ to have the following properties:

1. ρ is symmetric, has a continuous derivative, and $\rho(0) = 0$.
2. There exists a finite constant, $c_0 > 0$ such that ρ is strictly increasing on $[0, c_0]$ and constant on $[c_0, \infty)$.

Various choices can be made for ρ. An example is

$$\rho(y, c_0) = \begin{cases} \dfrac{y^2}{2} - \dfrac{y^4}{2c_0^2} + \dfrac{y^6}{6c_0^4}, & \text{if } |y| < c_0 \\ \dfrac{c_0^2}{6}, & \text{if } |y| \geqslant c_0 \end{cases},$$

which is an integral of Tukey's biweight function. Typically the constant b_0 in (3.39) is chosen assuming an underlying distribution for the model. When the family of elliptically contoured distributions is assumed, an appropriate choice for the constant would be $b_0 = E\left[\rho(d_k)\right]$ (Lopuhaä 1989). In this case, the constant c_0 can be chosen such that $0 < \frac{b_0}{\sup\ \rho} = r \leqslant \frac{n-p}{2n}$, which gives the finite breakdown point of $\epsilon_n^* = \frac{\lceil nr \rceil}{n}$, where $\lceil y \rceil$ indicates the smallest integer greater than or equal to y (Lopuhaä 1989). If $r = \frac{n-p}{2n}$, then the largest breakdown point of $\lfloor\frac{n-p+1}{2}\rfloor/n$ is obtained (Lopuhaä 1989), which is asymptotically equal to $\frac{1}{2}$. The constant c_0 affects the asymptotic efficiency of the estimators, as well as influencing the breakdown point. For a breakdown point near 50%, the asymptotic efficiency is low. Hence, it is not possible to achieve a 50% breakdown point and high asymptotic efficiency at the same time.

A generalization of the definition in (3.39) to the linear model and spectral model defined in (1.2) and (2.5) is the minimization of the loss function

$$w(\boldsymbol{\theta}) = n \log |\boldsymbol{\Sigma}| \text{ subject to } q(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{n} \rho(d_k) - b_0 = 0, \tag{3.40}$$

where $d_k = \left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})\right]^{\frac{1}{2}}$, $\boldsymbol{X}_k$ is given in (1.4), and $\boldsymbol{\beta}$ is defined in (1.5).

<u>Implicit Model for S-estimators</u>

The constraint in (3.40) can be accommodated by the usual theory of Lagrange Multipliers (Nocedal and Wright 1999, pg. 321). Second-order asymptotic distribution theory however, is made more complicated by introducing Lagrange parameters. Instead, if $w(\boldsymbol{\theta})$ is defined through an implicit parameter, then unconstrained optimization can be used; and asymptotic distribution results can easily be obtained. An implicit parameter for (3.40) is defined through the eigenvalue parameters, $\boldsymbol{\varphi}$. This implicit model uses a specific matrix $\boldsymbol{V}$. Properties of this matrix are given in Theorem 3.8.

THEOREM 3.8. *Let $\boldsymbol{V}$ be any $\nu_3 \times \nu_3$ invertible matrix that satisfies*

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{v}_2 \end{bmatrix} \text{ and } \boldsymbol{V}_1'\boldsymbol{v}_2 = \boldsymbol{0},$$

where $\boldsymbol{V}_1$ and $\boldsymbol{v}_2$ are $\nu_3 \times \nu_3 - 1$ and $\nu_3 \times 1$ matrices, respectively. Then *(a)* *The inverse of* $\boldsymbol{V}$ *can be specified in terms of* $\boldsymbol{V}_1$ *and* $\boldsymbol{v}_2$. *Let*

$$\boldsymbol{V}^* = \begin{bmatrix} (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\,\boldsymbol{V}_1' \\ (\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1}\,\boldsymbol{v}_2' \end{bmatrix}.$$

Then $\boldsymbol{V}^* = \boldsymbol{V}^{-1}$.

(b) $\text{ppo}(\boldsymbol{V}_1) + \text{ppo}(\boldsymbol{v}_2) = \boldsymbol{I}_{\nu_3}$,

where ppo$(\cdot)$ *is the perpendicular projection operator onto the column space of* $(\cdot)$.

PROOF. (a) The result can be shown by multiplying $\boldsymbol{V}^*\boldsymbol{V} = \boldsymbol{I}_{\nu_3}$. Accordingly,

$$\begin{aligned} \boldsymbol{V}^*\boldsymbol{V} &= \begin{bmatrix} (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\,\boldsymbol{V}_1' \\ (\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1}\,\boldsymbol{v}_2' \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{v}_2 \end{bmatrix} \\ &= \begin{bmatrix} (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\,\boldsymbol{V}_1'\boldsymbol{V}_1 & (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\,\boldsymbol{V}_1'\boldsymbol{v}_2 \\ (\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1}\,\boldsymbol{v}_2'\boldsymbol{V}_1 & (\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1}\,\boldsymbol{v}_2'\boldsymbol{v}_2 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I}_{\nu_3-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_1 \end{bmatrix} \\ &= \boldsymbol{I}_{\nu_3}. \end{aligned}$$

Hence $\boldsymbol{V}^* = \boldsymbol{V}^{-1}$.

(b) Multiplying $\boldsymbol{V}\boldsymbol{V}^{-1}$ yields

$$\begin{aligned} \boldsymbol{I}_{\nu_3} &= \boldsymbol{V}\boldsymbol{V}^{-1} \\ &= \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{v}_2 \end{bmatrix} \begin{bmatrix} (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\,\boldsymbol{V}_1' \\ (\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1}\,\boldsymbol{v}_2' \end{bmatrix} \\ &= \boldsymbol{V}_1\,(\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\,\boldsymbol{V}_1' + \boldsymbol{v}_2\,(\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1}\,\boldsymbol{v}_2' \\ &= \text{ppo}(\boldsymbol{V}_1) + \text{ppo}(\boldsymbol{v}_2). \end{aligned}$$

□

The model for the eigenvalue parameters is defined through the trivial equality $\boldsymbol{\varphi} = \boldsymbol{I}_{\nu_3}\boldsymbol{\varphi}$. Hence,

$$\begin{aligned}\boldsymbol{\varphi} = \boldsymbol{I}_{\nu_3}\boldsymbol{\varphi} &= \boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{\varphi} = \text{ppo}(\boldsymbol{V}_1)\boldsymbol{\varphi} + \text{ppo}(\boldsymbol{v}_2)\boldsymbol{\varphi} \\ &= \boldsymbol{V}_1 \underbrace{\left(\boldsymbol{V}_1'\boldsymbol{V}_1\right)^{-1}\boldsymbol{V}_1'\boldsymbol{\varphi}}_{\boldsymbol{\psi}} + \boldsymbol{v}_2 \underbrace{\left(\boldsymbol{v}_2'\boldsymbol{v}_2\right)^{-1}\boldsymbol{v}_2'\boldsymbol{\varphi}}_{\eta_\psi} \\ \boldsymbol{\varphi} &= \boldsymbol{V}_1\boldsymbol{\psi} + \boldsymbol{v}_2\eta_\psi, \end{aligned} \tag{3.41}$$

where $\boldsymbol{\psi}$ contains explicit parameters, and η_ψ is an implicit parameter defined through the constraint q in (3.40). The manner in which $\boldsymbol{V}_1$ and $\boldsymbol{v}_2$ are chosen is described in Theorem 3.9. The number of explicit parameters has decreased by one because η_ψ is an implicit parameter. The dimension is reduced because the constraint in (3.39) must be satisfied. Hence, the vector of explicit parameters for S-estimators can be written as

$$\underset{\nu-1\times 1}{\boldsymbol{\theta}_{s_1}} = \begin{bmatrix} \underset{\nu_1\times 1}{\boldsymbol{\beta}} \\ \underset{\nu_2\times 1}{\boldsymbol{\mu}} \\ \underset{\nu_3-1\times 1}{\boldsymbol{\psi}} \end{bmatrix}, \tag{3.42}$$

where $\nu_1 + \nu_2 + \nu_3 = \nu$. The S-estimating equation in (3.40) can then be rewritten in terms of the implicit parameter as minimizing the equation

$$w(\boldsymbol{\theta}_{s_1}), \tag{3.43}$$

where w is defined in (3.40), η_ψ is an implicit function of the explicit parameters defined through the constraint $E\left[\rho(d_k)\right] = b_0$, $\boldsymbol{X}_k$ is given in (1.4), $\boldsymbol{\beta}$ is defined in (1.5), and $d_k = \left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}\right)\right]^{\frac{1}{2}}$.

First Implicit Derivatives

Derivatives of $w(\boldsymbol{\theta}_{s_1})$ and $q(\boldsymbol{\theta}_{s_1})$ are required to find the S-estimates of $\boldsymbol{\theta}_{s_1}$ under the linear model in (1.2), the spectral model in (2.5), and model (3.41) for the eigenvalue parameters. Not all expressions for the required derivatives will be justified in this dissertation; however, expressions for all the derivatives can be found in the appendix. The derivatives that are obtained will illustrate the method by which remaining derivatives also can be obtained. In addition, numerical approximations to the derivatives were used to verify the accuracy of the analytical expressions for all the derivatives used in the dissertation.

Derivatives of $w(\boldsymbol{\theta}_{s_1})$ and $q(\boldsymbol{\theta}_{s_1})$ involving the model (3.41) require implicit derivatives (derivatives involving $\boldsymbol{\varphi}$ and η_ψ). If $\frac{\partial q}{\partial \eta_\psi} \neq 0$, then the implicit function theorem (Fulks 1979, pg. 352) guarantees the existence of the required derivatives. As a result, an implicit function exists relating η_ψ to the explicit parameters, $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$:

$$\eta_\psi = \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi}).$$

Two parameters are functionally dependent if one parameter is a function of the other parameter. If no such function exists, then the two parameters are functionally independent. To be clear on notation in the expressions of the derivatives, define

$$q^* \stackrel{\text{def}}{=} q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi}, \eta_\psi), \tag{3.44}$$

where η_ψ is treated as functionally independent of the other variables and $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ are functionally independent. Note that for some derivatives $\boldsymbol{\varphi}$ also is involved. In these cases, q^* is defined as

$$q^* \stackrel{\text{def}}{=} q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi}(\boldsymbol{\psi}, \eta_\psi)). \tag{3.45}$$

In both (3.44) and (3.45), η_ψ is treated as functionally independent of the other variables and $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ are functionally independent.

Further, define

$$q \stackrel{\text{def}}{=} q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})), \tag{3.46}$$

where η_ψ is a function of $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$, but $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ are functionally independent. Note that, as in (3.44), $\boldsymbol{\varphi}$ is involved in some of the derivatives. In these cases, q is defined as

$$q \stackrel{\text{def}}{=} q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})\}),$$

where η_ψ is a function of $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$, but $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ are functionally independent. Now the implicit derivatives can be obtained.

Theorem 3.9 (First Implicit Derivatives of η_ψ). *The first implicit derivatives of η_ψ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ are*

$$\begin{aligned}
\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}'} &= -\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \frac{\partial q^*}{\partial \boldsymbol{\beta}'}, \\
\frac{\partial \eta_\psi}{\partial \boldsymbol{\mu}'} &= -\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \frac{\partial q^*}{\partial \boldsymbol{\mu}'}, \quad \textit{and} \\
\frac{\partial \eta_\psi}{\partial \boldsymbol{\psi}'} &= -\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \boldsymbol{a}'\boldsymbol{V}_1,
\end{aligned}$$

provided that $\boldsymbol{a}'\boldsymbol{v}_2 \neq 0$*, where* $\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}$ *and* $\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$ *are given in Theorem 2.6,* $\boldsymbol{Y}$ *and* $\boldsymbol{X}$ *are given in* (1.2)*,*

$$\begin{aligned}
\rho^{(1)}(d_k) &= \frac{\partial \rho(d_k)}{\partial d_k}, \\
w_k^{(1)} &= \frac{1}{2}\frac{\rho^{(1)}(d_k)}{d_k}, \\
\boldsymbol{W}_1 &= \frac{1}{n}\,\mathrm{Diag}\left(w_k^{(1)}\right), \\
\boldsymbol{Z} &= \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}, \\
\boldsymbol{B} &= \mathrm{dvec}\big(\boldsymbol{\beta}, p, d\big), \\
\boldsymbol{W}_{xz} &= \boldsymbol{X}'\boldsymbol{W}_1\boldsymbol{Z}, \\
\boldsymbol{W}_{zz} &= \boldsymbol{Z}'\boldsymbol{W}_1\boldsymbol{Z}, \\
\frac{\partial q^*}{\partial \boldsymbol{\beta}} &= -2\,\mathrm{vec}\big(\boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\big), \\
\frac{\partial q^*}{\partial \boldsymbol{\mu}} &= -\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\,\mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big), \\
\boldsymbol{a} = \frac{\partial q^*}{\partial \boldsymbol{\varphi}} &= -\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\,\mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big).
\end{aligned}$$

PROOF. The proof for $\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}'}$ will given here. Expressions without proof for all other derivatives can be found in the appendix. Methods for obtaining the remaining derivatives are similar to those given below.

The implicit parameter, η_ψ, is defined by the constraint in (3.40). Using the notation in (3.46), the constraint in (3.40) can be written as

$$q = q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})) = 0. \tag{3.47}$$

Taking the derivative of (3.47) (using the total derivative) yields

$$\frac{\partial q}{\partial \boldsymbol{\beta}'} = \frac{\partial q^*}{\partial \boldsymbol{\beta}'} + \frac{\partial q^*}{\partial \eta_\psi}\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}'} = 0,$$

for all $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ in their respective parameter spaces. It follows that

$$\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}'} = -\left(\frac{\partial q^*}{\partial \eta_\psi}\right)^{-1}\frac{\partial q^*}{\partial \boldsymbol{\beta}'}.$$

Because q^* depends on η_ψ solely through $\boldsymbol{\varphi}$ in (3.45) and (3.41), it follows that

$$\frac{\partial q^*}{\partial \eta_\psi} = \frac{\partial q^*}{\partial \boldsymbol{\varphi}'}\frac{\partial \boldsymbol{\varphi}}{\partial \eta_\psi} = \boldsymbol{a}'\boldsymbol{v}_2.$$

Hence,

$$\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}'} = -\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1}\frac{\partial q^*}{\partial \boldsymbol{\beta}'}.$$

Note that the only requirement for $\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}'}$ to exist is that the inverse of $\frac{\partial q^*}{\partial \eta_\psi}$ exists. Considering that $\frac{\partial q^*}{\partial \eta_\psi}$ is a scalar, $\boldsymbol{v}_2$ must be chosen so that $\boldsymbol{a}'\boldsymbol{v}_2 \neq 0$. Any choice for $\boldsymbol{v}_2$ will do so long as $\boldsymbol{a}'\boldsymbol{v}_2 \neq 0$. Because $\boldsymbol{v}_2$ can be chosen as any vector, to ensure that $\boldsymbol{a}'\boldsymbol{v}_2 \neq 0$, $\boldsymbol{v}_2$ is chosen as

$$\boldsymbol{v}_2 = \frac{\boldsymbol{a}}{\sqrt{\boldsymbol{a}'\boldsymbol{a}}}. \tag{3.48}$$

Note that $||\boldsymbol{v}_2|| = 1$. Because $\boldsymbol{V}_1'\boldsymbol{v}_2 = \boldsymbol{0}$, it follows that $\boldsymbol{V}_1$ can be chosen to be any $\nu_3 \times \nu_3 - 1$ matrix whose columns span the null space of $\boldsymbol{v}_2'$. One explicit choice is

$$\boldsymbol{V}_1 = \left[\boldsymbol{I}_{\nu_3} - \boldsymbol{v}_2\left(\boldsymbol{v}_2'\boldsymbol{v}_2\right)^{-1}\boldsymbol{v}_2'\right]\boldsymbol{V}_0,$$

where $\boldsymbol{V}_0$ is a $\nu_3 \times \nu_3 - 1$ full column rank matrix whose columns are not in the column space of $\boldsymbol{v}_2$. Further note that for this choice for $\boldsymbol{v}_2$, $\boldsymbol{a}'\boldsymbol{V}_1 = \boldsymbol{0}$.

The derivative of q with respect to $\boldsymbol{\beta}'$ when η_ψ is constant requires the derivative of the square of d_k. Define $h_k = d_k^2$. Because $\boldsymbol{X}_k = (\boldsymbol{I}_p \otimes \boldsymbol{x}_k')$ by (1.4), it follows that (3.7) can be written as

$$
\begin{aligned}
\frac{\partial h_k}{\partial \boldsymbol{\beta}'} &= -2\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k = -2\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{I}_p \otimes \boldsymbol{x}_k'\right) \\
&= -2\left(\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k'\right) \\
&= -2\left[\operatorname{vec}\left(\boldsymbol{x}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)\right]'.
\end{aligned}
$$

The derivative $\frac{\partial q^*}{\partial \boldsymbol{\beta}}$, therefore can be expressed as

$$
\begin{aligned}
\frac{\partial q^*}{\partial \boldsymbol{\beta}} &= \frac{1}{n}\sum_{k=1}^{n}\frac{\partial \rho(d_k)}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial h_k}{\partial \boldsymbol{\beta}}\underbrace{\frac{\partial \rho(d_k)}{\partial h_k}}_{w_k^{(1)}} = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial h_k}{\partial \boldsymbol{\beta}}w_k^{(1)} \\
&= \frac{1}{n}\sum_{k=1}^{n} -2\operatorname{vec}\left(\boldsymbol{x}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)w_k^{(1)} = -\frac{2}{n}\sum_{k=1}^{n}\operatorname{vec}\left(\boldsymbol{x}_k w_k^{(1)}\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right) \\
&= -2\operatorname{vec}\left(\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{x}_k w_k^{(1)}\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right) = -2\operatorname{vec}\left(\boldsymbol{X}'\boldsymbol{W}_1\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\right) \\
&= -2\operatorname{vec}\left(\boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right).
\end{aligned}
$$

To finish the proof of $\frac{\partial \eta_\psi}{\partial \boldsymbol{\beta}}$, $w_k^{(1)}$ can be expressed as

$$
w_k^{(1)} = \frac{\partial \rho(d_k)}{\partial h_k} = \frac{\partial \rho(d_k)}{\partial d_k}\frac{\partial d_k}{\partial h_k} = \rho^{(1)}(d_k)\frac{\partial h_k^{\frac{1}{2}}}{\partial h_k} = \rho^{(1)}(d_k)\frac{1}{2}h_k^{-\frac{1}{2}} = \frac{1}{2}\frac{\rho^{(1)}(d_k)}{d_k}.
$$

The proof is similar for the other two derivatives. □

THEOREM 3.10 (FIRST IMPLICIT DERIVATIVES OF $\boldsymbol{\varphi}$). *The first implicit derivatives of $\boldsymbol{\varphi}$ evaluated at $\boldsymbol{\mu} = \boldsymbol{0}$ are*

$$
\boldsymbol{P}_{\boldsymbol{\beta}}^{(1)} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\beta}'} = -\boldsymbol{v}_2\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1}\frac{\partial q^*}{\partial \boldsymbol{\beta}'},
$$

$$\boldsymbol{P}_{\boldsymbol{\mu}}^{(1)} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{v}_2 \left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \frac{\partial q^*}{\partial \boldsymbol{\mu}'}, \text{ and}$$

$$\boldsymbol{P}_{\boldsymbol{\psi}}^{(1)} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} = \left[\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \boldsymbol{a}' \otimes \boldsymbol{I}_{\nu_3}\right] (2\boldsymbol{N}_{\nu_3}^{\perp}) \left(\boldsymbol{v}_2 \otimes \boldsymbol{V}_1\right),$$

where $\boldsymbol{a}$ *is defined in Theorem 3.9 and* $\boldsymbol{N}_{\nu_3}^{\perp} = \frac{1}{2}\left[\boldsymbol{I}_{\nu_3} - \boldsymbol{I}_{(\nu_3,\nu_3)}\right]$.

PROOF. Using model (3.41), the derivative of $\boldsymbol{\varphi}$ with respect to $\boldsymbol{\beta}'$ is

$$\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\beta}'} = \frac{\partial}{\partial \boldsymbol{\beta}'} \left[\boldsymbol{V}_1 \boldsymbol{\psi} + \boldsymbol{v}_2 \eta_{\psi}\right] = \boldsymbol{V}_1 \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\beta}'} + \boldsymbol{v}_2 \frac{\partial \eta_{\psi}}{\partial \boldsymbol{\beta}'}.$$

Because $\boldsymbol{\psi}$ and $\boldsymbol{\beta}$ are functionally independent, it follows that $\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\beta}'} = \boldsymbol{0}$. By Theorem 3.9,

$$\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\beta}'} = \boldsymbol{v}_2 \frac{\partial \eta_{\psi}}{\partial \boldsymbol{\beta}'} = -\boldsymbol{v}_2 \left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \frac{\partial q^*}{\partial \boldsymbol{\beta}'}.$$

The proof is similar for the derivative with respect to $\boldsymbol{\mu}$.

Using model (3.41), the derivative of $\boldsymbol{\varphi}$ with respect to $\boldsymbol{\psi}'$ is

$$\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} = \frac{\partial}{\partial \boldsymbol{\psi}'} \left[\boldsymbol{V}_1 \boldsymbol{\psi} + \boldsymbol{v}_2 \eta_{\psi}\right] = \boldsymbol{V}_1 + \boldsymbol{v}_2 \frac{\partial \eta_{\psi}}{\partial \boldsymbol{\psi}'}.$$

By Theorem 3.9,

$$\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} = \boldsymbol{V}_1 + \boldsymbol{v}_2 \frac{\partial \eta_{\psi}}{\partial \boldsymbol{\psi}'} = \boldsymbol{V}_1 - \boldsymbol{v}_2 \left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \boldsymbol{a}'\boldsymbol{V}_1 \tag{3.49}$$

The derivative in (3.49) can be simplified as follows:

$$\begin{aligned}\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} &= \boldsymbol{V}_1 - \boldsymbol{v}_2 \left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \boldsymbol{a}'\boldsymbol{V}_1 \\ &= \left[\left(\boldsymbol{a}'\boldsymbol{v}_2\right)^{-1} \otimes \boldsymbol{I}_{\nu_3}\right] \left[\boldsymbol{V}_1 \left(\boldsymbol{a}'\boldsymbol{v}_2 \otimes \boldsymbol{I}_{\nu_3 - 1}\right) - \boldsymbol{v}_2 \boldsymbol{a}'\boldsymbol{V}_1\right] \end{aligned} \tag{3.50}$$

Using the Kronecker properties properties $\boldsymbol{a}\boldsymbol{b}' = \boldsymbol{b}' \otimes \boldsymbol{a}$ and $\boldsymbol{A}\,(\boldsymbol{b} \otimes \boldsymbol{B}) = (\boldsymbol{b} \otimes \boldsymbol{A}\boldsymbol{B})$, (3.50) simplifies to

$$
\begin{aligned}
\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} &= \left[\,(\boldsymbol{a}'\boldsymbol{v}_2)^{-1} \otimes \boldsymbol{I}_{\nu_3}\right]\left[\,(\boldsymbol{a}'\boldsymbol{v}_2 \otimes \boldsymbol{V}_1) - (\boldsymbol{a}'\boldsymbol{V}_1 \otimes \boldsymbol{v}_2)\,\right] \\
&= \left[\,(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}\,\boldsymbol{a}' \otimes \boldsymbol{I}_{\nu_3}\right]\left[\,(\boldsymbol{v}_2 \otimes \boldsymbol{V}_1) - (\boldsymbol{V}_1 \otimes \boldsymbol{v}_2)\,\right] \\
&= \left[\,(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}\,\boldsymbol{a}' \otimes \boldsymbol{I}_{\nu_3}\right]\left[\,(\boldsymbol{v}_2 \otimes \boldsymbol{V}_1) - \boldsymbol{I}_{(\nu_3,\nu_3)}\,(\boldsymbol{v}_2 \otimes \boldsymbol{V}_1)\,\right] \\
&= \left[\,(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}\,\boldsymbol{a}' \otimes \boldsymbol{I}_{\nu_3}\right](2\boldsymbol{N}_{\nu_3}^{\perp})\,(\boldsymbol{v}_2 \otimes \boldsymbol{V}_1)\,,
\end{aligned}
$$

where $\boldsymbol{N}_{\nu_3}^{\perp} = \frac{1}{2}\left[\boldsymbol{I}_{\nu_3} - \boldsymbol{I}_{(\nu_3,\nu_3)}\right]$ is the perpendicular projection operator onto the space of skew-symmetric matrices. $\boldsymbol{N}_{\nu_3}^{\perp}$ is also the perpendicular projection operator onto the null space of symmetric matrices. □

Second Implicit Derivatives

The second implicit derivatives are needed for estimating the parameters using the Newton-Raphson iterative method. The implicit derivatives of η_ψ with respect to the parameters are given in Theorem 3.11.

THEOREM 3.11 (SECOND IMPLICIT DERIVATIVES OF $\boldsymbol{\varphi}$). *The second implicit derivatives of $\boldsymbol{\varphi}$ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ are*

$$
\begin{aligned}
\boldsymbol{P}_{\boldsymbol{\beta\beta}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} &= -\Bigg[\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{P}_{\boldsymbol{\beta}}^{(1)} + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'}\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \\
&\qquad + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'}\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\beta}}^{(1)}\Bigg]\,(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{r}')\,, \\
\boldsymbol{P}_{\boldsymbol{\mu\mu}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} &= -\Bigg[\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{P}_{\boldsymbol{\mu}}^{(1)} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'}\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \\
&\qquad + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'}\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\mu}}^{(1)}\Bigg]\,(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{r}')\,,
\end{aligned}
$$

$$\boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}} = -\left[\boldsymbol{P}_{\boldsymbol{\psi}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)}\right] (\boldsymbol{I}_{\nu_3 - 1} \otimes \boldsymbol{r}'),$$

$$\boldsymbol{P}_{\boldsymbol{\beta}\boldsymbol{\mu}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)}\right] (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{r}'),$$

$$\boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\mu}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\mu}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right] \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)} (\boldsymbol{I}_{\nu_3 - 1} \otimes \boldsymbol{r}'), \text{ and}$$

$$\boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\beta}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\beta}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right] \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)} (\boldsymbol{I}_{\nu_3 - 1} \otimes \boldsymbol{r}'),$$

where $\boldsymbol{r} = \boldsymbol{v}_2(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}$, $\boldsymbol{P}_{\cdot}^{(1)}$ *is given in Theorem 3.10,* $\boldsymbol{Y}$ *and* $\boldsymbol{X}$ *are given in* (1.2), $\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}$ *and* $\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}$ *are given in Theorem 2.6,* $\boldsymbol{F}_{\cdot\cdot}^{(11)}$ *is given in Appendix A,*

$$\boldsymbol{K}_{xx} = \left(\boldsymbol{X}' * \boldsymbol{X}'\right),$$

$$\boldsymbol{K}_{zx} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\right),$$

$$\boldsymbol{K}_{zz} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\right),$$

$$w_k^{(1)} = \frac{\partial \rho(d_k)}{\partial h_k} = \frac{1}{2}\frac{\rho^{(1)}(d_k)}{d_k},$$

$$w_k^{(2)} = \frac{\partial w_k^{(1)}}{\partial h_k} = \frac{1}{4}\left[\frac{\rho^{(2)}(d_k)}{d_k^2} - \frac{\rho^{(1)}(d_k)}{d_k^3}\right]$$

$$\boldsymbol{W}_i = \frac{1}{n}\,\mathrm{Diag}\left(w_k^{(i)}\right),$$

$$\boldsymbol{W}_{xx} = \boldsymbol{X}'\boldsymbol{W}_1\boldsymbol{X},$$

$$\boldsymbol{W}_{xz} = \boldsymbol{X}'\boldsymbol{W}_1\boldsymbol{Z} = \boldsymbol{W}_{zx}',$$

$$\boldsymbol{W}_{zz} = \boldsymbol{Z}'\boldsymbol{W}_1\boldsymbol{Z},$$

$$\boldsymbol{V}_{xx} = \boldsymbol{K}_{zx}\boldsymbol{W}_2\boldsymbol{K}_{zx}',$$

$$\boldsymbol{V}_{xz} = \boldsymbol{K}_{zx}\boldsymbol{W}_2\boldsymbol{K}_{zz}',$$

$$\boldsymbol{V}_{zz} = \boldsymbol{K}_{zz}\boldsymbol{W}_2\boldsymbol{K}_{zz}',$$

$$\boldsymbol{M}_{xx} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xx}\right) + 2\boldsymbol{V}_{xx},$$

$$\boldsymbol{M}_{xz} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{V}_{xz} = \boldsymbol{M}'_{zx},$$

$$\boldsymbol{M}_{zz} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right) + \frac{1}{2}\boldsymbol{V}_{zz}.$$

The derivatives of q^* *evaluated at* $\boldsymbol{\mu} = \mathbf{0}$ *are*

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xx},$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}},$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\boldsymbol{M}_{zx}$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}},$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zx},$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\mu}\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_2} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right)\right) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}},$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\mu}\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_2} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right)\right) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}},$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right)\right) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}, \text{ and}$$
$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right)\right) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}.$$

PROOF. The proof for $\frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}}$ will given here. Expressions without proof for all other derivatives can be found in Appendix C. Methods for obtaining the remaining derivatives are similar to those given below.

This theorem requires the use of the product rule for matrices with respect to the transpose of a vector. Suppose that $\boldsymbol{A}$ is a $r \times s$ matrix and $\boldsymbol{B}$ is a $s \times t$ matrix.

Further, suppose that the entries in $\boldsymbol{A}$ and $\boldsymbol{B}$ are functionally dependent on the $u \times 1$ vector $\boldsymbol{y}$. The derivative of the product $\boldsymbol{AB}$ with respect to the vector $\boldsymbol{y}'$ is

$$\frac{\partial \boldsymbol{AB}}{\partial \boldsymbol{y}'} = \frac{\partial \boldsymbol{A}}{\partial \boldsymbol{y}'}(\boldsymbol{I}_u \otimes \boldsymbol{B}) + \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial \boldsymbol{y}'}. \tag{3.51}$$

As in Theorem 3.10, the implicit parameter, η_ψ, is defined by the constraint in (3.40). Using the notation in (3.46), the constraint can be written as

$$q = q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})) = 0. \tag{3.52}$$

Taking the derivative of (3.52) (using the total derivative) with respect to $\boldsymbol{\mu}'$ yields

$$\frac{\partial q}{\partial \boldsymbol{\mu}'} = \frac{\partial q^*}{\partial \boldsymbol{\mu}'} + \frac{\partial q^*}{\partial \boldsymbol{\varphi}'}\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\mu}'} = \mathbf{0}, \tag{3.53}$$

for all $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ in their parameter spaces. The transpose of (3.53) yields

$$\frac{\partial q}{\partial \boldsymbol{\mu}} = \frac{\partial q^*}{\partial \boldsymbol{\mu}} + \frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\frac{\partial q^*}{\partial \boldsymbol{\varphi}} = \mathbf{0}. \tag{3.54}$$

Taking the derivative of (3.54) with respect to $\boldsymbol{\beta}'$ yields

$$\frac{\partial^2 q}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial}{\partial \boldsymbol{\beta}'}\left[\frac{\partial q^*}{\partial \boldsymbol{\mu}}\right] + \frac{\partial}{\partial \boldsymbol{\beta}'}\left[\frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\frac{\partial q}{\partial \boldsymbol{\varphi}'}\right] = \mathbf{0}. \tag{3.55}$$

Because $\frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}} = \frac{\partial \eta_\psi}{\partial \boldsymbol{\mu}}\boldsymbol{v}_2'$, (3.55) can be written as

$$\frac{\partial^2 q}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial}{\partial \boldsymbol{\beta}'}\left[\frac{\partial q^*}{\partial \boldsymbol{\mu}}\right] + \frac{\partial}{\partial \boldsymbol{\beta}'}\left[\frac{\partial \eta_\psi}{\partial \boldsymbol{\mu}}\boldsymbol{v}_2\frac{\partial q}{\partial \boldsymbol{\varphi}'}\right] = \mathbf{0}. \tag{3.56}$$

Using the total derivative on (3.56) and the product rule for matrices (3.51) yields

$$\begin{aligned}\frac{\partial^2 q}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\beta}'} + \frac{\partial^2 \eta_\psi}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}}(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{v}_2'\boldsymbol{a}) \\ + \frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\beta}'}\right] = \mathbf{0}.\end{aligned} \tag{3.57}$$

Solving (3.57) for $\frac{\partial^2 \eta_\psi}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}}$ yields

$$\frac{\partial^2 \eta_\psi}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime} \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\beta}'}^{(1)}\right] \big(\boldsymbol{I}_{\nu_1} \otimes (\boldsymbol{v}_2' \boldsymbol{a})^{-1}\big).$$

Finally, because $\dfrac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}} = \dfrac{\partial \eta_\psi}{\partial \boldsymbol{\mu}} \boldsymbol{v}_2'$, then

$$\begin{aligned} \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} &= \frac{\partial^2 \eta_\psi}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} \big(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{v}_2'\big) \\ &= -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime} \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \right. \\ &\quad \left. + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\beta}'}^{(1)}\right] \big(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{r}'\big). \end{aligned}$$

The proof is similar for the other five implicit derivatives of $\boldsymbol{\psi}$.

The second derivatives of q^* are used in each of the expressions for the implicit derivatives. The derivation of $\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}$ will be given. Expressions without proof for the other derivatives of q^* can be found in Appendix C. Methods for obtaining the remaining derivatives are similar to those given below.

The derivative of q^* with respect to $\boldsymbol{\mu}$ (without evaluating at $\boldsymbol{\mu} = \mathbf{0}$) is

$$\frac{\partial q^*}{\partial \boldsymbol{\mu}} = -\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right)' \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}_{zz} \boldsymbol{\Sigma}^{-1}\big). \tag{3.58}$$

The derivative of (3.58) with respect to $\boldsymbol{\varphi}'$ (without evaluating at $\boldsymbol{\mu} = \mathbf{0}$) is

$$\begin{aligned} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} &= \frac{\partial}{\partial \boldsymbol{\varphi}'}\left[\frac{\partial q^*}{\partial \boldsymbol{\mu}}\right] = \frac{\partial}{\partial \boldsymbol{\varphi}'}\left[-\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right)' \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}_{zz} \boldsymbol{\Sigma}^{-1}\big)\right] \\ &= -\left[\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}}\right]' \big[\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}_{zz} \boldsymbol{\Sigma}^{-1}\big)\big] - \left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right]' \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}_{zz} \boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}. \end{aligned}$$

Evaluating $\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}$ at $\boldsymbol{\mu} = \mathbf{0}$ yields

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}_{\boldsymbol{\mu}\boldsymbol{\varphi}}^{(11)\prime} \big[\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}_{zz} \boldsymbol{\Sigma}^{-1}\big)\big] - \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime} \left.\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1} \boldsymbol{W}_{zz} \boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}}. \tag{3.59}$$

The first step in finding the derivative of $\operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1})$ is

$$\begin{aligned}\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'} &= \frac{\partial(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\operatorname{vec}\boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'} \\ &= \frac{\partial(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{W}_{zz}\big) + (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\frac{\partial \operatorname{vec}\boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'}. \end{aligned} \tag{3.60}$$

For convenience, denote the first $\boldsymbol{\Sigma}^{-1}$ in (3.60) as $\boldsymbol{\Sigma}_1^{-1}$ and likewise for the second $\boldsymbol{\Sigma}^{-1}$. Using the identity

$$\big(\boldsymbol{ABC}\otimes\boldsymbol{D}\big)\boldsymbol{E} = \big(\boldsymbol{A}\otimes[\operatorname{vec}(\boldsymbol{C}')]'\otimes\boldsymbol{D}\big)\big(\operatorname{vec}(\boldsymbol{B}')\otimes\boldsymbol{E}\big), \tag{3.61}$$

then

$$\begin{aligned}\big(\boldsymbol{\Sigma}_1^{-1}\otimes\boldsymbol{\Sigma}_2^{-1}\big) &= \big(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Sigma}_2^{-1}\big)\big(\operatorname{vec}\big(\boldsymbol{\Sigma}_1^{-1}\big)\otimes\boldsymbol{I}_{p^2}\big) \\ &= \boldsymbol{I}_{(p,p)}\big(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Sigma}_1^{-1}\big)\big(\operatorname{vec}\big(\boldsymbol{\Sigma}_2^{-1}\big)\otimes\boldsymbol{I}_{(p,p)}\big).\end{aligned}$$

The derivative of $\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}$ with respect to $\boldsymbol{\varphi}'$ therefore is

$$\begin{aligned}\frac{\partial(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'} &= \left.\frac{\partial(\boldsymbol{\Sigma}_1^{-1}\otimes\boldsymbol{\Sigma}_2^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\Sigma}_2^{-1}\text{ fixed}} + \left.\frac{\partial(\boldsymbol{\Sigma}_1^{-1}\otimes\boldsymbol{\Sigma}_2^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\Sigma}_1^{-1}\text{ fixed}} \\ &= \big(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Sigma}_2^{-1}\big)\left(\frac{\partial \operatorname{vec}\big(\boldsymbol{\Sigma}_1^{-1}\big)}{\partial \boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{p^2}\right) \\ &\quad + \boldsymbol{I}_{(p,p)}\big(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Sigma}_1^{-1}\big)\left(\frac{\partial \operatorname{vec}\big(\boldsymbol{\Sigma}_2^{-1}\big)}{\partial \boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{(p,p)}\right).\end{aligned} \tag{3.62}$$

Using (3.16) and replacing $\boldsymbol{\Sigma}_1^{-1}$ and $\boldsymbol{\Sigma}_2^{-1}$ by $\boldsymbol{\Sigma}^{-1}$, (3.62) can be expressed as

$$\begin{aligned}\frac{\partial(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'} &= \big(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Sigma}^{-1}\big)\left[-\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\big)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{p^2}\right] \\ &\quad + \boldsymbol{I}_{(p,p)}\big(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Sigma}^{-1}\big)\left[-\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\big)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{(p,p)}\right].\end{aligned} \tag{3.63}$$

Suppose $\boldsymbol{A}$ is a $p^2 \times p^2$ matrix. Then

$$
\begin{aligned}
&\left(\boldsymbol{I}_p \otimes (\operatorname{vec} \boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{A}\right] \\
&\quad = \left(\boldsymbol{I}_p \otimes (\operatorname{vec} \boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \otimes \boldsymbol{I}_{p^2}\right]\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{A}\right] \\
&\quad = \left(\boldsymbol{I}_p \otimes (\operatorname{vec} \boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\boldsymbol{\Sigma}^{-1} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_p\right) \otimes \boldsymbol{I}_p\right]\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{A}\right] \\
&\quad = \left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{A}\right]. \qquad (3.64)
\end{aligned}
$$

By (3.64), (3.63) simplifies to

$$
\begin{aligned}
\frac{\partial(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'} &= -\left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{I}_{p^2}\right] \\
&\quad - \boldsymbol{I}_{(p,p)}\left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{I}_{(p,p)}\right]. \qquad (3.65)
\end{aligned}
$$

Evaluating (3.65) at $\boldsymbol{\mu} = \boldsymbol{0}$ yields

$$
\begin{aligned}
\left.\frac{\partial(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} &= \left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\right) \\
&\quad + \boldsymbol{I}_{(p,p)}\left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{(p,p)}\right). \qquad (3.66)
\end{aligned}
$$

The derivative in the last term of (3.60) can be written as

$$
\begin{aligned}
\frac{\partial \boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'} &= \frac{\partial \boldsymbol{Z}'\boldsymbol{W}_1\boldsymbol{Z}}{\partial \boldsymbol{\varphi}'} = \frac{\partial}{\partial \boldsymbol{\varphi}'}\left[\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{z}_k w_k^{(1)} \boldsymbol{z}_k'\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{z}_k \otimes \boldsymbol{z}_k)\frac{\partial w_k^{(1)}}{\partial \boldsymbol{\varphi}'} = \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{z}_k \otimes \boldsymbol{z}_k)\frac{\partial w_k^{(1)}}{\partial h_k}\frac{\partial h_k}{\partial \boldsymbol{\varphi}'} \\
&= \frac{1}{n}\sum_{k=1}^{n}(\boldsymbol{z}_k \otimes \boldsymbol{z}_k) w_k^{(2)}\frac{\partial h_k}{\partial \boldsymbol{\varphi}'}. \qquad (3.67)
\end{aligned}
$$

By equation (3.14), (3.67) simplifies to

$$
\begin{aligned}
\frac{\partial \boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'} &= -\frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{z}_k \otimes \boldsymbol{z}_k\right) w_k^{(2)}\left(\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} \\
&= -\frac{1}{n}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right)\sum_{k=1}^{n}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right) w_k^{(2)}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right)'\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}
\end{aligned}
$$

$$= -\big(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\big)\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\big)\boldsymbol{W}_2\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\big)'\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}$$

$$= -\big(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\big)\boldsymbol{K}_{zz}\boldsymbol{W}_2\boldsymbol{K}_{zz}'\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}$$

$$= -\big(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\big)\boldsymbol{V}_{zz}\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}. \tag{3.68}$$

Evaluating (3.68) at $\boldsymbol{\mu}=\mathbf{0}$ yields

$$\left.\frac{\partial \boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = -\big(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\big)\boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}. \tag{3.69}$$

Substituting (3.69) and (3.66) into (3.60) yields

$$\begin{aligned}\left.\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} &= -\Big[\big(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes\boldsymbol{I}_{p^2}\right)\\ &\quad + \boldsymbol{I}_{(p,p)}\big(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes\boldsymbol{I}_{(p,p)}\right)\Big]\big(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{W}_{zz}\big)\\ &\quad - (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\big(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\big)\boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\\ &= -2\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes\operatorname{vec}\boldsymbol{W}_{zz}\right) - \boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}. \end{aligned} \tag{3.70}$$

Using the identity

$$\big(\boldsymbol{A}\otimes\boldsymbol{BCD}\big)\boldsymbol{E} = \big(\boldsymbol{A}\otimes[\operatorname{vec}(\boldsymbol{D}')]'\otimes\boldsymbol{B}\big)\big(\boldsymbol{E}\otimes\operatorname{vec}\boldsymbol{C}\big),$$

(3.70) simplifies to

$$\left.\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = -2\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} - \boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}. \tag{3.71}$$

Because $\boldsymbol{N}_p\boldsymbol{V}_{zz} = \boldsymbol{V}_{zz}$, then (3.71) simplifies to

$$\begin{aligned}\left.\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} &= -2\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} - \boldsymbol{N}_p\boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\\ &= -2\boldsymbol{N}_p\Big[\big(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big) - \frac{1}{2}\boldsymbol{V}_{zz}\Big]\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\\ &= -2\boldsymbol{N}_p\boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}.\end{aligned}$$

Lastly, because $\boldsymbol{N}_p\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} = \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$, (3.59) simplifies to

$$\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}' \otimes \partial\boldsymbol{\mu}} = -\boldsymbol{F}^{(11)'}_{\boldsymbol{\mu\varphi}}\big[\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big] + 2\boldsymbol{F}^{(1)'}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}.$$

□

Solving for the Implicit Parameter η_ψ

According to the implicit function theorem, η_ψ is a function of $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ as long as $\frac{\partial q}{\partial \eta_\psi} \neq 0$. Accordingly, the Newton-Raphson method will be used to solve for the implicit parameter η_ψ.

Given values for $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$

$$q(\eta_\psi; \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi}) = \frac{1}{n}\sum_{k=1}^{n} \rho(d_k) - b_0 = 0, \tag{3.72}$$

can be solved for η_ψ. Therefore, for the purpose of solving the constraint, q can be thought of as a function of η_ψ, given fixed values of the explicit parameters. Accordingly, the notation in (3.44) will be used. The Newton-Raphson iterative equation for this situation can be derived as follows. Expand $q^*(\eta_\psi)$ in a Taylor series about the i^{th} iteration guess for η_ψ:

$$0 = q^*(\eta_\psi) \approx q^*(\eta_\psi^{(i)}) + \left.\frac{\partial q^*}{\partial \eta_\psi}\right|_{\eta_\psi = \eta_\psi^{(i)}} \big(\eta_\psi - \eta_\psi^{(i)}\big). \tag{3.73}$$

Solving (3.73) for η_ψ yields

$$\eta_\psi \approx \eta_\psi^{(i)} - \left[\left.\frac{\partial q^*}{\partial \eta_\psi}\right|_{\eta_\psi = \eta_\psi^{(i)}}\right]^{-1} q^*(\eta_\psi^{(i)}). \tag{3.74}$$

The right side of (3.74) becomes the new guess $\eta_\psi^{(i+1)}$, and the procedure is repeated. The Newton-Raphson procedure can be summarized as

$$\eta_\psi^{(i+1)} = \eta_\psi^{(i)} - \left[\left. \frac{\partial q^*}{\partial \eta_\psi} \right|_{\eta_\psi = \eta_\psi^{(i)}} \right]^{-1} q^*(\eta_\psi^{(i)}). \tag{3.75}$$

Because $\left. \frac{\partial q^*}{\partial \eta_\psi} \right|_{\eta_\psi = \eta_\psi^{(i)}} = \boldsymbol{a}'\boldsymbol{v}_2$, (3.75) simplifies to

$$\eta_\psi^{(i+1)} = \eta_\psi^{(i)} - (\boldsymbol{a}'\boldsymbol{v}_2)^{-1} q^*(\eta_\psi^{(i)}), \tag{3.76}$$

where $\boldsymbol{a}$ and $\boldsymbol{v}_2$ are given in Theorem 3.9. A desirable property of the Newton-Raphson method is local quadratic convergence in a neighborhood of the solution. However, there is no guarantee that the Newton-Raphson method will converge from an arbitrary starting point. A good starting point for η_ψ is necessary. An initial guess for η_ψ can be found from equation (3.41) as follows:

$$\begin{aligned} \boldsymbol{\varphi} = \boldsymbol{V}_1\boldsymbol{\psi} + \boldsymbol{v}_2\eta_\psi &\Longrightarrow \boldsymbol{v}_2'\boldsymbol{\varphi} = \boldsymbol{v}_2'\boldsymbol{V}_1\boldsymbol{\psi} + \boldsymbol{v}_2'\boldsymbol{v}_2\eta_\psi \\ &\Longrightarrow \boldsymbol{v}_2'\boldsymbol{\varphi} = \boldsymbol{v}_2'\boldsymbol{v}_2\eta_\psi \text{ because } \boldsymbol{v}_2'\boldsymbol{V}_1 = \boldsymbol{0} \\ &\Longrightarrow \eta_\psi^{(0)} = (\boldsymbol{v}_2'\boldsymbol{v}_2)^{-1} \boldsymbol{v}_2'\boldsymbol{\varphi} \text{ because } \boldsymbol{v}_2'\boldsymbol{v}_2 \neq 0 \end{aligned} \tag{3.77}$$

Even using the initial guess in (3.77), there is no guarantee that (3.76) will converge. In contrast, if the Bisection method (Burden and Faires 1997) is used, then convergence would be guaranteed. However, the rate of local convergence for the Bisection method is linear in contrast to the local quadratic convergence of the Newton-Raphson method. Because solving for the implicit parameter is required many times when solving for the S-estimators, a globally convergent and fast algorithm is desired. A hybrid of the Bisection and Newton-Raphson method will be used. First, a description of the Bisection method will be given.

The Bisection method uses two initial guesses $e_a^{(0)}$ and $e_b^{(0)}$, where $e_a^{(0)} < e_b^{(0)}$, and $q^*(e_a^{(0)})q^*(e_b^{(0)}) < 0$ (which means that $e_a^{(0)} < \eta_\psi < e_b^{(0)}$). On each successive iteration, the midpoint, $e_c^{(i)}$ of the interval $(e_a^{(i)}, e_b^{(i)})$ is found. A new interval, $(e_c^{(i)}, e_b^{(i)})$ or $(e_a^{(i)}, e_c^{(i)})$, is used depending on which interval contains the root. The procedure is then repeated until the length of the interval is very small. The iteration procedure can be summarized with the following steps:

Step 1. Compute $q_a^{*(i)} = q^*(e_a^{(i)})$

Step 2. Compute $e_c^{(i)} = \frac{e_a^{(i)}+e_b^{(i)}}{2}$ and $q_c^{*(i)} = q^*(e_c^{(i)})$.

Step 3. if $q_c^{*(i)}q_a^{*(i)} > 0$ then set $e_a^{(i+1)} = e_c^{(i)}$, $e_b^{(i+1)} = e_b^{(i)}$ and set $q_a^{*(i+1)} = q_c^{*(i)}$ else set $e_a^{(i+1)} = e_a^{(i)}$ and $e_b^{(i+1)} = e_c^{(i)}$.

Step 4. Repeat steps 1-3 until convergence.

A hybrid of the Newton-Raphson method and the Bisection that combines the global convergence of the Bisection method with the quadratic local convergence of the Newton-Raphson method can be constructed to solve for the implicit parameter η_ψ. The hybrid method uses the Newton-Raphson method when possible, and uses the Bisection method when the i^{th} Newton-Raphson guess $\eta_\psi^{(i+1)}$ is not in the interval $(e_a^{(i)}, e_b^{(i)})$.

Note that the choice for $\boldsymbol{v}_2$ in (3.48) ensures that $\frac{\partial q^*}{\partial \eta_\psi} > 0 \; \forall \; \boldsymbol{a}(\eta_\psi) \neq \mathbf{0}$. Hence $q^*(\eta_\psi)$ is an nondecreasing function and has at most one solution. However, a solution is always guaranteed for any ρ function since $0 < b_0 < \sup \rho$. Because (3.72) is a nondecreasing function of η_ψ and only one solution exists, the local quadratic convergence of the Newton-Raphson method will be guaranteed.

Although $\left.\frac{\partial q^*}{\partial \eta_\psi}\right|_{\eta_\psi=\eta_\psi^{(i)}} > 0$, it is sometimes very close to zero. This causes the next Newton-Raphson iteration $\eta_\psi^{(i+1)}$ to be very different from $\eta_\psi^{(i)}$. In this case, it is best to use a Bisection step instead of a Newton-Raphson step.

Initial guesses $e_a^{(0)}$ and $e_b^{(0)}$ are needed for the hybrid method, where $e_a^{(0)} < \eta_\psi < e_b^{(0)}$. A good place to start is the Newton-Raphson initial guess in (3.77). Because $q^*(\eta_\psi) < 0$ for negative values of η_ψ, it follows that if $q^*(\eta_\psi^{(0)}) < 0$, then $e_a^{(0)}$ can be set to $\eta_\psi^{(0)}$ to correctly place the initial guess $e_a^{(0)}$ less than the solution η_ψ. Likewise, if $q^*(\eta_\psi^{(0)}) > 0$, then set $e_b^{(0)} = \eta_\psi^{(0)}$. In either case, the sign of the other guess ($e_a^{(0)}$ or $e_b^{(0)}$) is the opposite of the sign of $q^*(\eta_\psi^{(0)})$. Without loss of generality, assume that the sign of $q^*(\eta_\psi^{(0)})$ is negative. Set $e_a^{(0)} = \eta_\psi^{(0)}$. The initial guess for $e_b^{(0)}$ can be found through the following steps:

Step 1. Set $e_b^{(0)}$ to any positive number.

Step 2. While $q^*(e_a^{(0)})q^*(e_b^{(0)}) > 0$ do steps 3-4

Step 3. Set $e_b^{(0)} = 2e_b^{(0)}$

Step 4. Evaluate $q^*(e_b^{(0)})$

A similar algorithm is employed if $q^*(\eta_\psi^{(0)}) > 0$. An algorithm for solving for η_ψ using the hybrid Newton-Raphson-Bisection method consists of the following twelve steps:

Step 1. Set the tolerance variable, TOL, to a small positive number. Set $step = 1$.

Step 2. Find the initial Newton-Raphson guess, $\eta_\psi^{(0)}$.

Step 3. Find the two initial endpoints of the Bisection interval $(e_a^{(0)}, e_b^{(0)})$ using the above algorithm.

Step 4. Set $q_a = q^*(e_a^{(0)})$. If $|q_a| < TOL$, then go to step 12.

Step 5. If $|q^*(e_a^{(0)})| < |q^*(e_b^{(0)})|$, then set $\eta_\psi^{(0)} = e_a^{(0)}$ else set $\eta_\psi^{(0)} = e_b^{(0)}$ (Replace the initial Newton iterate $\eta_\psi^{(0)}$ based on the minimum of $|q^*(e_a^{(0)})|$ or $|q^*(e_b^{(0)})|$.)

Step 6. While $|step| > TOL$, do steps 7-11.

Step 7. Find the next Newton iterate, $\eta_\psi^{(i+1)}$ using the formula (3.76). If the derivative is too small, then set the indicator variable *small* to 1 else set $small = 0$.

Step 8. If $\eta_\psi^{(i+1)} \in (e_a^{(i)}, e_b^{(i)})$ and $small = 0$, then set $e_c^{(i+1)} = \eta_\psi^{(i+1)}$ and set $step = (\boldsymbol{a}'\boldsymbol{v}_2)^{-1} q^*(\eta_\psi^{(i)})$ else set $e_c^{(i+1)} = \frac{e_a^{(i)} + e_b^{(i)}}{2}$ and $step = \frac{|e_b^{(i)} - e_c^{(i+1)}|}{2}$.

Step 9. If $q_a q^*(e_c^{(i+1)}) > 0$, then set $e_a^{(i+1)} = e_c^{(i+1)}$, $e_b^{(i+1)} = e_b^{(i)}$, and $q_a = q^*(e_c^{(i+1)})$ else set $e_a^{(i+1)} = e_a^{(i)}$ and $e_b^{(i+1)} = e_c^{(i+1)}$.

Step 10. Set $\eta_\psi^{(i+1)} = e_c^{(i+1)}$.

Step 11. Set $i = i + 1$.

Step 12. Return $\eta_\psi = e_c^{(i)}$

Solving the S-Equation

For a general choice of ρ, a modified Newton-Raphson method can be used to solve the S-estimating equation in (3.40) or (3.43). To be clear on notation in the expressions of the derivatives, define

$$w^* \stackrel{\text{def}}{=} w(\boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi\}),$$

where w is defined in (3.40), η_ψ is treated as functionally independent of the other variables, and $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ are functionally independent. Note that w^* is assumed to not depend on $\boldsymbol{\beta}$ and that the determinant of a matrix only depends on its eigenvalues.

Further, define

$$w \stackrel{\text{def}}{=} w(\boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})\}),$$

where η_ψ is a function of $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$, but $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$ are functionally independent. Now the first derivative of $\boldsymbol{w}$ with respect to $\boldsymbol{\theta}_{s_1}$ can be obtained.

Theorem 3.12 (First Derivatives of $w = n\log|\boldsymbol{\Sigma}|$). *The first derivatives of $w = n\log|\boldsymbol{\Sigma}|$ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ are*

$$\begin{aligned}
\frac{\partial w}{\partial \boldsymbol{\beta}} &= \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)\prime}\frac{\partial w^*}{\partial \boldsymbol{\varphi}}, \\
\frac{\partial w}{\partial \boldsymbol{\mu}} &= \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime}\frac{\partial w^*}{\partial \boldsymbol{\varphi}}, \quad \textit{and} \\
\frac{\partial w}{\partial \boldsymbol{\psi}} &= \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)\prime}\frac{\partial w^*}{\partial \boldsymbol{\varphi}},
\end{aligned}$$

where $\boldsymbol{P}_{\cdot}^{(1)}$ *is given in Theorem 3.10 and* $\dfrac{\partial w^*}{\partial \boldsymbol{\varphi}} = n\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big)$.

Proof. By equation (3.11), the derivative of w^* with respect to $\boldsymbol{\varphi}'$ is

$$\begin{aligned}\frac{\partial w^*}{\partial \boldsymbol{\varphi}'} &= n\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\varphi}'} = n\operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big)'\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\\ &= n\operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big)'\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \text{ (evaluated at } \boldsymbol{\mu}=\mathbf{0}).\end{aligned}$$

Hence,

$$\frac{\partial w^*}{\partial \boldsymbol{\varphi}} = n\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big).$$

Using the total derivative, the derivative of w with respect to $\boldsymbol{\psi}$ is

$$\frac{\partial w}{\partial \boldsymbol{\psi}} = \frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}}\frac{\partial w^*}{\partial \boldsymbol{\varphi}} = \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)\prime}\frac{\partial w^*}{\partial \boldsymbol{\varphi}}.$$

The proof is similar for the other two derivatives. □

Theorem 3.13 (Second Derivatives of $w = n\log|\boldsymbol{\Sigma}|$). *The second derivatives of* $w = n\log|\boldsymbol{\Sigma}|$ *evaluated at* $\boldsymbol{\mu}=\mathbf{0}$ *are*

$$\begin{aligned}
\frac{\partial^2 w}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} &= \boldsymbol{P}_{\boldsymbol{\beta\beta}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)\prime}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\beta}}^{(1)},\\
\frac{\partial^2 w}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} &= \boldsymbol{P}_{\boldsymbol{\mu\mu}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_2} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\mu}}^{(1)},\\
\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}} &= \boldsymbol{P}_{\boldsymbol{\psi\psi}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_3-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)\prime}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\psi}}^{(1)},\\
\frac{\partial^2 w}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} &= \boldsymbol{P}_{\boldsymbol{\beta\mu}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\beta}}^{(1)},\\
\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\mu}} &= \boldsymbol{P}_{\boldsymbol{\psi\mu}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_3-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)\prime}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\psi}}^{(1)}, \text{ and}\\
\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\beta}} &= \boldsymbol{P}_{\boldsymbol{\psi\beta}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_3-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)\prime}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\psi}}^{(1)},
\end{aligned}$$

where $\boldsymbol{P}^{(1)}_{\cdot}$ *is given in Theorem 3.10,* $\boldsymbol{P}^{(11)}_{\cdot\cdot}$ *is given in Theorem 3.11,* $\frac{\partial w^*}{\partial \boldsymbol{\varphi}}$ *is given in Theorem 3.12, and*

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = n \left[\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right) - \boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \right].$$

Proof. The derivative of w^* with respect to $\boldsymbol{\varphi}$ (but not evaluated at $\boldsymbol{\mu} = \mathbf{0}$) is

$$\frac{\partial w^*}{\partial \boldsymbol{\varphi}} = n \frac{\partial (\operatorname{vec} \boldsymbol{\Sigma})'}{\partial \boldsymbol{\varphi}} \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right). \tag{3.78}$$

Using the product rule in (3.51), the derivative of (3.78) with respect to $\boldsymbol{\varphi}'$ is

$$\begin{aligned}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} &= n \frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\frac{\partial (\operatorname{vec} \boldsymbol{\Sigma})'}{\partial \boldsymbol{\varphi}} \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \right] \\ &= n \frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\frac{\partial (\operatorname{vec} \boldsymbol{\Sigma})'}{\partial \boldsymbol{\varphi}} \right] \left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right) + n \frac{\partial (\operatorname{vec} \boldsymbol{\Sigma})'}{\partial \boldsymbol{\varphi}} \frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'}.\end{aligned}$$

Using (3.16), the second derivative of w^* with respect to $\boldsymbol{\varphi}'$ and $\boldsymbol{\varphi}$ is

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = n \left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right)' \left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right) - n \frac{\partial (\operatorname{vec} \boldsymbol{\Sigma})'}{\partial \boldsymbol{\varphi}} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}.$$

The second derivative of w^* with respect to $\boldsymbol{\varphi}'$ and $\boldsymbol{\varphi}$ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ is

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = n \left[\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\right) - \boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \right].$$

The derivative of w with respect to $\boldsymbol{\mu}$ is

$$\frac{\partial w}{\partial \boldsymbol{\mu}} = \frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}} \frac{\partial w^*}{\partial \boldsymbol{\varphi}}. \tag{3.79}$$

Using the product rule in (3.51), the derivative of (3.79) with respect to $\boldsymbol{\psi}'$ is obtained as follows:

$$\begin{aligned}
\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\mu}} &= \frac{\partial}{\partial \boldsymbol{\psi}'}\left[\frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right] \\
&= \frac{\partial}{\partial \boldsymbol{\psi}'}\left[\frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\right]\left(\boldsymbol{I}_{\nu_3-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\frac{\partial}{\partial \boldsymbol{\psi}'}\left[\frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right] \\
&= \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_3-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \frac{\partial \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} \\
&= \boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\mu}}^{(11)'}\left(\boldsymbol{I}_{\nu_3-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'}\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}_{\boldsymbol{\psi}}^{(1)}.
\end{aligned}$$

The proofs for the other five derivatives are similar. □

To solve the S-estimating equation in (3.40) or (3.43), a modified two-stage Newton-Raphson method was developed based on the algorithm in Boik (2002a). The structure of the problem naturally leads to a two stage algorithm. Define $\boldsymbol{\theta}_1 = \begin{pmatrix} \boldsymbol{\beta}' & \boldsymbol{\mu}' \end{pmatrix}$. The derivative of w with respect to the parameters is

$$\begin{aligned}
\frac{\partial w}{\partial \boldsymbol{\theta}_1} &= -\frac{\partial q^*}{\partial \boldsymbol{\theta}_1}\left(\boldsymbol{v}_2'\boldsymbol{a}\right)^{-1}\boldsymbol{v}_2'\frac{\partial w^*}{\partial \boldsymbol{\varphi}}, \text{ and} \\
\frac{\partial w}{\partial \boldsymbol{\psi}} &= \left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left[\boldsymbol{a}\left(\boldsymbol{v}_2'\boldsymbol{a}\right)^{-1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right].
\end{aligned}$$

Because $\left(\boldsymbol{v}_2'\boldsymbol{a}\right)^{-1}\boldsymbol{v}_2'\frac{\partial w^*}{\partial \boldsymbol{\varphi}}$ is a scalar, the solution to $\frac{\partial w}{\partial \boldsymbol{\theta}_1} = \mathbf{0}$ is the same as the solution to $\frac{\partial q^*}{\partial \boldsymbol{\theta}_1} = \mathbf{0}$. Solving $\frac{\partial q^*}{\partial \boldsymbol{\theta}_1} = \mathbf{0}$, followed by the minimization of w with respect to $\boldsymbol{\psi}$ will be the same as minimizing $\frac{\partial w}{\partial \boldsymbol{\theta}_{s_1}}$. Before the two stages can be used, an initial guess for the parameters must be found. Denote the guesses for the parameters after the i^{th} guess as $\hat{\boldsymbol{\beta}}_i$, $\hat{\boldsymbol{\Gamma}}_i$, and $\hat{\boldsymbol{\psi}}_i$. The two-stage modified Newton-Raphson algorithm has the following stages:

Stage 1. Fix the $(i-1)^{\text{th}}$ guess of $\boldsymbol{\psi}$. Find the minimizer $(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\Gamma}}_i)$ of q^* with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$.

Stage 2. Fix the i^{th} guesses for $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$. Find the minimizer $\hat{\boldsymbol{\psi}}_i$ of $w = n \log |\boldsymbol{\Sigma}|$ with respect to $\boldsymbol{\psi}$.

Stage 3. Iterate stages 1–2 until convergence

This two step algorithm has the advantage that in each step, the dimensions are reduced, and this allows it to be more efficient. The first stage of the two-stage modified Newton-Raphson algorithm includes the following steps:

Step 1. Fix the $(i-1)^{\text{th}}$ guess for $\boldsymbol{\psi}$. Denote the guesses for $\boldsymbol{\psi}$ and $\boldsymbol{\Lambda}$ as $\hat{\boldsymbol{\psi}}_{i-1}$ and $\hat{\boldsymbol{\Lambda}}_{i-1}$.

Step 2. Use the initial guesses for this sub-loop as the i^{th} guesses for $\boldsymbol{\Gamma}$ and $\boldsymbol{\beta}$ from the main loop. Denote these guesses by $\hat{\boldsymbol{\Gamma}}_{0,i,0}$, $\hat{\boldsymbol{\beta}}_{i,0}$.

Step 3. Denote the S-estimate of $\boldsymbol{\Sigma}$ after the j^{th} iteration by

$$\hat{\boldsymbol{\Sigma}}_{i,j} = \hat{\boldsymbol{\Gamma}}_{0,i,j} \hat{\boldsymbol{\Lambda}}_{i-1} \hat{\boldsymbol{\Gamma}}'_{0,i,j}.$$

Step 4. Set $\boldsymbol{\mu}_{i,j} = \mathbf{0}$, then $\hat{\boldsymbol{\theta}}_{1,i,j} = \begin{pmatrix} \hat{\boldsymbol{\beta}}'_{i,j} & \mathbf{0}' \end{pmatrix}'$. Update $\hat{\boldsymbol{\theta}}_{1,i,j}$ as

$$\hat{\boldsymbol{\theta}}_{1,i,j+1} = \hat{\boldsymbol{\theta}}_{1,i,j} + \alpha \hat{\boldsymbol{H}}_{\hat{\boldsymbol{\theta}}_{1,i,j}} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_{1,i,j}},$$

where $\alpha \in (0,1]$, $\hat{q}^*_{i,j}(\boldsymbol{\theta}_1) = q^*(\boldsymbol{\theta}_1; \hat{\boldsymbol{\beta}}_{i,j}, \hat{\boldsymbol{\Gamma}}_{0,i,j}, \hat{\boldsymbol{\Lambda}}_{i-1})$,

$$\boldsymbol{g}_{\hat{\boldsymbol{\theta}}_{1,i,j}} = \left. \frac{\partial \hat{q}^*_{i,j}(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} \right|_{\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_{1,i,j}}, \text{ and } \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_{1,i,j}} = -E \left[\left. \frac{\partial^2 \hat{q}^*_{i,j}(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}'_1 \otimes \partial \boldsymbol{\theta}_1} \right|_{\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_{1,i,j}} \right].$$

Step 5. Compute $\boldsymbol{G}(\hat{\boldsymbol{\mu}}_{i,j+1})$ using the algorithm from the last section of Chapter 2.

Step 6. Set $\hat{\boldsymbol{\Gamma}}_{0,i,j+1} = \hat{\boldsymbol{\Gamma}}_{0,i,j}\boldsymbol{G}(\hat{\boldsymbol{\mu}}_{i,j+1})$.

Step 7. Iterate steps 3–6 until convergence.

Step 8. Denote the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ after convergence as $\hat{\boldsymbol{\beta}}_i$ and $\hat{\boldsymbol{\Gamma}}_{0,i}$.

The second stage of the two-stage modified Newton-Raphson algorithm includes the following steps:

Step 1. Fix the i^{th} guesses for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\beta}$. Denote these guesses by $\hat{\boldsymbol{\Gamma}}_{0,i}$, $\hat{\boldsymbol{\beta}}_i$.

Step 2. Use the initial guesses for this sub-loop as the i^{th} guesses for $\boldsymbol{\psi}$ and $\boldsymbol{\Lambda}$ from the main loop. Denote these guesses by $\hat{\boldsymbol{\psi}}_{i,0}$, $\hat{\boldsymbol{\Lambda}}_{i,0}$.

Step 3. Denote the S-estimate of $\boldsymbol{\Sigma}$ after the j^{th} iteration by

$$\hat{\boldsymbol{\Sigma}}_{i,j} = \hat{\boldsymbol{\Gamma}}_{0,i}\hat{\boldsymbol{\Lambda}}_{i,j}\hat{\boldsymbol{\Gamma}}'_{0,i}.$$

Step 4. Update $\boldsymbol{\psi}$ as $\hat{\boldsymbol{\psi}}_{i,j}$ as

$$\hat{\boldsymbol{\psi}}_{i,j+1} = \hat{\boldsymbol{\psi}}_{i,j} + \alpha\hat{\boldsymbol{H}}_{\hat{\boldsymbol{\psi}}_{i,j}}\boldsymbol{g}_{\hat{\boldsymbol{\psi}}_{i,j}},$$

where $w_j(\boldsymbol{\psi}) = w(\boldsymbol{\psi}; \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\Gamma}}_{0,i}, \hat{\boldsymbol{\Lambda}}_{i,j})$,

$$\boldsymbol{g}_{\hat{\boldsymbol{\psi}}_{i,j}} = \left.\frac{\partial w_j(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{i,j}}, \boldsymbol{H}_{\hat{\boldsymbol{\psi}}_{i,j}} = -E\left[\left.\frac{\partial^2 w_j(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}_{i,j}}\right],$$

and $\alpha \in (0, 1]$.

Step 5. Set $\hat{\mathbf{\Lambda}}_{i,j+1} = \mathbf{\Lambda}(\hat{\boldsymbol{\psi}}_{i,j+1})$.

Step 6. Iterate steps 3–5 until convergence.

Step 7. Denote the estimates of $\boldsymbol{\psi}$ and $\mathbf{\Lambda}$ after convergence as $\hat{\boldsymbol{\psi}}_i$ and $\hat{\mathbf{\Lambda}}_i$.

CHAPTER 4

ASYMPTOTIC DISTRIBUTIONS AND ESTIMATING BIAS

In this chapter, asymptotic distributions of estimators of the location parameters, as well as estimators of the eigenvalues and eigenvectors of the scatter matrix are given. The scatter matrix is embedded within a multivariate fixed effects linear model. Together, the model for location and spectral parameters provide a framework within which inferences on Principal Components as well as on location parameters can be made. Further, second-order corrections are given for estimating the bias of the estimators.

Theoretical Background

In order to establish asymptotic distributions of M and S-estimators, some theorems need to be established. The first is the conditions for the asymptotic normality of the roots of an estimating equation, G_n, where G_n is defined in (3.1).

Theorem 4.1 (Conditions for Asymptotic Normality (Yuan and Jennrich 1998)). *Let $\boldsymbol{\theta}_0$ be the true value of the parameter $\boldsymbol{\theta}$. The following conditions lead to the existence, consistency, and asymptotic normality of a sequence $\hat{\boldsymbol{\theta}}_n$ of roots of $G_n(\boldsymbol{\theta})$:*

(a) *For each i, $\boldsymbol{g}_i(\boldsymbol{\theta}_0)$ has mean zero and covariance $\boldsymbol{V}_i$ and $\bar{\boldsymbol{V}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{V}_i \to \boldsymbol{V}$ which is positive definite.*

(b) *For the $\boldsymbol{V}_i$ above, there are positive numbers B and δ such that for all i,*

$$\boldsymbol{E}\left|\boldsymbol{g}_i(\boldsymbol{\theta}_0)'\left(\boldsymbol{I}_p + \boldsymbol{V}_i\right)^{-1}\boldsymbol{g}_i(\boldsymbol{\theta}_0)\right|^{1+\delta} \leqslant B.$$

Proof. The proof is found in Yuan and Jennrich (1998). □

Slutsky's Theorem is needed for the establishment of the asymptotic distributions of both the M and S estimators.

Theorem 4.2 (Slutsky's Theorem). *Suppose $\boldsymbol{t}_n$, $\boldsymbol{a}_n$ and $\boldsymbol{B}_n$ are random matrices of size $p \times 1$, $k \times 1$, and $k \times p$ respectively. Suppose $\boldsymbol{t}_n \xrightarrow{\text{dist}} \boldsymbol{t}$, $\boldsymbol{a}_n \xrightarrow{\text{prob}} \boldsymbol{a}$, $\boldsymbol{B}_n \xrightarrow{\text{prob}} \boldsymbol{B}$, where $\boldsymbol{a}$ and $\boldsymbol{B}$ are matrices of constants of size $k \times 1$ and $k \times p$, respectively, and $\boldsymbol{t}$ is a random $p \times 1$ matrix. Then the following are true:*

(a) $\boldsymbol{a}_n + \boldsymbol{B}_n\boldsymbol{t}_n \xrightarrow{\text{dist}} \boldsymbol{a} + \boldsymbol{B}\boldsymbol{t}$

(b) $\boldsymbol{a}_n + \boldsymbol{B}_n^{-1}\boldsymbol{t}_n \xrightarrow{\text{dist}} \boldsymbol{a} + \boldsymbol{B}^{-1}\boldsymbol{t}$, *if $k = p$ and $\boldsymbol{B}^{-1}$ exists.*

Proof. A proof may be found in Sen and Singer (1993, pg. 127). □

Asymptotic distributions for differentiable functions of asymptotically normal statistics can be found using the well known delta method.

Theorem 4.3 (Delta Method). *Let $\boldsymbol{t}_n$ be a random p-vector with asymptotic distribution $\sqrt{n}(\boldsymbol{t}_n - \boldsymbol{\theta}) \xrightarrow{\text{dist}} N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Suppose that $g(\boldsymbol{t}_n)$ is a differentiable function.*

Then

$$\sqrt{n}\,[\boldsymbol{g}(\boldsymbol{t}_n) - \boldsymbol{g}(\boldsymbol{\theta})] \xrightarrow{\text{dist}} N\big(\boldsymbol{0}, \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{\Sigma}\boldsymbol{G}(\boldsymbol{\theta})'\big),$$

where $\boldsymbol{G}(\boldsymbol{\theta}) = \left.\frac{\partial \boldsymbol{g}(\boldsymbol{t}_n)}{\partial \boldsymbol{t}_n}\right|_{\boldsymbol{t}_n=\boldsymbol{\theta}}$.

PROOF. A proof is given in Sen and Singer (1993, pg. 131). □

M-estimators

First Order Asymptotic Distributions

Define

$$\boldsymbol{\theta}_m = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\mu} \\ \boldsymbol{\varphi} \end{bmatrix}.$$

The M-estimation equations in (3.22) have an asymptotic normal distribution. They can be written as solutions to an estimating function, $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ that satisfies the conditions of Theorem 4.1. An expression for $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ is given in Theorem 4.4.

THEOREM 4.4 (M-ESTIMATING EQUATIONS). *Suppose* $v_3 = 1$. *Then the* M-*estimating equations* (3.22) *can be written as*

$$\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}} = n\ \text{vec}\big(\boldsymbol{R}_{xz}\boldsymbol{\Sigma}^{-1}\big)$$

$$\boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}} = n\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\text{vec}(\boldsymbol{R}_{zz} - \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{R}_0 = \frac{1}{n}\,\text{Diag}\left(v_1(d_k)\right),$$

$$\boldsymbol{R}_2 = \frac{1}{n}\,\mathrm{Diag}\left(v_2(d_k^2)\right),$$

$$\boldsymbol{R}_{xz} = \boldsymbol{X}'\boldsymbol{R}_0\boldsymbol{Z},\ \ \textit{and}$$

$$\boldsymbol{R}_{zz} = \boldsymbol{Z}'\boldsymbol{R}_2\boldsymbol{Z}.$$

Proof. The proof for $\boldsymbol{l}_{\boldsymbol{\beta}}^{(1)}$ is as follows:

$$\begin{aligned}
\boldsymbol{l}_{\boldsymbol{\beta}}^{(1)} &= \sum_{k=1}^{n} \boldsymbol{X}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k v_1(d_k) \\
&= \sum_{k=1}^{n} \left(\boldsymbol{I}_p \otimes \boldsymbol{x}_k\right)\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k v_1(d_k)\right) \\
&= \sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k v_1(d_k) \otimes \boldsymbol{x}_k\right)\mathrm{vec}(1) \\
&= \sum_{k=1}^{n} \mathrm{vec}\left(\boldsymbol{x}_k v_1(d_k)\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right) \\
&= n\,\mathrm{vec}\left(\boldsymbol{X}'\boldsymbol{R}_0\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\right) \\
&= n\,\mathrm{vec}\left(\boldsymbol{R}_{xz}\boldsymbol{\Sigma}^{-1}\right).
\end{aligned}$$

When $v_3 = 1$, $\boldsymbol{l}_{\boldsymbol{\zeta}}^{(1)}$ can be written as follows:

$$\begin{aligned}
\boldsymbol{l}_{\boldsymbol{\zeta}}^{(1)} &= \boldsymbol{F}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\sum_{k=1}^{n}\mathrm{vec}\left(\boldsymbol{z}_k v_2(d_k^2)\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right) \\
&= n\boldsymbol{F}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\boldsymbol{Z}'\boldsymbol{R}_0\boldsymbol{Z} - \boldsymbol{\Sigma}) \\
&= n\boldsymbol{F}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\boldsymbol{R}_{zz} - \boldsymbol{\Sigma}).
\end{aligned}$$

□

The expectation of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ under an elliptical distribution is $\mathbf{0}$. The proof is given in Theorem 4.6. First, a lemma is needed.

LEMMA 4.5. *Suppose that $\boldsymbol{z}_k$ follows a elliptical distribution with mean $\mathbf{0}$ and characteristic matrix $\boldsymbol{\Sigma}$. Let $d_k = \sqrt{\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k}$. Then $\boldsymbol{z}_k/d_k$ is independent of $h(d_k)$, where h is any function. Further, the expectation of $\boldsymbol{z}_k/d_k$ is zero and the covariance matrix is $\frac{1}{p}\boldsymbol{\Sigma}$.*

PROOF. Define $\boldsymbol{y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{z}_k$. Then $d_k = \sqrt{\boldsymbol{y}'\boldsymbol{y}}$. By Lemma 3.5, $\frac{\boldsymbol{y}}{\sqrt{\boldsymbol{y}'\boldsymbol{y}}}$ is independent of $\sqrt{\boldsymbol{y}'\boldsymbol{y}}$. As a result, $\frac{\boldsymbol{y}}{\sqrt{\boldsymbol{y}'\boldsymbol{y}}}$ is independent of $h(\sqrt{\boldsymbol{y}'\boldsymbol{y}})$. Equivalently, $\frac{\boldsymbol{z}_k}{d_k}$ is independent of $h(d_k)$.

The mean and covariance matrix of $\dfrac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{z}_k}{d_k}$ were given in Lopuhaä (1989) and are

$$E\left[\frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{z}_k}{d_k}\right] = \mathbf{0} \text{ and } \operatorname{Var}\left(\frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{z}_k}{d_k}\right) = \frac{1}{p}\boldsymbol{I}_p$$

Therefore,

$$E\left[\frac{\boldsymbol{z}_k}{d_k}\right] = \boldsymbol{\Sigma}^{\frac{1}{2}} E\left[\frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{z}_k}{d_k}\right] = \mathbf{0}$$

and

$$\begin{aligned}\operatorname{Var}\left(\frac{\boldsymbol{z}_k}{d_k}\right) &= E\left[\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k}\right] = \boldsymbol{\Sigma}^{\frac{1}{2}} E\left[\frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-\frac{1}{2}}}{d_k}\right]\boldsymbol{\Sigma}^{\frac{1}{2}} \\ &= \boldsymbol{\Sigma}^{\frac{1}{2}}\left[\frac{1}{p}\boldsymbol{I}_p\right]\boldsymbol{\Sigma}^{\frac{1}{2}} = \frac{1}{p}\boldsymbol{\Sigma}\end{aligned}$$

□

Theorem 4.6 (Expectation of the M-estimating Equations). *Under an elliptical distribution, the expectation of* $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ *is zero.*

Proof. First, the expectation of $\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}$:

$$\begin{aligned} E\left[\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}\right] &= E\left[n \operatorname{vec}\big(\boldsymbol{R}_{xz}\boldsymbol{\Sigma}^{-1}\big)\right] \\ &= E\left[\sum_{k=1}^{n} \operatorname{vec}\big(\boldsymbol{x}_k v_1(d_k)\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\big)\right] \\ &= \sum_{k=1}^{n} \operatorname{vec}\big(\boldsymbol{x}_k E\left[v_1(d_k)\boldsymbol{z}_k'\right]\boldsymbol{\Sigma}^{-1}\big) \\ &= \sum_{k=1}^{n} \operatorname{vec}\left(\boldsymbol{x}_k E\left[d_k v_1(d_k)\frac{\boldsymbol{z}_k'}{d_k}\right]\boldsymbol{\Sigma}^{-1}\right). \end{aligned} \tag{4.1}$$

By Lemma 4.5, the expectation in (4.1) can be factored since $\boldsymbol{z}_k'/d_k$ is independent of functions of d_k. Accordingly, (4.1) simplifies to

$$E\left[\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}\right] = \sum_{k=1}^{n} \operatorname{vec}\left(\boldsymbol{x}_k E\left[d_k v_1(d_k)\right] E\left[\frac{\boldsymbol{z}_k'}{d_k}\right]\boldsymbol{\Sigma}^{-1}\right) = \mathbf{0},$$

because the expectation of $\boldsymbol{z}_k'/d_k$ is zero.

The expectation of $\boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}}$ is

$$\begin{aligned} E\left[\boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}}\right] &= E\left[n\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\boldsymbol{R}_{zz}-\boldsymbol{\Sigma})\right] \\ &= n\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(E\left[\boldsymbol{R}_{zz}\right]-\boldsymbol{\Sigma}). \end{aligned}$$

The expectation of $\boldsymbol{R}_{zz}$ is

$$\begin{aligned} E\left[\boldsymbol{R}_{zz}\right] &= E\left[\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{z}_k v_2(d_k^2)\boldsymbol{z}_k'\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[\boldsymbol{z}_k v_2(d_k^2)\boldsymbol{z}_k'\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[d_k^2 v_2(d_k^2)\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k}\right]. \end{aligned} \tag{4.2}$$

By Lemma 4.5, the expectation in (4.2) can be factored since $\boldsymbol{z}_k/d_k$ is independent of functions of d_k. Further, because $E\left[\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k}\right] = \frac{1}{p}\boldsymbol{\Sigma}$ by Lemma 4.5, then the expectation of $\boldsymbol{R}_{zz}$ can be written as

$$E\left[\boldsymbol{R}_{zz}\right] = \frac{1}{n}\sum_{k=1}^{n} E\left[d_k^2 \upsilon_2(d_k^2)\right] E\left[\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k}\right] = \frac{c}{p}\boldsymbol{\Sigma},$$

where $c = \frac{1}{n}\sum_{k=1}^{n} E\left[d_k^2 \upsilon_2(d_k^2)\right]$.

Because $\boldsymbol{\zeta} = \begin{bmatrix}\boldsymbol{\mu}\\ \boldsymbol{\varphi}\end{bmatrix}$, then the expectation of $\boldsymbol{l}_{\boldsymbol{\zeta}}^{(1)}$ can be partitioned into two parts. First,

$$\begin{aligned}
E\left[\boldsymbol{l}_{\boldsymbol{\mu}}^{(1)}\right] &= n\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(E\left[\boldsymbol{R}_{zz}\right] - \boldsymbol{\Sigma})\\
&= n\left(\frac{c}{p}-1\right)\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\boldsymbol{\Sigma})\\
&= n\left(\frac{c}{p}-1\right)\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)'}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\\
&= n\left(\frac{c}{p}-1\right)2\boldsymbol{A}_1'\left(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right)\left(\boldsymbol{\Lambda}\boldsymbol{\Gamma}_0'\otimes\boldsymbol{\Gamma}_0'\right)\operatorname{vec}\left(\boldsymbol{\Gamma}_0\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}_0'\right)\\
&= n\left(\frac{c}{p}-1\right)2\boldsymbol{A}_1'\left(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right)\operatorname{vec}\boldsymbol{I}_p\\
&= n\left(\frac{c}{p}-1\right)2\boldsymbol{A}_1'\left(\operatorname{vec}\boldsymbol{I}_p - \operatorname{vec}\boldsymbol{I}_p\right)\\
&= \boldsymbol{0}.
\end{aligned}$$

Write $\boldsymbol{\Sigma}$ as $\kappa\boldsymbol{\Sigma}^*$, where $\boldsymbol{\Sigma}^*$ is the characteristic matrix and κ is the scalar that satisfies

$$E\left[d_k^2\upsilon_2(d_k^2)\right] = E\left[\boldsymbol{z}_k(\kappa\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{z}_k'\upsilon_2(\boldsymbol{z}_k(\kappa\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{z}_k')\right] = p.$$

The expectation of $\boldsymbol{l}_{\boldsymbol{\varphi}}^{(1)}$ is

$$\begin{aligned}
E\left[\boldsymbol{l}_{\boldsymbol{\varphi}}^{(1)}\right] &= n\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(E\left[\boldsymbol{R}_{zz}\right] - \boldsymbol{\Sigma})\\
&= n\left(\frac{c}{p}-1\right)\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)'}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)
\end{aligned}$$

$$
\begin{aligned}
&= n\left(\frac{c}{p}-1\right)\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}\left(\boldsymbol{\Gamma}_0 * \boldsymbol{\Gamma}_0\right)'\operatorname{vec}\bigl(\boldsymbol{\Sigma}^{-1}\bigr) \\
&= n\left(\frac{c}{p}-1\right)\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}\boldsymbol{L}_p'\left(\boldsymbol{\Gamma}_0' \otimes \boldsymbol{\Gamma}_0'\right)\operatorname{vec}\bigl(\boldsymbol{\Gamma}_0\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}_0'\bigr) \\
&= n\left(\frac{c}{p}-1\right)\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}\boldsymbol{L}_p'\operatorname{vec}\bigl(\boldsymbol{\Lambda}^{-1}\bigr) \\
&= n\left(\frac{c}{p}-1\right)\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}d(\boldsymbol{\Lambda}^{-1}),
\end{aligned}
$$

where d is defined in Theorem 2.1d. Note that by the definition of κ, $c = p$. Hence, the expectation of $\boldsymbol{l}_{\boldsymbol{\psi}}^{(1)}$ is

$$E\left[\boldsymbol{l}_{\boldsymbol{\varphi}}^{(1)}\right] = n\left(\frac{p}{p}-1\right)\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}d(\boldsymbol{\Lambda}^{-1}) = \mathbf{0}.$$

When the v_2 function corresponds to a particular underlying elliptical distribution, then

$$v_2 = \frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}.$$

In the case that v_2 corresponds to a particular underlying elliptical distribution, and data actually follow that distribution, then $\kappa = 1$. For example, if $v_2(d_k^2)$ corresponds to a multivariate-T distribution (3.26) and the data are sampled from this distribution, then the expectation $E\left[d_k^2 v_2(d_k^2)\right]$ can be written as

$$E\left[d_k^2 v_2(d_k^2)\right] = E\left[d_k^2\frac{\xi+p}{\xi+d_k^2}\right] = (\xi+p)E\left[\frac{d_k^2}{\xi+d_k^2}\right]. \tag{4.3}$$

Because $d_k^2 = \frac{\xi\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k}$ (see Lemma 3.6), it follows that the expectation in (4.3) can be written as

$$E\left[d_k^2 v_2(d_k^2)\right] = (\xi+p)E\left[\frac{d_k^2}{\xi+d_k^2}\right] = (\xi+p)E\left[\frac{\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k+\mathring{\boldsymbol{n}}_k'\mathring{\boldsymbol{n}}_k}\right].$$

Furthermore, because $\mathring{\boldsymbol{n}}_k \perp\!\!\!\perp \mathring{\boldsymbol{c}}_k$, it follows from Corollary 3.4 that $\frac{\mathring{\boldsymbol{n}}_k' \mathring{\boldsymbol{n}}_k}{\mathring{\boldsymbol{c}}_k + \mathring{\boldsymbol{n}}_k' \mathring{\boldsymbol{n}}_k} \sim \text{Beta}(p, \xi)$ and

$$E\left[d_k^2 v_2(d_k^2)\right] = (\xi + p)\frac{p}{p+\xi} = p.$$

More generally, the expectation is also p for any elliptical distribution. Suppose that differentiation and integration can be interchanged, and write $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = a\boldsymbol{\Sigma}^\circ$, for some positive constant a. Because the density of an elliptical distribution integrates to one, then

$$\begin{aligned}\int f(d_k^2)\; d\boldsymbol{y} &= \int \frac{g(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; d\boldsymbol{y} = 1 \\ &= \int \frac{g\left(\frac{\boldsymbol{z}_k' \boldsymbol{\Sigma}^{\circ -1} \boldsymbol{z}_k}{a}\right)}{a^{\frac{p}{2}} |\boldsymbol{\Sigma}^\circ|^{\frac{1}{2}}}\; d\boldsymbol{y} = 1. \end{aligned} \tag{4.4}$$

The derivative of both sides of (4.4) with respect to a yields

$$\begin{aligned}&\int \frac{-\frac{\boldsymbol{z}_k' \boldsymbol{\Sigma}^{\circ -1} \boldsymbol{z}_k}{a^2} g^{(1)}(d_k^2)}{a^{\frac{p}{2}} |\boldsymbol{\Sigma}^\circ|^{\frac{1}{2}}}\; d\boldsymbol{y} + \int \frac{g(d_k^2)}{|\boldsymbol{\Sigma}^\circ|^{\frac{1}{2}}} \left(-\frac{p}{2}\right) a^{-(p/2+1)}\; d\boldsymbol{y} = 0. \\ &\Longrightarrow \frac{1}{a} \int \frac{-d_k^2 g^{(1)}(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; d\boldsymbol{y} - \frac{p}{2a} \int \frac{g(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; d\boldsymbol{y} = 0.\end{aligned}$$

Hence

$$\int \frac{d_k^2 g^{(1)}(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; d\boldsymbol{y} = -\frac{p}{2} \tag{4.5}$$

Using (4.5), the expectation of $d_k^2 v_2(d_k^2)$ can be simplified to

$$E\left[d_k^2 v_2(d_k^2)\right] = \int d_k^2 \frac{-2g^{(1)}(d_k^2)}{g(d_k^2)} \frac{g(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; d\boldsymbol{y} = -2 \int \frac{d_k^2 g^{(1)}(d_k^2)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\; d\boldsymbol{y} = -2\left(-\frac{p}{2}\right) = p.$$

□

Because the expectation of the estimating equation, $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$, is zero, it follows that the roots of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ have an asymptotic normal distribution by Theorem 4.1. The derivatives of the estimating function are needed to specify the limiting distribution of the estimators of the parameters. The derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ are given in the Theorem 4.7.

Theorem 4.7 (First derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$). *The derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ with respect to the parameters are*

$$
\begin{aligned}
\boldsymbol{l}^{(2)}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'} = -\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{R}_{xx}\right) + 2\boldsymbol{K}_{zx}\boldsymbol{R}_1\boldsymbol{K}'_{zx}\right] \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta}\boldsymbol{\beta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\zeta}'} = -\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{R}_{xz}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{K}_{zx}\boldsymbol{R}_1\boldsymbol{K}'_{zz}\right]\boldsymbol{F}^{(1)} \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\beta}\boldsymbol{\zeta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}}}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{F}^{(1)\prime}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{zx}\right) + \boldsymbol{K}_{zz}\boldsymbol{R}_3\boldsymbol{K}'_{zx}\right] \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta}\boldsymbol{\zeta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}}}{\partial \boldsymbol{\zeta}'} = \boldsymbol{F}^{(11)\prime}\left[\boldsymbol{I}_{\nu_2+\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{zz}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\right)\right] \\
&\quad - 2\boldsymbol{F}^{(1)\prime}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{zz}\boldsymbol{\Sigma}^{-1}\right) + \frac{1}{2}\boldsymbol{K}_{zz}\boldsymbol{R}_3\boldsymbol{K}'_{zz}\right]\boldsymbol{F}^{(1)} \\
&\quad + \boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)},
\end{aligned}
$$

where $\boldsymbol{R}_0$, $\boldsymbol{R}_2$, $\boldsymbol{R}_{xz}$, and $\boldsymbol{R}_{zz}$ are given in Theorem 4.4, $\boldsymbol{K}_{zx}$ and $\boldsymbol{K}_{zz}$ are given in Theorem 3.11, $\boldsymbol{F}^{(1)}$ is given in Appendix B,

$$
\begin{aligned}
\boldsymbol{R}_1 &= \frac{1}{n}\operatorname{Diag}\left(\frac{v_1^{(1)}(d_k)}{2d_k}\right), \\
\boldsymbol{R}_3 &= \frac{1}{n}\operatorname{Diag}\left(v_2^{(1)}(d_k^2)\right), \\
\boldsymbol{R}_{xx} &= \boldsymbol{X}'\boldsymbol{R}_0\boldsymbol{Z}, \ \textit{and} \\
\boldsymbol{R}_{zx} &= \boldsymbol{Z}'\boldsymbol{R}_2\boldsymbol{X}.
\end{aligned}
$$

Note that $E\left[\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\zeta}}\right] = E\left[\boldsymbol{l}^{(2)\prime}_{\boldsymbol{\zeta\beta}}\right] = \mathbf{0}$. *Further, because* $E\left[R_{zz}\right] = \boldsymbol{\Sigma}$, *it follows that*

$$E\left[\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta\zeta}}\right] = -\boldsymbol{F}^{(1)\prime}\boldsymbol{K}_{zz}\boldsymbol{R}_3\boldsymbol{K}'_{zz}\boldsymbol{F}^{(1)} - \boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)\prime}.$$

PROOF. The derivative of $\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}$ with respect to $\boldsymbol{\zeta}'$ proceeds as follows. First, using the identity $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}'\otimes\boldsymbol{A})\,\text{vec}\,\boldsymbol{A}$, $\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}$ can be written two ways:

$$\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}} = \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{I}_p\otimes\boldsymbol{x}_k v_1(d_k)\boldsymbol{z}'_k\right)\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \quad \text{and}$$
$$\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}} = \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{x}_k\right)\boldsymbol{z}_k v_1(d_k).$$

The derivative of $\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}$ is therefore

$$\begin{aligned}\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta\beta}} = \frac{\partial\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial\boldsymbol{\zeta}'} &= \left.\frac{\partial\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial\boldsymbol{\zeta}'}\right|_{v_1(d_k)\text{ fixed}} + \left.\frac{\partial\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial\boldsymbol{\zeta}'}\right|_{\boldsymbol{\Sigma}^{-1}\text{ fixed}} \\ &= \frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{I}_p\otimes\boldsymbol{x}_k v_1(d_k)\boldsymbol{z}'_k\right)\frac{\partial\,\text{vec}(\boldsymbol{\Sigma}^{-1})}{\partial\boldsymbol{\zeta}'} + \frac{1}{n}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{x}_k\right)\boldsymbol{z}_k\frac{\partial v_1(d_k)}{\partial\boldsymbol{\zeta}'}.\end{aligned} \tag{4.6}$$

The derivative of $v_1(d_k)$ with respect to $\boldsymbol{\zeta}$ is

$$\frac{\partial v_1(d_k)}{\partial\boldsymbol{\zeta}'} = \frac{v_1^{(1)}(d_k)}{2d_k}\frac{\partial h_k}{\partial\boldsymbol{\zeta}'} = -\frac{v_1^{(1)}(d_k)}{2d_k}\,\text{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\boldsymbol{z}'_k\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}. \tag{4.7}$$

Using (3.16) and (4.7), (4.6) can be simplified to

$$\begin{aligned}\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta\beta}} &= -\frac{1}{n}\sum_{k=1}^{n}\Bigg[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{x}_k v_1(d_k)\boldsymbol{z}'_k\boldsymbol{\Sigma}^{-1}\right)\frac{\partial\,\text{vec}(\boldsymbol{\Sigma})}{\partial\boldsymbol{\zeta}'} \\ &\qquad + \frac{1}{n}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{x}_k\right)\frac{v_1^{(1)}(d_k)}{2d_k}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right)'\Bigg]\boldsymbol{F}^{(1)} \\ &= -\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{X}'\boldsymbol{R}_0\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{K}_{zx}\boldsymbol{R}_1\boldsymbol{K}_{zz}\right]\boldsymbol{F}^{(1)} \\ &= -\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{R}_{xz}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{K}_{zx}\boldsymbol{R}_1\boldsymbol{K}_{zz}\right]\boldsymbol{F}^{(1)}.\end{aligned}$$

The proof for the other derivatives is similar to that above. □

The first order asymptotic distributions of the estimators of the location parameters, linear functions of the identified eigenvectors, and the eigenvalue parameters for the M-estimators are given in the next four theorems.

Theorem 4.8 (First Order Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)$). *The asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)$ *is*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) \xrightarrow{\text{dist}} N\left(\mathbf{0}, \bar{\boldsymbol{I}}^{-1}_{\boldsymbol{\theta}_m,\infty}\bar{\boldsymbol{C}}_{m,\infty}\bar{\boldsymbol{I}}^{-1\prime}_{\boldsymbol{\theta}_m,\infty}\right),$$

where $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty} = \lim_{n\to\infty} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,n}$, $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,n} = -\frac{1}{n}E\left[\boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}\right]$, $\bar{\boldsymbol{C}}_{m,\infty} = \lim_{n\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}\right)$.

Proof. By the multivariate central limit theorem, the asymptotic distribution of $\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ is

$$\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m} \xrightarrow{\text{dist}} N\left(\mathbf{0}, \bar{\boldsymbol{C}}_{m,\infty}\right).$$

Expand $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}\big|_{\boldsymbol{\theta}_m=\hat{\boldsymbol{\theta}}_m}$ in a Taylor Series about $\hat{\boldsymbol{\theta}}_m = \boldsymbol{\theta}_m$. Then

$$0 = \boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m} + \boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) + \frac{1}{2}\frac{\partial \boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}}{\partial \boldsymbol{\theta}'_m}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m \otimes \hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) + O_p(n^{-\frac{1}{2}}).$$

Solving for $\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ yields

$$\begin{aligned}\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m} &= -\left[\frac{1}{n}\boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}\right]\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) \\ &\quad - \frac{1}{2n^{\frac{1}{2}}}\left[\frac{1}{n}\frac{\partial \boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}}{\partial \boldsymbol{\theta}'_m}\right]\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) \otimes \sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)\right] + O_p(n^{-1}). \qquad (4.8)\end{aligned}$$

It is assumed that $\frac{1}{n}\frac{\partial \boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}}{\partial \boldsymbol{\theta}'_m}$ is bounded. Hence, the second term of (4.8) is of size $O_p(n^{-\frac{1}{2}})$. Further,

$$\frac{1}{n}\boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m} = -\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,n} + O_p(n^{-\frac{1}{2}}) \xrightarrow{\text{prob}} -\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty},$$

by the law of large numbers. Accordingly, (4.8) simplifies to

$$\frac{1}{\sqrt{n}} \boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m} = \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,n} \sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big) + O_p(n^{-\frac{1}{2}}). \tag{4.9}$$

Solving (4.9) for $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big)$ yields

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big) = \bar{\boldsymbol{I}}^{-1}_{\boldsymbol{\theta}_m,n} \left[\frac{1}{\sqrt{n}} \boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}\right] + O_p(n^{-\frac{1}{2}}).$$

Therefore, by Theorem 4.2, the asymptotic distribution of $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big)$ is

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big) \xrightarrow{\text{dist}} N\left(\boldsymbol{0}, \bar{\boldsymbol{I}}^{-1}_{\boldsymbol{\theta}_m,\infty} \bar{\boldsymbol{C}}_{m,\infty} \bar{\boldsymbol{I}}^{-1\prime}_{\boldsymbol{\theta}_m,\infty}\right).$$

□

Asymptotic distributions of the location parameter estimators, linear functions of identified eigenvectors, and functions of eigenvalues can be obtained by using the Delta method (Theorem 4.3).

Theorem 4.9 (First Order Asymptotic Distribution of $\sqrt{n}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)$ under M-estimation). *The asymptotic distribution of* $\sqrt{n}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)$ *is*

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) \xrightarrow{\text{dist}} N\left(\boldsymbol{0}, \boldsymbol{E}_1' \bar{\boldsymbol{I}}^{-1}_{\boldsymbol{\theta}_m,\infty} \bar{\boldsymbol{C}}_{m,\infty} \bar{\boldsymbol{I}}^{-1\prime}_{\boldsymbol{\theta}_m,\infty} \boldsymbol{E}_1\right),$$

where $\boldsymbol{E}_1$ *is the* $\nu \times \nu_1$ *matrix*

$$\boldsymbol{E}_1 = \begin{bmatrix} \boldsymbol{I}_{\nu_1} \\ \boldsymbol{0}_{\nu_2+\nu_3\times\nu_1} \end{bmatrix}.$$

Proof. Because $\boldsymbol{E}_1'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big) = \sqrt{n}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big)$, it follows by Theorem 4.2 that

$$\sqrt{n}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) \xrightarrow{\text{dist}} N\left(\mathbf{0}, \boldsymbol{E}_1'\bar{\boldsymbol{I}}^{-1}_{\boldsymbol{\theta}_m,\infty}\bar{\boldsymbol{C}}_{m,\infty}\bar{\boldsymbol{I}}^{-1'}_{\boldsymbol{\theta}_m,\infty}\boldsymbol{E}_1\right).$$

□

Theorem 4.10 (First Order Asymptotic Distribution of Linear Functions of Identified Eigenvectors for M-estimators). *The asymptotic distribution of the estimators of linear functions of the identified eigenvectors is*

$$\sqrt{n}\,\text{vec}\left[(\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{U}\right] \xrightarrow{\text{dist}} N\left(\mathbf{0}, (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}^{(1)}_{\boldsymbol{G}}\boldsymbol{E}_2'\bar{\boldsymbol{I}}^{-1}_{\boldsymbol{\theta}_m,\infty}\bar{\boldsymbol{C}}_{m,\infty}\bar{\boldsymbol{I}}^{-1'}_{\boldsymbol{\theta}_m,\infty}\boldsymbol{E}_2\boldsymbol{D}^{(1)'}_{\boldsymbol{G}}(\boldsymbol{U}\otimes\boldsymbol{\Gamma}')\right),$$

where $\boldsymbol{U}$ is an elementary matrix chosen so that $\hat{\boldsymbol{\Gamma}}\boldsymbol{U}$ contains identified eigenvectors and $\boldsymbol{E}_2$ is the $\nu \times \nu_2$ matrix

$$\boldsymbol{E}_2 = \begin{bmatrix} \underset{\nu_1\times\nu_2}{\mathbf{0}} \\ \boldsymbol{I}_{\nu_2} \\ \underset{\nu_3\times\nu_2}{\mathbf{0}} \end{bmatrix}.$$

Proof. Note that $\hat{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}_0\boldsymbol{G}$, where $\boldsymbol{G} \approx \boldsymbol{I}_p$. If $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$, then $\hat{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\hat{\boldsymbol{G}}$. The Taylor series expansion of $\boldsymbol{g}(\boldsymbol{\mu}) = \text{vec}\big(\hat{\boldsymbol{\Gamma}}\boldsymbol{U}\big) = \text{vec}\big(\boldsymbol{\Gamma}\hat{\boldsymbol{G}}\boldsymbol{U}\big) = (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\text{vec}\,\hat{\boldsymbol{G}}$ about $\hat{\boldsymbol{\mu}} = \mathbf{0}$ is

$$\begin{aligned}\boldsymbol{g}(\hat{\boldsymbol{\mu}}) = \boldsymbol{g}(\boldsymbol{\mu}) &+ \frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}'}\hat{\boldsymbol{\mu}} + \frac{1}{2}\frac{\partial^2\boldsymbol{g}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\big(\hat{\boldsymbol{\mu}}\otimes\hat{\boldsymbol{\mu}}\big) \\ &+ \frac{1}{6}\frac{\partial^3\boldsymbol{g}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\big(\hat{\boldsymbol{\mu}}\otimes\hat{\boldsymbol{\mu}}\otimes\hat{\boldsymbol{\mu}}\big) + O_p(n^{-\frac{3}{2}}).\end{aligned} \tag{4.10}$$

Solving (4.10) for $\sqrt{n}[\boldsymbol{g}(\hat{\boldsymbol{\mu}}) - \boldsymbol{g}(\boldsymbol{\mu})]$ yields

$$\sqrt{n}[\boldsymbol{g}(\hat{\boldsymbol{\mu}}) - \boldsymbol{g}(\boldsymbol{\mu})] = \frac{\partial\boldsymbol{g}}{\partial\boldsymbol{\mu}'}\sqrt{n}\hat{\boldsymbol{\mu}} + \frac{1}{2\sqrt{n}}\frac{\partial^2\boldsymbol{g}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\big(\sqrt{n}\hat{\boldsymbol{\mu}}\otimes\sqrt{n}\hat{\boldsymbol{\mu}}\big) + O_p(n^{-1})$$

$$\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right] = (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\left.\frac{\partial\,\mathrm{vec}\,\hat{\boldsymbol{G}}}{\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}}\sqrt{n}\hat{\boldsymbol{\mu}} + \frac{1}{2\sqrt{n}}(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\left.\frac{\partial^2\,\mathrm{vec}\,\hat{\boldsymbol{G}}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}}\left(\sqrt{n}\hat{\boldsymbol{\mu}}\otimes\sqrt{n}\hat{\boldsymbol{\mu}}\right) + O_p(n^{-1}),$$

$$\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right] = (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\sqrt{n}\hat{\boldsymbol{\mu}} + \frac{1}{2\sqrt{n}}(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(2)}\left(\sqrt{n}\hat{\boldsymbol{\mu}}\otimes\sqrt{n}\hat{\boldsymbol{\mu}}\right) + O_p(n^{-1}). \tag{4.11}$$

Note that $\boldsymbol{E}_2'\sqrt{n}(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_m) = \sqrt{n}\hat{\boldsymbol{\mu}}$. To show first-order asymptotics, only the terms of size $O_p(1)$ must be kept. Accordingly,

$$\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right] = (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\sqrt{n}(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_m) + O_p(n^{-\frac{1}{2}}). \tag{4.12}$$

It follows from (4.12) and Theorem 4.2 that

$$\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right] \xrightarrow{\text{dist}} N\left(\boldsymbol{0}, (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty}^{-1}\bar{\boldsymbol{C}}_{m,\infty}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty}^{-1'}\boldsymbol{E}_2\boldsymbol{D}_{\boldsymbol{G}}^{(1)'}(\boldsymbol{U}\otimes\boldsymbol{\Gamma}')\right).$$

□

Theorem 4.11 (First Order Asymptotic Distribution of the Eigenvalues for M-estimators). *The asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$ *is*

$$\sqrt{n}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}) \xrightarrow{\text{dist}} N\left(0, \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_3'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty}^{-1}\bar{\boldsymbol{C}}_{m,\infty}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty}^{-1'}\boldsymbol{E}_3\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)'}\right),$$

where $\boldsymbol{E}_3$ *is the* $\nu\times\nu_3$ *matrix*

$$\boldsymbol{E}_3 = \begin{bmatrix} \underset{\nu_1+\nu_2\times\nu_3}{\boldsymbol{0}} \\ \boldsymbol{I}_{\nu_3} \end{bmatrix}.$$

Proof. Expand $\boldsymbol{g}(\hat{\boldsymbol{\varphi}}) = \boldsymbol{\lambda}(\hat{\boldsymbol{\varphi}})$ about $\hat{\boldsymbol{\varphi}} = \boldsymbol{\varphi}$.

$$\boldsymbol{g}(\hat{\boldsymbol{\varphi}}) = \boldsymbol{g}(\boldsymbol{\varphi}) + \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\varphi}'}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big) + \frac{1}{2}\frac{\partial^2 \boldsymbol{g}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \otimes \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big) + \frac{1}{6}\frac{\partial^3 \boldsymbol{g}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \otimes \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \otimes \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big) + O_p(n^{-\frac{3}{2}}).$$

Solving for $\sqrt{n}[\boldsymbol{g}(\hat{\boldsymbol{\varphi}}) - \boldsymbol{g}(\boldsymbol{\varphi})]$ yields

$$\sqrt{n}[\boldsymbol{g}(\hat{\boldsymbol{\varphi}}) - \boldsymbol{g}(\boldsymbol{\varphi})] = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\varphi}'}\sqrt{n}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big) + \frac{1}{2n^{\frac{1}{2}}}\frac{\partial^2 \boldsymbol{g}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'}\big(\sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \otimes \sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})\big) + \frac{1}{6n}\frac{\partial^3 \boldsymbol{g}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'}\big(\sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \otimes \sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \otimes \sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})\big) + O_p(n^{-\frac{3}{2}}).$$

Define $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}(\hat{\boldsymbol{\varphi}})$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\varphi})$. Then the expansion can be simplified as

$$\sqrt{n}[\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}] = \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\sqrt{n}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big) + \frac{1}{2n^{\frac{1}{2}}}\boldsymbol{D}_{\boldsymbol{\lambda}}^{(2)}\big(\sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \otimes \sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})\big) + \frac{1}{6n}\boldsymbol{D}_{\boldsymbol{\lambda}}^{(3)}\big(\sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \otimes \sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \otimes \sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})\big) + O_p(n^{-\frac{3}{2}}). \tag{4.13}$$

Keeping only the terms of size $O_p(1)$ simplifies (4.13) as

$$\sqrt{n}[\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}] = \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\sqrt{n}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big) + O_p(n^{-\frac{1}{2}}) \tag{4.14}$$

Because $\boldsymbol{E}_3'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big) = \sqrt{n}\big(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\big)$, then (4.14) becomes

$$\sqrt{n}[\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}] = \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_3'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m\big) + O_p(n^{-\frac{1}{2}}) \tag{4.15}$$

It follows from (4.15) and Theorem 4.2 that

$$\sqrt{n}\big(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\big) \xrightarrow{\text{dist}} N\left(0, \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_3'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty}^{-1}\bar{\boldsymbol{C}}_{m,\infty}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,\infty}^{-1\prime}\boldsymbol{E}_3\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}\right).$$

□

Sandwich Estimator

An estimator of the asymptotic variance, called the sandwich estimator, can be constructed from the data. First, write $\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ as

$$\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m} = \sum_{k=1}^{n} \boldsymbol{f}_k,$$

where $\boldsymbol{f}_k$ and $\boldsymbol{f}_j$ are independent when $j \neq k$. Note that $\boldsymbol{f}_k$ is the same as the $\boldsymbol{g}_k$ defined in (3.1). For the M-estimators, $\boldsymbol{f}_k$ can be written as

$$\boldsymbol{f}_k = \begin{bmatrix} \boldsymbol{f}_{k,\boldsymbol{\beta}} \\ \boldsymbol{f}_{k,\boldsymbol{\zeta}} \end{bmatrix} \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \begin{bmatrix} \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{x}_k\right) v_1(d_k) \\ \boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}} \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right) v_2(d_k^2) \end{bmatrix}$$

Because $E\left[\boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}'_{k,\boldsymbol{\zeta}}\right] = \boldsymbol{0}$, then a consistent estimator of the asymptotic variance of $\frac{1}{\sqrt{n}}\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_m}$ is

$$\hat{\hat{\boldsymbol{C}}}_{m,n} = \sum_{k=1}^{n} \begin{bmatrix} \boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}'_{k,\boldsymbol{\beta}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{f}_{k,\boldsymbol{\zeta}}\boldsymbol{f}'_{k,\boldsymbol{\zeta}} \end{bmatrix}.$$

The sum above can be rewritten as the matrix product

$$\hat{\hat{\boldsymbol{C}}}_{m,n} = \begin{bmatrix} \boldsymbol{F}_{m,\boldsymbol{\beta}}\boldsymbol{F}'_{m,\boldsymbol{\beta}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_{m,\boldsymbol{\zeta}}\boldsymbol{F}'_{m,\boldsymbol{\zeta}} \end{bmatrix}$$

where

$$\boldsymbol{F}_{m,\boldsymbol{\beta}} = \sqrt{n}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\boldsymbol{R}_0\right) \text{ and } \boldsymbol{F}_{m,\boldsymbol{\zeta}} = \sqrt{n}\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{R}_2\right)$$

The sandwich estimator of the variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)$ is therefore

$$\widehat{\text{Var}}\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)\right] = \hat{\boldsymbol{K}}_2^{-1}\hat{\hat{\boldsymbol{C}}}_{m,n}\hat{\boldsymbol{K}}_2^{\prime -1},$$

where $\boldsymbol{K}_2 = -\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_m,n} = \frac{1}{n}E\left[\boldsymbol{l}^{(3)}_{\boldsymbol{\theta}_m\boldsymbol{\theta}_m}\right]$. Note that many of the terms in $\hat{\boldsymbol{K}}_2$ have expectation zero. Hence, $\hat{\boldsymbol{K}}_2$ can be written as

$$\hat{\boldsymbol{K}}_2 = \frac{1}{n}\begin{bmatrix}\hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}\end{bmatrix}.$$

S-estimators

First Order Asymptotic Distributions

Define

$$\boldsymbol{\theta}_s \stackrel{\text{def}}{=} \begin{bmatrix}\boldsymbol{\theta}_{s_1} \\ \boldsymbol{\theta}_{s_2}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\theta}_{s_1} \\ \eta_\psi\end{bmatrix} = \begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\mu} \\ \boldsymbol{\psi} \\ \eta_\psi\end{bmatrix},$$

where $\boldsymbol{\theta}_{s_1}$ is given in (3.42). The solutions of $\frac{\partial w}{\partial \boldsymbol{\theta}_{s_1}} = \mathbf{0}$ also satisfy the equation $c\frac{\partial w}{\partial \boldsymbol{\theta}_{s_1}} = \mathbf{0}$, where c is a nonzero constant. Accordingly, the estimating equation for the solutions of $\frac{\partial w}{\partial \boldsymbol{\theta}_{s_1}}$ can be simplified. The simplification is given in Theorem 4.12.

Theorem 4.12 (S-estimating Equations). *The solutions of the estimating equations* $\frac{\partial w}{\partial \boldsymbol{\theta}_{s_1}} = \mathbf{0}$ *and* $q = 0$ *also satisfy the estimating equation*

$$\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s} = \begin{bmatrix}\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}} \\ \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}} \\ \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}} \\ \boldsymbol{l}^{(1)}_{\eta_\psi}\end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} n\frac{\partial w}{\partial \boldsymbol{\beta}}(\boldsymbol{a}'\boldsymbol{v}_2)(\boldsymbol{v}_2'\frac{\partial w^*}{\partial \boldsymbol{\varphi}})^{-1} \\ n\frac{\partial w}{\partial \boldsymbol{\mu}}(\boldsymbol{a}'\boldsymbol{v}_2)(\boldsymbol{v}_2'\frac{\partial w^*}{\partial \boldsymbol{\varphi}})^{-1} \\ \frac{\partial w}{\partial \boldsymbol{\psi}}(\boldsymbol{a}'\boldsymbol{v}_2) \\ nq^*\end{bmatrix} = \begin{bmatrix} -n\frac{\partial q^*}{\partial \boldsymbol{\beta}} \\ -n\frac{\partial q^*}{\partial \boldsymbol{\mu}} \\ (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^\perp_{\nu_3}\right)\left(\boldsymbol{a} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) \\ nq^*\end{bmatrix} = \mathbf{0},$$

where $\boldsymbol{a}$ *is defined in Theorem 3.9,* $\boldsymbol{N}^\perp_{\nu_3}$ *is defined in Theorem 3.10, derivatives of* w *and* w^* *are defined in Theorem 3.12,* q^* *is defined in* (3.40)*, and the derivatives of* q^* *are defined in Theorem 3.9.*

Proof. The first derivative of w with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial w}{\partial \boldsymbol{\beta}} = \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)\prime} \frac{\partial w^*}{\partial \boldsymbol{\varphi}} = -\frac{\partial q^*}{\partial \boldsymbol{\beta}} (\boldsymbol{v}_2' \boldsymbol{a})^{-1} \boldsymbol{v}_2' \frac{\partial w^*}{\partial \boldsymbol{\varphi}}.$$

Note that $(\boldsymbol{v}_2'\boldsymbol{a})^{-1}\boldsymbol{v}_2' \frac{\partial w^*}{\partial \boldsymbol{\varphi}}$ is a nonzero scalar. Hence, the estimating function for $\boldsymbol{\beta}$ can be defined as

$$\boldsymbol{l}_{\boldsymbol{\beta}}^{(1)} = n \frac{\partial w}{\partial \boldsymbol{\beta}} (\boldsymbol{a}'\boldsymbol{v}_2) \left(\boldsymbol{v}_2' \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right)^{-1} = -n \frac{\partial q^*}{\partial \boldsymbol{\beta}}.$$

The method is similar for the estimating equations for $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$. The estimating equation $\boldsymbol{l}_{\eta_\psi}^{(1)}$ is a multiple of the original constraint $q = 0$. □

The estimating function $\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$ satisfies the conditions of Theorem 4.1. In particular, the expectation of the function under an elliptical distribution is $\mathbf{0}$. The proof is given in Theorem 4.13.

Theorem 4.13 (Expectation of the S-estimating Equations). *Under an elliptical distribution, the expectation of* $\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$ *is zero.*

Proof. First, the expectation of $\boldsymbol{l}_{\boldsymbol{\beta}}^{(1)}$.

$$\begin{aligned}
E\left[\boldsymbol{l}_{\boldsymbol{\beta}}^{(1)}\right] &= 2\,\mathrm{vec}\big(E\left[\boldsymbol{W}_{xz}\right]\boldsymbol{\Sigma}^{-1}\big) \\
&= \frac{2}{n}\sum_{k=1}^{n} \mathrm{vec}\big(\boldsymbol{x}_k E\left[w_k^{(1)} \boldsymbol{z}_k'\right]\boldsymbol{\Sigma}^{-1}\big) \\
&= \frac{1}{n}\sum_{k=1}^{n} \mathrm{vec}\left(\boldsymbol{x}_k E\left[\frac{\rho^{(1)}(d_k)}{d_k}\boldsymbol{z}_k'\right]\boldsymbol{\Sigma}^{-1}\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} \mathrm{vec}\left(\boldsymbol{x}_k E\left[\rho^{(1)}(d_k)\frac{\boldsymbol{z}_k'}{d_k}\right]\boldsymbol{\Sigma}^{-1}\right) \qquad (4.16)
\end{aligned}$$

Because $\frac{\boldsymbol{z}_k'}{d_k}$ is independent of $\rho^{(1)}(d_k)$, then the expectation in (4.16) can be factored. Because $E\left[\frac{\boldsymbol{z}_k'}{d_k}\right] = \mathbf{0}$ by Lemma 4.5, it follows that (4.16) can be simplified as

$$E\left[\boldsymbol{l}_{\boldsymbol{\beta}}^{(1)}\right] = \frac{1}{n}\sum_{k=1}^{n} \operatorname{vec}\left(\boldsymbol{x}_k E\left[\rho^{(1)}(d_k)\right] E\left[\frac{\boldsymbol{z}_k'}{d_k}\right] \boldsymbol{\Sigma}^{-1}\right) = \mathbf{0},$$

The expectation of $\boldsymbol{l}_{\boldsymbol{\mu}}^{(1)}$ can be derived as follows. First, the expectation of $\boldsymbol{W}_{zz}$ is

$$\begin{aligned} E\left[\boldsymbol{W}_{zz}\right] &= \frac{1}{n}\sum_{k=1}^{n} E\left[\boldsymbol{z}_k w_k^{(1)} \boldsymbol{z}_k'\right] \\ &= \frac{1}{2n}\sum_{k=1}^{n} E\left[\boldsymbol{z}_k \frac{\rho^{(1)}(d_k)}{d_k} \boldsymbol{z}_k'\right] \\ &= \frac{1}{2n}\sum_{k=1}^{n} E\left[\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k} d_k \rho^{(1)}(d_k)\right]. \end{aligned} \tag{4.17}$$

Because $\frac{\boldsymbol{z}_k}{d_k}$ is independent of any function of d_k, the expectation in (4.17) can be factored. Because $E\left[\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k}\right] = \frac{1}{p}\boldsymbol{\Sigma}$, then (4.17) can be written as

$$\begin{aligned} E\left[\boldsymbol{W}_{zz}\right] &= \frac{1}{2n}\sum_{k=1}^{n} E\left[\frac{\boldsymbol{z}_k}{d_k}\frac{\boldsymbol{z}_k'}{d_k}\right] E\left[d_k \rho^{(1)}(d_k)\right] \\ &= \frac{1}{2np}\sum_{k=1}^{n} E\left[d_k \rho^{(1)}(d_k)\right] \boldsymbol{\Sigma} \\ &= c\boldsymbol{\Sigma}, \end{aligned}$$

where $c = \frac{1}{2np}\sum_{k=1}^{n} E\left[d_k \rho^{(1)}(d_k)\right]$. Hence,

$$\begin{aligned} E\left[\boldsymbol{l}_{\boldsymbol{\mu}}^{(1)}\right] &= \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime} \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1} E\left[\boldsymbol{W}_{zz}\right] \boldsymbol{\Sigma}^{-1}\right) \\ &= c\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime} \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \boldsymbol{v}_2' \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \\ &= 2c\boldsymbol{A}_1' \left(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right) \left(\boldsymbol{\Lambda}\boldsymbol{\Gamma}_0' \otimes \boldsymbol{\Gamma}_0'\right) \operatorname{vec}\left(\boldsymbol{\Gamma}_0 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma}_0'\right) \\ &= 2c\boldsymbol{A}_1' \left(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right) \operatorname{vec}\left(\boldsymbol{\Gamma}_0' \boldsymbol{\Gamma}_0 \boldsymbol{\Lambda}^{-1} \boldsymbol{\Gamma}_0' \boldsymbol{\Gamma}_0 \boldsymbol{\Lambda}\right) \end{aligned}$$

$$= 2c\boldsymbol{A}_1'\left(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right)\operatorname{vec}\boldsymbol{I}_p$$

$$= 2c\boldsymbol{A}_1'\left(\operatorname{vec}\boldsymbol{I}_p - \operatorname{vec}\boldsymbol{I}_p\right)$$

$$= \boldsymbol{0}.$$

The expectation of $\boldsymbol{a}$ can written as

$$E\left[\boldsymbol{a}\right] = E\left[-\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right)\right] = -\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}E\left[\boldsymbol{W}_{zz}\right]\boldsymbol{\Sigma}^{-1}\right)$$
$$= -c\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right) = -\frac{c}{n}\frac{\partial w^*}{\partial\boldsymbol{\varphi}}.$$

Hence, the expectation of $\boldsymbol{l}_{\boldsymbol{\psi}}^{(1)}$ can be written as

$$\begin{aligned}E\left[\boldsymbol{l}_{\boldsymbol{\psi}}^{(1)}\right] &= E\left[\left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left(\boldsymbol{a}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right)\right]\\ &= \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left(E\left[\boldsymbol{a}\right]\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right)\\ &= -\frac{c}{n}\left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left(\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right).\end{aligned}\tag{4.18}$$

Note that $\boldsymbol{N}_{\nu_3}^{\perp}$ is the ppo for the null space of symmetric matrices. Accordingly, when $\boldsymbol{A}$ is symmetric, $\boldsymbol{N}_{\nu_3}^{\perp}\operatorname{vec}(\boldsymbol{A}) = \boldsymbol{0}$. Hence, (4.18) simplifies to

$$\begin{aligned}E\left[\boldsymbol{l}_{\boldsymbol{\psi}}^{(1)}\right] &= -\frac{c}{n}\left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left(\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right)\\ &= -\frac{c}{n}\left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\operatorname{vec}\left(\frac{\partial w^*}{\partial\boldsymbol{\psi}'}\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right)\\ &= \boldsymbol{0}.\end{aligned}$$

Write $\boldsymbol{\Sigma}$ as $\kappa\boldsymbol{\Sigma}^*$, where κ is the positive scalar that satisfies

$$E\left[\rho\left\{\left[\boldsymbol{z}_k'(\kappa\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{z}_k\right]^{\frac{1}{2}}\right\}\right] = b_0.$$

Hence,

$$E\left[\boldsymbol{l}^{(1)}_{\eta_\psi}\right] = E\left[nq^*\right] = E\left[\sum_{k=1}^{n} \rho(d_k)\right] - nb_0 = \sum_{k=1}^{n} E\left[\rho(d_k)\right] - nb_0$$
$$= nE\left[\rho(d_k)\right] - nb_0 = nb_0 - nb_0$$
$$= 0.$$

If b_0 is chosen according to the underlying distribution from which the data have been sampled, then $\kappa = 1$, otherwise κ is a positive scalar. □

Because the expectation of the estimating equation, $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$, is zero, it follows that the roots of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$ have an asymptotic normal distribution by Theorem 4.1. The derivatives of the estimating function are needed to specify the limiting distribution of the estimators of the parameters. The derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$ are given in the following theorem:

THEOREM 4.14. *The derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$ with respect to the parameters are:*

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}},$$
$$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\mu}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}},$$
$$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\psi}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{V}_1,$$
$$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \eta_\psi} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{v}_2,$$
$$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\mu}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}},$$
$$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\mu}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}},$$

$$
\begin{aligned}
\boldsymbol{l}^{(2)}_{\boldsymbol{\psi}\boldsymbol{\mu}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\psi}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{V}_1, \\
\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\mu}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \eta_\psi} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{v}_2, \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\beta}\boldsymbol{\psi}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \boldsymbol{\beta}'} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right), \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\mu}\boldsymbol{\psi}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \boldsymbol{\mu}'} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right), \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\psi}\boldsymbol{\psi}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \boldsymbol{\psi}'} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right]\boldsymbol{V}_1, \\
\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\psi}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \eta_\psi} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right]\boldsymbol{v}_2, \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\beta}\eta_\psi} &= \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \boldsymbol{\beta}'} = n\frac{\partial q^*}{\partial \boldsymbol{\beta}'}, \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\mu}\eta_\psi} &= \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \boldsymbol{\mu}'} = n\frac{\partial q^*}{\partial \boldsymbol{\mu}'}, \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\psi}\eta_\psi} &= \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \boldsymbol{\psi}'} = n\boldsymbol{a}'\boldsymbol{V}_1, \text{ and} \\
\boldsymbol{l}^{(2)}_{\eta_\psi\eta_\psi} &= \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \eta_\psi} = n\boldsymbol{a}'\boldsymbol{v}_2,
\end{aligned}
$$

where $\boldsymbol{a}$ *and the derivatives of* q^* *are given in Theorem 3.9 and Theorem 3.11, the derivatives of* w^* *are given in Theorem 3.12 and Theorem 3.13, and* $\boldsymbol{N}^{\perp}_{\nu_3}$ *is given in Theorem 3.10.*

PROOF. The derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}$, $\boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}$, and $\boldsymbol{l}^{(1)}_{\eta_\psi}$ are trivial. The derivative of $\boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}$ with respect to η_ψ can be obtained as follows:

$$
\begin{aligned}
\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\psi}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \eta_\psi} &= \frac{\partial}{\partial \eta_\psi}(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\boldsymbol{a} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) \\
&= (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\frac{\partial}{\partial \boldsymbol{\varphi}'}\left[\boldsymbol{a} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right]\boldsymbol{v}_2 \\
&= (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{I}_{(\nu_3,\nu_3)}\left(\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \boldsymbol{a}\right)\right]\boldsymbol{v}_2 \\
&= (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right]\boldsymbol{v}_2
\end{aligned}
$$

The other derivatives of $\boldsymbol{l}_{\boldsymbol{\psi}}^{(1)}$ are obtained in a similar manner. □

The first order asymptotic distributions of the estimators of the location parameters, linear functions of the identified eigenvectors, and the eigenvalue parameters are given in the next three theorems.

Theorem 4.15 (First Order Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)$). *The asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)$ *is*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s) \xrightarrow{\text{dist}} N\left(\mathbf{0}, \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1} \bar{\boldsymbol{C}}_{s,\infty} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1\prime}\right),$$

where $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty} = \lim_{n\to\infty} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,n}$, $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,n} = -\frac{1}{n} E\left[\boldsymbol{l}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}^{(2)}\right]$, $\bar{\boldsymbol{C}}_{s,\infty} = \lim_{n\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{n}} \boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}\right)$.

Proof. The proof is similar to Theorem 4.8. □

Asymptotic distributions of the location parameter estimators, linear functions of identified eigenvectors, and functions of eigenvalues can be obtained by using the Delta method (Theorem 4.3).

Theorem 4.16 (First Order Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$). *The asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *is*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\text{dist}} N\left(\mathbf{0}, \boldsymbol{E}_1' \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1} \bar{\boldsymbol{C}}_{s,\infty} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1\prime} \boldsymbol{E}_1\right),$$

where $\boldsymbol{E}_1$ *is the* $\nu \times \nu_1$ *matrix*

$$\boldsymbol{E}_1 = \begin{bmatrix} \boldsymbol{I}_{\nu_1} \\ \mathbf{0}_{\nu_2+\nu_3\times\nu_1} \end{bmatrix}.$$

PROOF. The proof is similar to Theorem 4.9. □

THEOREM 4.17 (FIRST ORDER ASYMPTOTIC DISTRIBUTION OF LINEAR FUNCTIONS OF IDENTIFIED EIGENVECTORS). *The asymptotic distribution of the estimators of the linear functions of the identified eigenvectors* $\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right]$ *is*

$$\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right] \xrightarrow{\text{dist}} N\left(\mathbf{0}, (\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1}\bar{\boldsymbol{C}}_{s,\infty}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1\prime}\boldsymbol{E}_2\boldsymbol{D}_{\boldsymbol{G}}^{(1)\prime}\,(\boldsymbol{U}\otimes\boldsymbol{\Gamma}')\right),$$

where $\boldsymbol{U}$ *is an elementary matrix chosen so that* $\hat{\boldsymbol{\Gamma}}\boldsymbol{U}$ *contains identified eigenvectors and* $\boldsymbol{E}_2$ *is the* $\nu\times\nu_2$ *matrix*

$$\boldsymbol{E}_2 = \begin{bmatrix} \underset{\nu_1\times\nu_2}{\mathbf{0}} \\ \boldsymbol{I}_{\nu_2} \\ \underset{\nu_3\times\nu_2}{\mathbf{0}} \end{bmatrix}.$$

PROOF. The proof is similar to Theorem 4.10. □

THEOREM 4.18 (FIRST ORDER ASYMPTOTIC DISTRIBUTION OF THE EIGENVALUES). *The asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$ *is*

$$\sqrt{n}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}) \xrightarrow{\text{dist}} N\left(0, \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_{\boldsymbol{\lambda}}'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1}\bar{\boldsymbol{C}}_{s,\infty}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,\infty}^{-1\prime}\boldsymbol{E}_{\boldsymbol{\lambda}}\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)\prime}\right),$$

where $\boldsymbol{E}_{\boldsymbol{\lambda}}$ *is the* $\nu\times\nu_3$ *matrix*

$$\boldsymbol{E}_{\boldsymbol{\lambda}} = \begin{bmatrix} \underset{\nu_1+\nu_2\times\nu_3}{\mathbf{0}} \\ \boldsymbol{V}_1' \\ \boldsymbol{v}_2' \end{bmatrix}.$$

PROOF. The proof in Theorem 4.11 is similar with the exception that $\boldsymbol{E}_{\boldsymbol{\lambda}}'\sqrt{n}(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s)=\sqrt{n}(\hat{\boldsymbol{\varphi}}-\boldsymbol{\varphi})$. □

The coverage of $1-\alpha$ confidence intervals based on first order asymptotic distributions is $1-\alpha+O(n^{-\frac{1}{2}})$. Second-order corrections can be used to improve the coverage rate to $1-\alpha+O(n^{-1})$. These second-order corrections depend on the bias and skewness of the estimators. To estimate the bias and skewness, a higher-order expansion of the estimator can be employed.

Second Order Asymptotic Expansions

Theorem 4.19 (Second Order Asymptotic Expansion of $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)$). *The second order asymptotic expansion of* $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)$ *is*

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)=\hat{\boldsymbol{\delta}}_0+\frac{1}{\sqrt{n}}\hat{\boldsymbol{\delta}}_1+O_p(n^{-1}), \tag{4.19}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_0 &= -\boldsymbol{K}_2^{-1}\boldsymbol{Z}_1,\\
\hat{\boldsymbol{\delta}}_1 &= -\boldsymbol{K}_2^{-1}\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0+\frac{1}{2}\boldsymbol{K}_3\big(\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_0\big)\right],\\
\boldsymbol{x}^{\otimes i} &= \underbrace{x\otimes x\otimes\cdots\otimes \boldsymbol{x}}_{i\ \mathit{times}},\\
\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(i)} &= \frac{\partial^i \boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}}{(\partial\boldsymbol{\theta}_s')^{\otimes i-1}},\\
\boldsymbol{K}_i &= \frac{1}{n}E\left[\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(i)}\right],\ \ \mathit{and}\\
\boldsymbol{Z}_i &= \sqrt{n}\left(\frac{1}{n}\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(i)}-\boldsymbol{K}_i\right).
\end{aligned}$$

Proof. Note that $\boldsymbol{Z}_i = O_p(1)$, $\boldsymbol{K}_i = O(1)$, $\hat{\boldsymbol{\delta}}_0 = O_p(1)$, and $\hat{\boldsymbol{\delta}}_1 = O_p(1)$. Expanding $\boldsymbol{l}^{(1)}_{\hat{\boldsymbol{\theta}}_s}$ in a Taylor series about $\hat{\boldsymbol{\theta}}_s = \boldsymbol{\theta}_s$ yields

$$\boldsymbol{l}^{(1)}_{\hat{\boldsymbol{\theta}}_s} = 0 = \boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s} + \boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_s}\big(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\big) + \frac{1}{2}\boldsymbol{l}^{(3)}_{\boldsymbol{\theta}_s}\big(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s \otimes \hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\big) + O_p(n^{-\frac{1}{2}}). \tag{4.20}$$

Define

$$\omega \stackrel{\text{def}}{=} \sqrt{n}\big(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\big). \tag{4.21}$$

Because $\boldsymbol{Z}_i = \sqrt{n}\left(\frac{1}{n}\boldsymbol{l}^{(i)}_{\boldsymbol{\theta}_s} - \boldsymbol{K}_i\right)$, it follows that

$$\boldsymbol{l}^{(i)}_{\boldsymbol{\theta}_s} = \sqrt{n}\boldsymbol{Z}_i + n\boldsymbol{K}_i. \tag{4.22}$$

By using (4.21) and (4.22), (4.20) can be simplified as

$$0 = \sqrt{n}\boldsymbol{Z}_1 + n\boldsymbol{K}_1 + \big(\sqrt{n}\boldsymbol{Z}_2 + n\boldsymbol{K}_2\big)\frac{1}{\sqrt{n}}\omega + \frac{1}{2}\big(\sqrt{n}\boldsymbol{Z}_3 + n\boldsymbol{K}_3\big)\frac{\omega\otimes\omega}{n} + O_p(n^{-\frac{1}{2}}). \tag{4.23}$$

It follows that dividing (4.23) by $\sqrt{n}$ yields

$$0 = \boldsymbol{Z}_1 + \left(\frac{\boldsymbol{Z}_2}{\sqrt{n}} + \boldsymbol{K}_2\right)\omega + \frac{1}{2}\left(\frac{\boldsymbol{Z}_3}{n} + \frac{\boldsymbol{K}_3}{\sqrt{n}}\right)\big(\omega\otimes\omega\big) + O_p(n^{-1}), \tag{4.24}$$

because $\boldsymbol{K}_1 = \boldsymbol{0}$ (by Theorem 4.13). Write ω as

$$\omega = \hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\hat{\boldsymbol{\delta}}_1 + O_p(n^{-1}). \tag{4.25}$$

Substituting (4.25) into (4.24) yields

$$\begin{aligned} 0 = \boldsymbol{Z}_1 &+ \left(\frac{\boldsymbol{Z}_2}{\sqrt{n}} + \boldsymbol{K}_2\right)\left(\hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\hat{\boldsymbol{\delta}}_1 + O_p(n^{-1})\right) \\ &+ \frac{1}{2}\left(\frac{\boldsymbol{Z}_3}{n} + \frac{\boldsymbol{K}_3}{\sqrt{n}}\right)\left(\hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\hat{\boldsymbol{\delta}}_1 + O_p(n^{-1}) \otimes \hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\hat{\boldsymbol{\delta}}_1 + O_p(n^{-1})\right) + O_p(n^{-1}). \end{aligned} \tag{4.26}$$

Multiplying (4.26) out and keeping only the terms of size $O_p(n^{-\frac{1}{2}})$ and larger yields

$$0 = \boldsymbol{Z}_1 + \boldsymbol{K}_2\hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0 + \boldsymbol{K}_2\hat{\boldsymbol{\delta}}_1 + \frac{1}{2}\boldsymbol{K}_3\big(\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\big)\right] + O_p(n^{-1}). \tag{4.27}$$

Because (4.27) is written in the form of

$$\mathbf{0} = \boldsymbol{C}_1 + \frac{1}{\sqrt{n}}\boldsymbol{C}_2 + \frac{1}{n}\boldsymbol{C}_3 + \cdots,$$

then it must be that

$$\boldsymbol{C}_1 = \boldsymbol{C}_2 = \boldsymbol{C}_3 = \cdots = \mathbf{0}.$$

Hence,

$$\boldsymbol{Z}_1 + \boldsymbol{K}_2\hat{\boldsymbol{\delta}}_0 = \mathbf{0} \Longrightarrow \hat{\boldsymbol{\delta}}_0 = -\boldsymbol{K}_2^{-1}\boldsymbol{Z}_1, \text{ and}$$

$$\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0 + \boldsymbol{K}_2\hat{\boldsymbol{\delta}}_1 + \frac{1}{2}\boldsymbol{K}_3\big(\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\big) = \mathbf{0} \Longrightarrow \hat{\boldsymbol{\delta}}_1 = -\boldsymbol{K}_2^{-1}\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0 + \frac{1}{2}\boldsymbol{K}_3\big(\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\big)\right].$$

□

THEOREM 4.20 (SECOND ORDER EXPANSIONS FOR THE ESTIMATORS). *Second order expansions for the estimators in Theorem 4.16, Theorem 4.17, and Theorem 4.18 are:*

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \boldsymbol{E}_1'\hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\boldsymbol{E}_1'\hat{\boldsymbol{\delta}}_1 + O_p(n^{-1}) \\
\sqrt{n}\operatorname{vec}\left[(\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{U}\right] &= (\boldsymbol{U}' \otimes \boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}(\boldsymbol{U}' \otimes \boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\hat{\boldsymbol{\delta}}_1 \\
&\quad + \frac{1}{2\sqrt{n}}(\boldsymbol{U}' \otimes \boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(2)}\big(\boldsymbol{E}_2' \otimes \boldsymbol{E}_2'\big)\Big(\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\Big) + O_p(n^{-1}), \textit{ and} \\
\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) &= \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_{\boldsymbol{\lambda}}'\hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}}\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_{\boldsymbol{\lambda}}'\hat{\boldsymbol{\delta}}_1 \\
&\quad + \frac{1}{2\sqrt{n}}\boldsymbol{D}_{\boldsymbol{\lambda}}^{(2)}\big(\boldsymbol{E}_{\boldsymbol{\lambda}}' \otimes \boldsymbol{E}_{\boldsymbol{\lambda}}'\big)\Big(\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\Big) + O_p(n^{-1}),
\end{aligned}$$

where $\hat{\boldsymbol{\delta}}_0$ *and* $\hat{\boldsymbol{\delta}}_1$ *are defined in Theorem 4.19.*

Proof. Because $\boldsymbol{E}_1'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)=\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, it follows that multiplying (4.19) on the left by $\boldsymbol{E}_1'$ yields

$$\begin{aligned}\boldsymbol{E}_1'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)&=\boldsymbol{E}_1'\left[\hat{\boldsymbol{\delta}}_0+\frac{1}{\sqrt{n}}\hat{\boldsymbol{\delta}}_1+O_p(n^{-1})\right],\\ \Longrightarrow\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})&=\boldsymbol{E}_1'\hat{\boldsymbol{\delta}}_0+\frac{1}{\sqrt{n}}\boldsymbol{E}_1'\hat{\boldsymbol{\delta}}_1+O_p(n^{-1}).\end{aligned}$$

Because $\boldsymbol{E}_2'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)=\sqrt{n}\hat{\boldsymbol{\mu}}$, it follows that (4.11) simplifies to

$$\begin{aligned}\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right]&=(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)\\&+\frac{1}{2\sqrt{n}}(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(2)}\big(\boldsymbol{E}_2'\otimes\boldsymbol{E}_2'\big)\Big(\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)\otimes\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)\Big)+O_p(n^{-1}).\end{aligned}\tag{4.28}$$

Substituting (4.19) into (4.28) and keeping only the terms that are of order $O_p(n^{-1})$ or larger yields

$$\begin{aligned}\sqrt{n}\,\mathrm{vec}\left[(\hat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\boldsymbol{U}\right]&=(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\hat{\boldsymbol{\delta}}_0+\frac{1}{\sqrt{n}}(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(1)}\boldsymbol{E}_2'\hat{\boldsymbol{\delta}}_1\\&+\frac{1}{2\sqrt{n}}(\boldsymbol{U}'\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_{\boldsymbol{G}}^{(2)}\big(\boldsymbol{E}_2'\otimes\boldsymbol{E}_2'\big)\Big(\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_0\Big)+O_p(n^{-1}).\end{aligned}$$

Lastly, since $\boldsymbol{E}_{\boldsymbol{\lambda}}'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)=\sqrt{n}(\hat{\boldsymbol{\varphi}}-\boldsymbol{\varphi})$, it follows that (4.13) simplifies to

$$\begin{aligned}\sqrt{n}[\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}]&=\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)}\boldsymbol{E}_{\boldsymbol{\lambda}}'\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)\\&+\frac{1}{2n^{\frac{1}{2}}}\boldsymbol{D}_{\boldsymbol{\lambda}}^{(2)}\big(\boldsymbol{E}_{\boldsymbol{\lambda}}'\otimes\boldsymbol{E}_{\boldsymbol{\lambda}}'\big)\big(\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)\otimes\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s-\boldsymbol{\theta}_s\big)\big)+O_p(n^{-1}).\end{aligned}\tag{4.29}$$

Substituting (4.19) into (4.29) and keeping only the terms that are of order $O_p(n^{-1})$ or larger yields

$$\sqrt{n}[\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}] = \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)} \boldsymbol{E}_{\boldsymbol{\lambda}}' \hat{\boldsymbol{\delta}}_0 + \frac{1}{\sqrt{n}} \boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)} \boldsymbol{E}_{\boldsymbol{\lambda}}' \hat{\boldsymbol{\delta}}_1 + \frac{1}{2\sqrt{n}} \boldsymbol{D}_{\boldsymbol{\lambda}}^{(2)} \left(\boldsymbol{E}_{\boldsymbol{\lambda}}' \otimes \boldsymbol{E}_{\boldsymbol{\lambda}}' \right) \left(\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0 \right) + O_p(n^{-1}).$$

□

To obtain second order expansions of $\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)$, expressions for the second derivative of $\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$ (like third derivatives of w) with respect to the parameters must be obtained. These derivatives are also needed for estimating the bias of the estimators.

The second derivatives of $\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$ with respect to the parameters are given in Theorem 4.21.

Theorem 4.21 (Second derivatives of $\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$). *The second derivatives of the estimating function $\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ are*

$$\boldsymbol{l}_{\boldsymbol{\beta\beta\beta}}^{(3)} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}},$$

$$\boldsymbol{l}_{\boldsymbol{\beta\mu\beta}}^{(3)} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}},$$

$$\boldsymbol{l}_{\boldsymbol{\mu\beta\beta}}^{(3)} = \boldsymbol{l}_{\boldsymbol{\beta\mu\beta}}^{(3)} \boldsymbol{I}_{(\nu_2,\nu_1)},$$

$$\boldsymbol{l}_{\boldsymbol{\beta\psi\beta}}^{(3)} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} \left(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{V}_1 \right),$$

$$\boldsymbol{l}_{\boldsymbol{\psi\beta\beta}}^{(3)} = \boldsymbol{l}_{\boldsymbol{\beta\psi\beta}}^{(3)} \boldsymbol{I}_{(\nu_3-1,\nu_1)},$$

$$\boldsymbol{l}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\beta}}^{(3)} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} \left(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{v}_2 \right),$$

$$\boldsymbol{l}_{\eta_\psi\boldsymbol{\beta\beta}}^{(3)} = \boldsymbol{l}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\beta}}^{(3)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\mu\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\mu\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\beta}}\boldsymbol{I}_{(\nu_3-1,\nu_2)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\beta}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\psi\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{V}_1\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{V}_1\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\beta}},$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{v}_2\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\beta\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta\mu}}\boldsymbol{I}_{(\nu_1,\nu_2)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\beta\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\mu}}\boldsymbol{I}_{(\nu_3-1,\nu_1)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\mu}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\mu\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\mu\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\mu}}\boldsymbol{I}_{(\nu_3-1,\nu_2)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu}\boldsymbol{\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\mu}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{V}_1\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{V}_1\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi}\boldsymbol{\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\mu}},$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{v}_2\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\beta}\boldsymbol{\psi}} = \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\boldsymbol{\psi}} = \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\beta}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\boldsymbol{\psi}}\boldsymbol{I}_{(\nu_2,\nu_1)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\boldsymbol{\psi}} = \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right) + \left(\frac{\partial^2 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial^2 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)\right]\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\beta}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\boldsymbol{\psi}}\boldsymbol{I}_{(\nu_3-1,\nu_1)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\psi}} = \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right) + \left(\frac{\partial^2 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial^2 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)\right]\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\psi}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\mu}\boldsymbol{\psi}} = \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\boldsymbol{\psi}} = \left(\boldsymbol{v}_2'\otimes\boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial w^*}{\partial\boldsymbol{\varphi}}\right) + \left(\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}}\otimes\frac{\partial^2 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)\right]\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\mu}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\boldsymbol{\psi}} \boldsymbol{I}_{(\nu_3-1,\nu_2)},$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{v}_2),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu}\boldsymbol{\psi}},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) (2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{V}_1 \otimes \boldsymbol{V}_1),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) (2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{V}_1 \otimes \boldsymbol{v}_2),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\psi}},$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) (2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{v}_2 \otimes \boldsymbol{v}_2),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\beta}\eta_\psi} = n \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}'},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\eta_\psi} = n \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}'},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\beta}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\eta_\psi} \boldsymbol{I}_{(\nu_2,\nu_1)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\eta_\psi} = n \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{V}_1),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\beta}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\eta_\psi}\boldsymbol{I}_{(\nu_3-1,\nu_1)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\eta_\psi},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\mu}\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\mu}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\eta_\psi}\boldsymbol{I}_{(\nu_3-1,\nu_2)},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\eta_\psi},$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\psi}\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{V}_1\otimes\boldsymbol{V}_1\right),$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{V}_1\otimes\boldsymbol{v}_2\right),$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\eta_\psi}, \ \textit{and}$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{v}_2\otimes\boldsymbol{v}_2\right),$$

where $\boldsymbol{a}$ *is given in Theorem 3.9, second derivatives of* q^*, $\boldsymbol{K}_{..}$, $\boldsymbol{W}_{.}$, $\boldsymbol{W}_{..}$, $\boldsymbol{V}_{..}$, *and* $\boldsymbol{M}_{..}$, *are given in Theorem 3.11,* $\boldsymbol{F}^{(1)}_{.}$, $\boldsymbol{F}^{(11)}_{..}$, *and* $\boldsymbol{F}^{(111)}_{...}$ *are given in Appendix A,* $\frac{\partial w^*}{\partial\boldsymbol{\varphi}}$ *and* $\frac{\partial^2 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}$ *are given in Theorem 3.13,*

$$\begin{aligned}\frac{\partial^3 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}} = {} & \boldsymbol{F}^{(111)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3^2}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right) - 2\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}\left[\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right]\boldsymbol{N}_{\nu_3} \\ & + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\left[\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right]\left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right) \\ & - \boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}},\end{aligned}$$

$$\boldsymbol{K}_{xzx} = \left(\boldsymbol{X} * \boldsymbol{K}_{zx}\right),$$

$$\boldsymbol{K}_{xzz} = \left(\boldsymbol{X} * \boldsymbol{K}_{zz}\right),$$

$$\boldsymbol{K}_{zzx} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} * \boldsymbol{K}_{zx}\right),$$

$$\boldsymbol{K}_{zzz} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} * \boldsymbol{K}_{zz}\right),$$

$$\boldsymbol{K}_{zxzx} = \left(\boldsymbol{K}_{zx} * \boldsymbol{K}_{zx}\right),$$

$$\boldsymbol{K}_{zxzz} = \left(\boldsymbol{K}_{zx} * \boldsymbol{K}_{zz}\right),$$

$$\boldsymbol{K}_{zzzx} = \left(\boldsymbol{K}_{zz} * \boldsymbol{K}_{zx}\right),$$

$$\boldsymbol{K}_{zzzz} = \left(\boldsymbol{K}_{zz} * \boldsymbol{K}_{zz}\right),$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{K}_{xx}\boldsymbol{W}_2\boldsymbol{K}'_{zx},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{K}_{xx}\boldsymbol{W}_2\boldsymbol{K}'_{zz}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\varphi}'} = -\boldsymbol{K}_{xx}\boldsymbol{W}_2\boldsymbol{K}'_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\beta}'} = -\boldsymbol{I}_{(p,d)}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_d\right)\boldsymbol{M}_{xx},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{I}_{(p,d)}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_d\right)\boldsymbol{V}_{xz}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} = -\boldsymbol{I}_{(p,d)}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_d\right)\boldsymbol{V}_{xz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zz}}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{N}_p\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right)\boldsymbol{M}_{zx},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zz}}{\partial \boldsymbol{\mu}'} = -\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right)\boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)},$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'} = -\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right)\boldsymbol{V}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)},$$

$$\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\beta}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{xzx}\right) + 2\boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zxzx} + \boldsymbol{K}_{zx}\boldsymbol{W}_2\left[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{xx}\right]'\left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(d,p)} \otimes \boldsymbol{I}_d\right)\Big],$$

$$\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\mu}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzx}\right) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzx}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zx}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_d\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{\nu_1}\right),$$

$$\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\varphi}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzx}\right) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzx}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zx}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_d\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\nu_1}\right),$$

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\beta}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{xzz}\right) + 2\boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zxzz}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{(p,d)}\boldsymbol{K}_{zx}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(d,p)} \otimes \boldsymbol{I}_p\right) \left(\boldsymbol{I}_{\nu_1} \otimes 2\boldsymbol{N}_p\right)\Big],$$

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\mu}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\right) \left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\right),$$

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\right) \left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\right),$$

$$\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\beta}'} = -\Big[2\boldsymbol{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_2\boldsymbol{K}'_{xzz}\right) + 2\boldsymbol{K}_{zz}\boldsymbol{W}_3\boldsymbol{K}'_{zxzz}$$
$$+ \boldsymbol{K}_{zz}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{(p,d)}\boldsymbol{K}_{zx}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(d,p)} \otimes \boldsymbol{I}_p\right) \left(\boldsymbol{I}_{\nu_1} \otimes 2\boldsymbol{N}_p\right)\Big],$$

$$\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\mu}'} = -\Big[2\boldsymbol{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zz}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz}$$
$$+ \boldsymbol{K}_{zz}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\right) \left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\right),$$

$$\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\varphi}'} = -\Big[2\boldsymbol{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zz}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz}$$
$$+ \boldsymbol{K}_{zz}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\right) \left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\right),$$

$$\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\mu}'} = -\left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{p^2}\right)$$
$$- \boldsymbol{I}_{(p,p)}\left(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\right)\left(2\boldsymbol{N}_p\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{(p,p)}\right) + \frac{1}{2}\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\mu}'},$$

$$\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\varphi}'} = -\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\big)$$

$$- \boldsymbol{I}_{(p,p)}\big(\boldsymbol{I}_p \otimes (\operatorname{vec} \boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(2\boldsymbol{N}_p\boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{(p,p)}\big) + \frac{1}{2}\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\varphi}'},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec} \boldsymbol{I}_d)' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{I}_{(p,d)}\right) + 4\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\beta}'},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}_{\boldsymbol{\mu\mu\mu}}^{(111)\prime}\Big[\boldsymbol{I}_{\nu_2^2} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\Big] + 4\boldsymbol{F}_{\boldsymbol{\mu\mu}}^{(11)\prime}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\big)\boldsymbol{N}_{\nu_2}$$

$$+ 2\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime}\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\mu}'}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\big) + 2\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime}\boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\mu\mu}}^{(2)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}_{\boldsymbol{\varphi\varphi\varphi}}^{(111)\prime}\Big[\boldsymbol{I}_{\nu_3^2} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\Big] + 4\boldsymbol{F}_{\boldsymbol{\varphi\varphi}}^{(11)\prime}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\big)\boldsymbol{N}_{\nu_3}$$

$$+ 2\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\big) + 2\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\varphi\varphi}}^{(2)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\right)$$

$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\beta}'}\big(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\big),$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{(\nu_2,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}\right), \nu_2, \nu_1^2\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{I}_{(\nu_2,\nu_1)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = -2\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{W}_{xx}\big)\big(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{\nu_1}\big)$$

$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec} \boldsymbol{I}_d)' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{I}_{(p,d)}\right) + 4\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\varphi}'},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{I}_{(\nu_1,\nu_2)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{(\nu_3,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\right), \nu_3, \nu_1^2\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = -4\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)} \otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\big)$$

$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\mu}'} \otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\right)$$

$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\mu}'}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\big) + 2\boldsymbol{M}_{xz}\boldsymbol{F}_{\boldsymbol{\mu\mu}}^{(2)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{I}_{(\nu_2,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}\right), \nu_2, \nu_1\nu_2\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} \boldsymbol{I}_{(\nu_1,\nu_2)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = -4\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big)$$
$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec}\boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right)$$
$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big) + 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{I}_{(\nu_3,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\right), \nu_3, \nu_1\nu_3\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{I}_{(\nu_1,\nu_3)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}^{(111)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\mu}}\big[\boldsymbol{I}_{\nu_3^2} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big] + 4\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\mu}\boldsymbol{\varphi}}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big)\boldsymbol{N}_{\nu_3}$$
$$+ 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{I}_{(\nu_3,\nu_2)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\right), \nu_3, \nu_2\nu_3\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{I}_{(\nu_2,\nu_3)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}^{(111)\prime}_{\boldsymbol{\mu}\boldsymbol{\mu}\boldsymbol{\varphi}}\big[\boldsymbol{I}_{\nu_2^2} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big] + 4\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\mu}}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)\boldsymbol{N}_{\nu_2}$$
$$+ 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\mu}'}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(2)}_{\boldsymbol{\mu}\boldsymbol{\mu}},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{I}_{(\nu_2,\nu_3)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}}\right), \nu_2, \nu_2\nu_3\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} \boldsymbol{I}_{(\nu_3,\nu_2)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = -4\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)$$
$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec}\boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$
$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big) + 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\mu}},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} \boldsymbol{I}_{(\nu_2,\nu_3)},$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{I}_{(\nu_2,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}\right), \nu_2, \nu_1\nu_3\right],$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} \boldsymbol{I}_{(\nu_1,\nu_3)},$$
$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = \operatorname{dvec}\left[\left(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{I}_{(\nu_3,\nu_1)} \right) \operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} \right), \nu_3, \nu_1\nu_2 \right], \textit{ and}$$
$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{I}_{(\nu_1,\nu_2)}.$$

Proof. The proof for $\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\psi}}$ will given here. Expressions without proof for all other derivatives can be found in Appendix C. Methods for the remaining derivatives are similar to that given below.

The first derivative of $\boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}$ with respect to η_ψ is

$$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3} \right) \left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] \boldsymbol{v}_2. \tag{4.30}$$

The derivative of (4.30) with respect to $\boldsymbol{\psi}'$ is

$$\begin{aligned}
\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\psi}} &= \frac{\partial}{\partial \boldsymbol{\psi}'} \left[(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3} \right) \left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] \boldsymbol{v}_2 \right] \\
&= (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3} \right) \frac{\partial}{\partial \boldsymbol{\psi}'} \left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) \right. \\
&\qquad \left. + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{I}_{\nu_3 - 1} \otimes \boldsymbol{v}_2) \\
&= (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3} \right) \frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{V}_1 \otimes \boldsymbol{v}_2).
\end{aligned} \tag{4.31}$$

The derivative in (4.31) can be separated into two parts. The first part is

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right] \\
&\quad = \left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \boldsymbol{I}_{(\nu_3,\nu_3)} \left(\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \\
&\quad = \left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \boldsymbol{I}_{(\nu_3,\nu_3)}.
\end{aligned} \tag{4.32}$$

The derivative of the second part is

$$\frac{\partial}{\partial \boldsymbol{\varphi}'}\left[\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right] = \frac{\partial}{\partial \boldsymbol{\varphi}'}\left[\frac{\partial q^*}{\partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right]$$
$$= \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right) + \boldsymbol{I}_{(\nu_3,\nu_3)}\left(\frac{\partial^3 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}} \otimes \boldsymbol{a}\right)$$
$$= \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}}\right). \tag{4.33}$$

Adding (4.32) and (4.33) yields

$$\frac{\partial}{\partial \boldsymbol{\varphi}'}\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right] = \left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right)$$
$$+ \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)(2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}}\right). \tag{4.34}$$

Using (4.34), (4.31) simplifies to

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right)\right.$$
$$+ \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)(2\boldsymbol{N}_{\nu_3})$$
$$\left. + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right](\boldsymbol{V}_1 \otimes \boldsymbol{v}_2).$$

The other derivatives of $\boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_s}$ are similar to that above. The third derivative of w^* with respect to $\boldsymbol{\varphi}'$, $\boldsymbol{\varphi}'$, and $\boldsymbol{\varphi}$ is given below.

The second derivative of w^* with respect to $\boldsymbol{\varphi}'$ and $\boldsymbol{\varphi}$ (without being evaluated at $\boldsymbol{\mu} = \boldsymbol{0}$) is

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \left(\frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)'(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec} \boldsymbol{\Sigma}^{-1}) - \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\right)'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\right). \tag{4.35}$$

The derivative of (4.35) with respect to $\boldsymbol{\varphi}'$ is

$$\frac{\partial^3 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}} = \frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right)\right] - \frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\left(\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right)'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right)\right]. \tag{4.36}$$

The first derivative in (4.36) can be obtained as follows:

$$\frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right)\right] = = \frac{\partial^3(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3^2}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right) + \left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\frac{\partial}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right). \tag{4.37}$$

Note that since $\left(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right) = \boldsymbol{I}_{(p^2,\nu_3)}\left(\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_{\nu_3}\right)$, then

$$\frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\left(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right)\right] = \boldsymbol{I}_{(p^2,\nu_3)}\left(\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}^{-1}}{\partial\boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{\nu_3}\right) = \left(\boldsymbol{I}_{\nu_3}\otimes\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}^{-1}}{\partial\boldsymbol{\varphi}'}\right)\boldsymbol{I}_{(\nu_3,\nu_3)} = -\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right)\boldsymbol{I}_{(\nu_3,\nu_3)}.$$

Accordingly, (4.37) simplifies to

$$\frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right)\right] = = \frac{\partial^3(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3^2}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right) - \left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial\operatorname{vec}(\boldsymbol{\Sigma})}{\partial\boldsymbol{\varphi}'}\right)\boldsymbol{I}_{(\nu_3,\nu_3)}. \tag{4.38}$$

The second derivative in (4.36) can be simplified as follows

$$\frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\frac{\partial(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right] = \frac{\partial^2(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right) + \left[\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3}\otimes\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right) + \left[\frac{\partial\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial^2\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}. \tag{4.39}$$

Because $\boldsymbol{I}_{(p,p)}\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} = \frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}$ and by (3.65), it follows that

$$\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3}\otimes\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right) = -\Bigg[\left(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{p^2}\right)$$
$$+\boldsymbol{I}_{(p,p)}\left(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{(p,p)}\right)\Bigg]\left(\boldsymbol{I}_{\nu_3}\otimes\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right)$$
$$= -2\boldsymbol{N}_p\left(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\otimes\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]. \tag{4.40}$$

Because $\boldsymbol{N}_p\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'} = \frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}$, it follows that (4.40) simplifies (4.39) to

$$\frac{\partial}{\partial\boldsymbol{\varphi}'}\left[\frac{\partial(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right] = \frac{\partial^2(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right)$$
$$-2\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\otimes\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]$$
$$+\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}. \tag{4.41}$$

Together, (4.41) and (4.38) simplifies (4.36) to

$$\begin{aligned}\frac{\partial^3 w^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}} &= \frac{\partial^3(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3^2}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right)\\
&-\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial\boldsymbol{\varphi}'}\right)\boldsymbol{I}_{(\nu_3,\nu_3)}\\
&-\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right)\\
&+2\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\otimes\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]\\
&-\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\\
&= \frac{\partial^3(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\left(\boldsymbol{I}_{\nu_3^2}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\right)\\
&-\left(\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right)'\left(\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma})}{\partial\boldsymbol{\varphi}'}\right)(2\boldsymbol{N}_{\nu_3})\\
&+2\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\otimes\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]\\
&-\left[\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}.\end{aligned} \tag{4.42}$$

Evaluating (4.42) at $\boldsymbol{\mu} = \mathbf{0}$ yields

$$\frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \boldsymbol{F}^{(111)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3^2} \otimes \operatorname{vec} \boldsymbol{\Sigma}^{-1}\right) - 2\boldsymbol{F}^{(11)'}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right) \boldsymbol{N}_{\nu_3}$$
$$+ 2\boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}'} \left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right) \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right)$$
$$- \boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}.$$

The third derivatives of q^* are used in each of the expressions. The derivation of $\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}$ will be given. Expressions without proof for the other derivatives of q can be found in Appendix C. Methods for the remaining derivatives are similar to that given.

The second derivative of q^* with respect to $\boldsymbol{\mu}'$ and $\boldsymbol{\beta}$ is

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xz} \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}. \tag{4.43}$$

Taking the derivative of (4.43) with respect to $\boldsymbol{\beta}'$ and evaluating at $\boldsymbol{\mu} = \mathbf{0}$ yields

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = 2\frac{\partial \boldsymbol{M}_{xz}}{\partial \boldsymbol{\varphi}'} \left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right) + 2\boldsymbol{M}_{xz} \boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\mu}}. \tag{4.44}$$

The derivative of $\boldsymbol{M}_{xz}$ with respect to $\boldsymbol{\varphi}'$ can be obtained as follows:

$$\begin{aligned} \frac{\partial \boldsymbol{M}_{xz}}{\partial \boldsymbol{\varphi}'} &= \frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xz} \boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{V}_{xz}\right] \\ &= \frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xz} \boldsymbol{\Sigma}^{-1}\right)\right] + \frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}. \end{aligned} \tag{4.45}$$

Note that $\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xz} \boldsymbol{\Sigma}^{-1}\right)$ can be expressed as

$$\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xz} \boldsymbol{\Sigma}^{-1}\right) = \left(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\right) \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \tag{4.46}$$

or, by using (3.61), it also can be expressed as

$$\begin{aligned}\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right) &= \boldsymbol{I}_{(d,p)}\left(\boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{I}_{(p,p)}\\ &= \boldsymbol{I}_{(d,p)}\left(\boldsymbol{I}_d\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\operatorname{vec}\boldsymbol{W}_{zx}\otimes \boldsymbol{I}_{(p,p)}\right). \qquad (4.47)\end{aligned}$$

The derivative of $\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right)$ can be obtained using the total derivative in conjunction with (4.46) and (4.47). The result is

$$\begin{aligned}\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'} &= \left.\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'}\right|_{\boldsymbol{W}_{xz}\text{ fixed}} + \left.\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'}\right|_{\text{both }\boldsymbol{\Sigma}^{-1}\text{ fixed}}\\ &= \left(\boldsymbol{I}_p\otimes \boldsymbol{W}_{xz}\right)\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'}\\ &\quad + \boldsymbol{I}_{(d,p)}\left(\boldsymbol{I}_d\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{zx}}{\partial\boldsymbol{\varphi}'}\otimes \boldsymbol{I}_{(p,p)}\right). \qquad (4.48)\end{aligned}$$

Using (3.66), (4.48) simplifies to

$$\begin{aligned}\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right)}{\partial\boldsymbol{\varphi}'} &= -\left(\boldsymbol{I}_p\otimes \boldsymbol{W}_{xz}\right)\left(\boldsymbol{\Sigma}^{-1}\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes \boldsymbol{I}_{p^2}\right)\\ &\quad - \left(\boldsymbol{I}_p\otimes \boldsymbol{W}_{xz}\right)\boldsymbol{I}_{(p,p)}\left(\boldsymbol{\Sigma}^{-1}\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes \boldsymbol{I}_{(p,p)}\right)\\ &\quad + \boldsymbol{I}_{(d,p)}\left(\boldsymbol{I}_d\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{zx}}{\partial\boldsymbol{\varphi}'}\otimes \boldsymbol{I}_{(p,p)}\right). \qquad (4.49)\end{aligned}$$

Finally, by using (4.45) and (4.49), (4.44) simplifies to

$$\begin{aligned}\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}} &= 2\frac{\partial\boldsymbol{M}_{xz}}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3}\otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\right) + 2\boldsymbol{M}_{xz}\boldsymbol{F}_{\boldsymbol{\varphi\mu}}^{(2)}\\ &= 2\Bigg[-\left(\boldsymbol{I}_p\otimes \boldsymbol{W}_{xz}\right)\left(\boldsymbol{\Sigma}^{-1}\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes \boldsymbol{I}_{p^2}\right)\\ &\quad - \left(\boldsymbol{I}_p\otimes \boldsymbol{W}_{xz}\right)\boldsymbol{I}_{(p,p)}\left(\boldsymbol{\Sigma}^{-1}\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes \boldsymbol{I}_{(p,p)}\right)\\ &\quad + \boldsymbol{I}_{(d,p)}\left(\boldsymbol{I}_d\otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes \boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{zx}}{\partial\boldsymbol{\varphi}'}\otimes \boldsymbol{I}_{(p,p)}\right)\\ &\quad + \frac{\partial\boldsymbol{V}_{xz}}{\partial\boldsymbol{\varphi}'}\Bigg]\left(\boldsymbol{I}_{\nu_3}\otimes \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\right) + 2\boldsymbol{M}_{xz}\boldsymbol{F}_{\boldsymbol{\varphi\mu}}^{(2)}\end{aligned}$$

$$= 2\Big[-\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)(2\boldsymbol{N}_p)\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$

$$+ \boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec}\boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$

$$+ \frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)\Big] + 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\mu}}$$

$$= -4\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$

$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial \operatorname{vec}\boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$

$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right) + 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\mu}}.$$

Rearrangements of third derivatives of q^* are required in the expressions for the second derivatives of $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$. To find a rearrangement of a third derivative of q^*, see Theorem 4.22.

The derivative of $\boldsymbol{V}_{xz}$ with respect to $\boldsymbol{\varphi}'$ will be given. The other derivatives of $\boldsymbol{V}_{\cdot\cdot}$ and $\boldsymbol{W}_{\cdot\cdot}$ are similar to the method used below. Define $\boldsymbol{s}_k \stackrel{\text{def}}{=} \operatorname{vec}(\boldsymbol{x}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}) = (\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{x}_k)$ and $\boldsymbol{t}_k \stackrel{\text{def}}{=} \operatorname{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}) = (\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k)$. Because

$$\boldsymbol{V}_{xz} = \boldsymbol{K}_{xz}\boldsymbol{W}_2\boldsymbol{K}_{zz} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{s}_k w^{(2)}_k \boldsymbol{t}_k',$$

then the derivative of $\boldsymbol{V}_{xz}$ with respect to $\boldsymbol{\varphi}'$ is

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'} = \sum_{k=1}^{n}\left[\frac{\partial \boldsymbol{s}_k}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3} \otimes w^{(2)}_k \boldsymbol{t}_k'\right) + \boldsymbol{s}_k \frac{\partial w^{(2)}_k}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{t}_k'\right) + \boldsymbol{s}_k w^{(2)}_k \frac{\partial \boldsymbol{t}_k'}{\partial \boldsymbol{\varphi}'}\right]. \tag{4.50}$$

The derivatives in each part of (4.50) should be simplified separately. The derivative of $\boldsymbol{s}_k$ with respect to $\boldsymbol{\varphi}'$ can be expressed as

$$\frac{\partial \boldsymbol{s}_k}{\partial \boldsymbol{\varphi}'} = \frac{\partial \operatorname{vec}(\boldsymbol{x}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'} = \left(\boldsymbol{I}_p \otimes \boldsymbol{x}_k\boldsymbol{z}_k'\right)\frac{\partial \operatorname{vec}(\boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\varphi}'} = -\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}.$$

The derivative of $w_k^{(2)}$ with respect to $\boldsymbol{\varphi}'$ evaluated at $\boldsymbol{\mu} = \mathbf{0}$ can be shown to equal

$$\frac{\partial w_k^{(2)}}{\partial \boldsymbol{\varphi}'} = -w_k^{(3)} \boldsymbol{t}_k' \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}.$$

Finally, note that because $\text{vec}(\boldsymbol{e}_k' \otimes \boldsymbol{b}) = \boldsymbol{e}_k \otimes \boldsymbol{b}$, where $\boldsymbol{b}$ is any vector, the vec of $\frac{\partial \boldsymbol{t}_k}{\partial \boldsymbol{\varphi}'}$ equals

$$\text{vec}\left(\frac{\partial \boldsymbol{t}_k}{\partial \boldsymbol{\varphi}'}\right) = \sum_{k=1}^{n} \text{vec}\left(\boldsymbol{e}_k' \otimes \frac{\partial \boldsymbol{t}_k}{\partial \varphi_k}\right) = \sum_{k=1}^{n} \boldsymbol{e}_k \otimes \frac{\partial \boldsymbol{t}_k}{\partial \varphi_k} = \frac{\partial \boldsymbol{t}_k}{\partial \boldsymbol{\varphi}}.$$

It can be shown that

$$\frac{\partial \boldsymbol{t}_k}{\partial \boldsymbol{\varphi}'} = -2\boldsymbol{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}.$$

Hence,

$$\frac{\partial \boldsymbol{t}_k'}{\partial \boldsymbol{\varphi}'} = \text{vec}\left(\frac{\partial \boldsymbol{t}_k}{\partial \boldsymbol{\varphi}'}\right)' = \text{vec}\left[-2\boldsymbol{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\right]'.$$

The first term of (4.50) can be simplified as

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} \frac{\partial \boldsymbol{s}_k}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3} \otimes w_k^{(2)} \boldsymbol{t}_k'\right) &= -\frac{1}{n}\sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \left(\boldsymbol{I}_{\nu_3} \otimes w_k^{(2)} \boldsymbol{t}_k'\right) \\
&= -\frac{1}{n}\sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1}\right) \left(\boldsymbol{I}_{p^2} \otimes w_k^{(2)} \boldsymbol{t}_k'\right) \left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\right) \\
&= -\frac{1}{n}\sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1} \otimes w_k^{(2)} \boldsymbol{t}_k'\right) \left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\right) \\
&= -\frac{1}{n}\sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k w_k^{(2)} \left[\boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \otimes \boldsymbol{t}_k\right]'\right) \left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\right) \\
&= -\frac{1}{n}\sum_{k=1}^{n} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{x}_k w_k^{(2)} \left[\boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k\right]'\right) \left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\right) \\
&= -\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}' \boldsymbol{W}_2 \boldsymbol{K}_{zzz}'\right) \left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \otimes \boldsymbol{I}_{p^2}\right). \qquad (4.51)
\end{aligned}$$

The second term of (4.50) can be written as

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{s}_k\frac{\partial w_k^{(2)}}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_3}\otimes\boldsymbol{t}_k'\right) &= -\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{s}_k w_k^{(3)}\boldsymbol{t}_k'\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\left(\boldsymbol{I}_{\nu_3}\otimes\boldsymbol{t}_k'\right)\\
&= -\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{s}_k w_k^{(3)}\left(\boldsymbol{t}_k\otimes\boldsymbol{t}_k\right)'\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes\boldsymbol{I}_{p^2}\right)\\
&= -\frac{1}{n}\sum_{k=1}^{n}\left(\boldsymbol{x}_k\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right)w_k^{(3)}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right)'\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes\boldsymbol{I}_{p^2}\right)\\
&= -\boldsymbol{K}_{xz}\boldsymbol{W}_3\boldsymbol{K}_{zzzz}'\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes\boldsymbol{I}_{p^2}\right). \qquad (4.52)\end{aligned}$$

To simplify the third term of (4.50), an identity will be used. Suppose $\boldsymbol{A}$ is an $a\times c$ matrix and $\boldsymbol{B}$ is an $d\times b$ matrix. Then

$$\operatorname{vec}(\boldsymbol{A}\otimes\boldsymbol{B}) = \left(\boldsymbol{I}_c\otimes\boldsymbol{I}_{(a,b)}\otimes\boldsymbol{I}_d\right)\left(\operatorname{vec}\boldsymbol{A}\otimes\operatorname{vec}\boldsymbol{B}\right).$$

The vec part of the third term can then be written as

$$\begin{aligned}\operatorname{vec}\left[-2\boldsymbol{N}_p\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\right] &= -\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\otimes 2\boldsymbol{N}_p\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)\\
&= -\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\otimes 2\boldsymbol{N}_p\right)\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\otimes\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\right)\right)\\
&= -\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\otimes 2\boldsymbol{N}_p\right)\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\otimes\boldsymbol{I}_{p^2}\right)\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right)\\
&= -\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\otimes 2\boldsymbol{N}_p\right)\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\otimes\boldsymbol{I}_{p^2}\right)\boldsymbol{t}_k.\end{aligned}$$

Hence, the third term of (4.50) can be simplified as

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{s}_k w_k^{(2)}\operatorname{vec}\left(\frac{\partial\boldsymbol{t}_k}{\partial\boldsymbol{\varphi}'}\right)'\\
&\quad= -\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{s}_k w_k^{(2)}\boldsymbol{t}_k'\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\otimes\boldsymbol{I}_{p^2}\right)'\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes 2\boldsymbol{N}_p\right)\\
&\quad= -\boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}_{zz}'\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\otimes\boldsymbol{I}_{p^2}\right)'\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes 2\boldsymbol{N}_p\right)\\
&\quad= -\boldsymbol{K}_{zx}\boldsymbol{W}_3\left(\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\otimes\boldsymbol{K}_{zz}\right)'\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\otimes 2\boldsymbol{N}_p\right). \qquad (4.53)\end{aligned}$$

Together, (4.51), (4.52), and (4.53) can simplify (4.50). The result can be written as

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz} + \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\right)\left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\right)$$

□

THEOREM 4.22. *Suppose $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are vectors of size $a \times 1$, $b \times 1$, and $c \times 1$, respectively. Suppose $\frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{y}' \otimes \partial \boldsymbol{z}}$ is known. Then*

(a) $$\frac{\partial^3 q^*}{\partial \boldsymbol{y}' \otimes \partial \boldsymbol{x}' \otimes \partial \boldsymbol{z}} = \frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{y}' \otimes \partial \boldsymbol{z}}\boldsymbol{I}_{(b,a)}.$$

(b) $$\frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{z}' \otimes \partial \boldsymbol{y}} = \operatorname{dvec}\left[\left(\boldsymbol{I}_a \otimes \boldsymbol{I}_{(b,c)}\right)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{y}' \otimes \partial \boldsymbol{z}}\right), b, ca\right].$$

PROOF. The definition of the third derivative of $\frac{\partial^3 q^*}{\partial \boldsymbol{y}' \otimes \partial \boldsymbol{x}' \otimes \partial \boldsymbol{z}}$ is

$$\begin{aligned}
\frac{\partial^3 q^*}{\partial \boldsymbol{y}' \otimes \partial \boldsymbol{x}' \otimes \partial \boldsymbol{z}} &= \sum_{i=1}^{b}\sum_{j=1}^{a}\sum_{k=1}^{c} \boldsymbol{e}_i^{b'} \otimes \boldsymbol{e}_j^{a'} \otimes \boldsymbol{e}_k^{c} \otimes \frac{\partial^3 q^*}{\partial y_i \partial x_j \partial z_k} \\
&= \sum_{i=1}^{b}\sum_{j=1}^{a}\sum_{k=1}^{c} \left(\boldsymbol{e}_j^{a'} \otimes \boldsymbol{e}_i^{b'}\right)\boldsymbol{I}_{(b,a)} \otimes \boldsymbol{e}_k^{c} \otimes \frac{\partial^3 q^*}{\partial y_i \partial x_j \partial z_k} \\
&= \sum_{i=1}^{b}\sum_{j=1}^{a}\sum_{k=1}^{c} \boldsymbol{e}_j^{a'} \otimes \boldsymbol{e}_i^{b'} \otimes \boldsymbol{e}_k^{c} \otimes \frac{\partial^3 q^*}{\partial y_i \partial x_j \partial z_k}\boldsymbol{I}_{(b,a)}
\end{aligned}$$

If interchanging the order of differentiation is allowed, then

$$\begin{aligned}
\frac{\partial^3 q^*}{\partial \boldsymbol{y}' \otimes \partial \boldsymbol{x}' \otimes \partial \boldsymbol{z}} &= \sum_{j=1}^{a}\sum_{i=1}^{b}\sum_{k=1}^{c} \boldsymbol{e}_j^{a'} \otimes \boldsymbol{e}_i^{b'} \otimes \boldsymbol{e}_k^{c} \otimes \frac{\partial^3 q^*}{\partial x_j \partial y_i \partial z_k}\boldsymbol{I}_{(b,a)} \\
&= \frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{y}' \otimes \partial \boldsymbol{z}}\boldsymbol{I}_{(b,a)}.
\end{aligned}$$

To prove the second part, suppose $\boldsymbol{A}$ is a $c \times b$ matrix, and suppose $\boldsymbol{x}$ is a $a \times 1$ vector. Then the vec of $\frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}'}$ is

$$\operatorname{vec}\left(\frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}'}\right) = \sum_{i=1}^{a} \operatorname{vec}\left(\boldsymbol{e}_i^{a\prime} \otimes \frac{\partial \boldsymbol{A}}{\partial x_i}\right) = \sum_{i=1}^{a} \boldsymbol{e}_i^{a} \otimes \frac{\partial \operatorname{vec} \boldsymbol{A}}{\partial x_i} = \frac{\partial \operatorname{vec} \boldsymbol{A}}{\partial \boldsymbol{x}}.$$

Further, the vec of $\frac{\partial \boldsymbol{A}'}{\partial \boldsymbol{x}'}$ is

$$\begin{aligned}
\operatorname{vec}\left(\frac{\partial \boldsymbol{A}'}{\partial \boldsymbol{x}'}\right) &= \operatorname{vec}\left[\sum_{i=1}^{a} \boldsymbol{e}_i^{a\prime} \otimes \frac{\partial \boldsymbol{A}'}{\partial x_i}\right] = \sum_{i=1}^{a} \boldsymbol{e}_i^{a} \otimes \frac{\partial \operatorname{vec} \boldsymbol{A}'}{\partial x_i} \\
&= \sum_{i=1}^{a} \boldsymbol{e}_i^{a} \otimes \frac{\partial \boldsymbol{I}_{(b,c)} \operatorname{vec} \boldsymbol{A}}{\partial x_i} = \sum_{i=1}^{a} \boldsymbol{e}_i^{a} \otimes \boldsymbol{I}_{(b,c)} \frac{\partial \operatorname{vec} \boldsymbol{A}}{\partial x_i} \\
&= \left(\boldsymbol{I}_a \otimes \boldsymbol{I}_{(b,c)}\right) \sum_{i=1}^{a} \boldsymbol{e}_i^{a} \otimes \frac{\partial \operatorname{vec} \boldsymbol{A}}{\partial x_i} = \left(\boldsymbol{I}_a \otimes \boldsymbol{I}_{(b,c)}\right) \frac{\partial \operatorname{vec} \boldsymbol{A}}{\partial \boldsymbol{x}} \\
&= \left(\boldsymbol{I}_a \otimes \boldsymbol{I}_{(b,c)}\right) \operatorname{vec}\left(\frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}'}\right).
\end{aligned}$$

Hence

$$\frac{\partial \boldsymbol{A}'}{\partial \boldsymbol{x}'} = \operatorname{dvec}\left[\left(\boldsymbol{I}_a \otimes \boldsymbol{I}_{(b,c)}\right) \operatorname{vec}\left(\frac{\partial \boldsymbol{A}}{\partial \boldsymbol{x}'}\right), b, ca\right].$$

By noting that $\frac{\partial^2 q^*}{\partial \boldsymbol{y}' \otimes \partial \boldsymbol{z}} = \boldsymbol{A}$, then

$$\frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{z}' \otimes \partial \boldsymbol{y}} = \operatorname{dvec}\left[\left(\boldsymbol{I}_a \otimes \boldsymbol{I}_{(b,c)}\right) \operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{x}' \otimes \partial \boldsymbol{y}' \otimes \partial \boldsymbol{z}}\right), b, ca\right].$$

□

<u>Sandwich Estimator</u>

An estimator of the asymptotic variance, called the sandwich estimator, can be constructed from the data. First, write $\frac{1}{\sqrt{n}} \boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$ as

$$\frac{1}{\sqrt{n}} \boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s} = \sum_{k=1}^{n} \boldsymbol{f}_k,$$

where $\boldsymbol{f}_k$ and $\boldsymbol{f}_j$ are independent when $j \neq k$. Hence, for the S-estimators, $\boldsymbol{f}_k$ is equal to

$$\boldsymbol{f}_k = \begin{bmatrix} \boldsymbol{f}_{k,\boldsymbol{\beta}} \\ \boldsymbol{f}_{k,\boldsymbol{\mu}} \\ \boldsymbol{f}_{k,\boldsymbol{\psi}} \\ \boldsymbol{f}_{k,\eta_\psi} \end{bmatrix} \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \begin{bmatrix} 2\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{x}_k\right) w_k^{(1)} \\ \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right) w_k^{(1)} \\ -\frac{1}{n}\left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)'} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right)\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{z}_k\right) w_k^{(1)} \\ \rho(d_k) - b_0 \end{bmatrix}$$

Define $\boldsymbol{f}_{k,\boldsymbol{\vartheta}} = \begin{bmatrix} \boldsymbol{f}_{k,\boldsymbol{\mu}}' & \boldsymbol{f}_{k,\boldsymbol{\psi}}' & \boldsymbol{f}_{k,\eta_\psi}' \end{bmatrix}'$. Because $E\left[\boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}_{k,\boldsymbol{\vartheta}}'\right] = \boldsymbol{0}$, then a consistent estimator of the asymptotic variance of $\frac{1}{\sqrt{n}}\boldsymbol{l}_{\boldsymbol{\theta}_s}^{(1)}$ is

$$\hat{\hat{\boldsymbol{C}}}_{s,n} = \sum_{k=1}^{n} \begin{bmatrix} \boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}_{k,\boldsymbol{\beta}}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{f}_{k,\boldsymbol{\vartheta}}\boldsymbol{f}_{k,\boldsymbol{\vartheta}}' \end{bmatrix}.$$

The sum above can be rewritten as the matrix product

$$\hat{\hat{\boldsymbol{C}}}_{s,n} = \begin{bmatrix} \boldsymbol{F}_{s,\boldsymbol{\beta}}\boldsymbol{F}_{s,\boldsymbol{\beta}}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_{s,\boldsymbol{\vartheta}}\boldsymbol{F}_{s,\boldsymbol{\vartheta}}' \end{bmatrix},$$

where

$$\boldsymbol{F}_{s,\boldsymbol{\beta}} = 2\sqrt{n}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\boldsymbol{W}_1\right) \text{ and}$$

$$\boldsymbol{F}_{s,\boldsymbol{\vartheta}} = \frac{1}{\sqrt{n}} \begin{bmatrix} n\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)'}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_1\right) \\ -\left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}_{\nu_3}^{\perp}\right)\left(\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)'} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right)\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_1\right) \\ \rho(\tilde{\boldsymbol{d}}_k) - b_0 \end{bmatrix},$$

and

$$\rho(\tilde{\boldsymbol{d}}_k) - b_0 = \begin{bmatrix} \rho(d_1) - b_0 & \rho(d_2) - b_0 & \rho(d_3) - b_0 & \cdots & \rho(d_2) - b_0. \end{bmatrix}$$

The sandwich estimator of the variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)$ is therefore

$$\widehat{\text{Var}}\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)\right] = \hat{\boldsymbol{K}}_2^{-1}\hat{\hat{\boldsymbol{C}}}_{s,n}\hat{\boldsymbol{K}}_2'^{-1},$$

where $\boldsymbol{K}_2 = -\bar{\boldsymbol{I}}_{\boldsymbol{\theta}_s,n} = \frac{1}{n}E\left[\boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_s\boldsymbol{\theta}_s}\right]$. Note that many of the terms in $\hat{\boldsymbol{K}}_2$ have expectation zero. Hence, $\hat{\boldsymbol{K}}_2$ can be written as

$$\hat{\boldsymbol{K}}_2 = \frac{1}{n}\begin{bmatrix} \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\beta\beta}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\mu\mu}} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\psi\mu}} & \hat{\boldsymbol{l}}^{(1)}_{\eta_\psi\boldsymbol{\mu}} \\ \mathbf{0} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\mu\psi}} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\psi\psi}} & \hat{\boldsymbol{l}}^{(1)}_{\eta_\psi\boldsymbol{\psi}} \\ \mathbf{0} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\mu}\eta_\psi} & \hat{\boldsymbol{l}}^{(1)}_{\boldsymbol{\psi}\eta_\psi} & \hat{\boldsymbol{l}}^{(1)}_{\eta_\psi\eta_\psi} \end{bmatrix}.$$

The second order expansions can be used to obtain corrections to the bias. The expected bias of the estimators can be found by taking the expectation of (4.19). Because the expectation of $\hat{\boldsymbol{\delta}}_0$ is zero, it follows that

$$\begin{aligned} E\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)\right] &= \frac{1}{\sqrt{n}}E\big[\hat{\boldsymbol{\delta}}_1\big] + O_p(n^{-1}) \\ &= -\frac{1}{\sqrt{n}}\boldsymbol{K}_2^{-1}\left(E\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0\right] + \frac{1}{2}\boldsymbol{K}_3 E\big[\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\big]\right) + O_p(n^{-1}) \end{aligned}$$

Because

$$E\big[\hat{\boldsymbol{\delta}}_0 \otimes \hat{\boldsymbol{\delta}}_0\big] = E\big[\operatorname{vec}\big(\hat{\boldsymbol{\delta}}_0\hat{\boldsymbol{\delta}}_0'\big)\big] = \operatorname{vec}\big(E\big[\hat{\boldsymbol{\delta}}_0\hat{\boldsymbol{\delta}}_0'\big]\big) = \operatorname{vec}\Big(\widehat{\operatorname{Var}}\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)\right]\Big),$$

an estimator of the expected bias is

$$E\big[\hat{\boldsymbol{\delta}}_1\big] = -\hat{\boldsymbol{K}}_2^{-1}\left(\hat{E}\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0\right] + \frac{1}{2}\hat{\boldsymbol{K}}_3\operatorname{vec}\Big(\hat{\boldsymbol{K}}_2^{-1}\hat{\hat{\boldsymbol{C}}}_{s,n}\hat{\boldsymbol{K}}_2'^{-1}\Big)\right).$$

An estimator of the expectation of $\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0$ must be constructed to produce an estimator of the bias. Form the sum

$$\hat{E}\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0\right] = \sum_{j=1}^{n}\sum_{k=1}^{n}\boldsymbol{Z}_{2j}\hat{\boldsymbol{\delta}}_{0k}.$$

The expectation above will be zero whenever $j \neq k$ because $E\big[\hat{\boldsymbol{\delta}}_0\big] = \mathbf{0}$. Hence, an estimator for the expectation of $\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0$ is

$$\hat{E}\left[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0\right] = \sum_{k=1}^{n}\boldsymbol{Z}_{2k}\hat{\boldsymbol{\delta}}_{0k}.$$

Hence, an estimator of the bias $E\big[\hat{\boldsymbol{\delta}}_1\big]$ is

$$E\big[\hat{\boldsymbol{\delta}}_1\big] = -\hat{\boldsymbol{K}}_2^{-1}\left(\sum_{k=1}^{n} \boldsymbol{Z}_{2k}\hat{\boldsymbol{\delta}}_{0k} + \frac{1}{2}\hat{\boldsymbol{K}}_3 \operatorname{vec}\Big(\hat{\boldsymbol{K}}_2^{-1}\hat{\hat{\boldsymbol{C}}}_{s,n}\hat{\boldsymbol{K}}_2'^{-1}\Big)\right).$$

Because the expectation of many expressions was zero, the sandwich estimator and the estimator of the bias were adjusted to reflect that fact. A description of the adjustment for the sandwich estimator, as well as $\hat{\boldsymbol{K}}_2$, has already been given. All expressions involving the sandwich estimator or the bias that have expectation zero are given in Table 4 and 5.

$E[\hat{\boldsymbol{K}}_2]$	$E[\boldsymbol{F}_s\boldsymbol{F}_s']$	$E\big[\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0\big]$
$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\beta}}$	$\boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}'_{k,\boldsymbol{\mu}}$	$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\beta}}\boldsymbol{f}_{k,\boldsymbol{\beta}}$
$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi\beta}}$	$\boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}'_{k,\boldsymbol{\psi}}$	$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\beta}}\boldsymbol{f}_{k,\boldsymbol{\mu}}$
$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\beta}}$	$\boldsymbol{f}_{k,\boldsymbol{\beta}}\boldsymbol{f}'_{k,\eta_\psi}$	$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi\beta}}\boldsymbol{f}_{k,\boldsymbol{\psi}}$
$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\mu}}$	$\boldsymbol{f}_{k,\boldsymbol{\mu}}\boldsymbol{f}'_{k,\boldsymbol{\beta}}$	$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\beta}}\boldsymbol{f}_{k,\eta_\psi}$
$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\psi}}$	$\boldsymbol{f}_{k,\boldsymbol{\psi}}\boldsymbol{f}'_{k,\boldsymbol{\beta}}$	$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\mu}}\boldsymbol{f}_{k,\boldsymbol{\beta}}$
$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta}\eta_\psi}$	$\boldsymbol{f}_{k,\eta_\psi}\boldsymbol{f}'_{k,\boldsymbol{\beta}}$	
$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu}\eta_\psi}$		

Table 4. Expressions having an expectation of $\mathbf{0}$.

$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\beta\beta}}$	$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta\mu}}$	$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta\mu}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta}\eta_\psi}$
$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\mu\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\psi\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu\mu}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu\psi}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi}\eta_\psi}$
$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\mu}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\psi}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\beta}\eta_\psi}$
$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\mu\beta}}$	$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\beta\mu}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\psi}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\eta_\psi}$
$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\beta}}$	$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\boldsymbol{\beta}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\mu}}$	$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu}\eta_\psi}$	$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta}\eta_\psi}$

Table 5. Expressions having zero expectation in $E[\hat{\boldsymbol{K}}_3]$.

The zero expectations show that the location parameters are asymptotically independent of the other parameters. In addition, the zero expectations also show that the asymptotic bias of the location parameters is zero.

CHAPTER 5

ILLUSTRATION

To illustrate the techniques from this dissertation, two steps are taken. First, for the saturated model, the new algorithm is shown to give the same results as Ruppert's SURREAL algorithm (Ruppert 1992). Second, a simple structure is imposed on the scatter matrix and the subsequent estimates are found.

Two different data sets will be used. These data sets are from Mardia et al. (1979, Table 1.2.1) and Bollen (1989, Table 2.5).

The data from Mardia et al. (1979, Table 1.2.1) consists of examination scores achieved by 88 students in areas of Mechanics, Vectors, Algebra, Analysis, and Statistics. Two of the tests were closed book tests (Mechanics and Vectors), whereas the other tests were open book tests. The data consists of discrete values, and not necessarily continuous values. The data are displayed in Table 6.

Table 6: Scores in Open-Book and Closed-Book Examinations (out of 100)

Mechanics(C)	Vectors(C)	Algebra(O)	Analysis(O)	Statistics(O)
77	82	67	67	81
63	78	80	70	81
75	73	71	66	81
55	72	63	70	68
63	63	65	70	63
53	61	72	64	73
51	67	65	65	68
59	70	68	62	56
62	60	58	62	70

Table 6: continued

Mechanics(C)	Vectors(C)	Algebra(O)	Analysis(O)	Statistics(O)
64	72	60	62	45
52	64	60	63	54
55	67	59	62	44
50	50	64	55	63
65	63	58	56	37
31	55	60	57	73
60	64	56	54	40
44	69	53	53	53
42	69	61	55	45
62	46	61	57	45
31	49	62	63	62
44	61	52	62	46
49	41	61	49	64
12	58	61	63	67
49	53	49	62	47
54	49	56	47	53
54	53	46	59	44
44	56	55	61	36
18	44	50	57	81
46	52	65	50	35
32	45	49	57	64
30	69	50	52	45
46	49	53	59	37
40	27	54	61	61
31	42	48	54	68
36	59	51	45	51
56	40	56	54	35
46	56	57	49	32
45	42	55	56	40
42	60	54	49	33
40	63	53	54	25
23	55	59	53	44
48	48	49	51	37
41	63	49	46	34
46	52	53	41	40
46	61	46	38	41
40	57	51	52	31
49	49	45	48	39
22	58	53	56	41
35	60	47	54	33

Table 6: continued

Mechanics(C)	Vectors(C)	Algebra(O)	Analysis(O)	Statistics(O)
48	56	49	42	32
31	57	50	54	34
17	53	57	43	51
49	57	47	39	26
59	50	47	15	46
37	56	49	28	45
40	43	48	21	61
35	35	41	51	50
38	44	54	47	24
43	43	38	34	49
39	46	46	32	43
62	44	36	22	42
48	38	41	44	33
34	42	50	47	29
18	51	40	56	30
35	36	46	48	29
59	53	37	22	19
41	41	43	30	33
31	52	37	27	40
17	51	52	35	31
34	30	50	47	36
46	40	47	29	17
10	46	36	47	39
46	37	45	15	30
30	34	43	46	18
13	51	50	25	31
49	50	38	23	9
18	32	31	45	40
8	42	48	26	40
23	38	36	48	15
30	24	43	33	25
3	9	51	47	40
7	51	43	17	22
15	40	43	23	18
15	38	39	28	17
5	30	44	36	18
12	30	32	35	21
5	26	15	20	20
0	40	21	9	14

The S-estimators of location and scatter can be computed by using SURREAL provided that a saturated model is employed. In a saturated model, the design matrix for location parameters consists of a column of ones ($\boldsymbol{X} = \mathbf{1}_n$), and no modeling of eigenvalues occurs. The results of the SURREAL algorithm produce the following estimates:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 39.71 \\ 51.26 \\ 50.99 \\ 47.42 \\ 42.14 \end{bmatrix}, \text{ and}$$

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 297.00 & 120.14 & 99.36 & 108.75 & 117.16 \\ 120.14 & 170.05 & 84.77 & 96.44 & 100.94 \\ 99.36 & 84.77 & 109.64 & 108.59 & 120.52 \\ 108.75 & 96.44 & 108.60 & 216.30 & 154.24 \\ 117.16 & 100.94 & 120.52 & 154.24 & 301.45 \end{bmatrix}.$$

The results for the algorithms in this dissertation produced very similar results. The estimate for $\boldsymbol{\beta}$ differed in the eighth significant figure, whereas the estimate for $\boldsymbol{\Sigma}$ differed in the sixth significant figure. The differences between the estimates are negligible. However, the determinant of the estimate of the scatter matrix for the SURREAL algorithm is a little larger than that of the algorithm of this dissertation. In addition, the constraint q is a bit smaller for the estimates computed from the programs of this dissertation than for surreal. The determinant and value of the constraint q for the estimates from each algorithm are given in Table 7.

	$\det \hat{\mathbf{\Sigma}}$	q
SURREAL	36,554,699,827.17853	-3.4×10^{-8}
New	36,554,696,521.87241	-3.1×10^{-15}

Table 7. Determinant of the scatter estimate and value of the constraint q for each algorithm for the Scores data.

The sandwich estimate of the asymptotic variance of the location parameters under the saturated model is

$$\begin{bmatrix} 298.03 & 119.60 & 99.78 & 113.01 & 119.47 \\ 119.60 & 174.83 & 86.35 & 100.08 & 104.69 \\ 99.78 & 86.35 & 109.58 & 108.39 & 121.91 \\ 113.01 & 100.08 & 108.39 & 217.86 & 156.18 \\ 119.47 & 104.69 & 121.91 & 156.18 & 310.37 \end{bmatrix}$$

and the estimate for the asymptotic variance of the $\boldsymbol{\varphi}$ parameters is

$$\begin{bmatrix} 1.57 & -0.68 & -1.12 & 0.19 & 0.69 \\ -0.68 & 5.19 & -14.08 & -0.40 & -0.05 \\ -1.12 & -14.08 & 119.22 & -13.65 & -5.85 \\ 0.19 & -0.40 & -13.65 & 17.53 & -3.06 \\ 0.69 & -0.05 & -5.85 & -3.06 & 6.14 \end{bmatrix}.$$

With these estimates, confidence intervals for the parameters can be constructed. Note that the sandwich estimator for $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is very similar to the estimate for $\mathbf{\Sigma}$. A possible explanation for this can be seen in the variance for the least squares estimator of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. When $\boldsymbol{X} = \mathbf{1}_n$, the variance of the least squares estimator is

$$\operatorname{Var}\left(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right) = n\left[(\boldsymbol{X}'\boldsymbol{X})^{-1} \otimes \mathbf{\Sigma}\right] = \mathbf{\Sigma}.$$

Because $\boldsymbol{X} = \mathbf{1}_n$, it would make sense that $\hat{\mathbf{\Sigma}}$ would be similar to the sandwich estimator for the location parameters.

To illustrate the flexibility of the new methods, two models on the scatter structure will be used. Model 1 is motivated by the eigenvalues in Table 8.

Eigenvalue	Value
$\hat{\lambda}_1$	680.9062
$\hat{\lambda}_2$	193.9235
$\hat{\lambda}_3$	103.6743
$\hat{\lambda}_4$	84.2216
$\hat{\lambda}_5$	31.7052

Table 8. Eigenvalues for the S-estimate of scatter under a saturated model.

The third and fourth eigenvalues of the S-estimate of scatter are relatively similar. Accordingly, model the third and fourth eigenvalues as one eigenvalue with a multiplicity of two. The design matrix for the location parameters is still $\boldsymbol{X} = \mathbf{1}_n$. The matrices $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ for Model 1 can be chosen to give this simple eigenvalue multiplicity structure. They are

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{T}_2 = \boldsymbol{I}_4.$$

The algorithms of this dissertation produce the following result:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 39.67 \\ 51.23 \\ 50.99 \\ 47.43 \\ 42.12 \end{bmatrix}, \quad \det \hat{\boldsymbol{\Sigma}} = 36{,}912{,}486{,}807.82695, \ q = 5.3 \times 10^{-15}, \text{ and}$$

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 297.34 & 120.77 & 100.27 & 113.03 & 115.62 \\ 120.77 & 173.81 & 84.23 & 89.93 & 104.95 \\ 100.27 & 84.23 & 109.54 & 107.93 & 121.52 \\ 113.03 & 89.93 & 107.93 & 216.29 & 156.70 \\ 115.62 & 104.95 & 121.52 & 156.70 & 299.10 \end{bmatrix}.$$

The eigenvalues of the estimate of the scatter matrix are displayed in Table 9. Note that the third and fourth eigenvalues have the same value.

Eigenvalue	Value
$\hat{\lambda}_1$	683.1045
$\hat{\lambda}_2$	193.8024
$\hat{\lambda}_3$	93.7104
$\hat{\lambda}_4$	93.7104
$\hat{\lambda}_5$	31.7505

Table 9. Eigenvalues for the S-estimate of scatter under Model 1.

The sandwich estimate of the asymptotic variance of the location parameters under Model 1 is

$$\begin{bmatrix} 299.52 & 120.45 & 100.42 & 114.77 & 120.70 \\ 120.45 & 175.53 & 86.50 & 100.64 & 104.75 \\ 100.42 & 86.50 & 109.72 & 108.63 & 122.18 \\ 114.77 & 100.64 & 108.63 & 217.49 & 156.87 \\ 120.70 & 104.75 & 122.18 & 156.87 & 310.93 \end{bmatrix}$$

and the estimate for the asymptotic variance of the $\boldsymbol{\varphi}$ parameters is

$$\begin{bmatrix} 1.57 & -0.74 & 0.05 & 0.68 \\ -0.74 & 2.82 & -1.53 & -0.91 \\ 0.05 & -1.53 & 12.93 & -3.32 \\ 0.68 & -0.91 & -3.32 & 6.07 \end{bmatrix}.$$

Note that the variance for the third eigenvalue has decreased quite a bit from the variance of the third and fourth eigenvalues of the saturated model.

Model 2 allows the largest eigenvalue to be unrestricted, and models the last four eigenvalues under the exponential model $\exp\{\varphi_1 + k\varphi_2\}$. Hence the structure of $\boldsymbol{\lambda}$

would be

$$\boldsymbol{\lambda} = \begin{bmatrix} e^{\varphi_1+3\varphi_2} + e^{\varphi_3} \\ e^{\varphi_1+3\varphi_2} \\ e^{\varphi_1+2\varphi_2} \\ e^{\varphi_1+1\varphi_2} \\ e^{\varphi_1} \end{bmatrix}.$$

The matrices $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ corresponding to Model 2 are

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{T}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 3 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

The results of the dissertation's algorithm produce the following estimates:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 39.74 \\ 51.29 \\ 50.98 \\ 47.41 \\ 42.17 \end{bmatrix}, \ \det \hat{\boldsymbol{\Sigma}} = 38{,}375{,}399{,}536.44630, \ q = 1.3 \times 10^{-15}, \text{ and}$$

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 300.31 & 122.71 & 98.51 & 101.81 & 116.46 \\ 122.71 & 161.06 & 84.78 & 107.40 & 95.92 \\ 98.51 & 84.78 & 115.45 & 108.78 & 119.35 \\ 101.81 & 107.40 & 108.78 & 215.31 & 153.38 \\ 116.46 & 95.92 & 119.35 & 153.38 & 307.16 \end{bmatrix}.$$

The eigenvalues of the estimate for the scatter matrix under Model 2 are given in Table 10

Eigenvalue	Value
$\hat{\lambda}_1$	680.6703
$\hat{\lambda}_2$	201.1009
$\hat{\lambda}_3$	114.7248
$\hat{\lambda}_4$	65.4486
$\hat{\lambda}_5$	37.3374

Table 10. Eigenvalues for the S-estimate of scatter under Model 2.

The exponential model can be verified by taking the differences between the logarithm of the eigenvalues. The difference between the log of any two successive eigenvalues should be constant for eigenvalues 2 through 5. That is the case, as can be seen in Table 11.

$\log(\hat{\lambda}_1) - \log(\hat{\lambda}_2)$	1.21927
$\log(\hat{\lambda}_2) - \log(\hat{\lambda}_3)$	0.56127
$\log(\hat{\lambda}_3) - \log(\hat{\lambda}_4)$	0.56127
$\log(\hat{\lambda}_4) - \log(\hat{\lambda}_5)$	0.56127

Table 11. Difference between log of successive eigenvalues.

The sandwich estimate of the asymptotic variance of the location parameters under the model 2 is

$$\begin{bmatrix} 297.38 & 119.02 & 100.23 & 111.12 & 118.27 \\ 119.02 & 174.93 & 87.56 & 99.63 & 105.39 \\ 100.23 & 87.56 & 110.40 & 109.01 & 122.47 \\ 111.12 & 99.63 & 109.01 & 218.78 & 155.91 \\ 118.27 & 105.39 & 122.47 & 155.91 & 310.99 \end{bmatrix}$$

and the estimate for the asymptotic variance of the φ parameters is

$$\begin{bmatrix} 1.41 & -0.67 & 0.85 \\ -0.67 & 0.53 & -0.88 \\ 0.85 & -0.88 & 6.38 \end{bmatrix}.$$

Note the lower asymptotic variance for the eigenvalues, as compared to the saturated model or Model 1.

The data set from Bollen (1989, pg. 30) comes from a study that assesses the reliability and validity of human perceptions versus physical measures of cloud cover. The data in Table 12 contain the perception estimates of three judges of the percent of the visible sky containing clouds in each of 60 slides. The rows of the table are not independent of each other since the same judge is perceiving each slide. Nonetheless, it is useful to compare the results of SURREAL to the results from the programs in this dissertation. In addition, observations 40, 51, and 52 vary widely between judges. Some of the slides that were hazy were interpreted as very cloudy by some judges, but almost clear by other judges. Hence, a robust estimate of the location, as well as the scatter matrix would be justified.

Table 12: Three Estimates of Percent Cloud Cover for 60 Slides

Observation	Cover1	Cover2	Cover3
1	0	5	0
2	20	20	20
3	80	85	90
4	50	50	70
5	5	2	5
6	1	1	2
7	5	5	2
8	0	0	0
9	10	15	5
10	0	0	0
11	0	0	0
12	10	30	10
13	0	2	2
14	10	10	5
15	0	0	0
16	0	0	0
17	5	0	20
18	10	20	20
19	20	45	15
20	35	75	60
21	90	99	100
22	50	90	80
23	35	85	70
24	25	15	40
25	0	0	0
26	0	0	0
27	10	10	20

Table 12: continued

Observation	Cover1	Cover2	Cover3
28	40	75	30
29	35	70	20
30	55	90	90
31	35	95	80
32	0	0	0
33	0	0	0
34	5	1	2
35	20	60	50
36	0	0	0
37	0	0	0
38	0	0	0
39	15	55	50
40	95	0	40
41	40	35	30
42	40	50	40
43	15	60	5
44	30	30	15
45	75	85	75
46	100	100	100
47	100	90	85
48	100	95	100
49	100	95	100
50	100	99	100
51	100	30	95
52	100	5	95
53	0	0	0
54	5	5	5
55	80	90	85
56	80	95	80
57	80	90	70
58	40	55	50
59	20	40	5
60	1	0	0

The results of the SURREAL algorithm produce the following estimates:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 27.01 \\ 34.30 \\ 30.08 \end{bmatrix}, \text{ and}$$

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 1566.53 & 1651.15 & 1602.35 \\ 1651.15 & 2037.37 & 1821.21 \\ 1602.35 & 1821.21 & 1832.99 \end{bmatrix}.$$

Again, the difference between the estimates is small. The estimates for $\boldsymbol{\beta}$ differed in the fifth decimal place, whereas the estimate for $\boldsymbol{\Sigma}$ differed in the third decimal place. Once again, the determinant of the estimate of the scatter matrix for the SURREAL algorithm is a little larger than that of the results of the algorithm in this dissertation. Also, the constraint q is slightly better satisfied by the estimates computed from the programs of this dissertation. The determinant and value of the constraint q for the estimates from each algorithm are given in Table 13.

	$\det \hat{\boldsymbol{\Sigma}}$	q
SURREAL	62,849,646.427	-8.6×10^{-8}
New	62,849,627.008	2.2×10^{-16}

Table 13. Determinant of the scatter estimate and value of the constraint q for each algorithm for the Cloud Cover data.

The sandwich estimate of the asymptotic variance of the location parameters is

$$\begin{bmatrix} 1382.98 & 1476.44 & 1421.90 \\ 1476.44 & 1791.83 & 1614.74 \\ 1421.90 & 1614.74 & 1606.18 \end{bmatrix}$$

and the estimate for the asymptotic variance of the $\boldsymbol{\varphi}$ parameters is

$$\begin{bmatrix} 4.85 & 3.96 & -0.17 \\ 3.96 & 13.22 & 1.13 \\ -0.17 & 1.13 & 2.36 \end{bmatrix}.$$

The flexibility of the models allows not only modeling of the scatter matrix, but also allows for modeling covariates of the location parameters. The data illustrated in this chapter were not appropriate for a model different than $\boldsymbol{Y} = \mathbf{1}_n\boldsymbol{\beta}' + \boldsymbol{E}$, because no covariates associated with the rows of $\boldsymbol{Y}$ were available. However, if covariates exist, the algorithms of this thesis can provide the appropriate estimates.

First order confidence intervals can be constructed for the estimators using the theory from Chapter 4. The next chapter presents simulations verifying the asymptotic distribution results from Chapter 4.

CHAPTER 6

SIMULATION

A simulation was conducted by generating data from a multivariate-T distribution with 13 degrees of freedom. Thirteen degrees of freedom was the fewest degrees of freedom allowed in order for all the necessary expectations involving the ρ function and its derivatives to exist.

The assumed model for the location parameters was with $\boldsymbol{X} = \mathbf{1}_n$. The dimension of the data was $p = 4$ with multiplicity vector $m = \begin{pmatrix} 2 & 1 & 1 \end{pmatrix}'$. The largest eigenvalue is modeled to have a multiplicity of two, and the three distinct eigenvalues are ordered, but otherwise unstructured. The $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ corresponding to this model is

$$\boldsymbol{T}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \boldsymbol{T}_2 = \boldsymbol{I}_3.$$

Five thousand trials were conducted with a sample size of 200 for each trial. For each trial, a data set was generated. Next, from 80 different random starting points, the smallest local minimizer was chosen as the S-estimator of multivariate location and scatter. The sandwich estimator, as well as the expected value of $\hat{\boldsymbol{\delta}}_1$ were computed.

The empirical estimate of the asymptotic variance of the location parameters is

$$n\frac{\hat{\boldsymbol{\beta}}' \left(\boldsymbol{I}_N - \mathbf{1}_N \left(\mathbf{1}_N'\mathbf{1}_N\right)^{-1} \mathbf{1}_N'\right) \boldsymbol{\beta}}{N-1} = \begin{bmatrix} 33.88 & 0.79 & -0.24 & 0.07 \\ 0.79 & 33.65 & 0.42 & -0.10 \\ -0.24 & 0.42 & 11.58 & 0.09 \\ 0.07 & -0.10 & 0.09 & 1.11 \end{bmatrix},$$

where N indicates the number of trials (5000). The average sandwich estimator for the location parameters over all 5000 trials is

$$\overline{\widehat{\text{Sand}}} = \begin{bmatrix} 33.64 & -0.02 & -0.04 & -0.01 \\ -0.02 & 33.73 & -0.03 & 0.00 \\ -0.04 & -0.03 & 11.22 & 0.00 \\ -0.01 & 0.00 & 0.00 & 1.12 \end{bmatrix}.$$

As can be seen by both of the estimates above, the sandwich estimator does an effective job in approximating the asymptotic variance of the location parameters.

For the eigenvalue parameters, the empirical estimate of the asymptotic variance is

$$n\frac{\hat{\boldsymbol{\varphi}}'\left(\boldsymbol{I}_N - \mathbf{1}_N\left(\mathbf{1}_N'\mathbf{1}_N\right)^{-1}\mathbf{1}_N'\right)\boldsymbol{\varphi}}{N-1} = \begin{bmatrix} 2.53 & -0.12 & 0.14 \\ -0.12 & 3.13 & -1.13 \\ 0.14 & -1.13 & 3.32 \end{bmatrix}$$

The average sandwich estimator for the eigenvalue parameters over all 5000 trials is

$$\overline{\widehat{\text{Sand}}} = \begin{bmatrix} 2.49 & -0.11 & 0.15 \\ -0.11 & 3.08 & -1.13 \\ 0.15 & -1.13 & 3.38 \end{bmatrix}$$

The sandwich estimator does an effective job in approximating the asymptotic variance of the eigenvalue parameters.

The actual bias was also compared to the theoretical bias. Keep in mind, though, that the actual scatter matrix being estimated is $\kappa\boldsymbol{\Sigma}^*$, where $\boldsymbol{\Sigma}^*$ is the characteristic matrix of the underlying elliptical distribution. Because the data were multivariate-T with 13 degrees of freedom and b_0 was computed under a Gaussian distribution, then the value for κ is about 1.14286. The empirical bias is computed by averaging over all 5000 trials the difference $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s\big)$. The empirical and theoretical estimates for the bias are given in Table 14.

Empirical bias for $\boldsymbol{\beta}$	Empirical bias for $\boldsymbol{\varphi}$	Theoretical bias for $\boldsymbol{\beta}$	Theoretical bias for $\boldsymbol{\varphi}$
-0.0027	-0.0252	0	-0.0027
0.0032	-0.0268	0	0.0268
0.0019	0.0009	0	0.0518
0.0012		0	

Table 14. Empirical and Theoretical bias for the parameters.

The simulations reflect the validity of the theory presented in Chapter 4. The bias equations still need to be verified.

CHAPTER 7

SUMMARY AND CONCLUSION

The first chapter of the dissertation presented the linear model, as well as models for scatter matrices. Maximum likelihood theory was reviewed. A motivation was given for the robust estimation of multivariate regression parameters and scatter matrices. Various robust estimators were introduced and their basic properties were described.

The second chapter introduced the parameterization of the multivariate regression model. The spectral model of Boik (2002a), along with the parameterization of the eigenvalue and eigenvector parameters was given. An algorithm for solving for the implicit parameter η was given (Boik 2002a).

Chapter 3 presented methods for estimating the parameters using either the M-estimating equations or the S-estimating equations. A modified Fisher-Scoring algorithm was presented for solving for the M-estimates using the likelihood from the multivariate-T distribution. First and second implicit derivatives for S-estimators were given. A two-step modified Newton-Raphson algorithm to solve for the S-estimates of the parameters was described, which included a hybrid Newton-Raphson-Bisection method to solve for the implicit parameter.

In chapter 4, first order asymptotic distributions for estimators of the parameters were given. Second order expansions of the estimators, along with a correction for the bias of the estimators were described. The computation of the sandwich estimator was described, as well as how to compute the estimated bias from the data.

In chapter 5, two examples were presented. First, the new algorithms was compared to Ruppert's SURREAL algorithm for each data set showing that they produce nearly identical results. Second, models were fit to illustrate the flexibility of the new estimators. In addition, the sandwich estimate of the asymptotic variance was given.

In chapter 6, simulations were given confirming the validity of the asymptotic theory.

Conclusion

Structured M and S-estimators of multivariate location and scatter can now be computed. A multivariate linear model can be fit, with modeling of the location parameters, as well as a spectral model for the eigenvalues. The focus for the model of the scatter matrix was a Principal Components model. However, the techniques can be applied to other scatter matrix models, as long as the scatter matrix can be written as a differentiable function of parameters. The spectral model can be of use in reducing the number of parameters in the covariance matrix, such as in the case of longitudinal studies. If covariates exist, then a linear model for the location parameters can be employed.

Most methods of modeling robust estimators involve a two step process. For instance, to do a robust Principal Components analysis, a robust estimate for $\mathbf{\Sigma}$ is found first. Then a standard Principal Component analysis is done, assuming that the robust estimator of $\mathbf{\Sigma}$ is the MLE. The validity of a two step procedure is not as clear as a one step procedure. However, this dissertation presents a one-step procedure for robust Principal Components. If we have a parsimonious model, then a one step estimator can be more efficient than a two step estimator.

For a saturated model (a special case of the linear model), a new method for the computation of S-estimators has been developed. The new method is derivative based, whereas SURREAL is an ad hoc search method (Rocke and Woodruff 1997). Because it is derivative based, local convergence is faster than SURREAL (e.g. once the iterations are close to the solution, the methods converge quickly to the solution). The results from the new method and SURREAL were nearly identical to each other. The new method satisfied the constraint a little better than SURREAL. In addition, the determinant of the covariance matrix under the new methods was a bit smaller than that of the estimate from SURREAL.

Examples were shown illustrating the flexibility of the modeling for the eigenvalues. In addition, the computation of the asymptotic variance of the estimators was given. Further, second order corrections for the bias of the estimators were obtained.

Future Research

Future research may include applying the linear model to other kinds of robust estimators. These robust estimators include Kent and Tyler's Redescending M-estimators and CM-estimators (Kent and Tyler 1991, 1996). The specification of a multivariate objective function for these estimators makes them an ideal candidate for the linear model.

Second order asymptotic expansions of the estimators can be used to find second order corrections for the skewness of the estimators. The second order bias and skewness can then be used to find second order corrections for confidence intervals, thus providing better coverage for the parameters.

The linear model may be extended to cases in which there is a functional relationship between the location parameters and the spectral model parameters. This dissertation did not allow such a functional relationship.

Another area of research would be to employ different models for scatter matrices, such as a Toeplitz model, or a Function Analysis model.

To improve the global optimization of the problem, the algorithm can be mixed with a Line-Search method or a Trust Region method (Nocedal and Wright 1999). Other methods that may be worth investigating are steepest descent or other optimization procedures.

REFERENCES CITED

Abramowitz, M. and Stegun, I. A. (eds.) (1992), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, New York: Dover Publications Inc., reprint of the 1972 edition.

Balestra, P. (1976), in *La Derivation Matricielle*, Paris: Sirey, no. 12 in Collection De L'IME.

Boik, R. J. (1998), "A Local Parameterization of Orthogonal and Semi-orthogonal Matrices With Applications," *Journal of Multivariate Analysis*, 67, 244–276.

— (2002a), "Spectral models for covariance matrices," *Biometrika*, 89, 159–182.

— (2002b), "Spectral Models For Covariance Matrices: Supplement," Department of Mathematical Sciences, Montana State University, Bozeman MT 59715. Available from `http://www.math.montana.edu/~rjboik/spectral/`.

— (2003), "Principal component models for correlation matrices," *Biometrika*, 90, 679–701.

Bollen, K. A. (1989), *Structural equations with latent variables*, John Wiley & Sons.

Burden, R. and Faires, J. D. (1997), *Numerical Analysis*, Pacific Grove, CA: Brooks/Cole Publishing Co., sixth ed., with 1 IBM-PC floppy disk (3.5 inch; HD).

Casella, G. and Berger, R. L. (1990), *Statistical inference*, Duxbury Press.

Chu, K. C. (1973), "Estimation and decision for linear systems with elliptical random processes," *IEEE Trans. Automatic Control*, AC-18, 499–505.

Coakley, C. W. and Hettmansperger, T. P. (1993), "A bounded influence, high breakdown, efficient regression estimator," *Journal of the American Statistical Association*, 88, 872–880.

Croux, C., Rousseeuw, P. J., and Hössjer, O. (1994), "Generalized S-estimators," *Journal of the American Statistical Association*, 89, 1271–1281.

Davies, L. (1992), "The asymptotics of Rousseeuw's minimum volume ellipsoid estimator," *The Annals of Statistics*, 20, 1828–1843.

Davies, P. L. (1987), "Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices," *The Annals of Statistics*, 15, 1269–1292.

Donoho, D. L. and Huber, P. J. (1983), "The notion of breakdown point," in *A Festschrift for Erich L. Lehmann*, pp. 157–184.

Edgeworth, F. Y. (1887), "On Observations Relating to Several Quantities," *Hermathena*, 6, 279–285.

Fang, K.-T., Kotz, S., and Ng, K. W. (1990), *Symmetric multivariate and related distributions*, Chapman & Hall Ltd.

Fulks, W. (1979), *Advanced Calculus, An Introduction to Analysis, 3rd Ed.*, John Wiley & Sons.

Hampel, F. R. (1968), "Contributions to the theory of robust estimation," Ph.D. thesis, University of California, Berkeley.

Huber, P. J. (1964), "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, 35, 73–101.

— (1981), *Robust statistics*, John Wiley & Sons.

Kent, J. T. and Tyler, D. E. (1991), "Redescending M-estimates of multivariate location and scatter," *The Annals of Statistics*, 19, 2102–2119.

— (1996), "Constrained M-estimation for multivariate location and scatter," *The Annals of Statistics*, 24, 1346–1370.

Khatri, C. G. and Rao, C. R. (1968), "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhyā, Series A, Indian Journal of Statistics*, 30, 167–180.

Lopuhaä, H. P. (1989), "On the relation between S-estimators and M-estimators of multivariate location and covariance," *The Annals of Statistics*, 17, 1662–1683.

— (1992), "Highly efficient estimators of multivariate location with high breakdown point," *The Annals of Statistics*, 20, 398–413.

MacRae, E. C. (1974), "Matrix derivatives with an application to an adaptive linear decision problem," *The Annals of Statistics*, 2, 337–346.

Magnus, J. R. (1988), *Linear structures*, Oxford University Press.

Magnus, J. R. and Neudecker, H. (1979), "The commutation matrix: Some properties and applications," *The Annals of Statistics*, 7, 381–394.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate analysis*, Academic Press.

Maronna, R. A. (1976), "Robust M-estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 4, 51–67.

Nocedal, J. and Wright, S. J. (1999), *Numerical Optimization*, New York: Springer-Verlag.

Pawitan, Y. (2001), *In all likelihood: statistical modelling and inference using likelihood*, Oxford University Press.

Press, S. J. (1982), *Applied multivariate analysis: Using Bayesian and Frequentist methods of inference (Second edition)*, Krieger Publishing Company, Inc.

Rocke, D. M. and Woodruff, D. L. (1997), "Robust estimation of multivariate location and shape," *Journal of Statistical Planning and Inference*, 57, 245–255.

Rousseeuw, P. and Yohai, V. (1984), “Robust regression by means of S-estimators,” in *Robust and nonlinear time series analysis*, pp. 256–272.

Rousseeuw, P. J. (1983), “Multivariate estimation with high breakdown point,” in *Fourth Pannonian Symposium on Mathematical Statistics*, Bad Tatzmanndorf, Austria.

— (1984), “Least median of squares regression,” *Journal of the American Statistical Association*, 79, 871–880.

— (1985), “Multivariate estimation with high breakdown point,” in *Mathematical Statistics with Applications*, pp. 283–297.

Ruppert, D. (1992), “Computing S estimators for regression and multivariate location/dispersion,” *Journal of Computational and Graphical Statistics*, 1, 253–270.

SAS Institute, I. (1997), *SAS/STAT Software: Changes and Enhancements through release 6.12*, Cary, NC: SAS Institute, Inc.

Searle, S. R. (1982), *Matrix algebra useful for statistics*, John Wiley & Sons.

Sen, P. K. and Singer, J. d. M. (1993), *Large sample methods in statistics: an introduction with applications*, Chapman & Hall Ltd.

Siegel, A. F. (1982), “Robust regression using repeated medians,” *Biometrika*, 69, 242–244.

Yohai, V. J. (1987), “High breakdown-point and high efficiency robust estimates for regression,” *The Annals of Statistics*, 15, 642–656.

You, J. (1999), “A Monte Carlo comparison of several high breakdown and efficient estimators,” *Computational Statistics and Data Analysis*, 30, 205–219.

Yuan, K.-H. and Bentler, P. M. (1998), “Robust mean and covariance structure analysis,” *The British Journal of Mathematical and Statistical Psychology*, 51, 63–88.

Yuan, K.-H. and Jennrich, R. I. (1998), “Asymptotics of estimating equations under natural conditions,” *Journal of Multivariate Analysis*, 65, 245–260.

APPENDICES

APPENDIX A

Spectral Model derivatives

Derivatives of $\boldsymbol{\lambda}$

The models for the eigenvalues are found in (2.14), (2.15), and (2.16). The derivatives for Model (2.15) are

$$\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\lambda}}^{(1)} &= \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}'} = \left(\exp\{\odot \boldsymbol{T}_2\boldsymbol{\varphi}\}' * \boldsymbol{T}_1\right)\boldsymbol{T}_2, \\
\boldsymbol{D}_{\boldsymbol{\lambda}}^{(2)} &= \frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \left(\exp\{\odot \boldsymbol{T}_2\boldsymbol{\varphi}\}' * \boldsymbol{T}_1\right)\left(\boldsymbol{T}_2' * \boldsymbol{T}_2'\right)', \text{ and} \\
\boldsymbol{D}_{\boldsymbol{\lambda}}^{(3)} &= \frac{\partial^3 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \left(\exp\{\odot \boldsymbol{T}_2\boldsymbol{\varphi}\}' * \boldsymbol{T}_1\right)\left(\boldsymbol{T}_2' * \boldsymbol{T}_2' * \boldsymbol{T}_2'\right)',
\end{aligned}$$

where $*$ denotes the Khatri-Rao column wise product.

Derivatives of $\operatorname{vec} \boldsymbol{G}$

The implicit derivatives of $\operatorname{vec} \boldsymbol{G}$ are used in the derivatives of $\operatorname{vec} \boldsymbol{\Sigma}$. The matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are indicator matrices that place $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ in the correct places in $\operatorname{vec} \boldsymbol{G}$. They satisfy the following properties:

$$\begin{aligned}
&\boldsymbol{A}_1' \operatorname{vec} \boldsymbol{G} = \boldsymbol{\mu}, && \boldsymbol{A}_1'\boldsymbol{A}_1 = \boldsymbol{I}_{\nu_2}, && \boldsymbol{D}_p'\boldsymbol{A}_2 = \boldsymbol{I}_{\frac{p(p+1)}{2}}, \\
&\boldsymbol{A}_2' \operatorname{vec} \boldsymbol{G} = \boldsymbol{\eta}, && \boldsymbol{A}_2'\boldsymbol{A}_2 = \boldsymbol{I}_{\frac{p(p+1)}{2}}, \text{ and} && \boldsymbol{A}_2'\boldsymbol{A}_1 = \boldsymbol{0},
\end{aligned}$$

where $\boldsymbol{D}_p$ is the duplication matrix (Magnus 1988). The derivatives of $\operatorname{vec} \boldsymbol{G}$ are

$$\begin{aligned}
\boldsymbol{D}_{\boldsymbol{G}}^{(1)} &= \left.\frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} && = \boldsymbol{A}_1 - \boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{A}_1 = \left[\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)}\right]\boldsymbol{A}_1 \\
\boldsymbol{D}_{\boldsymbol{G}}^{(2)} &= \left.\frac{\partial^2 \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} && = \boldsymbol{A}_2\boldsymbol{D}_p'\left(\boldsymbol{I}_p \otimes \operatorname{vec}(\boldsymbol{I}_p)' \otimes \boldsymbol{I}_p\right)\left(\boldsymbol{D}_{\boldsymbol{G}}^{(1)} \otimes \boldsymbol{D}_{\boldsymbol{G}}^{(1)}\right) \\
\boldsymbol{D}_{\boldsymbol{G}}^{(3)} &= \left.\frac{\partial^3 \operatorname{vec}(\boldsymbol{G})}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} && = -\boldsymbol{A}_2\boldsymbol{D}_p'\left(\boldsymbol{I}_p \otimes \operatorname{vec}(\boldsymbol{I}_p)' \otimes \boldsymbol{I}_p\right)\Big[\left(\boldsymbol{D}_{\boldsymbol{G}}^{(1)} \otimes \boldsymbol{I}_{(p,p)}\boldsymbol{D}_{\boldsymbol{G}}^{(2)}\right) \\
& && \quad - \left(\boldsymbol{D}_{\boldsymbol{G}}^{(2)} \otimes \boldsymbol{D}_{\boldsymbol{G}}^{(1)}\right)\Big(\boldsymbol{I}_{\nu_2^3} + \left\{\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{I}_{(\nu_2,\nu_2)}\right\}\Big)\Big].
\end{aligned}$$

Derivatives of vec $\mathbf{\Sigma}$

The derivatives of vec $\mathbf{\Sigma}$ with respect to the parameters are used in both the M-estimators and the S-estimators. The derivatives that pertain to both M-estimation and S-estimation are given here and are special cases of derivatives from Boik (2002b). Those specific to S-estimation are given in Appendix C.

First Derivatives.

$$\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} = \left.\frac{\partial \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = 2\boldsymbol{N}_p\left(\mathbf{\Gamma}_0\mathbf{\Lambda} \otimes \mathbf{\Gamma}_0\right)\boldsymbol{D}^{(1)}_{\boldsymbol{G}}$$

$$\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} = \left.\frac{\partial \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \left(\mathbf{\Gamma}_0 \otimes \mathbf{\Gamma}_0\right)\boldsymbol{L}_p\boldsymbol{D}^{(1)}_{\boldsymbol{\lambda}} = \left(\mathbf{\Gamma}_0 * \mathbf{\Gamma}_0\right)\boldsymbol{D}^{(1)}_{\boldsymbol{\lambda}}$$

Second Derivatives.

$$\boldsymbol{F}^{(2)}_{\boldsymbol{\mu}\boldsymbol{\mu}} = \left.\frac{\partial^2 \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = 2\boldsymbol{N}_p\Big\{\left(\mathbf{\Gamma}_0\mathbf{\Lambda} \otimes \mathbf{\Gamma}_0\right)\boldsymbol{D}^{(2)}_{\boldsymbol{G}} - \big(\mathbf{\Gamma}_0 \otimes \operatorname{vec}(\mathbf{\Lambda})' \otimes \mathbf{\Gamma}_0\big)\big(\boldsymbol{D}^{(1)}_{\boldsymbol{G}} \otimes \boldsymbol{D}^{(1)}_{\boldsymbol{G}}\big)\Big\}$$

$$\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = \left.\frac{\partial^2 \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \left(\mathbf{\Gamma}_0 \otimes \mathbf{\Gamma}_0\right)\boldsymbol{L}_p\boldsymbol{D}^{(2)}_{\boldsymbol{\lambda}} = \left(\mathbf{\Gamma}_0 * \mathbf{\Gamma}_0\right)\boldsymbol{D}^{(2)}_{\boldsymbol{\lambda}}$$

$$\boldsymbol{F}^{(2)}_{\boldsymbol{\mu}\boldsymbol{\varphi}} = \left.\frac{\partial^2 \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = -2\boldsymbol{N}_p\big(\mathbf{\Gamma}_0 \otimes \operatorname{vec}(\boldsymbol{I}_p)' \otimes \mathbf{\Gamma}_0\big)\big(\boldsymbol{D}^{(1)}_{\boldsymbol{G}} \otimes \boldsymbol{L}_p\boldsymbol{D}^{(1)}_{\boldsymbol{\lambda}}\big)$$

$$\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\mu}} = \left.\frac{\partial^2 \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \boldsymbol{F}^{(2)}_{\boldsymbol{\mu}\boldsymbol{\varphi}}\boldsymbol{I}_{(\nu_2,\nu_1)}$$

Third Derivatives.

$$\begin{aligned}\boldsymbol{F}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\mu}\boldsymbol{\mu}} = \left.\frac{\partial^3 \operatorname{vec}(\mathbf{\Sigma})}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} &= 2\boldsymbol{N}_p\Big[\left(\mathbf{\Gamma}_0\mathbf{\Lambda} \otimes \mathbf{\Gamma}_0\right)\boldsymbol{D}^{(3)}_{\boldsymbol{G}} - \big(\mathbf{\Gamma}_0 \otimes \operatorname{vec}(\mathbf{\Lambda})' \otimes \mathbf{\Gamma}_0\big) \\ &\quad \times \Big\{\big(\boldsymbol{D}^{(1)}_{\boldsymbol{G}} \otimes \boldsymbol{D}^{(2)}_{\boldsymbol{G}}\big) + 2\big(\boldsymbol{I}_{(p,p)}\boldsymbol{D}^{(2)}_{\boldsymbol{G}} \otimes \boldsymbol{D}^{(1)}_{\boldsymbol{G}}\big)\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{N}_{\nu_2}\big)\Big\}\Big]\end{aligned}$$

$$\boldsymbol{F}^{(3)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\varphi}} = \left.\frac{\partial^3 \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \left(\mathbf{\Gamma}_0 \otimes \mathbf{\Gamma}_0\right)\boldsymbol{L}_p\boldsymbol{D}^{(3)}_{\boldsymbol{\lambda}} = \left(\mathbf{\Gamma}_0 * \mathbf{\Gamma}_0\right)\boldsymbol{D}^{(3)}_{\boldsymbol{\lambda}}$$

$$\boldsymbol{F}^{(3)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\mu}} = \left.\frac{\partial^3 \operatorname{vec} \mathbf{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = 2\boldsymbol{N}_p\big(\mathbf{\Gamma}_0 \otimes \operatorname{vec}(\boldsymbol{I}_p)' \otimes \mathbf{\Gamma}_0\big)\big(\boldsymbol{L}_p\boldsymbol{D}^{(2)}_{\boldsymbol{\lambda}} \otimes \boldsymbol{D}^{(1)}_{\boldsymbol{G}}\big)$$

$$\boldsymbol{F}^{(3)}_{\boldsymbol{\varphi\mu\mu}} = \left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = 2\boldsymbol{N}_p\Big[\left(\boldsymbol{\Gamma}_0\otimes\operatorname{vec}(\boldsymbol{I}_p)'\otimes\boldsymbol{\Gamma}_0\right)\left(\boldsymbol{L}_p\boldsymbol{D}^{(1)}_{\boldsymbol{\lambda}}\otimes\boldsymbol{D}^{(2)}_{\boldsymbol{G}}\right)$$
$$+\left(\operatorname{vec}\left(\boldsymbol{L}_p\boldsymbol{D}^{(1)}_{\boldsymbol{\lambda}}\right)'\otimes\boldsymbol{\Gamma}_0\otimes\boldsymbol{\Gamma}_0\right)$$
$$\times\left\{\boldsymbol{I}_{\nu_3}\otimes\left(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{D}^{(1)}_{\boldsymbol{G}}\otimes\boldsymbol{D}^{(1)}_{\boldsymbol{G}}\right)\right\}\Big]$$

Rearrangements of Derivatives

Rearrangements of derivatives are necessary for second and third order derivatives. The rearrangements relate the expressions for $\boldsymbol{F}^{(2)}_{..}$ and $\boldsymbol{F}^{(3)}_{...}$ to $\boldsymbol{F}^{(11)}_{..}$ and $\boldsymbol{F}^{(111)}_{...}$ through the dvec operator. The relationship between $\boldsymbol{F}^{(11)}_{..}$, $\boldsymbol{F}^{(2)}_{..}$ and $\boldsymbol{F}^{(111)}_{...}$, $\boldsymbol{F}^{(3)}_{...}$ is

$$\boldsymbol{F}^{(11)}_{\boldsymbol{\mu\mu}} = \left.\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(2)}_{\boldsymbol{\mu\mu}}, p^2\nu_2, \nu_2\right],$$
$$\boldsymbol{F}^{(11)}_{\boldsymbol{\varphi\varphi}} = \left.\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi\varphi}}, p^2\nu_3, \nu_3\right],$$
$$\boldsymbol{F}^{(11)}_{\boldsymbol{\mu\varphi}} = \left.\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(2)}_{\boldsymbol{\mu\varphi}}, p^2\nu_3, \nu_2\right],$$
$$\boldsymbol{F}^{(11)}_{\boldsymbol{\varphi\mu}} = \left.\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi\mu}}, p^2\nu_3, \nu_2\right],$$
$$\boldsymbol{F}^{(111)}_{\boldsymbol{\mu\mu\mu}} = \left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}\otimes\partial\boldsymbol{\mu}\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(3)}_{\boldsymbol{\mu\mu\mu}}, p^2\nu_2^2, \nu_2\right],$$
$$\boldsymbol{F}^{(111)}_{\boldsymbol{\varphi\varphi\varphi}} = \left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}\otimes\partial\boldsymbol{\varphi}\otimes\partial\boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(3)}_{\boldsymbol{\varphi\varphi\varphi}}, p^2\nu_3^2, \nu_3\right],$$
$$\boldsymbol{F}^{(111)}_{\boldsymbol{\varphi\varphi\mu}} = \left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}\otimes\partial\boldsymbol{\varphi}\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(3)}_{\boldsymbol{\varphi\varphi\mu}}\boldsymbol{I}_{(\nu_2,\nu_3^2)}, p^2\nu_3^2, \nu_2\right], \text{ and}$$
$$\boldsymbol{F}^{(111)}_{\boldsymbol{\mu\mu\varphi}} = \left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}\otimes\partial\boldsymbol{\mu}\otimes\partial\boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \operatorname{dvec}\left[\boldsymbol{F}^{(3)}_{\boldsymbol{\varphi\mu\mu}}, p^2\nu_2^2, \nu_3\right].$$

APPENDIX B

M-estimator derivatives (evaluated at $\boldsymbol{\mu} = \mathbf{0}$)

Matrix Definitions

The following table gives expressions for commonly used matrices in M-estimation. The $*$ symbol denotes the Khatri-Rao column-wise product (Khatri and Rao 1968).

$$\boldsymbol{K}_{zx} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\right) \qquad \boldsymbol{R}_3 = \frac{1}{n}\,\mathrm{Diag}\left(v_2^{(1)}(d_k^2)\right)$$

$$\boldsymbol{K}_{zz} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\right) \qquad \boldsymbol{R}_{xx} = \boldsymbol{X}'\boldsymbol{R}_0\boldsymbol{Z}$$

$$\boldsymbol{R}_0 = \frac{1}{n}\,\mathrm{Diag}\left(v_1(d_k)\right) \qquad \boldsymbol{R}_{xz} = \boldsymbol{X}'\boldsymbol{R}_0\boldsymbol{Z}$$

$$\boldsymbol{R}_1 = \frac{1}{n}\,\mathrm{Diag}\left(\frac{v_1^{(1)}(d_k)}{2d_k}\right) \qquad \boldsymbol{R}_{zx} = \boldsymbol{Z}'\boldsymbol{R}_2\boldsymbol{X}$$

$$\boldsymbol{R}_2 = \frac{1}{n}\,\mathrm{Diag}\left(v_2(d_k^2)\right) \qquad \boldsymbol{R}_{zz} = \boldsymbol{Z}'\boldsymbol{R}_2\boldsymbol{Z}$$

First Derivatives

Under the linear and spectral models defined in (1.2) and (2.5), the first derivatives of the log likelihood function of the family of elliptically contoured distributions with respect to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\zeta} = \begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\varphi}\end{pmatrix}$ are

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = \sum_{k=1}^{n} \frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\boldsymbol{z}_k'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_k = \mathrm{vec}\big(\boldsymbol{X}'\boldsymbol{T}_g\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\big)' \text{ and}$$

$$\begin{aligned}\left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\mu=\mathbf{0}} &= \frac{n}{2}\,\mathrm{vec}\left[\sum_{k=1}^{n}\frac{-2g^{(1)}(d_k^2)}{ng(d_k^2)}\boldsymbol{z}_k\boldsymbol{z}_k' - \boldsymbol{\Sigma}\right]'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}\\ &= \frac{n}{2}\,\mathrm{vec}\left[\frac{1}{n}\boldsymbol{Z}'\boldsymbol{T}_g\boldsymbol{Z} - \boldsymbol{\Sigma}\right]\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)},\end{aligned}$$

where $g^{(1)}(d_k^2) = \frac{\partial g(d_k^2)}{\partial d_k^2}$, $\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}$, $\boldsymbol{X}_k$ is defined in (1.4), $\boldsymbol{F}^{(1)} = \begin{bmatrix}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}\\ \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\end{bmatrix}$, and $\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}$, $\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}$ are defined in Theorem 2.6, and $\boldsymbol{T}_g = \mathrm{Diag}\left(\frac{-2g^{(1)}(d_k^2)}{g(d_k^2)}\right)$.

The first derivatives of the log-likelihood function of the Multivariate-T distribution with respect to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\varphi} \end{bmatrix}$ are

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}'} = (\xi + p) \sum_{k=1}^{n} \frac{z_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_k}{\xi + d_k^2} = \operatorname{vec}\big(\boldsymbol{X}' \boldsymbol{T}_g \boldsymbol{Z} \boldsymbol{\Sigma}^{-1}\big)',$$

$$\begin{aligned} \left.\frac{\partial \ell}{\partial \boldsymbol{\zeta}'}\right|_{\mu=\mathbf{0}} &= \frac{n}{2} \operatorname{vec}\left[\frac{1}{n} \sum_{k=1}^{n} \frac{\xi + p}{\xi + d_k^2} \boldsymbol{z}_k \boldsymbol{z}_k' - \boldsymbol{\Sigma}\right]' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)} \\ &= \frac{n}{2} \operatorname{vec}\left[\frac{1}{n} \boldsymbol{Z}' \boldsymbol{T}_g \boldsymbol{Z} - \boldsymbol{\Sigma}\right] \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}, \end{aligned}$$

where $\boldsymbol{T}_g = \operatorname{Diag}\left(\frac{\xi+p}{\xi+d_k^2}\right)$, $\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta}$, $\boldsymbol{X}_k$ is defined in (1.4), $\boldsymbol{F}^{(1)} = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\mu}}^{(1)} \\ \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \end{bmatrix}$, and $\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}$ and $\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}$ are defined in Theorem 2.6.

Estimating Function

The estimating function for the M-estimators is

$$\boldsymbol{l}_{\boldsymbol{\theta}_m}^{(1)} = \begin{bmatrix} \boldsymbol{l}_{\boldsymbol{\beta}}^{(1)} \\ \boldsymbol{l}_{\boldsymbol{\zeta}}^{(1)} \end{bmatrix} = \begin{bmatrix} \operatorname{vec}(\boldsymbol{R}_{xz} \boldsymbol{\Sigma}^{-1}) \\ \boldsymbol{F}^{(1)\prime} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}(\boldsymbol{R}_{zz} - \boldsymbol{\Sigma}) \end{bmatrix}$$

Fisher's Information Matrix

The expected value of the negative Hessian matrix for the Multivariate-T distribution is

$$E\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right] = \frac{\xi + p}{\xi + p + 2} \begin{bmatrix} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}' \boldsymbol{X} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{F}^{(1)\prime} \boldsymbol{T}_{\boldsymbol{\Sigma}} \boldsymbol{F}^{(1)} \end{bmatrix},$$

where $\boldsymbol{F}^{(1)}$ is given in Theorem 3.1, and

$$\boldsymbol{T}_{\boldsymbol{\Sigma}} = \frac{n}{2}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \frac{1}{\xi + p} \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big) \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\big)'\right].$$

Second Derivatives

The first derivatives of the estimating function (like second derivatives of the likelihood equation) are

$$
\begin{aligned}
\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\beta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'} = -\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{R}_{xx}\right) + 2\boldsymbol{K}_{zx}\boldsymbol{R}_1\boldsymbol{K}'_{zx}\right] \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta\beta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\zeta}'} = -\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{R}_{xz}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{K}_{zx}\boldsymbol{R}_1\boldsymbol{K}'_{zz}\right]\boldsymbol{F}^{(1)} \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\zeta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}}}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{F}^{(1)'}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{zx}\right) + \boldsymbol{K}_{zz}\boldsymbol{R}_3\boldsymbol{K}'_{zx}\right] \\
\boldsymbol{l}^{(2)}_{\boldsymbol{\zeta\zeta}} &= \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\zeta}}}{\partial \boldsymbol{\zeta}'} = \boldsymbol{F}^{(11)'}\left[\boldsymbol{I}_{\nu_2+\nu_3} \otimes \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{zz}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\right)\right] \\
&\quad - 2\boldsymbol{F}^{(1)'}\left[(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{zz}\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}\boldsymbol{K}_{zz}\boldsymbol{R}_3\boldsymbol{K}'_{zz}\right]\boldsymbol{F}^{(1)} \\
&\quad + \boldsymbol{F}^{(1)'}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)}.
\end{aligned}
$$

APPENDIX C

S-estimator derivatives (evaluated at $\boldsymbol{\mu} = \mathbf{0}$)

Scalar and vector definitions

$$\boldsymbol{a} = \frac{\partial q^*}{\partial \boldsymbol{\varphi}} \qquad w = n \log |\boldsymbol{\Sigma}|$$

$$\boldsymbol{\beta} = \operatorname{vec}(\boldsymbol{B}) \qquad w_k^{(1)} = \frac{\partial \rho(d_k)}{\partial h_k} = \frac{1}{2} \frac{\rho^{(1)}(d_k)}{d_k}$$

$$h_k = \boldsymbol{z}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_k = d_k^{\frac{1}{2}} \qquad w_k^{(2)} = \frac{\partial w_k^{(1)}}{\partial h_k} = \frac{1}{4} \left[\frac{\rho^{(2)}(d_k)}{d_k^2} - \frac{\rho^{(1)}(d_k)}{d_k^3} \right]$$

$$\rho^{(i)}(d_k) = \frac{\partial^i \rho(d_k)}{\partial d_k^i} \qquad w_k^{(3)} = \frac{\partial w_k^{(2)}}{\partial h_k} = \frac{1}{8} \left[\frac{\rho^{(3)}(d_k)}{d_k^3} - 3\frac{\rho^{(2)}(d_k)}{d_k^4} + 3\frac{\rho^{(1)}(d_k)}{d_k^5} \right]$$

$$q = \frac{1}{n} \sum_{k=1}^{n} \rho(d_k) - b_0 \qquad \boldsymbol{x}_k = \boldsymbol{X}' \boldsymbol{e}_k$$

$$\boldsymbol{r} = \boldsymbol{v}_2 \left(\boldsymbol{a}' \boldsymbol{v}_2 \right)^{-1} \qquad \boldsymbol{y}_k = \boldsymbol{Y}' \boldsymbol{e}_k$$

$$\boldsymbol{z}_k = \boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta}$$

Derivatives of ρ

The ρ function is chosen by the investigator. One choice for ρ is the integral of Tukey's biweight function. This function and its first three derivatives are:

$$\rho(y, c_0) = \begin{cases} \frac{y^2}{2} - \frac{y^4}{2c_0^2} + \frac{y^6}{6c_0^4} & \text{if } |y| \leqslant c_0; \\ \frac{c_0^2}{6} & \text{if } |y| > c_0. \end{cases}$$

$$\rho^{(1)}(y, c_0) = y \left[1 - \left(\frac{y}{c_0} \right)^2 \right]^2 I_{[-c_0, c_0]}(y),$$

$$\rho^{(2)}(y, c_0) = \left[1 - \left(\frac{y}{c_0} \right)^2 \right] \left[1 - 5\left(\frac{y}{c_0} \right)^2 \right] I_{[-c_0, c_0]}(y), \text{ and}$$

$$\rho^{(3)}(y, c_0) = \frac{-4y}{c_0^2} \left[3 - 5\left(\frac{y}{c_0} \right)^2 \right] I_{[-c_0, c_0]}(y),$$

where

$$I_{[-c_0, c_0]}(y) = \begin{cases} 1 & \text{if } y \in [-c_0, c_0]; \\ 0 & \text{if } y \notin [-c_0, c_0]. \end{cases}$$

Matrix definitions

The following table gives expressions for commonly used matrices. The $*$ symbol denotes the Khatri-Rao column-wise product (Khatri and Rao 1968). The $\otimes$ symbol signifies the right Kronecker product.

$$\boldsymbol{K}_{xx} = \left(\boldsymbol{X}' * \boldsymbol{X}'\right) \qquad \boldsymbol{M}_{xx} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xx}\right) + 2\boldsymbol{V}_{xx}$$

$$\boldsymbol{K}_{zx} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\right) \qquad \boldsymbol{M}_{xz} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{V}_{xz} = \boldsymbol{M}'_{zx}$$

$$\boldsymbol{K}_{zz} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\right) \qquad \boldsymbol{M}_{zz} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\right) + \frac{1}{2}\boldsymbol{V}_{zz}$$

$$\boldsymbol{K}_{xzx} = \left(\boldsymbol{X}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\right) \qquad \boldsymbol{V}_{xx} = \boldsymbol{K}_{zx}\boldsymbol{W}_2\boldsymbol{K}'_{zx}$$

$$\boldsymbol{K}_{xzz} = \left(\boldsymbol{X}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\right) \qquad \boldsymbol{V}_{xz} = \boldsymbol{K}_{zx}\boldsymbol{W}_2\boldsymbol{K}'_{zz}$$

$$\boldsymbol{K}_{zzx} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{X}'\right) \qquad \boldsymbol{V}_{zz} = \boldsymbol{K}_{zz}\boldsymbol{W}_2\boldsymbol{K}'_{zz}$$

$$\boldsymbol{K}_{zzz} = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}' * \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\right) \qquad \boldsymbol{W}_i = \frac{1}{n}\,\mathrm{Diag}\left(w_k^{(i)}\right)$$

$$\boldsymbol{K}_{zxzx} = \left(\boldsymbol{K}_{zx} * \boldsymbol{K}_{zx}\right) \qquad \boldsymbol{W}_{xx} = \boldsymbol{X}'\boldsymbol{W}_1\boldsymbol{X}$$

$$\boldsymbol{K}_{zxzz} = \left(\boldsymbol{K}_{zx} * \boldsymbol{K}_{zz}\right) \qquad \boldsymbol{W}_{xz} = \boldsymbol{X}'\boldsymbol{W}_1\boldsymbol{Z} = \boldsymbol{W}'_{zx}$$

$$\boldsymbol{K}_{zzzx} = \left(\boldsymbol{K}_{zz} * \boldsymbol{K}_{zx}\right) \qquad \boldsymbol{W}_{zz} = \boldsymbol{Z}'\boldsymbol{W}_1\boldsymbol{Z}$$

$$\boldsymbol{K}_{zzzz} = \left(\boldsymbol{K}_{zz} * \boldsymbol{K}_{zz}\right) \qquad \boldsymbol{X}_k = \boldsymbol{I}_p \otimes \boldsymbol{x}'_k$$

$$\boldsymbol{N}_p = \frac{1}{2}\left[\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right] \qquad \boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}$$

$$\boldsymbol{N}^{\perp}_{\nu_3} = \frac{1}{2}\left[\boldsymbol{I}_{\nu_3^2} - \boldsymbol{I}_{(\nu_3,\nu_3)}\right]$$

Derivatives with η_ψ constant

The notation w^* and q^* is used for the derivatives of w and q when η_ψ is constant. Specifically, when η_ψ is functionally independent of the other variables,

$$w^* = w(\boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi\}), \; q^* = q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi\}).$$

First derivatives of w and q.

$$\frac{\partial w^*}{\partial \boldsymbol{\beta}} = \mathbf{0} \text{ (by construction)} \qquad \frac{\partial q^*}{\partial \boldsymbol{\beta}} = -2\,\text{vec}\big(\boldsymbol{W}_{xz}\boldsymbol{\Sigma}^{-1}\big)$$

$$\frac{\partial w^*}{\partial \boldsymbol{\mu}} = \mathbf{0} \qquad \frac{\partial q^*}{\partial \boldsymbol{\mu}} = -\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime}\,\text{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)$$

$$\frac{\partial w^*}{\partial \boldsymbol{\varphi}} = n\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\,\text{vec}\big(\boldsymbol{\Sigma}^{-1}\big) \qquad \frac{\partial q^*}{\partial \boldsymbol{\varphi}} = -\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\,\text{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)$$

Second derivatives of w.

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = \mathbf{0} \qquad \frac{\partial^2 w^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = \mathbf{0} \qquad \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = \mathbf{0}$$

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \mathbf{0} \qquad \frac{\partial^2 w^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = \mathbf{0} \qquad \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = \mathbf{0}$$

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = \mathbf{0} \qquad \frac{\partial^2 w^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = \mathbf{0}$$

$$\frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = n\left[\boldsymbol{F}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_3} \otimes \text{vec}\big(\boldsymbol{\Sigma}^{-1}\big)\right) - \boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)}\right]$$

Second derivatives of q.

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xx} \qquad \frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xz}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)} \qquad \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = 2\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime}\boldsymbol{M}_{zx}$$

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = 2\boldsymbol{M}_{xz}\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)} \qquad \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = 2\boldsymbol{F}_{\boldsymbol{\varphi}}^{(1)\prime}\boldsymbol{M}_{zx}$$

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}_{\boldsymbol{\mu}\boldsymbol{\mu}}^{(11)\prime}\left(\boldsymbol{I}_{\nu_2} \otimes \text{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\right) + 2\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)\prime}\boldsymbol{M}_{zz}\boldsymbol{F}_{\boldsymbol{\mu}}^{(1)}$$

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}^{(11)'}_{\boldsymbol{\varphi}\boldsymbol{\mu}} \left(\boldsymbol{I}_{\nu_2} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\right) + 2\boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}$$

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}^{(11)'}_{\boldsymbol{\mu}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\right) + 2\boldsymbol{F}^{(1)'}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$$

$$\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}^{(11)'}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3} \otimes \operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\right) + 2\boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$$

Third derivatives of w. All third order derivatives of w when η_ψ is functionally independent of the other variables are zero with one exception:

$$\frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \boldsymbol{F}^{(111)'}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3^2} \otimes \operatorname{vec} \boldsymbol{\Sigma}^{-1}\right) - 2\boldsymbol{F}^{(11)'}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \left(\boldsymbol{I}_{\nu_3} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right) \boldsymbol{N}_{\nu_3}$$
$$+ 2\boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}} \left(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec} \boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\right) \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right)$$
$$- \boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}.$$

Preliminary derivatives for q.

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{K}_{xx}\boldsymbol{W}_2\boldsymbol{K}'_{zx}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{K}_{xx}\boldsymbol{W}_2\boldsymbol{K}'_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{xx}}{\partial \boldsymbol{\varphi}'} = -\boldsymbol{K}_{xx}\boldsymbol{W}_2\boldsymbol{K}'_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\beta}'} = -\boldsymbol{I}_{(p,d)} \left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_d\right) \boldsymbol{M}_{xx}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{I}_{(p,d)} \left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_d\right) \boldsymbol{V}_{xz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} = -\boldsymbol{I}_{(p,d)} \left(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_d\right) \boldsymbol{V}_{xz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zz}}{\partial \boldsymbol{\beta}'} = -2\boldsymbol{N}_p \left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right) \boldsymbol{M}_{zx}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zz}}{\partial \boldsymbol{\mu}'} = -\big(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\big)\boldsymbol{V}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{W}_{zz}}{\partial \boldsymbol{\varphi}'} = -\big(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\big)\boldsymbol{V}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}$$

$$\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\beta}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{xzx}\right) + 2\boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zxzx}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{xx}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(d,p)} \otimes \boldsymbol{I}_d\right) \Big]$$

$$\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\mu}'} = -\Big[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzx}\right) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzx}$$
$$+ \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zx}\big]' \left(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_d\right) \Big] \left(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{\nu_1}\right)$$

$$\frac{\partial \boldsymbol{V}_{xx}}{\partial \boldsymbol{\varphi}'} = -\Big[\big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzx}\big) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzx} + \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zx}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_d\big)\Big]\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\nu_1}\big)$$

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\beta}'} = -\Big[\big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{xzz}\big) + 2\boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zxzz} + \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{(p,d)}\boldsymbol{K}_{zx}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(d,p)} \otimes \boldsymbol{I}_p\big)\left(\boldsymbol{I}_{\nu_1} \otimes 2\boldsymbol{N}_p\right)\Big]$$

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\mu}'} = -\Big[\big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\big) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz} + \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\big)\left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big]\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\big)$$

$$\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'} = -\Big[\big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\big) + \boldsymbol{K}_{zx}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz} + \boldsymbol{K}_{zx}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\big)\left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big]\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\big)$$

$$\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\beta}'} = -\Big[2\boldsymbol{N}_p\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_2\boldsymbol{K}'_{xzz}\right) + 2\boldsymbol{K}_{zz}\boldsymbol{W}_3\boldsymbol{K}'_{zxzz} + \boldsymbol{K}_{zz}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{(p,d)}\boldsymbol{K}_{zx}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(d,p)} \otimes \boldsymbol{I}_p\big)\left(\boldsymbol{I}_{\nu_1} \otimes 2\boldsymbol{N}_p\right)\Big]$$

$$\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\mu}'} = -\Big[2\boldsymbol{N}_p\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zz}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz} + \boldsymbol{K}_{zz}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\big)\left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big]\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\big)$$

$$\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\varphi}'} = -\Big[2\boldsymbol{N}_p\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{Z}'\boldsymbol{W}_2\boldsymbol{K}'_{zzz}\right) + \boldsymbol{K}_{zz}\boldsymbol{W}_3\boldsymbol{K}'_{zzzz} + \boldsymbol{K}_{zz}\boldsymbol{W}_2\big[\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{K}_{zz}\big]'\big(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p\big)\left(\boldsymbol{I}_{p^2} \otimes 2\boldsymbol{N}_p\right)\Big]\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{\boldsymbol{p}^2}\big)$$

$$\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\mu}'} = -\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{p^2}\big) - \boldsymbol{I}_{(p,p)}\big(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(2\boldsymbol{N}_p\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}} \otimes \boldsymbol{I}_{(p,p)}\big) + \frac{1}{2}\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\mu}'}$$

$$\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\varphi}'} = -\big(\boldsymbol{\Sigma}^{-1} \otimes (\operatorname{vec}\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{p^2}\big) - \boldsymbol{I}_{(p,p)}\big(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(2\boldsymbol{N}_p\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{I}_{(p,p)}\big) + \frac{1}{2}\frac{\partial \boldsymbol{V}_{zz}}{\partial \boldsymbol{\varphi}'}$$

Third derivatives of q.

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}} = 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d\otimes(\operatorname{vec}\boldsymbol{I}_d)'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{xx}}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{I}_{(p,d)}\right)+4\frac{\partial\boldsymbol{V}_{xx}}{\partial\boldsymbol{\beta}'}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}} = -\boldsymbol{F}^{(111)'}_{\boldsymbol{\mu\mu\mu}}\Big[\boldsymbol{I}_{\nu_2^2}\otimes\operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\Big]+4\boldsymbol{F}^{(11)'}_{\boldsymbol{\mu\mu}}\big(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)\boldsymbol{N}_{\nu_2}$$
$$+2\boldsymbol{F}^{(1)'}_{\boldsymbol{\mu}}\frac{\partial\boldsymbol{M}_{zz}}{\partial\boldsymbol{\mu}'}\big(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)+2\boldsymbol{F}^{(1)'}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(2)}_{\boldsymbol{\mu\mu}}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}} = -\boldsymbol{F}^{(111)'}_{\boldsymbol{\varphi\varphi\varphi}}\Big[\boldsymbol{I}_{\nu_3^2}\otimes\operatorname{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\Big]+4\boldsymbol{F}^{(11)'}_{\boldsymbol{\varphi\varphi}}\big(\boldsymbol{I}_{\nu_3}\otimes\boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big)\boldsymbol{N}_{\nu_3}$$
$$+2\boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}}\frac{\partial\boldsymbol{M}_{zz}}{\partial\boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big)+2\boldsymbol{F}^{(1)'}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi\varphi}}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}} = 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{zx}}{\partial\boldsymbol{\beta}'}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$
$$+2\frac{\partial\boldsymbol{V}_{xz}}{\partial\boldsymbol{\beta}'}\big(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{I}_{(\nu_2,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}}\right),\nu_2,\nu_1^2\right]$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}} = \frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}}\boldsymbol{I}_{(\nu_2,\nu_1)}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}} = -2\big(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{W}_{xx}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\otimes\boldsymbol{I}_{\nu_1}\big)$$
$$+2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d\otimes(\operatorname{vec}\boldsymbol{I}_d)'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{xx}}{\partial\boldsymbol{\varphi}'}\otimes\boldsymbol{I}_{(p,d)}\right)+4\frac{\partial\boldsymbol{V}_{xx}}{\partial\boldsymbol{\varphi}'}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}} = \frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}}\boldsymbol{I}_{(\nu_1,\nu_2)}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{I}_{(\nu_3,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\right),\nu_3,\nu_1^2\right]$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}} = -4\big(\boldsymbol{I}_p\otimes\boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1}\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)$$
$$+2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d\otimes(\operatorname{vec}\boldsymbol{\Sigma}^{-1})'\otimes\boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial\operatorname{vec}\boldsymbol{W}_{zx}}{\partial\boldsymbol{\mu}'}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$
$$+2\frac{\partial\boldsymbol{V}_{xz}}{\partial\boldsymbol{\mu}'}\big(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)+2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\mu\mu}}$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}} = \operatorname{dvec}\left[\big(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{I}_{(\nu_2,\nu_1)}\big)\operatorname{vec}\left(\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}}\right),\nu_2,\nu_1\nu_2\right]$$

$$\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}} = \frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}}\boldsymbol{I}_{(\nu_1,\nu_2)}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = -4\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1} \otimes (\mathrm{vec}\,\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big)$$
$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\mathrm{vec}\,\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\right)$$
$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big) + 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = \mathrm{dvec}\left[\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{I}_{(\nu_3,\nu_1)}\big)\,\mathrm{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\right), \nu_3, \nu_1\nu_3\right]$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{I}_{(\nu_1,\nu_3)}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = -\boldsymbol{F}^{(111)\prime}_{\boldsymbol{\varphi}\boldsymbol{\varphi}\boldsymbol{\mu}}\big[\boldsymbol{I}_{\nu_3^2} \otimes \mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big] + 4\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\mu}\boldsymbol{\varphi}}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big)\boldsymbol{N}_{\nu_3}$$
$$+ 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}}\big) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\mu}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = \mathrm{dvec}\left[\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{I}_{(\nu_3,\nu_2)}\big)\,\mathrm{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\right), \nu_3, \nu_2\nu_3\right]$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{I}_{(\nu_2,\nu_3)}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = -\boldsymbol{F}^{(111)\prime}_{\boldsymbol{\mu}\boldsymbol{\mu}\boldsymbol{\varphi}}\big[\boldsymbol{I}_{\nu_2^2} \otimes \mathrm{vec}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_{zz}\boldsymbol{\Sigma}^{-1}\big)\big] + 4\boldsymbol{F}^{(11)\prime}_{\boldsymbol{\varphi}\boldsymbol{\mu}}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{M}_{zz}\boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)\boldsymbol{N}_{\nu_2}$$
$$+ 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\frac{\partial \boldsymbol{M}_{zz}}{\partial \boldsymbol{\mu}'}\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big) + 2\boldsymbol{F}^{(1)\prime}_{\boldsymbol{\varphi}}\boldsymbol{M}_{zz}\boldsymbol{F}^{(2)}_{\boldsymbol{\mu}\boldsymbol{\mu}}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = \mathrm{dvec}\left[\big(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{I}_{(\nu_2,\nu_3)}\big)\,\mathrm{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}}\right), \nu_2, \nu_2\nu_3\right]$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{I}_{(\nu_3,\nu_2)}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}} = -4\big(\boldsymbol{I}_p \otimes \boldsymbol{W}_{xz}\big)\boldsymbol{N}_p\big(\boldsymbol{\Sigma}^{-1} \otimes (\mathrm{vec}\,\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\big(\boldsymbol{F}^{(1)}_{\boldsymbol{\varphi}} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big)$$
$$+ 2\boldsymbol{I}_{(d,p)}\big(\boldsymbol{I}_d \otimes (\mathrm{vec}\,\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\big)\left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{W}_{zx}}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\right)$$
$$+ 2\frac{\partial \boldsymbol{V}_{xz}}{\partial \boldsymbol{\varphi}'}\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{F}^{(1)}_{\boldsymbol{\mu}}\big) + 2\boldsymbol{M}_{xz}\boldsymbol{F}^{(2)}_{\boldsymbol{\varphi}\boldsymbol{\mu}}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{I}_{(\nu_2,\nu_3)}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \mathrm{dvec}\left[\big(\boldsymbol{I}_{\nu_3} \otimes \boldsymbol{I}_{(\nu_2,\nu_1)}\big)\,\mathrm{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}\right), \nu_2, \nu_1\nu_3\right]$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{I}_{(\nu_1,\nu_3)}$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} = \operatorname{dvec}\left[\left(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{I}_{(\nu_3,\nu_1)}\right) \operatorname{vec}\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\right), \nu_3, \nu_1\nu_2\right]$$

$$\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} = \frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{I}_{(\nu_1,\nu_2)}$$

Implicit Derivatives

In this section, expressions for the derivatives are given assuming that η_ψ is a function of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$. In this case, w and q are

$$w = w(\boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})\}),\; q = q(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi}\{\boldsymbol{\psi}, \eta_\psi(\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi})\}).$$

First derivatives.

$$\boldsymbol{P}^{(1)}_{\boldsymbol{\beta}} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\beta}'} = -\boldsymbol{v}_2(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}\frac{\partial q^*}{\partial \boldsymbol{\beta}'} = -\boldsymbol{r}\frac{\partial q^*}{\partial \boldsymbol{\beta}'}$$

$$\boldsymbol{P}^{(1)}_{\boldsymbol{\mu}} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\mu}'} = -\boldsymbol{v}_2(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}\frac{\partial q^*}{\partial \boldsymbol{\mu}'} = -\boldsymbol{r}\frac{\partial q^*}{\partial \boldsymbol{\mu}'}$$

$$\boldsymbol{P}^{(1)}_{\boldsymbol{\psi}} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\psi}'} = \left[(\boldsymbol{a}'\boldsymbol{v}_2)^{-1}\boldsymbol{a}' \otimes \boldsymbol{I}_{\nu_3}\right](2\boldsymbol{N}^{\perp}_{\nu_3})(\boldsymbol{v}_2 \otimes \boldsymbol{V}_1)$$

Second derivatives.

$$\boldsymbol{P}^{(11)\prime}_{\boldsymbol{\beta\beta}} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{P}^{(1)}_{\boldsymbol{\beta}} + \boldsymbol{P}^{(1)\prime}_{\boldsymbol{\beta}}\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} + \boldsymbol{P}^{(1)\prime}_{\boldsymbol{\beta}}\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}^{(1)}_{\boldsymbol{\beta}}\right](\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{r}')$$

$$\boldsymbol{P}^{(11)\prime}_{\boldsymbol{\mu\mu}} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{P}^{(1)}_{\boldsymbol{\mu}} + \boldsymbol{P}^{(1)\prime}_{\boldsymbol{\mu}}\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} + \boldsymbol{P}^{(1)\prime}_{\boldsymbol{\mu}}\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}^{(1)}_{\boldsymbol{\mu}}\right](\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{r}')$$

$$\boldsymbol{P}^{(11)\prime}_{\boldsymbol{\psi\psi}} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}} = -\boldsymbol{P}^{(1)\prime}_{\boldsymbol{\psi}}\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}^{(1)}_{\boldsymbol{\psi}}(\boldsymbol{I}_{\nu_3-1} \otimes \boldsymbol{r}')$$

$$\boldsymbol{P}^{(11)\prime}_{\boldsymbol{\beta\mu}} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} + \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{P}^{(1)}_{\boldsymbol{\beta}} + \boldsymbol{P}^{(1)\prime}_{\boldsymbol{\mu}}\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} + \boldsymbol{P}^{(1)\prime}_{\boldsymbol{\mu}}\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\boldsymbol{P}^{(1)}_{\boldsymbol{\beta}}\right](\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{r}')$$

$$\boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\mu}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\mu}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right] \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)} \left(\boldsymbol{I}_{\nu_3 - 1} \otimes \boldsymbol{r}'\right)$$

$$\boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\beta}}^{(11)'} = \frac{\partial^2 \boldsymbol{\varphi}'}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\beta}} = -\left[\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}} + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'} \frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right] \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)} \left(\boldsymbol{I}_{\nu_3 - 1} \otimes \boldsymbol{r}'\right)$$

Derivatives of $\boldsymbol{w} = n \ln |\boldsymbol{\Sigma}|$

The derivatives in this section are given assuming that η_ψ is a function of the other parameters.

First derivatives.

$$\frac{\partial w}{\partial \boldsymbol{\beta}} = \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'} \frac{\partial w^*}{\partial \boldsymbol{\varphi}}$$

$$\frac{\partial w}{\partial \boldsymbol{\mu}} = \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial w^*}{\partial \boldsymbol{\varphi}}$$

$$\frac{\partial w}{\partial \boldsymbol{\psi}} = \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)'} \frac{\partial w^*}{\partial \boldsymbol{\varphi}}$$

Second derivatives.

$$\frac{\partial^2 w}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}} = \boldsymbol{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(11)'} \left(\boldsymbol{I}_{\nu_1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'} \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)}$$

$$\frac{\partial^2 w}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}} = \boldsymbol{P}_{\boldsymbol{\mu}\boldsymbol{\mu}}^{(11)'} \left(\boldsymbol{I}_{\nu_2} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)}$$

$$\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\psi}} = \boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{(11)'} \left(\boldsymbol{I}_{\nu_3 - 1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)'} \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)}$$

$$\frac{\partial^2 w}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}} = \boldsymbol{P}_{\boldsymbol{\beta}\boldsymbol{\mu}}^{(11)'} \left(\boldsymbol{I}_{\nu_1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)}$$

$$\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\mu}} = \boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\mu}}^{(11)'} \left(\boldsymbol{I}_{\nu_3 - 1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\mu}}^{(1)'} \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)}$$

$$\frac{\partial^2 w}{\partial \boldsymbol{\psi}' \otimes \partial \boldsymbol{\beta}} = \boldsymbol{P}_{\boldsymbol{\psi}\boldsymbol{\beta}}^{(11)'} \left(\boldsymbol{I}_{\nu_3 - 1} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \boldsymbol{P}_{\boldsymbol{\beta}}^{(1)'} \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \boldsymbol{P}_{\boldsymbol{\psi}}^{(1)}$$

Estimating Function and derivatives

In this section, the definition of the estimating function for the asymptotic distribution for the S-estimators is given, as well as the first and section derivatives of the estimating function. The first derivatives are like the second derivatives of the function w. The second derivatives are like the third derivatives of the function w.

Estimating Function. The definition of the estimating function is like the first derivatives of w.

$$\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s} = \begin{bmatrix} \boldsymbol{l}^{(1)}_{\beta} \\ \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}} \\ \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}} \\ \boldsymbol{l}^{(1)}_{\eta_\psi} \end{bmatrix} = \begin{bmatrix} -n\frac{\partial q^*}{\partial \boldsymbol{\beta}} \\ -n\frac{\partial q^*}{\partial \boldsymbol{\mu}} \\ (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1')\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\boldsymbol{a} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) \\ nq^* \end{bmatrix}.$$

Second Derivatives (first derivatives of $\boldsymbol{l}^{(1)}$).

The first derivatives of the estimating function (like second derivatives of w) are

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\mu}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\psi}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{V}_1$$

$$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\beta}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\beta}}}{\partial \eta_\psi} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\boldsymbol{v}_2$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\mu}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\mu}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi\mu}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\psi}'} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{V}_1$$

$$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\mu}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\mu}}}{\partial \eta_\psi} = -n\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}}\boldsymbol{v}_2$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta\psi}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \boldsymbol{\beta}'} = \left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right)$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu\psi}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \boldsymbol{\mu}'} = \left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right)$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi\psi}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \boldsymbol{\psi}'} = \left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right]\boldsymbol{V}_1$$

$$\boldsymbol{l}^{(2)}_{\eta_\psi\boldsymbol{\psi}} = \frac{\partial \boldsymbol{l}^{(1)}_{\boldsymbol{\psi}}}{\partial \eta_\psi} = \left(\boldsymbol{v}_2' \otimes \boldsymbol{V}_1'\right)\left(2\boldsymbol{N}^{\perp}_{\nu_3}\right)\left[\left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}}\right) + \left(\boldsymbol{a} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}}\right)\right]\boldsymbol{v}_2$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\beta}\eta_\psi} = \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \boldsymbol{\beta}'} = n\frac{\partial q^*}{\partial \boldsymbol{\beta}'}$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\mu}\eta_\psi} = \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \boldsymbol{\mu}'} = n\frac{\partial q^*}{\partial \boldsymbol{\mu}'}$$

$$\boldsymbol{l}^{(2)}_{\boldsymbol{\psi}\eta_\psi} = \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \boldsymbol{\psi}'} = n\boldsymbol{a}'\boldsymbol{V}_1$$

$$\boldsymbol{l}^{(2)}_{\eta_\psi\eta_\psi} = \frac{\partial \boldsymbol{l}^{(1)}_{\eta_\psi}}{\partial \eta_\psi} = n\boldsymbol{a}'\boldsymbol{v}_2$$

Third Derivatives (second derivatives of $\boldsymbol{l}^{(1)}$). The second derivatives of the estimating function (like third derivatives of w) are

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\beta\beta}} = -n\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu\beta}} = -n\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu\beta}}\boldsymbol{I}_{(\nu_2,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\beta}} = -n\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\left(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\beta\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\beta}}\boldsymbol{I}_{(\nu_3-1,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\beta}}\left(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\mu\beta}} = -n\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\mu\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\beta}}\boldsymbol{I}_{(\nu_3-1,\nu_2)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\psi\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{V}_1\otimes\boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{V}_1\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi\beta}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\boldsymbol{\beta}}$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\boldsymbol{\beta}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\beta}}\left(\boldsymbol{v}_2\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\beta\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\mu\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu\beta\mu}}\boldsymbol{I}_{(\nu_1,\nu_2)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\beta\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta\psi\mu}}\boldsymbol{I}_{(\nu_3-1,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\mu\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi\mu\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu\psi\mu}}\boldsymbol{I}_{(\nu_3-1,\nu_2)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\boldsymbol{\mu}} = -n\frac{\partial^3 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \boldsymbol{\mu}\boldsymbol{\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi \boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\mu}} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} (\boldsymbol{V}_1 \otimes \boldsymbol{V}_1)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi \boldsymbol{\mu}} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} (\boldsymbol{V}_1 \otimes \boldsymbol{v}_2)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \boldsymbol{\psi}\boldsymbol{\mu}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi \boldsymbol{\mu}}$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \eta_\psi \boldsymbol{\mu}} = -n \frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}} (\boldsymbol{v}_2 \otimes \boldsymbol{v}_2)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\beta}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left(\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left(\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\beta}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\boldsymbol{\psi}} \boldsymbol{I}_{(\nu_2,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{V}_1)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\beta}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\boldsymbol{\psi}} \boldsymbol{I}_{(\nu_3-1,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi \boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{v}_2)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \boldsymbol{\beta}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi \boldsymbol{\psi}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\mu}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}^{\perp}_{\nu_3}\right) \left[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \right] (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{V}_1)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\mu}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\boldsymbol{\psi}} \boldsymbol{I}_{(\nu_3-1,\nu_2)}$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \boldsymbol{\mu}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}_{\nu_3}^{\perp}\right) \Bigg[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \Bigg] (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{v}_2)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi \boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\eta_\psi \boldsymbol{\mu}\boldsymbol{\psi}}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\psi}\boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}_{\nu_3}^{\perp}\right) \Bigg[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) (2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \Bigg] (\boldsymbol{V}_1 \otimes \boldsymbol{V}_1)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi \boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}_{\nu_3}^{\perp}\right) \Bigg[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) (2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \Bigg] (\boldsymbol{V}_1 \otimes \boldsymbol{v}_2)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \boldsymbol{\psi}\boldsymbol{\psi}} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi \boldsymbol{\psi}}$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi \eta_\psi \boldsymbol{\psi}} = (\boldsymbol{v}_2' \otimes \boldsymbol{V}_1') \left(2\boldsymbol{N}_{\nu_3}^{\perp}\right) \Bigg[\left(\frac{\partial^3 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial w^*}{\partial \boldsymbol{\varphi}} \right) + \left(\frac{\partial^2 q^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \otimes \frac{\partial^2 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) (2\boldsymbol{N}_{\nu_3}) + \left(\boldsymbol{a} \otimes \frac{\partial^3 w^*}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}} \right) \Bigg] (\boldsymbol{v}_2 \otimes \boldsymbol{v}_2)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\beta}\eta_\psi} = n \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\beta}'}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\eta_\psi} = n \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\mu}'}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\beta}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\mu}\eta_\psi} \boldsymbol{I}_{(\nu_2,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\eta_\psi} = n \frac{\partial^2 q^*}{\partial \boldsymbol{\beta}' \otimes \partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{V}_1)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\beta}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\boldsymbol{\psi}\eta_\psi} \boldsymbol{I}_{(\nu_3-1,\nu_1)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\beta}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\beta}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\beta}\eta_\psi\eta_\psi}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\mu}\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\mu}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\boldsymbol{\psi}\eta_\psi}\boldsymbol{I}_{(\nu_3-1,\nu_2)}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\mu}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\mu}\eta_\psi\eta_\psi}$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\boldsymbol{\psi}\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{V}_1\otimes\boldsymbol{V}_1\right)$$

$$\boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{V}_1\otimes\boldsymbol{v}_2\right)$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\boldsymbol{\psi}\eta_\psi} = \boldsymbol{l}^{(3)}_{\boldsymbol{\psi}\eta_\psi\eta_\psi}$$

$$\boldsymbol{l}^{(3)}_{\eta_\psi\eta_\psi\eta_\psi} = n\frac{\partial^2 q^*}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\left(\boldsymbol{v}_2\otimes\boldsymbol{v}_2\right)$$

APPENDIX D

Summary of Matlab Programs

Common Programs

1. `commute.m`: Computes the commutation matrix, $\boldsymbol{I}_{(m,n)}$.

2. `constantmatrices.m`: Computes the matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{L}_p$, $\boldsymbol{D}_p$, and $\boldsymbol{D}_{\boldsymbol{G}}^{(1)}$.

3. `dup.m`: Computes the duplication matrix, $\boldsymbol{D}_p$.

4. `dvec.m`: Given an matrix $\boldsymbol{A}$ with ab elements, $\text{dvec}(\boldsymbol{A}, a, b)$ is the $a \times b$ matrix such that $\text{vec}(\boldsymbol{A}) = \text{vec}\left\{\text{dvec}(\boldsymbol{A}, a, b)\right\}$.

5. `khatrirao.m`: Computes the Khatri-Rao column-wise product.

6. `kron3.m`: Computes the Kronecker product of three or more matrices.

7. `Lp.m`: Computes the matrix $\boldsymbol{L}_p$.

8. `ppo.m`: Computes the perpendicular projection operator of the input matrix.

9. `solve_eta.m`: Given the parameter $\boldsymbol{\mu}$, solves for the implicit parameter $\boldsymbol{\eta}$ under the spectral model.

10. `spdiag.m`: Either converts a vector into a sparse diagonal matrix, or extracts the diagonal elements of a matrix.

11. `vec.m`: Stacks the columns of a matrix into a column vector.

M-Estimators

1. `inform.m`: Computes the information matrix of the multivariate-T distribution under the linear model and spectral model.

2. `lntlikelihood.m`: Computes the log likelihood function of the multivariate-T distribution under the linear model and spectral model.

3. `mestim.m`: Computes the M-estimators of location and scatter under the linear model and spectral model using the multivariate-T likelihood

4. `scoref.m`: Computes the Score function for the multivariate-T distribution under the linear model and spectral model.

5. `solve_mles.m`: Solves for the mles of the multivariate-T distribution under the linear model and spectral model.

S-Estimators

1. `compute_a.m`: Computes the vector $\boldsymbol{a}$.

2. `deriv1_w.m`: Computes the first derivative of the function $w = n \log |\boldsymbol{\Sigma}|$.

3. `deriv_w.m`: Computes the first derivative of w, as well as $\boldsymbol{l}^{(1)}_{\boldsymbol{\theta}_s}$, $\boldsymbol{l}^{(2)}_{\boldsymbol{\theta}_s}$, $\boldsymbol{l}^{(3)}_{\boldsymbol{\theta}_s}$, the sandwich estimator, and an estimator of the bias.

4. `exprho.m`: Computes the expected value of $\rho(d_k)$ when the distribution of the data is the multivariate-T distribution.

5. `generate_data.m`: Generates a random set of data from a multivariate-T distribution.

6. `initial_guess.m`: Computes an initial guess for $\boldsymbol{V}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{\psi}$.

7. `kappa.m`: Computes the constant κ.

8. `make_E.m`: Constructs the elementary matrix, $\boldsymbol{E}_{k,v}$, which is used to put $\boldsymbol{K}_3$ together.

9. `make_F11_mumu.m`: Computes $\boldsymbol{F}^{(11)}_{\boldsymbol{\mu\mu}}$.

10. `make_F11_phiphi.m`: Computes $\boldsymbol{F}^{(11)}_{\boldsymbol{\varphi\varphi}}$.

11. `model_lam.m`: Given $\boldsymbol{\Gamma}$ and $\boldsymbol{\varphi}$, computes $\boldsymbol{D}^{(1)}_{\boldsymbol{\lambda}}$, $\boldsymbol{D}^{(2)}_{\boldsymbol{\lambda}}$, $\boldsymbol{D}^{(3)}_{\boldsymbol{\lambda}}$, $\boldsymbol{\Sigma}^{-1}$, and $\boldsymbol{\Lambda}^{-1}$.

12. `newguess.m`: Generates a subsample of $\boldsymbol{Y}$.

13. `qeval.m`: Evaluates the constraint q and returns a vector containing d_k.

14. `rho.m`: Compute the ρ function. For this program, ρ is the integral of Tukey's bi-weight function.

15. `robust_load.m`: Loads constant matrices and true values of the parameters used when performing simulations.

16. `sample_wor.m`: Generate a sample without replacement.

17. `sandwich_sim.m`: Compute the average sandwich estimator and the average bias from the simulation.

18. `satisfy_constraint.m`: Given a value of $\boldsymbol{\psi}$, solves for η_ψ and updates various quantities so they satisfy the constraint q.

19. `simulation.m`: Performs a simulation.

20. `Smultivar.m`: Computes the S-estimators of multivariate location and scatter using the algorithm of Ruppert. Program received from Christophe Croux.

21. `solve_LS.m`: Computes a "local" solution to the S-estimating equations using the Least Squares estimate as the initial guess.

22. `solve_c0.m`: Solves for the constants b_0 and c_0.

23. `solve_dq_betamu.m`: Solves $\frac{\partial q^*}{\partial \boldsymbol{\mu}} = \mathbf{0}$ and $\frac{\partial q^*}{\partial \boldsymbol{\beta}} = \mathbf{0}$. This is the first stage of the two-stage modified Newton-Raphson algorithm.

24. `solve_dw_psi.m`: Solves $\frac{\partial w}{\partial \boldsymbol{\psi}} = \mathbf{0}$. This is the second stage of the two-stage modified Newton-Raphson algorithm.

25. `solve_eta_psi.m`: Given values for $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\psi}$, solves for the implicit parameter η_ψ. That is, solves $q(\eta_\psi; \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\psi}) = 0$ using a hybrid Newton-Raphson-Bisection algorithm.

26. `solve_newguess.m`: Computes a "local" solution to the S-estimating equations using a random subsample of $\boldsymbol{Y}$ as a guess. Command is preceded by `newguess.m`.

27. `solve_s_estimate_sim.m`: Solve for the S-estimator using 80 random starting points. Used in the simulation program.

28. `solve_sest.m`: The two-stage modified Newton-Raphson algorithm for computing a "local" solution to the S-estimating equations.

29. `solve_sestimator.m`: Solves for the S-estimators of multivariate location and scatter using the new algorithms.

30. `solve_surreal.m`: Computes a "local" solution to the S-estimating equations using the SURREAL estimate as the initial guess.

APPENDIX E

Matlab Programming Code

Common Code

commute.m.

```
function Imn = commute(m,n)
% COMMUTE   Commutation matrix
%
%   Imn = commute(m,n) computes the commutation matrix which switches
%   vec(A') with vec(A).  In particular, if A:m x n, then vec(A) =
%   I(m,n)vec(A').  Written by S. K. Hyde
%
%   Alternate notation is K(n,m) (from Magnus and Neudecker 1979).
%   K(n,m) = I(m,n) = commute(m,n).
%
% These times are for a SunFire 880, 8 processor, 900 Mhz
% UltraSparcIII w/8MB Cache/cpu, and 16384 MB of main memory (Using
% one processor):
%   m    n   for loop   kron def  this one
% ---- ----  --------   --------  --------
%   10   10    0.0100    0.0900    0.0200
%   20   20    0.0100    0.1400    0.0100
%   50   50    0.0600    0.4200         0
%  100  100    0.6300    1.1200    0.0100
%  200  200   10.0500    4.5600    0.0600
%   10 1000    0.6300    0.1300    0.0100
% 1000   10    0.6000    0.1200    0.0200
%  200  300   22.4700    6.2600    0.0900
%  200 1000  249.0200   19.3500    0.3100

mn = m*n;
i = 1:n;
j = 1:m;

% Specify the location of the ones in the commutation matrix.
%  There is one "1" per row.
%  The row locations are from 1:mn
%  The column location can be specified by:
%    c = kron(ones(1,n),j*n) + kron(i-n,ones(1,m));
%  However, the one below will be cheaper.
r = 1:mn;
c = repmat(j*n,1,n) + reshape(repmat(i-n,m,1),1,mn);
Imn = sparse(r,c,1,mn,mn);
```

---

constantmatrices.m.

```
function  [Dg,A1,A2,L,Dp] = constantmatrices(m);
% CONSTANTMATRICES    Construct constant matrices
%
%   Construct Dg, A1, A2, L, and Dp for a given multiplicity vector m.
%      Routines written by S. K. Hyde
%

p = sum(m);

% Construct A1
%   find location of zeros in G.
%   index1 contains the locations of the zeros in G.
upper = (p^2-m'*m)/2;
cm = [0; cumsum(m) ];
index1 = [];
for i = 1:length(m)
  index0 = find(tril(ones(p,m(i)))==0);
  index0 = (p+1)*cm(i) + index0;
  index1 = [ index1; index0 ];
end
temp = ones(p,p);
temp(index1) = 0;           %corresponds to zeros in G.
row1 = find(triu(temp,1)); %extract row locations for ones in A1
col1 = 1:upper;            %column locations for ones in A1
A1 = sparse(row1,col1,1,p^2,upper);

%
%  Construct A2
%
if p == 1,
  A2 = sparse([1]);
else
  rank = p*(p+1)/2;
  k = 2:p;
  index(1) = rank;
  index(k) = -k;
  index = cumsum(index);
  row2 = ones(1,rank);
  row2(index) = p:-1:1;
  row2 = cumsum(row2);   %Row locations of the ones
  col2 = 1:rank;         %Column locations of the ones.
  A2 = sparse(row2,col2,1,p^2,rank);
end;

%
% Use other routines for the rest
```

```
%
L = Lp(p);
Dp = dup(p);
Dg = A1 - A2*Dp'*A1;
```

---

dup.m.

```
function D = dup(p);
% DUP    Duplication matrix
%
% D = dup(p) generates the Duplication matrix of order p.  Very fast
% for large p when compared to a double loop written from the
% definition.  Written by S. K. Hyde
%
%time =
% These times are for a SunFire 880, 8 processor, 900 Mhz
% UltraSparcIII w/8MB Cache/cpu, and 16384 MB of main memory (Using
% one processor):
%    p  dup(old)  dup_old_mine    dup
% ---- --------  ------------  -------
%    2     0.01       0.0200    0.0300
%    4     0.01            0    0.0100
%    8     0.01            0         0
%   16     0.04            0         0
%   32     0.17       0.0100         0
%   64     0.96       0.0200    0.0100
%  128     6.73       0.1100    0.0200
%  256    69.93       0.5000    0.1000
%  512  1089.11       4.0200    0.4400
% 1024   ~           46.6600    1.9600
%

rank = p*(p+1)/2;

if p == 1,
  D = sparse(1);
  return;
end

% These next few lines perform the equivalent of
%
%index(1) = rank;
%for k = 2:p
%  index(k) = index(k-1) - k;
```

```
%end
%
k = 2:p;
index(1) = rank;
index(k) = -k;
index = cumsum(index);

j = 2:(p-1);
index2(1) = p^2-rank;
index2(j) = -j;
index2 = cumsum(index2);

% Get row indicies
row1 = ones(1,rank);
row1(index) = p:-1:1;
row1 = cumsum(row1);

row2 = repmat(p,1,p^2-rank);
row2(index2) = [ 1-p*(0:1:length(index2)-2) p+1];
row2 = cumsum(row2);

% Get column indicies
col1 = 1:rank;
col2 = col1;
col2(index) = [];

% Collect indicies and place 1's where they should go.
first = [row1 row2];
second = [col1 col2];
D = sparse(first,second,1,p^2,rank);
```

---

dvec.m.

```
function Y = dvec(y,n,d);
% DVEC  De-vec command
%
%   Y = dvec(y,n,d) performs the inverse of the vec function and
%   reshapes the result as the n x d matrix Y.
%
Y = reshape(y,n,d);
```

---

khatrirao.m.

```
function C = khatrirao(A,B);
% KHATRIRAO   Khatri-Rao column-wise product
%
%   C = khatrirao(A,B) computes the Khatri-Rao column-wise product of
%   the matrix A and B.  The Khatri-Rao product is the column-wise
%   kronecker product of each column of A with each column of B.  For
%   example,
%
%   if   A = [ a1 a2 a3 ...  an ],
%        B = [ b1 b2 b3 ...  bn ], then
%
%        C = [ a1 x b1    a2 x b2   a3 x b3  ...  an x bn ],
%   where x is the kronecker product.
%
%   Note that C has the same number of columns as A and B
%
%   When the number of nonzero elements is less than (density) percent
%   of the number of total elements in both matrices, uses the
%   definition of the product.  (A * B) = (A x B)*L.
%
%   For not very sparse matrices, the normal routine is much faster.
%   Written by S. K. Hyde, using the matlab kron function as a
%   template.
%

[ma,na] = size(A);
[mb,nb] = size(B);

if (na ~= nb)
  error('both matrices must have the same number of columns');
  return;
end;

density = .04;
densA = nnz(A)/prod(size(A));
densB = nnz(A)/prod(size(B));
dens = max(densA,densB);

if dens > density,
  [ia,ib] = meshgrid(1:ma,1:mb);
  C = A(ia,1:end).*B(ib,1:end);
else
  C = kron(sparse(A),sparse(B))*Lp(na);
end
```

---

kron3.m.

```
function A = kron3(varargin);
% KRON3    Kroenecker product of 2 or more matrices
%
%   KRON(A1,A2,...,An) is the Kronecker tensor product of A1, A2,
%   ..., An.
%
if nargin == 1,
  error('Too few arguments');
else
  A = varargin{1};
  for i = 2:length(varargin);
    A = kron(A,varargin{i});
  end
end
```

Lp.m.

```
function L = Lp(n);
% LP     Computes the Khatri-Rao product of Ip and Ip.
%
%   L = Lp(n) computes the Khatri-Rao product of Ip and Ip.  Much
%   faster than khatrirao(Ip,Ip) since Ip is sparse.  Written by
%   S. K. Hyde
%

% Specify the location of the ones in the matrix.
%  The column and row location can be specified by:
col = 1:n;
row = (n+1)*col - n;
L = sparse(row,col,1,n^2,n);
```

---

ppo.m.

```
function P = ppo(X);
```

```
% PPO   Perpendicular projection operator
%
%   P = ppo(X) finds the ppo of X.
%
P = X*pinv(X'*X)*X';
```

---

solve_eta.m.

```
function eta = solve_eta(A1,A2,mu);
% SOLVE_ETA    Solve for the implicit parameters in G
%
%   Given a solution for mu, eta = solve_eta(A1,A2,mu) solves for the
%   implicit parameters eta.  Written by R. J. Boik.
%

[p2,nu1] = size(A1);
[p2,s] = size(A2);
p = round(sqrt(p2));
Ip = speye(p);
check = reshape(A2*ones(s,1),p,p);
Amu = reshape(A1*mu,p,p);
for k = 1:p
  m = Amu(1:p-k,p-k+1);
  if k == 1;
    e = sqrt(1-m'*m);
    check(p,p) = e;
  else
    T = check(:,p-k+2:p) + Amu(:,p-k+2:p);
    T1 = T(1:p-k,:)';
    T2 = T(p-k+1:p,:)';
    [U,D,V] = svd(T2);
%
%   T2 is (k-1) x k
%
    d = diag(D);
    T2p = V(:,1:k-1)*diag(1./d(1:k-1))*U(:,1:k-1)';
    u = V(:,k);
    w0 = -T2p*T1*m;
    a2 = u'*u;
    a1 = 2*u'*w0;
    a0 = w0'*w0 - 1+m'*m;
    d = a1^2 - 4*a0*a2;
    w1 = w0 + u*(-a1+sqrt(d))/(2*a2);
    w2 = w0 + u*(-a1-sqrt(d))/(2*a2);
```

```
    w = w1;
    if 1-w1(1) > 1-w2(1)
      w = w2;
    end
    check(p-k+1:p,p-k+1) = w;
  end
end
eta = A2'*vec(check);
```

---

spdiag.m.

```
function D = spdiag(d);
% SPDIAG   sparse diag command
%
%   D = spdiag(d) constructs a sparse diagonal matrix of a vector
%   input without constructing the zero off diagonal elements.  For
%   a matrix argument, extracts the diagonal of the matrix and
%   converts to a sparse matrix.
%
D = diag(sparse(d));
```

---

vec.m.

```
function v = vec(A);
% VEC   vec command
%
%   v = vec(A) computes the vec of the matrix A, defined as stacking
%   all the columns of A on top of each other.  That is, if
%   A = [ a1 a2 .. ap ], then vec(A) = [ a1' a2' ... ap' ]'.
%
[a,b] = size(A);
v = reshape(A,a*b,1);
```

M-estimators

inform.m.

```
function [iIb,Ig] = inform(v,X,S,iS,Dgam);
%
% [iIb,Ig] = inform(v,X,S,iS,Dgam);
%
%   This computes the information matrix for the multivariate t
%   distribution.  Because Ibg and Igb are both zeros, they are not
%   computed.
%
%   Inputs
%      n    = sample size
%      p    = dimension of data
%      v    = df for the chi-square
%      X    = design matrix
%      S    = current guess for Sigma
%      iS   = inverse of current guess for Sigma
%      Dgam = Derivative of G wrt mu'
%
%    Outputs
%      iIb = inverse of the information matrix for beta
%      Ig  = the information matrix for gamma.
%
n = size(X,1);
p = size(S,1);

vp = (v+p)/(v+p+2);
iIb = kron(S,pinv(X'*X))/vp;

veciSDgam = vec(iS)'*Dgam;
Ig = (n*vp)/2*(Dgam'*kron(iS,iS)*Dgam - 1/(v+p)*veciSDgam'*veciSDgam);
```

---

lntlikelihood.m.

```
function lnL = lntlikelihood(nu,X,beta,iSmle,T);
%
% This calculates the log of the Multivariate T likelihood function
% (except the constants).
%
[n,p] = size(T);
d = size(X,2);

part = 0;
B = dvec(beta,d,p);

for k=1:n
```

```
  Z = (T(k,:) - X(k,:)*B)';
  part = part + log(nu + Z'*iSmle*Z);
end

lnL = n/2*log(det(iSmle)) - (nu+p)/2*part;
```

---

mestim.m.

```
function mest = mestim(Y,m,X);
% MESTIM    M-estimators using Multivariate-T likelihood
%
%   This estimates the scatter matrix and the location vector/matrix
%   satisfying the model Y = XB + E using the multivariate T
%   distribution and the underlying distribution.  Output is similar
%   to Smultivar.m
%
%   Inputs:
%     Y : data matrix.  Columns represent variables and rows represent
%         units or cases
%     m : multiplcity vector (if not used, then vector of ones is
%         used)
%     X : Design matrix
%
%   Outputs:
%     mest.mean : location parameters (the matrix B)
%     mest.scat : scatter matrix      (Sigma)
%     mest.md   : mahalanobis distances. (dk's)
%

[n,p] = size(Y);            %sample size and dimension of the
                            %problem (size of Sigma).
if nargin==1
  % Both of these below are used for Multivariate Location/Scale
  m = ones(p,1);
  X = ones(n,1);
end

d = 1;                      %Dimension of beta (1 if multiv loc/scale)
nu = 5;                     %df for chi-square.
dimu=(p^2-m'*m)/2;          %Dimension of mu.
[Dg,A1,A2,L,Dp]= constantmatrices(m); %Constant matrices.

M=[];
avg=[];
```

```
for i=1:length(m);
  M = [ M repmat(i,1,m(i)) ];
  avg = [ avg repmat(1/m(i),1,m(i)) ];
end;
M = sparse(dummyvar(M));    %Simple multiplicity structure.
avg = (spdiag(avg)*M)';     %inv(M'*M)*M' (Used to calculate estimates
                            %  for eigenvalues. (phi0)

S = (nu-2)/nu*Y'*(eye(n)-ppo(X))*Y/(n-rank(X)); %Estimate for Sigma
beta = vec((X'*X)\X'*Y);    %Initial extimate for beta.

[G0,L0,dumm] = svd(S);
phi0 = avg*diag(sparse(L0)); %Avg multiple eigenvalues.

[Smle,Gam,Lam,Bmle] = solve_mles(nu,m,Dg,A1,A2,L,M,X,Y,G0,phi0,beta);

mest.mean = Bmle;
mest.scat = Smle;
Z = Y-X*Bmle;
mest.md = sqrt(sum(Z*inv(Smle).*Z,2));
```

---

scoref.m.

```
function [scorebeta, scoregamma] = scoref(v,X,beta,iS,Dgam,T);
%
% [scorebeta, scoregamma] = scoref(n,p,v,d,X,beta,iS,Dgam,T);
%
%   This computes the score function for the mutivariate t
%   distribution.
%
%   Inputs
%      v    = df for the chi-square
%      X    = design matrix
%      beta = current guess for beta
%      iS   = inverse of current guess for Sigma
%      Dgam = Derivative of G wrt mu'
%      T    = Data matrix
%
%   Outputs
%      scorebeta  = score function for beta
%      scoregamma = score function for gamma
%
[n,p] = size(T);
d = size(X,2);
```

```
A = 0;
lbeta = sparse(1,p*d);
vp = (v+p)/2;
B = dvec(beta,d,p);

for k=1:n
  Z = (T(k,:) - X(k,:)*B)';
  zS = Z'*iS;
  vzSz = v+zS*Z;
  A = A + vec(zS'*zS)'/vzSz;
  lbeta = lbeta + kron(zS,X(k,:))/vzSz;
end

A = -n/2*vec(iS)' + vp*A;
scorebeta = 2*vp*lbeta';
scoregamma = (A*Dgam)';
```

---

solve_mles.m.

```
function [Smle,Gam,Lam,Bmle] = solve_mles(nu,m,Dg,A1,A2,L,M,X,Y,...
                                         G0,phi0,beta);
%=====================================================================
%
%  [Smle,Gam,Lam,Bmle] = solve_mles(nu,m,Dg,A1,A2,L,M,X,Y,...
%                                   G0,phi0,beta);
%
%  This gives maximum likelihood estimates for the scatter matrix, the
%  location parameters, the eigenvalue matrix, and the eigenvector
%  matrix under a multivariate-t distribution.   Using Model 1 from
%  the dissertation (Where M = T3).
%
%  Inputs
%     nu   = dimension of parameter vectors
%     m    = multiplicity vector.
%     Dg   = Derivative of vecG wrt mu
%     A1   = Matrix associated with mu
%     A2   = Matrix associated with eta
%     L    = khatrirao(Ip,Ip)
%     M    = Average matrix (computed in robust_load.m)
%     X    = Design Matrix (location parameters)
%     Y    = Data Matrix Y
%     G0   = Initial guess for Eigenvector matrix
```

```
%     phi0 = Initial guess for phi.
%     beta = Current guess for the location parameters
%
%  Outputs
%     Smle  = MLE of S
%     Gam   = MLE of Gam
%     Lam   = MLE of Lam
%     Bmle  = MLE of B
%
%======================================================================
[n,p] = size(Y);
d = size(X,2);

dimbeta = p*d;
dimu = (p^2-m'*m)/2;
dimlam = length(m);
Ip = speye(p);
Ipp = commute(p,p);
tNp = speye(p^2)+Ipp;
LM = .5*L*M;
stop = 1;

% Initial Guesses for each
G = sparse(p,p);
Gam = G0;
Lam = M*phi0;    %Duplicate multiple ones.
Smle = G0*diag(Lam)*G0';
iSmle = G0*diag(1./Lam)*G0';
gamma = [ zeros(dimu,1) ; phi0 ];
k=0;
lnL = lntlikelihood(nu,X,beta,iSmle,Y);

while stop > 1e-7,
  k=k+1;
  gamma(1:dimu) = zeros(dimu,1);
  Dgam =  tNp*kron(Gam,Gam)*[ kron(spdiag(Lam),Ip)*Dg  LM ];

  [iIb,Ig] = inform(nu,X,Smle,iSmle,Dgam);
  [scoreb,scoreg] = scoref(nu,X,beta,iSmle,Dgam,Y);

  %Update mu and phi
  updateg = Ig\scoreg;
  updateb = iIb*scoreb;
  gamma_new = gamma + updateg;
  mu = gamma_new(1:dimu);
  stop = sqrt(abs([scoreb; scoreg]'*[ updateb ; updateg ]));
```

```
%
% If max(abs(mu))<.2 continue normally, otherwise force max of mu to
% be .2. (Force updateg and updateb with the same constant.)
%
maxmu = max(abs(mu));
alpha = 1;
if maxmu > .2
  alpha = .2/maxmu;
  updateg = alpha*updateg;
  gamma_new = gamma + updateg;
  mu = gamma_new(1:dimu);
end;
updateb = alpha*updateb;
beta_new = beta + updateb;

%Form current estimates for G and Lm
phi = gamma_new(1+dimu:end);
eta = solve_eta(A1,A2,mu);
G = dvec(A1*mu+A2*eta,p,p);

%Estimate matrices.
Lam = M*phi;              %note: diag(M*phi) = dvec(L*M*phi,p,p);
Gam_new = Gam*G;
Smle = Gam_new*diag(Lam)*Gam_new';
iSmle = Gam_new*diag(1./Lam)*Gam_new';

% Check for a reduction in the log likelihood before continuing.
lnL_new = lntlikelihood(nu,X,beta_new,iSmle,Y);
while (lnL_new < lnL) & (k>5) & norm(lnL_new-lnL)>1e-12,
  lnL_new < lnL
  fprintf('lnL %15.10g lnL_new %15.10g updateg %9.6g\n',lnL, ...
          lnL_new,norm(updateg));
  updateb = updateb/2;
  updateg = updateg/2;

  gamma_new = gamma + updateg;
  beta_new = beta + updateb;
  mu = gamma_new(1:dimu);

  %Form current estimates for G and Lm
  phi = gamma_new(1+dimu:end);
  eta = solve_eta(A1,A2,mu);
  G = dvec(A1*mu+A2*eta,p,p);

  %Estimate matrices.
  Lam = M*phi;          %note: diag(M*phi) = dvec(L*M*phi,p,p);
```

```
    Gam_new = Gam*G;
    Smle = Gam_new*diag(Lam)*Gam_new';
    iSmle = Gam_new*diag(1./Lam)*Gam_new';

    lnL_new = lntlikelihood(nu,X,beta_new,iSmle,Y);
%    pause;
  end;

%    fprintf('lnL %9.6g lnL_new %9.6g updateg %9.6g\n',...
%             lnL,lnL_new,norm(updateg));
  fprintf('It#%3.0f, Max(mu)= %6.4e, stop= %6.4e\n', k, maxmu, stop)
  gamma = gamma_new;
  beta = beta_new;
  lnL = lnL_new;
  Gam = Gam_new;
end
Gammle = Gam;
Lammle = Lam;
Bmle = dvec(beta,d,p);
```

S-estimators

compute_a.m.

```
function [a,W1,Wzz] = compute_a(b0,c0,Z,Gam,dk,Laminv,D1_lam);
%======================================================================
%
%  compute_a   Compute the vector a.
%
%   [a,W1,Wzz] = compute_a(b0,c0,Z,Gam,dk,Laminv,D1_lam) computes the
%   current value of a given latest phi (for the newest variable defs)
%
%   Inputs
%     b0     = Constant associated with constraint q
%     c0     = Constant associated with breakdown point
%     Z      = Centered Data Matrix (Y-XB)
%     Gam    = Current estimate for the Eigenvector matrix
%     dk     = Current estimate for the distances dk
%     Laminv = Current inverse of the eigenvalue matrix
%     D1_lam = Current first derivative of lambda wrt phi
%
%   Outputs
%     a      = Current estimate of derivative of q wrt phi
%     W1     = Current matrix containing (1/n)*Diag(rho(1)(dk)/dk)
%     Wzz    = Current matrix Z'*W1*Z
```

```
%
%==============================================================

n = size(Z,1);
W1 = (1/n)*spdiag(.5*rho(1.1,dk,c0));         %(1/2)*diag(rho(dk)/dk)
Wzz = Z'*W1*Z;
LaminvGam = Laminv*Gam';
%
%  Any of the following representations will work.  However, the last
%  one used is the fastest.  Note that L'*vec(W) = diag(W) (the
%  diagonal of ANY matrix W);
%    a = -D1_lam'*L'*kron(Gam',Gam')*kron(Siginv,Siginv)*vec(Wzz)
%      = -D1_lam'*L'*vec(Gam'*Siginv*Wzz*Siginv*Gam)
%      = -D1_lam'*L'*vec(Laminv*Gam'*Wzz*Gam*Laminv)
%      = -D1_lam'*diag(Laminv*Gam'*Wzz*Gam*Laminv);
%
a = -D1_lam'*diag(LaminvGam*Wzz*LaminvGam');
```

---

deriv1_w.m.

```
function [dwvec] = ...
    deriv1_w(T1,T2,A1,A2,Dg,Y,X,d,b0,c0,nu,Gam,beta,mu,psi,phi);
%==============================================================
%
%  deriv1_w   Compute the first derivative of w
%
%   [dwvec] =
%   deriv1_w(T1,T2,A1,A2,Dg,Y,X,d,b0,c0,nu,Gam,beta,mu,psi,phi)
%   outputs the first derivative of w wrt the parameters
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     A1   = Matrix associated with mu
%     A2   = Matrix associated with eta
%     Dg   = Derivative of vecG wrt mu
%     Y    = Data Matrix Y
%     X    = Design Matrix (location parameters)
%     d    = number of rows of B
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     nu   = vector containing sizes of beta, mu, phi
%     lam  = Current guess eigenvalues
%     Gam  = Current guess for Eigenvector matrix
```

```
%     beta = Current guess for the location parameters
%     psi  = Current value for psi.
%     phi  = Current value for phi.
%
%   Outputs
%     dwvec = vector containing dw/dbeta dw/dmu dw/dpsi
%
%======================================================================

[n,p] = size(Y);

tNp = commute(p,p)+speye(p^2);
B_est = dvec(beta,d,p);
Z = Y - X*B_est;

[V1,V2,psi]=initial_guess(T1,T2,Z,b0,c0,Gam,beta,phi);
[a,V1,V2,phi,psi,eta_psi,lam,D1_lam,dk,W1,Wzz]= ...
    satisfy_constraint(T1,T2,V1,V2,Z,b0,c0,Gam,beta,phi,psi);

Siginv = Gam*diag(1./lam)*Gam';
F1{2} = tNp*kron(Gam*diag(lam),Gam)*Dg;
F1{3} = khatrirao(Gam,Gam)*D1_lam;

Wxz = X'*W1*Z;
Wzz = Z'*W1*Z;
SWzzS = Siginv*Wzz*Siginv;

dq1{1} = -2*vec(Wxz*Siginv);
dq1{2} = -F1{2}'*vec(SWzzS);
%dq1{3} = -F1{3}'*vec(SWzzS);
dep{1} = -inv(a'*V2)*dq1{1}';
dep{2} = -inv(a'*V2)*dq1{2}';
%dep{3} = -inv(a'*V2)*a'*V1;

P1{1} = V2*dep{1};
P1{2} = V2*dep{2};
P1{3} = V1;   %Actually is P1{3} = V1+V2*dep{3};
              %(Choice of a makes a'*V1=0)
dw1{3} = F1{3}'*vec(Siginv);

for i=1:3
  dw{i} = P1{i}'*dw1{3};
end;

dwvec = [ dw{1}; dw{2}; dw{3} ];
```

deriv_w.m.

```
function [dw,dw1,d2w,d3w,K2hat,Cmeat,sandwich,EZ2D0,ED1] = ...
    deriv_w(T1,T2,A1,A2,Dg,Y,X,b0,c0,nu,Gam,beta,phi);

%======================================================================
%
% deriv_w    Compute various quantities from the solution of the
%            S-estimating equation.
%
%  [dw,dw1,d2w,d3w,K2hat,Cmeat,sandwich,EZ0D0,ED1]
%    = deriv_w(T1,T2,A1,A2,Dg,L,Y,X,b0,c0,nu,Gam,beta,phi);
%
%   Evaluate the derivatives of w, as well as approximate the bias and
%   asymptotic variance (sandwich estimator).  Note that dw must be
%   zero for the bias and asymptotic variance to be correct.  That is,
%   the estimates of Gam, beta, phi must be the S-estimates.
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     A1   = Matrix associated with mu
%     A2   = Matrix associated with eta
%     Dg   = Derivative of vecG wrt mu
%     Y    = Data Matrix Y
%     X    = Design Matrix (location parameters)
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     nu   = vector containing sizes of beta, mu, phi
%     Gam  = S-estimate for Eigenvector matrix
%     beta = S-estimate for the location parameters
%     phi  = S-estimate for phi.
%
%   Ouputs
%     dw       = dw/dtheta
%     dw1      = dw/dtheta given eta_psi is constant
%     d2w      = d2w/(dtheta' x dtheta)
%     d3w      = d3w/(dtheta' x dtheta' x dtheta)
%     K2hat    = Approx to Fisher's Information matrix
%     Cmeat    = Meat of sandwich estimator
%     sandwich = sandwich estimator
%     EZ0D0    = Expected value of Z0D0
%     ED1      = Expected value of D1
```

```
%
%=====================================================================




%/////////////////////////////////////////////////////////////////////
%
% This consists of the derivatives of w and q (and others) assuming
% that eta_psi is constant and also assuming eta_psi is a function of
% the parameters beta, mu, and psi.
%
%/////////////////////////////////////////////////////////////////////

%
% Matrices and constants
%
[n,p] = size(Y);
d = size(X,2);

Dp  = dup(p);
Ipp = commute(p,p);
Ip2 = speye(p^2);
tNp = Ipp + Ip2;
Id  = speye(d);
Ip  = speye(p);
Idp = commute(d,p);
Ipd = Idp';
dims = [ nu(1) nu(2) nu(3)-1 1 ];
for i=1:3
  tNnu{i} = speye(nu(i)^2) + commute(nu(i),nu(i));
  Inu{i} = speye(nu(i));
  Inu_2{i} = speye(nu(i)^2);
end
for i=1:4
  tNdim{i} = speye(dims(i)^2) + commute(dims(i),dims(i));
  Idim{i} = speye(dims(i));
  Idim_2{i} = speye(dims(i)^2);
end;



%
% Data and current estimates.
%
B_est = dvec(beta,d,p);
```

```
Z = Y - X*B_est;

[V1,V2,psi]=initial_guess(T1,T2,Z,b0,c0,Gam,beta,phi);
[a,V1,V2,phi,psi,eta_psi,lam,D1lam,dk,W1,Wzz]= ...
    satisfy_constraint(T1,T2,V1,V2,Z,b0,c0,Gam,beta,phi,psi);

[Siginv,Laminv,D1lam,lam,D2lam,D3lam] = model_lam(T1,T2,V1,V2,Gam, ...
                                                  psi,eta_psi);
Sig = Gam*diag(lam)*Gam';
iSiS = kron(Siginv,Siginv);
w = log(det(Sig));


%/////////////////////////////////////////////////////////////////////
%
%
% Derivatives of vecSig and lambda
%
%
%/////////////////////////////////////////////////////////////////////

% First derivatives of vecSig
%   (Compute F1's);
F1{2} = tNp*kron(Gam*diag(lam),Gam)*Dg;
F1{3} = khatrirao(Gam,Gam)*D1lam;

% Second derivatives of vecSig
%   (Compute F11's and F2's)
kGamGam = kron(Gam,Gam);
kDgDg = kron(Dg,Dg);
K3 = kron3(Ip,(vec(Ip))',Ip);
K4 = kron3(Ip,diag(lam),Ip2);
Dg2part = A2*Dp'*K3;
L = khatrirao(Ip,Ip);
LD1lam = L*D1lam;
Dg2 = Dg2part*kDgDg;
Dg3 = -A2*Dp'*K3*(kron(Dg,Ipp*Dg2) - kron(Dg2,Dg)*(...
    speye(nu(2)^3) + kron(Inu{2},commute(nu(2),nu(2))))) ;

%
% Specify each part of F2.  F{2,2} is F2_{mu',mu'}, F{2,3} is
% F2_{mu',phi'} and so forth for the rest.
% ============================================================
%

F2{2,2} = tNp*kGamGam*(kron(diag(lam),Ip)*Dg2part-K3*K4)*kDgDg;
```

```
F2{2,3} = -tNp*kGamGam*K3*sparse(kron(Dg,LD1lam));
F2{3,2} = F2{2,3}*commute(nu(3),nu(2));
F2{3,3} = khatrirao(Gam,Gam)*D2lam;

%
% Form F11's
% Since vec(F11) = vec(F2), then this is an easy way to get F11.
% ================================================================
for i=2:3
  for j=2:3
    F11{i,j} = dvec(F2{i,j},p^2*nu(j),nu(i));
  end;
end;


% Third derivatives of vecSig
%   (Compute F111's and F3's)

Lam = diag(lam);
F3{2,2,2} = tNp*(kron(Gam*Lam,Gam)*Dg3 - kron3(Gam,vec(Lam)',Gam)*...
            (kron(Dg,Dg2)-kron(Ipp*Dg2,Dg)*kron(Inu{2},tNnu{2})));
F111{2,2,2} = dvec(F3{2,2,2},p^2*nu(2)^2,nu(2));

F3{3,3,2} = tNp*kron3(Gam,vec(Ip)',Gam)*kron(L*D2lam,Dg);
F111{3,3,2} = dvec(F3{3,3,2}*commute(nu(2),nu(3)^2), ...
                  p^2*nu(3)^2,nu(2));

F3{3,2,2} = tNp*(kron3(Gam,vec(Ip)',Gam)*kron(LD1lam,Dg2) + ...
            kron3(vec(LD1lam)',Gam,Gam)*kron(Inu{3},kron3(Ip, ...
            Ipp,Ip)*kron(Dg,Dg)));
F111{2,2,3} = dvec(F3{3,2,2},p^2*nu(2)^2,nu(3));

F3{3,3,3} = kron(Gam,Gam)*L*D3lam;
F111{3,3,3} = dvec(F3{3,3,3},p^2*nu(3)^2,nu(3));



%//////////////////////////////////////////////////////////////////////
%
%
% Miscellaneous derivatives
%
%
%//////////////////////////////////////////////////////////////////////

w2k = (1/4)*(rho(2,dk,c0)-rho(1.1,dk,c0))./(dk.^2);
```

```
W2 = spdiag(w2k/n);
W3 = spdiag((1/8)*(rho(3,dk,c0)./dk-12*w2k)./(n*dk.^2));

%
% Form Khatri-Rao product matrices  (cols are kronecker products)
%
ZSiginv = Z*Siginv;   % (form rows of zk'*Siginv)

Kxx = khatrirao(X',X');
Kzx = khatrirao(ZSiginv',X');
Kzz = khatrirao(ZSiginv',ZSiginv');
Kxzx = khatrirao(X',Kzx);
Kxzz = khatrirao(X',Kzz);
Kzzx = khatrirao(ZSiginv',Kzx);
Kzzz = khatrirao(ZSiginv',Kzz);
Kzxzx = khatrirao(Kzx,Kzx);
Kzxzz = khatrirao(Kzx,Kzz);
Kzzzx = khatrirao(Kzz,Kzx);
Kzzzz = khatrirao(Kzz,Kzz);

%
% Form Wxx, Wxz, Wzz, Wzx
%
Wxx = X'*W1*X;
Wzz = Z'*W1*Z;
Wxz = X'*W1*Z;
Wzx = Wxz';

%
% Form Vxx, Vxz, Vzz
%
W2Kzz = W2*Kzz';
W2Kzx = W2*Kzx';

Vzz = Kzz*W2Kzz;
Vxz = Kzx*W2Kzz;
Vxx = Kzx*W2Kzx;

%
% Form Mxx, Mxz, Mzz
%
SWzzS = Siginv*Wzz*Siginv;
Mxx = kron(Siginv,Wxx) + 2*Vxx;
Mxz = kron(Siginv,Wxz*Siginv) + Vxz;
Mzz = kron(Siginv,SWzzS) + .5*Vzz;
```

```
%
% Derivatives of vecWxx
%
% not needed because E[] is zero
%dvecWxx{1} = -2*Kxx*W2Kzx;
dvecWxx{2} = -Kxx*W2Kzz*F1{2};
dvecWxx{3} = -Kxx*W2Kzz*F1{3};

%
% Derivatives of vecWzx, vecWzz
%
dvecWzx{1} = -Ipd*kron(Sig,Id)*Mxx;
% not needed because E[] is zero
%dvecWzz{1} = -tNp*kron(Sig,Sig)*Mxz';

for i=2:3
  % not needed because E[] is zero
  %  dvecWzx{i} = -khatrirao(X',Z')*W2Kzz*F1{i};
  dvecWzz{i} = -khatrirao(Z',Z')*W2Kzz*F1{i};
end;

%
%  Form dVxx
%


% Not needed because E[] is zero
%dVxx{1}= -kron(Siginv,X'*W2*Kxzx') - 2*Kzx*W3*Kzxzx' + ...
%          -Kzx*W2*Kxx'*kron(vec(Siginv)',speye(d^2))*kron3(Ip,Idp,Id);

temp1 = -(kron(Siginv,X'*W2*Kzzx') + Kzx*W3*Kzzzx' + ...
        Kzx*W2Kzx*kron(vec(Siginv)',Inu{1})*...
        kron3(Ip,Ipp,Id));
for i=2:3
  dVxx{i} = temp1*kron(F1{i},Inu{1});
end;

%
%  Form dVxz
%

dVxz{1} = -kron(Siginv,X'*W2*Kxzz') - 2*Kzx*W3*Kzxzz' - Kzx*W2Kzx ...
     *Idp*kron(vec(Siginv)',Inu{1})*kron3(Ip,Ipp,Id)*kron(Inu{1},tNp);

% Not needed (E[]=0)
common1 = kron(vec(Siginv)',Ip2)*kron3(Ip,Ipp,Ip)*kron(Ip2,tNp);
```

```
%temp1 = -(kron(Siginv,X'*W2*Kzzz')+Kzx*W3*Kzzzz' + ...
%         Kzx*W2Kzz*common1);
%for i=2:3
%  dVxz{i} = temp1*kron(F1{i},Ip2);
%end

%
%  Form dVxz
%

% not needed (E[]=0)
%dVzz{1} = -tNp*kron(Siginv,Siginv*Z'*W2*Kxzz') - 2*Kzz*W3*Kzxzz' ...
%      - Kzz*W2Kzx*Idp*kron(vec(Siginv)',Inu{1})*kron3(Ip,Idp,Ip)* ...
%      kron(Inu{1},tNp);

temp1 = -(tNp*kron(Siginv,Siginv*Z'*W2*Kzzz') + Kzz*W3*Kzzzz' + ...
       Kzz*W2Kzz*common1);
for i=2:3
  dVzz{i} = temp1*kron(F1{i},Ip2);
end




%//////////////////////////////////////////////////////////////////////
%
%
% Derivatives of q when ep is constant
%
%
%//////////////////////////////////////////////////////////////////////

for i=2:3
dMzz{i} = - kron3(Siginv,vec(Siginv)',SWzzS)*kron(F1{i},Ip2) ...
      - Ipp*kron3(Ip,vec(Ip)',Siginv)*kron(tNp*Mzz*F1{i},Ipp) ...
      + .5*dVzz{i};
end;

% First derivatives
dq1{1} = -2*vec(Wxz*Siginv);
dq1{2} = -F1{2}'*vec(SWzzS);
dq1{3} = -F1{3}'*vec(SWzzS);

% Second derivatives
```

```
d2q11{1,1} = 2*Mxx;
d2q11{2,1} = 2*Mxz*F1{2};
d2q11{3,1} = 2*Mxz*F1{3};
d2q11{1,2} = d2q11{2,1}';
d2q11{1,3} = d2q11{3,1}';
for i=2:3
  for j=2:3
    d2q11{j,i} = -F11{i,j}'*kron(Inu{j},vec(SWzzS)) + ...
       2*F1{i}'*Mzz*F1{j};
  end;
end;

I31 = commute(nu(3),nu(1));
I32 = commute(nu(3),nu(2));
I21 = commute(nu(2),nu(1));
SvSS = kron3(Siginv,vec(Siginv)',Siginv);
IdvSS = kron3(Id,vec(Siginv)',Siginv);

% Third Derivatives
% -------A-------
%d3q111{1,1,1} = 2*Idp*kron3(Id,vec(Id)',Siginv)* ...
%      kron(dvecWxx{1},Ipd) + 4*dVxx{1};
% -------B and C-------
for i=2:3
  d3q111{i,i,i} = -F111{i,i,i}'*kron(Inu_2{i},vec(SWzzS)) ...
      + 2*F11{i,i}'*kron(Inu{i},Mzz*F1{i})*tNnu{i} ...
      + 2*F1{i}'*dMzz{i}*kron(Inu{i},F1{i}) + 2*F1{i}'*Mzz*F2{i,i};
end;
% -------D-------
d3q111{1,2,1} = 2*Idp*IdvSS*kron(dvecWzx{1},F1{2}) ...
      + 2*dVxz{1}*kron(Inu{1},F1{2});
d3q111{1,1,2} = dvec(kron(Inu{1},I21)*vec(d3q111{1,2,1}),...
                    nu(2),nu(1)^2);
d3q111{2,1,1} = d3q111{1,2,1}*I21;

% -------E-------
d3q111{3,1,1} = -2*kron3(Siginv,vec(Siginv)',Wxx)* ...
    kron(F1{3},Inu{1}) + 2*Idp*kron(Id,vec(Id)'*Siginv)* ...
    kron(dvecWxx{3},Ipd) + 4*dVxx{3};
d3q111{1,3,1} = d3q111{3,1,1}*I31';
d3q111{1,1,3} = dvec(kron(Inu{1},I31)*vec(d3q111{1,3,1}),...
                    nu(3),nu(1)^2);
% -------F-------
%d3q111{2,2,1} = 2*(...
%                 -kron(Ip,Wxz)*tNp*SvSS*kron(F1{2},F1{2}) ...
%                 + Idp*IdvSS*kron(dvecWzx{2},F1{2}) ...
```

```
%                  + dVxz{2}*kron(Inu{2},F1{2}) + Mxz*F2{2,2});
%d3q111{2,1,2} = dvec(kron(Inu{2},I21)*vec(d3q111{2,2,1}),...
%                      nu(2),nu(1)*nu(2));
%d3q111{1,2,2} = d3q111{2,1,2}*I21';

% -------G-------
%d3q111{3,3,1} = 2*(Idp*IdvSS*kron(dvecWzx{3},F1{3}) ...
%                - kron(Ip,Wxz)*tNp*SvSS*kron(F1{3},F1{3}) ...
%                + dVxz{3}*kron(Inu{3},F1{3}) + Mxz*F2{3,3});
%d3q111{3,1,3} = dvec(kron(Inu{3},I31)*vec(d3q111{3,3,1}),...
%                      nu(3),nu(1)*nu(3));
%d3q111{1,3,3} = d3q111{3,1,3}*I31';
% -------H-------
d3q111{3,3,2} = -F111{3,3,2}'*kron(Inu_2{3},vec(SWzzS)) ...
               + 2*F11{2,3}'*kron(Inu{3},Mzz*F1{3})*tNnu{3} ...
               + 2*F1{2}'*dMzz{3}*kron(Inu{3},F1{3}) ...
               + 2*F1{2}'*Mzz*F2{3,3};
d3q111{3,2,3} = dvec(kron(Inu{3},I32)*vec(d3q111{3,3,2}),...
                     nu(3),nu(3)*nu(2));
d3q111{2,3,3} = d3q111{3,2,3}*I32';
% -------I-------
d3q111{2,2,3} = -F111{2,2,3}'*kron(Inu_2{2},vec(SWzzS)) ...
               + 2*F11{3,2}'*kron(Inu{2},Mzz*F1{2})*tNnu{2} ...
               + 2*F1{3}'*dMzz{2}*kron(Inu{2},F1{2}) ...
               + 2*F1{3}'*Mzz*F2{2,2};
d3q111{2,3,2} = dvec(kron(Inu{2},I32')*vec(d3q111{2,2,3}),...
                     nu(2),nu(3)*nu(2));
d3q111{3,2,2} = d3q111{2,3,2}*I32;

% -------J-------
%d3q111{3,2,1} = 2*(-kron(Ip,Wxz)*tNp*SvSS*kron(F1{3},F1{2}) + ...
%                 Idp*IdvSS*kron(dvecWzx{3},F1{2}) + ...
%                 dVxz{3}*kron(Inu{3},F1{2}) + Mxz*F2{3,2});
%d3q111{2,3,1} = d3q111{3,2,1}*I32';
%d3q111{3,1,2} = dvec(kron(Inu{3},I21)*vec(d3q111{3,2,1}),...
%                      nu(2),nu(1)*nu(3));
%d3q111{1,3,2} = d3q111{3,1,2}*I31';
%d3q111{2,1,3} = dvec(kron(Inu{2},I31)*vec(d3q111{2,3,1}),...
%                      nu(3),nu(1)*nu(2));
%d3q111{1,2,3} = d3q111{2,1,3}*I21';

%///////////////////////////////////////////////////////////////////////
%
%
% Derivatives of w when ep is constant
%
```

```
%
%//////////////////////////////////////////////////////////////////////


% First derivatives
dw1{1} = zeros(nu(1),1);
dw1{2} = zeros(nu(2),1);
dw1{3} = n*F1{3}'*vec(Siginv);

% Second derivatives
d2w11{3,3} = n*(F11{3,3}'*kron(Inu{3},vec(Siginv)) - ...
                F1{3}'*iSiS*F1{3});

% Third derivatives
d3w111{3,3,3} = n*(F111{3,3,3}'*kron(Inu_2{3},vec(Siginv)) ...
                - F11{3,3}'*kron(Inu{3},iSiS*F1{3})*tNnu{3} ...
                + 2*F1{3}'*SvSS*kron(F1{3},F1{3}) ...
                - F1{3}'*iSiS*F2{3,3});


%//////////////////////////////////////////////////////////////////////
%
%
% Implicit derivatives (derivs of phi and ep)
%
%
%//////////////////////////////////////////////////////////////////////

% First derivatives
dep{1} = -inv(a'*V2)*dq1{1}';
dep{2} = -inv(a'*V2)*dq1{2}';
dep{3} = -inv(a'*V2)*a'*V1;

P1{1} = V2*dep{1};
P1{2} = V2*dep{2};
P1{3} = V1;          %Actually is P1{3} = V1+V2*dep{3}; (Choice of a
                     %makes a'*V1=0);

% Second derivatives
% do 2nd ders of beta, mu
%dummy = [ 1 1; 1 2; 2 2 ];
%for i=1:3
%  i1 = dummy(i,1); i2 = dummy(i,2);
%  d2ep{i1,i2} = -(d2q11{i1,i2} + d2q11{3,i2}*P1{i1} + ...
%             P1{i2}'*d2q11{i1,3} ...
%             + P1{i2}'*d2q11{3,3}*P1{i1})*kron(Idim{i1},inv(V2'*a));
```

```
%end;
%
%dumend = kron(Idim{3},inv(V2'*a));
%d2ep{3,3} = -(             P1{3}'*d2q11{3,3})*P1{3}*dumend;
%d2ep{3,2} = -(d2q11{3,2} + P1{2}'*d2q11{3,3})*P1{3}*dumend;
%d2ep{3,1} = -(d2q11{3,1} + P1{1}'*d2q11{3,3})*P1{3}*dumend;
%
%dummy = [ 1 1; 1 2; 2 2; 3 3; 3 2; 3 1] ;
%for i=1:size(dummy,1)
%  i1 = dummy(i,1);
%  i2 = dummy(i,2);
%  p1 = dims(i1);
%  p2 = dims(i2);
%  Ip1p2 = commute(p1,p2);
%  P11{i1,i2} = (d2ep{i1,i2}*kron(Idim{i1},V2'))';
%  P2{i2,i1} = dvec(vec(P11{i1,i2}),nu(3),p1*p2);
%  P11{i2,i1} = dvec(vec(P2{i2,i1}*Ip1p2),p2*nu(3),p1);
%end
%P2{1,2} = P2{2,1}*I21';
%

%///////////////////////////////////////////////////////////////////
%
%
% Derivatives of w (ep is a function of beta, mu, and psi)
%
%
%///////////////////////////////////////////////////////////////////

% First Derivatives
for i=1:3
  dw{i} = P1{i}'*dw1{3};
end;

%% Second Derivatives
%dummy = [ 1 1; 2 2; 3 3; 1 2; 3 2; 3 1];
%for i=1:6
%  i1 = dummy(i,1);  i2 = dummy(i,2);
%  d2w{i1,i2} = P11{i1,i2}'*kron(Idim{i1},dw1{3}) + ...
%               P1{i2}'*d2w11{3,3}*P1{i1};
%end;
%


%-------------------------------------------------------------------
%
```

```
%  Derivatives for Asymptotic Distribution expansions
%    Using current choice for l1
%
%----------------------------------------------------------------------

tNperp_nu3 = eye(nu(3)^2)-commute(nu(3),nu(3));
first = kron(V2',V1')*tNperp_nu3;

% def of l1
l1 = [ -n*dq1{1};
       -n*dq1{2};
       first*kron(a,dw1{3});
       n*(mean(rho(0,dk,c0)-b0))];

% derivs of l1_beta and l1_mu
for j=1:2
  for i=1:2
    l2{i,j} = - n*d2q11{i,j};
  end
  l2{3,j} = -n*d2q11{3,j}*V1;
  l2{4,j} = -n*d2q11{3,j}*V2;
end;

% derivs of l1_psi
for i=1:2
  l2{i,3} = first*kron(d2q11{i,3},dw1{3});
end
common = first*(kron(d2q11{3,3},dw1{3}) + kron(a,d2w11{3,3}));
l2{3,3} = common*V1;
l2{4,3} = common*V2;

% derivs of l1_etapsi
for i=1:2
  l2{i,4} = n*dq1{i}';
end
l2{3,4} = n*a'*V1;
l2{4,4} = n*a'*V2;

%
%Second derivatives (like 3rd derivs of w) (no implicit derivatives)
%   (Expectations that are zero are zeros out, but real derivs are
%   commented out)

I3n1 = commute(nu(3)-1,nu(1));
I3n2 = commute(nu(3)-1,nu(2));
```

```
%------1--------
%l3{1,1,1} = -n*d3q111{1,1,1};
l3{1,1,1} = sparse(nu(1),nu(1)*nu(1));
%------2--------
l3{1,2,1} = -n*d3q111{1,2,1};
l3{2,1,1} = l3{1,2,1}*I21;
%------3--------
l3{1,3,1} = -n*d3q111{1,3,1}*kron(Idim{1},V1);
l3{3,1,1} = l3{1,3,1}*I3n1;
%------4--------
l3{1,4,1} = -n*d3q111{1,3,1}*kron(Idim{1},V2);
l3{4,1,1} = l3{1,4,1};
%------5--------
%l3{2,2,1} = -n*d3q111{2,2,1};
l3{2,2,1} = sparse(nu(1),nu(2)*nu(2));
%------6--------
%l3{2,3,1} = -n*d3q111{2,3,1}*kron(Idim{2},V1);
l3{2,3,1} = sparse(nu(1),nu(2)*(nu(3)-1));
l3{3,2,1} = l3{2,3,1}*I3n2;
%------7--------
%l3{2,4,1} = -n*d3q111{2,3,1}*kron(Idim{2},V2);
l3{2,4,1} = sparse(nu(1),nu(2));
l3{4,2,1} = l3{2,4,1};
%------8--------
%l3{3,3,1} = -n*d3q111{3,3,1}*kron(V1,V1);
l3{3,3,1} = sparse(nu(1),(nu(3)-1)^2);
%------9--------
%l3{3,4,1} = -n*d3q111{3,3,1}*kron(V1,V2);
l3{3,4,1} = sparse(nu(1),nu(3)-1);
l3{4,3,1} = l3{3,4,1};
%-----10--------
%l3{4,4,1} = -n*d3q111{3,3,1}*kron(V2,V2);
l3{4,4,1} = sparse(nu(1),1);
%-----11--------
l3{1,1,2} = -n*d3q111{1,1,2};
%-----12--------
%l3{2,1,2} = -n*d3q111{2,1,2};
l3{2,1,2} = sparse(nu(2),nu(2)*nu(1));
l3{1,2,2} = l3{2,1,2}*I21';
%-----13--------
%l3{1,3,2} = -n*d3q111{1,3,2}*kron(Idim{1},V1);
l3{1,3,2} = sparse(nu(2),nu(1)*(nu(3)-1));
l3{3,1,2} = l3{1,3,2}*I3n1;
%-----14--------
%l3{1,4,2} = -n*d3q111{1,3,2}*kron(Idim{1},V2);
l3{1,4,2} = sparse(nu(2),nu(1));
```

```
l3{4,1,2} = l3{1,4,2};
%-----15--------
l3{2,2,2} = -n*d3q111{2,2,2};
%-----16--------
l3{2,3,2} = -n*d3q111{2,3,2}*kron(Idim{2},V1);
l3{3,2,2} = l3{2,3,2}*I3n2;
%-----17--------
l3{2,4,2} = -n*d3q111{2,3,2}*kron(Idim{2},V2);
l3{4,2,2} = l3{2,4,2};
%-----18--------
l3{3,3,2} = -n*d3q111{3,3,2}*kron(V1,V1);
%-----19--------
l3{3,4,2} = -n*d3q111{3,3,2}*kron(V1,V2);
l3{4,3,2} = l3{3,4,2};
%-----20--------
l3{4,4,2} = -n*d3q111{3,3,2}*kron(V2,V2);
%-----21--------
l3{1,1,3} = first*kron(d3q111{1,1,3},dw1{3});
%-----22--------
%l3{1,2,3} = first*kron(d3q111{1,2,3},dw1{3});
l3{1,2,3} = sparse(nu(3)-1,nu(1)*nu(2));
l3{2,1,3} = l3{1,2,3}*I21;
%-----23--------
%l3{1,3,3} = first*(kron(d3q111{1,3,3},dw1{3}) + ...
%                    kron(d2q11{1,3},d2w11{3,3}))*kron(Idim{1},V1);
l3{1,3,3} = sparse(nu(3)-1,nu(1)*(nu(3)-1));
l3{3,1,3} = l3{1,3,3}*I3n1;
%-----24--------
%l3{1,4,3} = first*(kron(d3q111{1,3,3},dw1{3}) + ...
%                    kron(d2q11{1,3},d2w11{3,3}))*kron(Idim{1},V2);
l3{1,4,3} = sparse(nu(3)-1,nu(1));
l3{4,1,3} = l3{1,4,3};
%-----25--------
l3{2,2,3} = first*kron(d3q111{2,2,3},dw1{3});
%-----26--------
l3{2,3,3} = first*(kron(d3q111{2,3,3},dw1{3}) + ...
                   kron(d2q11{2,3},d2w11{3,3}))*kron(Idim{2},V1);
l3{3,2,3} = l3{2,3,3}*I3n2;
%-----27--------
l3{2,4,3} = first*(kron(d3q111{2,3,3},dw1{3}) + ...
                   kron(d2q11{2,3},d2w11{3,3}))*kron(Idim{2},V2);
l3{4,2,3} = l3{2,4,3};
%-----28a-------
l3phiphipsi = first*(kron(d3q111{3,3,3},dw1{3}) + ...
                     kron(d2q11{3,3},d2w11{3,3})*tNnu{3} + ...
                     kron(a,d3w111{3,3,3}));
```

```
%-----28--------
l3{3,3,3} = l3phiphipsi*kron(V1,V1);
%-----29--------
l3{3,4,3} = l3phiphipsi*kron(V1,V2);
l3{4,3,3} = l3{3,4,3};
%-----30--------
l3{4,4,3} = l3phiphipsi*kron(V2,V2);
%-----31--------
l3{1,1,4} = n*dvec(d2q11{1,1},1,nu(1)^2);
%-----32--------
%l3{1,2,4} = n*dvec(d2q11{1,2},1,nu(1)*nu(2));
l3{1,2,4} = sparse(1,nu(1)*nu(2));
l3{2,1,4} = l3{1,2,4}*I21;
%-----33--------
%l3{1,3,4} = n*dvec(d2q11{1,3},1,nu(1)*nu(3))*kron(Idim{1},V1);
l3{1,3,4} = sparse(1,nu(1)*(nu(3)-1));
l3{3,1,4} = l3{1,3,4}*I3n1;
%-----34--------
%l3{1,4,4} = n*dvec(d2q11{1,3},1,nu(1)*nu(3))*kron(Idim{1},V2);
l3{1,4,4} = sparse(1,nu(1));
l3{4,1,4} = l3{1,4,4};
%-----35--------
l3{2,2,4} = n*dvec(d2q11{2,2},1,nu(2)^2);
%-----36--------
l3{2,3,4} = n*dvec(d2q11{2,3},1,nu(2)*nu(3))*kron(Idim{2},V1);
l3{3,2,4} = l3{2,3,4}*I3n2;
%-----37--------
l3{2,4,4} = n*dvec(d2q11{2,3},1,nu(2)*nu(3))*kron(Idim{2},V2);
l3{4,2,4} = l3{2,4,4};
%-----38--------
l3{3,3,4} = n*dvec(d2q11{3,3},1,nu(3)^2)*kron(V1,V1);
%-----39--------
l3{3,4,4} = n*dvec(d2q11{3,3},1,nu(3)^2)*kron(V1,V2);
l3{4,3,4} = l3{3,4,4};
%-----40--------
l3{4,4,4} = n*dvec(d2q11{3,3},1,nu(3)^2)*kron(V2,V2);


% Put them together to get d2l1_theta wrt theta' theta'
%   Note that K3 = (1/n)*l3
df = sum(dims);
K3hat = sparse(df,df^2);

for i=1:3
  Ei = make_E(i,dims);
  for j=1:3
```

```
    Ej = make_E(j,dims);
    for k=1:3
      Ek = make_E(k,dims);
      K3hat = K3hat + (1/n)*Ek*l3{i,j,k}*kron(Ei',Ej');
    end;
  end;
end;


%////////////////////////////////////////////////////////////////////
%
%  Calculate asymptotic quantities:
%
%    The sandwich estimator
%    Expectation of delta0, delta1
%
%////////////////////////////////////////////////////////////////////
%
% Sandwich estimator
%

% fbeta fmu fpsi fep are so that the meat part of the sandwich
% estimator can be found using a matrix multiplication, rather than a
% sum of stuff.  Note that diag(w1) = n*W1
temp = khatrirao(ZSiginv',ZSiginv'*n*W1);
fbeta = (2/sqrt(n))*khatrirao(ZSiginv',X'*n*W1);
fmu = (1/sqrt(n))*F1{2}'*temp;
fpsi = (-1/n^1.5)*first*kron(F1{3}',dw1{3})*temp;
fep = (1/sqrt(n))*(rho(0,dk,c0) - b0);
F = [ fbeta; fmu; fpsi; fep' ];

%
% The following is K2hat = E(l2)
%   zero out parts that have expectation zero.
dumtwo = size(l2{2,1},2) + size(l2{3,1},2) + size(l2{4,1},2);
K2hat = (1/n)*[
    l2{1,1} zeros(size(l2{2,1},1),dumtwo);
    zeros(size(l2{1,2})) l2{2,2} l2{3,2} l2{4,2};
    zeros(size(l2{1,3})) l2{2,3} l2{3,3} l2{4,3};
    zeros(size(l2{1,4})) zeros(size(l2{2,4})) l2{3,4} l2{4,4}];

iK2hat = inv(K2hat);
%
% Zero out the parts of Cmeat that have expectation zero.
%
Cmeat = F*F';
```

```
nna = size(fbeta,1);
Cmeat(1:nna,nna+1:end) = zeros(nna,sum(nu)-nna);
Cmeat(nna+1:end,1:nna) = zeros(sum(nu)-nna,nna);

sandwich_psi_ep = iK2hat*Cmeat*iK2hat';

% Change over to phi
change = [ speye(nu(1)) sparse(nu(1),nu(2)+nu(3));
           sparse(nu(2),nu(1)) speye(nu(2)) sparse(nu(2),nu(3));
           sparse(nu(3),nu(1)+nu(2)) V1 V2 ];
sandwich = change*sandwich_psi_ep*change';

d2w=l2;
d3w=l3;

%////////////////////////////////////////////////////////////////////
%
%  Expectation of Z2*delta0
%     EZD stands for Expectation of Z2 D0
%
%////////////////////////////////////////////////////////////////////

GG = inv(K2hat)*F;                     % inv(K2)*f_k (cols in GG)
G{1} = GG(1:dims(1),:);                % G for beta
G{2} = GG(dims(1)+1:sum(dims(1:2)),:);  % G for mu
G{3} = GG(sum(dims(1:2))+1:sum(dims(1:3)),:);    % G for psi
G{4} = GG(sum(dims(1:3))+1:end,:);      % G for psi

for ii=1:4;
  for jj=1:4;
    EZD{ii,jj} = zeros(dims(jj),1);
  end
end

for j=1:n
  zj = Z(j,:)';
  xj = X(j,:)';

% Wxx_j = xj*W1(j,j)*xj';
  Wxz_j = xj*W1(j,j)*zj';
  Wzz_j = zj*W1(j,j)*zj';
  SWzzjS = Siginv*Wzz_j*Siginv;

  W2Kzz_j = W2(j,j)*Kzz(:,j)';
  W2Kzx_j = W2(j,j)*Kzx(:,j)';
```

```
  Vzz_j = Kzz(:,j)*W2Kzz_j;
  Vxz_j = Kzx(:,j)*W2Kzz_j;
  Vxx_j = Kzx(:,j)*W2Kzx_j;

% Mxx_j = kron(Siginv,Wxx_j) + 2*Vxx_j;
  Mxz_j = kron(Siginv,Wxz_j*Siginv) + Vxz_j;
  Mzz_j = kron(Siginv,SWzzjS) + .5*Vzz_j;

% First Derivatives
  dq1_j{1} = -2*vec(Wxz_j*Siginv);
  dq1_j{2} = -F1{2}'*vec(SWzzjS);
  dq1_j{3} = -F1{3}'*vec(SWzzjS);
  a_j = dq1_j{3};

% Second derivatives
% d2q11_j{1,1} = 2*Mxx_j;
% d2q11_j{2,1} = 2*Mxz_j*F1{2};
  d2q11_j{3,1} = 2*Mxz_j*F1{3};
% d2q11_j{1,2} = d2q11_j{2,1}';
  d2q11_j{1,3} = d2q11_j{3,1}';
  for ii=2:3
    for jj=2:3
      d2q11_j{jj,ii} = -F11{ii,jj}'*kron(Inu{jj},vec(SWzzjS)) + ...
                         2*F1{ii}'*Mzz_j*F1{jj};
    end;
  end;

  % Add parts that only depend on j.  Separate for easier error
  % checking:  Parts that depend on k will be added in next loop

  % all with beta at end have expectation of zero.
  % Beta' beta
%  EZD{1,1} = EZD{1,1} + d2q11_j{1,1}*G{1}(:,j);    %(1) (Nj+Nj)*Zg

  % mu' beta
%  EZD{2,1} = EZD{2,1} + d2q11_j{2,1}*G{2}(:,j);

  % psi' beta
%  EZD{3,1} = EZD{3,1} + d2q11_j{3,1}*V1*G{3}(:,j);

  % eta_psi' beta
%  EZD{4,1} = EZD{4,1} + d2q11_j{3,1}*V2*G{4}(:,j);

  % This one also has expectation of zero.
  % beta' mu
%  EZD{1,2} = EZD{1,2} + d2q11_j{1,2}*G{1}(:,j);
```

```
  % mu' mu
  EZD{2,2} = EZD{2,2} + d2q11_j{2,2}*G{2}(:,j);

  % psi' mu
  EZD{3,2} = EZD{3,2} + d2q11_j{3,2}*V1*G{3}(:,j);

  % eta_psi' mu
  EZD{4,2} = EZD{4,2} + d2q11_j{3,2}*V2*G{4}(:,j);

  firbet = -1/n*first;
  % beta' psi
  EZD{1,3} = EZD{1,3} + firbet*kron(d2q11_j{1,3},dw1{3})*G{1}(:,j);

  % mu' psi
  EZD{2,3} = EZD{2,3} + firbet*kron(d2q11_j{2,3},dw1{3})*G{2}(:,j);

  common_j = -1/n*first*(kron(d2q11_j{3,3},dw1{3}) + ...
                         kron(a_j,d2w11{3,3}));
  % psi' psi
  EZD{3,3} = EZD{3,3} + common_j*V1*G{3}(:,j);

  % eta_psi' psi
  EZD{4,3} = EZD{4,3} + common_j*V2*G{4}(:,j);

  % beta' psi
  EZD{1,4} = EZD{1,4} + -dq1_j{1}'*G{1}(:,j);

  % mu' psi
  EZD{2,4} = EZD{2,4} + -dq1_j{2}'*G{2}(:,j);

  % psi' psi
  EZD{3,4} = EZD{3,4} + -a_j'*V1*G{3}(:,j);

  % eta_psi' psi
  EZD{4,4} = EZD{4,4} + -a_j'*V2*G{4}(:,j);
end

for i=1:4
  SumEZD{i} = zeros(size(EZD{1,i}));
end

for i=1:4
  SumEZD{1} = SumEZD{1} + EZD{i,1}; %This should be zero.
  SumEZD{2} = SumEZD{2} + EZD{i,2};
  SumEZD{3} = SumEZD{3} + EZD{i,3};
```

```
  SumEZD{4} = SumEZD{4} + EZD{i,4};
end

EZ2D0 = sqrt(n)*[ SumEZD{1}; SumEZD{2}; SumEZD{3}; SumEZD{4}];

ED1 = -inv(K2hat)*(EZ2D0 + .5*K3hat*vec(sandwich));
```

---

exprho.m.

```
function erhodk = exprho(p,v,b0,c0,k);
%
% EXPRHO   Expected value of the integral of tukey's bi-weight
%          function assuming an underlying multivariate-t
%          distribution
%
%   Needed for the computation of kappa when the underlying
%   distribution is actually multivariate-t, but assumed to
%   something else.
%
%   Inputs
%     p = dimension of Sigma
%     v = degrees of freedom of the multivariate t
%     b0 = S-estimator constant
%     c0 = S-estimator constant
%     lpt = initial left point
%     rpt = initial right point
%
%   Outputs
%     zero = value of kappa
%
%

c2 = c0^2;

erhodk = 1/(2*k)*exphk(1,p,v,k*c2) - ...
         1/(2*k^2*c2)*exphk(2,p,v,k*c2) + ...
         1/(6*k^3*c0^4)*exphk(3,p,v,k*c2) + ...
         c2/6*(1-exphk(0,p,v,k*c2))-b0;

%-----------------------------------------------------------------------

function ehk = exphk(a,p,v,c);
  front = exp(a*log(v) + gammaln(a+p/2) + gammaln(v/2-a) - ...
```

```
               gammaln(p/2) - gammaln(v/2));
  ehk = front*betacdf(c/(v+c),a+p/2,v/2-a);
return;
```

---

generate_data.m.

```
%======================================================================
%
% generate_data.m
%
%  This routine generates a data set following a multivariate T
%  distribution and contaminated with a few points from another
%  multivariate T distribution with a different location, but the same
%  variance
%
% Generate the data according to the value of kinddata.
%    if kinddata =
%       1 =  a mixture of two multivariate T distributions.
%       2 =  a mixtures of three multivariate T distributions.
%       3 =  a multivariate T distribution.
%
%======================================================================

kinddata = 3;

%---------------------------------------------------
% Generate random multivariate T distribution data.
%---------------------------------------------------
Y = (sqrt(v)*randn(n,p)*Sigmah)./sqrt(chi2rnd(v,n,1)*ones(1,p))+X*B_t;

if kinddata == 1,
  %-------------------------------------------------------------------
  % Add a number of data points with location moved, but variance
  % same in each case
  %-------------------------------------------------------------------
  % A mixture of two distributions with the location shifted in one.
  %-------------------------------------------------------------------
  points = 1;
  J = sample_wor(points,n);
  Y(J,:) = Y(J,:) + 4*X(J,:)*B_t;

elseif kinddata == 2,
  %------------------------------------------------------------------
  % A mixture of three distributions with the location shifted in
```

```
  %  two of them.
  %---------------------------------------------------------------
  points1 = 5;
  points2 = 5;
  J = sample_wor(points1+points2,n);
  J1 = J(1:points1);
  J2 = J(points1+1:end);
  Y(J1,:) = Y(J1,:) +  4*X(J1,:)*B_t;
  Y(J2,:) = Y(J2,:) + 40*X(J2,:)*B_t;
end
```

---

initial_guess.m.

```
function [V1,V2,psi] = initial_guess(T1,T2,Z,b0,c0,Gam,beta,phi);
%===================================================================
%
%  initial_guess
%
%   [V1,V2,psi] = initial_guess(T1,T2,Z,b0,c0,Gam,beta,phi) computes
%   initial guesses for V1, V2, and psi.
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     Z    = Centered Data Matrix (Y-XB)
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     Gam  = Current guess for Eigenvector matrix
%     beta = Current guess for the location parameters
%     phi  = Current value for phi.
%
%   Outputs
%     V1  = initial guess for Implicit model matrix
%     V2  = initial guess for Implicit model matrix
%     psi = initial guess for psi
%
%
%====================================================================

%--------------------------------------------------------------------
%
% This computes the initial guesses for V1, V2, and psi
%
%--------------------------------------------------------------------
```

```
[Siginv,Laminv,D1_lam] = model_lam(T1,T2,1,1,Gam,phi);
[q,dk] = qeval(Z,Siginv,b0,c0);                              %get dk
a = compute_a(b0,c0,Z,Gam,dk,Laminv,D1_lam);

%----------------------------------------------------------------------
% Compute V1 and V2.  Let V2 be in R(a).  That way a'*V2 != 0.
%   and get initial guess for psi.
%----------------------------------------------------------------------
V2 = a/norm(a);
V1 = null(V2');
psi = (V1'*V1)\(V1'*phi);
```

<u>kappa.m</u>.

```
function zero = kappa(p,v,b0,c0,lpt,rpt);
%
% This solves for kappa given values for p, v, b0, c0, given that the
% root is bound between lpt and rpt.
%
%   Inputs
%     p = dimension of Sigma
%     v = degrees of freedom of the multivariate t
%     b0 = S-estimator constant
%     c0 = S-estimator constant
%     lpt = initial left point
%     rpt = initial right point
%
%   Outputs
%     zero = value of kappa
%

i=1;
n0 = 1000;
tol = 1e-14;
fa = exprho(p,v,b0,c0,lpt);
step = 1;

while (i < n0) & (step > tol);
  mid = (lpt+rpt)/2;
  fmid = exprho(p,v,b0,c0,mid);
  i = i + 1;
  if fa*fmid > 0,
    lpt = mid;
  else
```

```
    rpt = mid;
  end
  step = abs(((lpt-rpt)/2)/mid);
end

zero = mid;
```

---

make_E.m.

```
function E=make_E(j,m);
% MAKE_E   Make an elementary matrix.
%
%   Constructs an elementary matrix needed to put a matrix derivative
%   together.  This matrix is the transpose of E in equation 14 of
%   (Boik 2002).
%
a = sum(m(1:j-1));
b = sum(m(j+1:end));
r = m(j);
E = [ sparse(a,r); speye(r); sparse(b,r) ];
```

---

make_F11_mumu.m.

```
function F11_mumu = make_F11_mumu(tNp,A2,Dp,Dg,Lam,Gam);
%===================================================================
%
%  make_F11_mumu
%
%   F11_mumu = make_F11_mumu(tNp,A2,Dp,Dg,Lam,Gam) computes the F11
%   matrix d^2 vec Sig/(d\mu' \kron d\mu)
%
%   Inputs
%     tNp      = 2*Np
%     A2       = A2 matrix
%     Dp       = Duplication matrix of order p
%     Dg       = Derivative of G wrt mu
%     Lam      = Current eigenvalue matrix
%     Gam      = Current eigenvector matrix
%
%   Outputs
%     F11_mumu = deriv of vecSigma wrt mu' x mu
```

```
%
%==============================================================

% ============================================================
% Computation involving K3 and K4 are quite fast for Matlab 6.1 since
% both are very sparse.  If instead, vec(Sig) were in the middle of
% the triple kronecker or Sig instead of Lam, then it would be better
% to just form the triple kroenecker instead of multiplying.
% ============================================================

p = size(Gam,1);
nu2 = size(Dg,2);
Ip = speye(p);
Ip2 = speye(p^2);

kGamGam = sparse(kron(Gam,Gam));
kDgDg = kron(Dg,Dg);
K3 = kron3(Ip,(vec(Ip))',Ip);
K4 = kron3(Ip,Lam,Ip2);
Dg2part = A2*Dp'*K3;

%
% Compute F11_mumu
%   Since vec(F11_mumu) = vec(F2_mumu), then
%
F2_mumu = (tNp*kGamGam)*((kron(Lam,Ip)*Dg2part-K3*K4)*kDgDg);% mu' mu'
F11_mumu = dvec(vec(F2_mumu),p^2*nu2,nu2);                   % mu mu'
```

---

make_F11_phiphi.m.

```
function F11_phiphi = make_F11_phiphi(T1,T2,Gam,phi);
%==============================================================
%
%  make_F11_phiphi
%
%   F11_phiphi = make_F11_phiphi(T1,T2,Gam,phi) computes the
%   F11_phiphi matrix d^2 vec Sig/(d\phi' \kron d\phi)
%
%   Inputs
%     T1  = Design matrix (models linear relations btwn eigvals)
%     T2  = Design matrix (models exponential relations btwn eigvals)
%     Gam = Current eigenvector matrix
%     phi = Current value of phi
%
```

```
%   Outputs
%     F11_phiphi = deriv of vecSigma wrt phi' x phi
%
%======================================================================

[p,q1] = size(T1);
nu3 = size(T2,2);

% Grab D2lam from model_lam
[A,B,C,D,D2lam,E] = model_lam(T1,T2,1,1,Gam,phi);
F2_phiphi = sparse(khatrirao(Gam,Gam))*D2lam;

%
%Get F11_phiphi using the relation vec(F11phi) = vec(F2phi)
%
F11_phiphi = dvec(vec(F2_phiphi),p^2*nu3,nu3);    % phi phi'
```

---

model_lam.m.

```
function [Siginv,Laminv,D1lam,lam,D2lam,D3lam] = model_lam(T1,T2,...
                                                   V1,V2,Gam,psi,ep);
%======================================================================
%
%  model_lam
%
%   [Siginv,Laminv,D1lam,lam,D2lam,D3lam] = model_lam(T1,T2,V1,...
%                                                      V2,Gam,psi,ep)
%   models lam, D1lam, Laminv, Siginv, given a particular psi and
%   eta_psi (ep) or just phi.
%
%   If eta_psi (ep) is not passed to the function, then the function
%   assumes that psi is actually phi.  Otherwise it assumes that psi
%   and eta_psi (ep) are given.
%
%   D2lam and D3lam are not computed unless they are specified as
%   return arguments.
%
%   Inputs
%     T1  = Design matrix (models linear relations btwn eigvals)
%     T2  = Design matrix (models exponential relations btwn eigvals)
%     V1  = Implicit model matrix
%     V2  = Implicit model matrix
%     Gam = Current estimate for Gamma
%     psi = Current estimate for psi
```

```
%     ep  = Current estimate for eta_psi
%
%   Outputs
%     Siginv  = New estimate of the inverse of Sigma
%     Laminv  = New estimate of the inverse of eigenvalue matrix
%     D1lam   = New derivative of lambda wrt phi
%     lam     = New estimate for lambda
%     D2lam   = New second derivative of lambda wrt phi
%     D3lam   = New third derivative of lambda wrt phi
%
%===================================================================

if nargin == 7,
  phi = V1*psi + V2*ep;
else
  phi = psi;
end


% Derivatives of lambda with respect to phi
%
%-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-
%  All derivatives below assume Model 2:
%
%   lam = T1*exp(T2*phi)
expphi = exp(T2*phi);
lam = T1*expphi;
Ddum  = khatrirao(expphi',T1);
D1lam = sparse(Ddum*T2);

if nargout == 6,
  Ddum2 = sparse(khatrirao(T2',T2'));
  D2lam = sparse(Ddum*Ddum2');
  D3lam = sparse(Ddum*khatrirao(T2',Ddum2)');
end
%-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-

% Form Siginv(ep) and Laminv_new(ep)
Laminv = spdiag(1./lam);
Siginv = Gam*Laminv*Gam';
```

---

newguess.m.

```
%
```

```
% Get a new subsample (a new guess)
%
J = sample_wor(n/2,n);
subsample = Y(J,:);
```

---

qeval.m.

```
function [q,dk] = qeval(Z,Siginv,b0,c0);
%======================================================================
%
%  qeval    Evaluate the constraint.
%
%   [q,dk] = qeval(Z,Siginv,b0,c0) Computes the current value of the
%   function q, along with dk.
%
%   Inputs
%     Z      = Centered Data Matrix (Y-XB)
%     Siginv = Current estimate of the inverse of Sigma
%     b0     = Constant associated with constraint q
%     c0     = Constant associated with breakdown point
%
%   Outputs
%     q      = the function q, evaluated at dk
%     dk     = mahalanobis distances of each data point
%
%======================================================================
%
%   This computes sqrt(diag(Z*Siginv*Z')) without computing the off
%   diagonal elements.  A simple for loop was sufficiently fast for
%   small n, but was not as fast for large n.  Here are the timings
%   (in seconds) for 100 iterations:
%
%                                         code
%             n     loop      diag      below
%            ---   ------    ------    ------
%              2   0.1200    0.0800    0.0900
%              4   0.1100    0.0900    0.0900
%              8   0.1400    0.0800    0.0900
%             16   0.2000    0.0900    0.0900
%             32   0.3100    0.0900    0.1000
%             64   0.5700    0.1200    0.1000
%            128   1.0800    0.1700    0.1200
%            256   2.1700    0.4400    0.1500
```

```
%           512   4.2500    1.1800    0.2200
%          1024   8.8800    5.5300    0.3400
%          2048  18.9900   30.2500    0.6000
%          4096  44.0200  118.2100    1.1100
%
%=====================================================================

dk = sqrt(sum(Z*Siginv.*Z,2));         % fast sqrt(diag(Z*Siginv*Z'))
q = sum(rho(0,dk,c0))/size(Z,1) - b0;
```

---

rho.m.

```
function p = rho(n,t,c0)
%=====================================================================
%
%  rho     integral of Tukey's biweight function
%
%   p = rho(n,t,c0) returns the rho function and its derivatives, plus
%   the first derivative divided by t, depending on the value given
%   for n.  In this file, the rho function is the integral of Tukey's
%   biweight function.
%
%   Inputs
%     n  =  Determines which derivative (or function) to use.
%     t  =  Data vector
%     c0 =  Constant associated with the breakdown point.
%
%     Choices for n include
%       0.  rho function,
%       1.  derivative of the rho function,
%       1.1 derivative of the rho function divided by t,
%       2.  second derivative of the rho function,
%       3.  third derivative of the rho function.
%
%   Output
%     p  =  The value of the function determined by n evaluated at t.
%
%=====================================================================

% First fill everything with zeros
p = zeros(size(t));

if n == 0,
  w = abs(t);
```

```
  % Fill in spots in p where w<c0
  wlc = find(w<c0);
  tlc = t(wlc);
  tc = (tlc./c0).^2;
  p(wlc) = tlc.^2/2.*(1 - tc + tc.^2/3);
  cc = c0^2/6;
  p(find(w>=c0)) = cc;

else
  %Fill in spots where p is not zero (when t is in the interval[-c0,c0])
  notzero = find(t>=-c0 & t<=c0);

  switch n
   case 1,
     p(notzero) = t(notzero).*(1-(t(notzero)/c0).^2).^2;
   case 1.1,
     p(notzero) = (1-(t(notzero)/c0).^2).^2;
   case 2,
     tcnz = (t(notzero)/c0).^2;
     p(notzero) = (1-tcnz).*(1-5*tcnz);
   case 3,
     tcnz = (t(notzero)/c0).^2;
     p(notzero) = (-4*t(notzero)/c0^2).*(3-5*tcnz);
   otherwise
     error('n not valid');
  end;
end;
```

---

robust_load.m.

```
%-----------------------------------------------------------------------
% robust_load.m
%     Loads all constant matices and the true parameters for the
%     model in preparation to running generate_data.m  Used for
%     simulations.
%
%-----------------------------------------------------------------------


m = [ 2 1 1 ]';               %Multiplicity vector.
p = sum(m);                   %Dimension of the problem (size of Sigma).
d = 1;                        %Number of cols in X.
v = 13;                       %df for chi-square.
```

```
n = 200;                    %sample size.
[Dg,A1,A2,L,Dp]= constantmatrices(m); %Constant matrices.

if s_estim==1,
  bdp = .25;                %Breakdown point (Used for s-estimators)
  [b0,c0] = solve_c0(p,bdp);
end

%-----------------------------------------------------------------------
% M is a matrix that will duplicate rows of T1 according to the
% specified multiplicity vector m.  avg is the moore penrose
% inverse of M.  M*inv(M'*M)*M'*lam = lam with correct multiplicities.
%
M=[];
avgi=[];
for i=1:length(m);
  M = [ M repmat(i,1,m(i)) ];
  avgi = [ avgi repmat(1/m(i),1,m(i)) ];
end;
M = sparse(dummyvar(M));%Simple multiplicity structure.
avg = (spdiag(avgi)*M)';%inv(M'*M)*M' (Used to calculate estimates
                        %  for eigenvalues.  It averages eigenvalues)
%-----------------------------------------------------------------------
% Model 2 for eigenvalues lam = T1*exp(T2*phi)
%
% Construct T1 to order the distinct eigenvalues (Assumes for the
% simulations that nu(3) is the same as the number of distinct
% eigenvalues)
nu = [ p*d; (p^2-m'*m)/2; length(m)];  %For nu(3) = # of distinct
                                       %eigenvalues

T1 = M*fliplr(triu(ones(nu(3),nu(3))));
q1 = size(T1,2);
T2 = speye(q1);


%-----------------------------------------------------------------------
% Calculate true parameters below:
% ============================================
% Calculate true values for the eigenvalues
% ============================================
lam_t = [ 30 10 1 ]';        %true values for eigenvalues.
phi_t = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*M*lam_t); %(Under Model 2)

% =================================================================
% Generate true values for the eigenvectors and covariance matrix.
```

```
% ================================================================
%G = randn(p,p);             %Generalized eigenvectors.
G = eye(p);                  %Diagonal Sigma.
G = orth(G);                 %Orthogonalize the eigenvectors.
Sigma_t = G*diag(M*lam_t)*G';%true values for Sigma.
Sigmah = sqrtm(Sigma_t);     %Symmetric square root.

% ================================================
% Generate true values for location parameters and
%    generate a design matrix
% ================================================
%X = unidrnd(4,n,d);         %Design matrix.
%B_t = unidrnd(4,d,p);       %true values of the location parameters.
%B_t = diag(1:d)*ones(d,p);
X = ones(n,1);               %means model
B_t = 1:p;                   %means model

d = size(X,1);

%--------------------------------------------------------------------
```

sample_wor.m.

```
function Sample = sample_wor(ss,ps);
% SAMPLE_WOR    Sample without replacement
%
%   Sample = samplewor(ss,ps) selects a sample without
%   replacement. Written by R. J. Boik
%
%   Inputs
%     ss = sample size
%     ps = population size
%
%   Outputs
%     Sample = random sample of size ss from a population of ss
%              (without replacement)
%
clear Sample;
r = [1:ps]';
out = [];
for j = 1:ss;
  kk = unidrnd(length(r));
  out = [out; r(kk)];
  r(kk) = [];
```

```
end;
Sample = sort(out);
return
```

---

sandwich_sim.m.

```
%-----------------------------------------------------
%
% Find the sandwich estimator for each data set in
%   the simulation and the estimator of the bias...
%
%-----------------------------------------------------

dwvec = zeros(nsamp,nu(1)+nu(2)+nu(3)-1);
detsandwich = zeros(nsamp,1);

if onlyindex == 1,
  endindex = length(indx);
else
  endindex = nsamp;
end


for ii=1:endindex
  if onlyindex == 1,
    iii = indx(ii);
  else
    iii = ii;
  end
  [dw,dw1,d2w,d3w,K2{iii},FF{iii},sand{iii},EZ2D0{iii},ED1{iii}] = ...
     deriv_w(T1,T2,A1,A2,Dg,YY{iii},X,b0,c0,nu,Gam_sim{iii}, ...
             beta_sim{iii},phi_sim{iii});
 dwvec(iii,:) = [ dw{1}' dw{2}' dw{3}' ];
 detsandwich(iii) = det(sand{iii});
end;

%
% Check to see if a minimizer was actually found
%   If not, run the simulation again.
%
norms = [];
for jj=1:nsamp
  norms = [norms; norm(dwvec(jj,:))];
```

```
end

indx = find(norms/n > 1e-6);

%-----------------------------------------------------------------------
%
% This takes the simulated data finds empirical results which will be
% comapred to theoretical results
%

betavec = zeros(nsamp,nu(1));
muvec = zeros(nsamp,nu(2));
psivec = zeros(nsamp,nu(3)-1);
phivec = zeros(nsamp,nu(3));

for iii=1:nsamp
  betavec(iii,:) = beta_sim{iii}';
  muvec(iii,:) = mu_sim{iii}';
  phivec(iii,:) = phi_sim{iii}';
  psivec(iii,:) = psi_sim{iii}';
end

thetavec = [betavec  muvec  phivec];  %muvec should be zeros.

covarSim = n*(thetavec'*thetavec - thetavec'*ppo(ones(nsamp,1))* ...
            thetavec)/(nsamp-1);

%-----------------------------------------------------------------------
%
% Find the average sandwich estimator.
%
%-----------------------------------------------------------------------

sandavg = zeros(size(covarSim));
Ed1 = zeros(size(ED1{1}));

for ii=1:nsamp
  sandavg = sandavg + sand{ii};
  Ed1 = Ed1 + ED1{ii};
end
sandavg = sandavg/nsamp;
Ed1 = Ed1/nsamp;


kap = kappa(p,v,b0,c0,.1,20);
Sigstar = kap*Sigma_t;
```

```
[G0,L0,dum] = svd(Sigstar);
kaplam_t =  M*avg*diag(sparse(L0));
kapphi_t = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*kaplam_t);

thetavec_t = [ vec(B_t); zeros(nu(2),1); kapphi_t ];

emp_bias = (mean(thetavec)' - thetavec_t);
theo_bias = 1/n*Ed1;

sandavg
covarSim

emp_bias
theo_bias
```

---

satisfy_constraint.m.

```
function [a,V1,V2,phi,psi,eta_psi,lam,D1_lam,dk,W1,Wzz]= ...
    satisfy_constraint(T1,T2,V1,V2,Z,b0,c0,Gam,beta,phi,psi);
%======================================================================
%
%  satisfy_constraint
%
%    [a,V1,V2,phi,psi,eta_psi,lam,D1_lam,dk,W1,Wzz]= ...
%    satisfy_constraint(T1,T2,V1,V2,Z,b0,c0,Gam,beta,phi,psi) takes a
%    given value of psi, solves for eta_psi, and updates all values
%    returned.
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     V1   = Implicit model matrix
%     V2   = Implicit model matrix
%     Z    = Centered Data Matrix (Y-XB)
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     Gam  = Current guess for Eigenvector matrix
%     beta = Current guess for the location parameters
%     phi  = Current value for phi.
%     psi  = Current value for psi.
%
%   Outputs
%     a       = current value for dq/dphi
%     V1      = estimate for Implicit model matrix
```

```
%     V2      = estimate for Implicit model matrix
%     phi     = estimate for phi (That satisfies the constraint)
%     psi     = estimate for psi (That satisfies the constraint)
%     eta_psi = estimate for eta_psi (That satisfies the constraint)
%     lam     = estimate for lambda
%     D1_lam  = derivative of phi wrt lambda.
%     dk      = current value for dk's
%     W1      = current value for (1/n)*diag(rho(1)(dk)/dk)
%     Wzz     = Z'*W1*Z
%
%======================================================================

eta_psi = solve_eta_psi(T1,T2,V1,V2,Gam,b0,c0,Z,phi,psi);
phi = sparse(V1*psi + V2*eta_psi);  %phi that satisfies the constraint

%----------------------------------------------------------------------
% Now update a, psi, eta_psi, V1, and V2 for this particular phi.
%----------------------------------------------------------------------
[Siginv,Laminv,D1_lam,lam] = model_lam(T1,T2,V1,V2,Gam,psi,eta_psi);
[q,dk] = qeval(Z,Siginv,b0,c0);
[a,W1,Wzz] = compute_a(b0,c0,Z,Gam,dk,Laminv,D1_lam);

%----------------------------------------------------------------------
% Recompute V1 and V2 so that V2 is the unit vector in the same
%   direction of a.  Also V1'*V2 = 0 and V1'*V1 = I.
%   Recompute psi and eta_psi using the new V1 and V2;
%----------------------------------------------------------------------
V2 = a/norm(a);
V1 = null(V2');
psi = sparse((V1'*V1)\(V1'*phi));
eta_psi = (V2'*V2)\(V2'*phi);  %solve_eta_psi is only used when phi
                               %does not satisfy the constraint.
```

---

simulation.m.

```
%
% Perform a simulation.
%
% This generates nsamp data sets, and finds the S-estimators.  It
% saves the data in the _sim variables below.
%
% To solve for the $S$-estimators for particular data sets, specify
% their indicies in the vector indx.  Further, specify onlyindex = 1
% to use this.
```

```
%

nsamp = 1000;

if onlyindex == 1,
  endindex = length(indx);
else
  endindex = nsamp;
end


for iiii=1:endindex
  if onlyindex == 1,
    ii = indx(iiii);
    Y = YY{ii};
  else
    ii = iiii;
    generate_data
  end

  solve_s_estimate_sim

  % Keep all these
  YY{ii} = Y;
  S_sim{ii} = S_min;
  Gam_sim{ii} = Gam_min;
  lam_sim{ii} = lam_min;
  beta_sim{ii} = beta_min;
  mu_sim{ii} = mu_min;
  psi_sim{ii} = psi_min;
  phi_min = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam_min);
  phi_sim{ii} = phi_min;
end;
```

Smultivar.m.

```
function res= Smultivar(x,nsamp,bdp);

%--------------------------------------------------------------------
%
% Smultivar   S-estimates of multivariate location and scatter
%
%   res = Smultivar(x,nsamp,bdp) computes biweight S-estimator for
```

```
%   location/scatter with algorithm of Ruppert.  Recevied by
%   S.K. Hyde from Christophe Croux.
%
%   Inputs
%     x     :a data matrix. Rows of the matrix represent observations,
%            columns represent variables.
%     nsamp :The number of random p-subsets considered.
%     bdp   :Breakdon value of the S-estimator.
%
%   Outputs
%     mean  :vector of estimates for the center of the data.
%     cov   :matrix of estimates for the scatter of the data.
%     mah   :vector of robust distances versus the mean and covariance.
%     scale :distance scale estimate.
%
%------------------------------------------------------------------

tol=1e-5;
seed=now;
s=10e10;
[n,p]=size(x);
[c,k]=multivarpar(p,bdp);
la=1;

for loop=1:nsamp
    fprintf('Iter= %5.0f\b\b\b\b\b\b\b\b\b\b\b',loop);
    [ranset,seed]=randomset(n,p+1,seed);
    xj=x(ranset,:);
    mu=mean(xj);
    xjcenter=xj-repmat(mu,p+1,1);
    cov=(xjcenter'*xjcenter)/(p+1);
    determ=det(cov);
    if determ > 1e-15
       if determ^(1/p)> 1e-5
          cov=(determ^(-1/p)).*cov;
          if loop > ceil(nsamp/5)
             if loop==ceil(nsamp/2)
                la=2;
             end
             if loop==ceil(nsamp*(0.8))
                la=4;
             end
          [random,seed]=uniran(seed);
          random=random^la;
          mu=random*mu+(1-random)*muopt;
          cov=random*cov+(1-random)*covopt;
```

```
      end
      determ=det(cov);
      cov=(determ^(-1/p)).*cov;
      md=mahalanobis(x,mu,cov,n,p);
      md=md.^(1/2);
      if mean(rhobiweight(md/s,c))<k
         if s<5e10
            s=sestck(md,s,c,k,tol);
         else
            s=sestck(md,0,c,k,tol);
         end
         muopt=mu;
         covopt=cov;
         mdopt=md;
         psi=psibiweight(md,s*c);
         u=psi./md;
         ubig=repmat(u',1,p);
         mu=mean(ubig.*x)./mean(u);
         xcenter=x-repmat(mu,n,1);
         cov=((ubig.*xcenter)'*xcenter);
         cov=(det(cov)^(-1/p)).*cov;
         okay=0;
         jj=1;
         while ((jj<3)&(okay~=1))
               jj=jj+1;
               md=mahalanobis(x,mu,cov,n,p);
               md=md.^(1/2);
               if mean(rhobiweight(md/s,c))<k
                  muopt=mu;
                  covopt=cov;
                  mdopt=md;
                  okay=1;
                  if s<5e10
                     s=sestck(md,s,c,k,tol);
                  else
                     s=sestck(md,0,c,k,tol);
                  end
               else
                  mu=(mu+muopt)/2;
                  cov=(cov+covopt)./2;
                  cov=(determ^(-1/p)).*cov;
               end
         end
      end
   end
end
```

```
end
covopt=s^2*covopt;
mdopt=mdopt/(s^2);
res.mean=muopt;
res.cov=covopt;
res.mah=mdopt;
res.scale=s;

% --------------------------------------------------------------------

function rho=rhobiweight(x,c)

% Computes Tukey's biweight rho function met constante c voor alle
% waarden in de vector x.

hulp=x.^2/2-x.^4/(2*c^2)+x.^6/(6*c^4);
rho=hulp.*(abs(x)<c)+c^2/6.*(abs(x)>=c);

% --------------------------------------------------------------------

function psi=psibiweight(x,c)

% Computes Tukey's biweight psi function met constante c voor alle
% waarden in de vector x.

hulp=x-2.*x.^3/(c^2)+x.^5/(c^4);
psi=hulp.*(abs(x)<c);

% --------------------------------------------------------------------

function scale=sestck(x,start,c,k,tol)

% Computes Tukey's biweight objectief function (schaal)
% corresponderend met de mahalanobis distances x.

if start>0
   s=start;
else
   s=madnoc(x);
end
crit=2*tol;
rhoold=mean(rhobiweight(x/s,c))-k;
while crit>=tol
      delta=rhoold/mean(psibiweight(x/s,c).*(x/(s^2)));
      isqu=1;
      okay=0;
```

```
      while ((isqu<10) & (okay~=1))
            rhonew=mean(rhobiweight(x/(s+delta),c))-k;
            if abs(rhonew) < abs(rhoold)
               s=s+delta;
               okay=1;
            else
               delta=delta/2;
               isqu=isqu+1;
            end
      end
      if isqu==10;
         crit=0;
      else
         crit=(abs(rhoold)-abs(rhonew))/(max([abs(rhonew),tol]));
      end
      rhoold=rhonew;
end
scale=abs(s);

% --------------------------------------------------------------------
function mad=madnoc(y)

y=abs(y);
[n,p]=size(y);
if ((n==1) & (p>1))
   y=y';
   [n,p]=size(y);
end
if floor(n/2)==n/2
   odd=0;
else
   odd=1;
end
y=sort(y);
if odd==1
   mad=y((n+1)/2,:);
else
   mad=(y(n/2,:)+y(n/2+1,:))/2;
end
mad=mad/0.6745;



% --------------------------------------------------------------------

function mah=mahalanobis(dat,meanvct,covmat,n,p)
```

```
% Computes the mahalanobis distances.

for k=1:p
   d=covmat(k,k);
   covmat(k,:)=covmat(k,:)/d;
   rows=setdiff(1:p,k);
   b=covmat(rows,k);
   covmat(rows,:)=covmat(rows,:)-b*covmat(k,:);
   covmat(rows,k)=-b/d;
   covmat(k,k)=1/d;
end

hlp=dat-repmat(meanvct,n,1);
mah=sum(hlp*covmat.*hlp,2)';

% --------------------------------------------------------------------

function [random,seed]=uniran(seed)

% The random generator.

seed=floor(seed*5761)+999;
quot=floor(seed/65536);
seed=floor(seed)-floor(quot*65536);
random=seed/65536.D0;

% --------------------------------------------------------------------


function [ranset,seed] = randomset(tot,nel,seed)

for j = 1:nel
   [random,seed]=uniran(seed);
   num=floor(random*tot)+1;
   if j > 1
      while any(ranset==num)
         [random,seed]=uniran(seed);
         num=floor(random*tot)+1;
      end
   end
   ranset(j)=num;
end

% --------------------------------------------------------------------
```

```
function [c,k]=multivarpar(p,bdp)
if p>51
   error('p too big')
end
if ((bdp~=0.15)&(bdp~=0.25)&(bdp~=0.5))
   error('bdp is not 0.15, 0.25 or 0.5')
end
x1=[   1.0000000,         4.0962619,        0.41948293;
       2.0000000,         5.9814734,        0.89444920;
       3.0000000,         7.3996298,         1.3688621;
       4.0000000,         8.5863179,         1.8431200;
       5.0000000,         9.6276998,         2.3173143;
       6.0000000,         10.566886,         2.7914766;
       7.0000000,         11.429124,         3.2656205;
       8.0000000,         12.230705,         3.7397528;
       9.0000000,         12.982879,         4.2138775;
       10.000000,         13.693790,         4.6879967;
       11.000000,         14.369570,         5.1621121;
       12.000000,         15.014961,         5.6362244;
       13.000000,         15.633727,         6.1103345;
       14.000000,         16.228918,         6.5844428;
       15.000000,         16.803037,         7.0585497;
       16.000000,         17.358176,         7.5326554;
       17.000000,         17.896101,         8.0067601;
       18.000000,         18.418323,         8.4808640;
       19.000000,         18.926139,         8.9549674;
       20.000000,         19.420681,         9.4290700;
       21.000000,         19.902938,         9.9031723;
       22.000000,         20.373782,         10.377274;
       23.000000,         20.833987,         10.851375;
       24.000000,         21.284245,         11.325476;
       25.000000,         21.725172,         11.799577;
       26.000000,         22.157327,         12.273678;
       27.000000,         22.581213,         12.747778;
       28.000000,         22.997287,         13.221878;
       29.000000,         23.405964,         13.695978;
       30.000000,         23.807628,         14.170078;
       31.000000,         24.202628,         14.644178;
       32.000000,         24.591281,         15.118277;
       33.000000,         24.973889,         15.592377;
       34.000000,         25.350722,         16.066476;
       35.000000,         25.722034,         16.540575;
       36.000000,         26.088063,         17.014674;
       37.000000,         26.449026,         17.488774;
       38.000000,         26.805129,         17.962872;
       39.000000,         27.156563,         18.436971;
```

```
        40.000000,          27.503506,          18.911070;
        41.000000,          27.846127,          19.385169;
        42.000000,          28.184583,          19.859268;
        43.000000,          28.519024,          20.333366;
        44.000000,          28.849586,          20.807465;
        45.000000,          29.176405,          21.281564;
        46.000000,          29.499602,          21.755662;
        47.000000,          29.819296,          22.229760;
        48.000000,          30.135600,          22.703859;
        49.000000,          30.448618,          23.177957;
        50.000000,          30.758450,          23.652055;
        51.000000,          31.065193,          24.126154;];
x2=[    1.0000000,         2.9370207,         0.35941939;
        2.0000000,         4.4274441,         0.81676061;
        3.0000000,         5.5280753,          1.2733169;
        4.0000000,         6.4426144,          1.7294679;
        5.0000000,         7.2422688,          2.1854354;
        6.0000000,         7.9618713,          2.6413078;
        7.0000000,         8.6215425,          3.0971245;
        8.0000000,         9.2341642,          3.5529061;
        9.0000000,         9.8085671,          4.0086639;
        10.000000,         10.351124,          4.4644050;
        11.000000,         10.866613,          4.9201340;
        12.000000,         11.358720,          5.3758537;
        13.000000,         11.830369,          5.8315664;
        14.000000,         12.283917,          6.2872737;
        15.000000,         12.721300,          6.7429765;
        16.000000,         13.144134,          7.1986756;
        17.000000,         13.553779,          7.6543716;
        18.000000,         13.951402,          8.1100654;
        19.000000,         14.337999,          8.5657571;
        20.000000,         14.714441,          9.0214468;
        21.000000,         15.081488,          9.4771351;
        22.000000,         15.439811,          9.9328219;
        23.000000,         15.790003,          10.388508;
        24.000000,         16.132596,          10.844193;
        25.000000,         16.468063,          11.299877;
        26.000000,         16.796829,          11.755559;
        27.000000,         17.119283,          12.211242;
        28.000000,         17.435774,          12.666924;
        29.000000,         17.746620,          13.122605;
        30.000000,         18.052117,          13.578286;
        31.000000,         18.352525,          14.033966;
        32.000000,         18.648098,          14.489646;
        33.000000,         18.939056,          14.945325;
        34.000000,         19.225612,          15.401005;
```

```
        35.000000,      19.507958,      15.856683;
        36.000000,      19.786276,      16.312362;
        37.000000,      20.060733,      16.768040;
        38.000000,      20.331484,      17.223718;
        39.000000,      20.598678,      17.679396;
        40.000000,      20.862450,      18.135074;
        41.000000,      21.122927,      18.590751;
        42.000000,      21.380232,      19.046428;
        43.000000,      21.634477,      19.502106;
        44.000000,      21.885767,      19.957782;
        45.000000,      22.134205,      20.413459;
        46.000000,      22.379886,      20.869136;
        47.000000,      22.622899,      21.324813;
        48.000000,      22.863327,      21.780489;
        49.000000,      23.101255,      22.236165;
        50.000000,      23.336757,      22.691841;
        51.000000,      23.569906,      23.147518;];
x3= [   1.0000000,      1.5476490,     0.19960105;
        2.0000000,      2.6608073,     0.58999036;
        3.0000000,      3.4528845,     0.99353327;
        4.0000000,      4.0965660,      1.3984860;
        5.0000000,      4.6520272,      1.8034444;
        6.0000000,      5.1476889,      2.2082229;
        7.0000000,      5.5994636,      2.6128313;
        8.0000000,      6.0172838,      3.0173070;
        9.0000000,      6.4078219,      3.4216813;
        10.000000,      6.7758232,      3.8259800;
        11.000000,      7.1247909,      4.2302194;
        12.000000,      7.4574119,      4.6344139;
        13.000000,      7.7757879,      5.0385717;
        14.000000,      8.0816116,      5.4427012;
        15.000000,      8.3762589,      5.8468067;
        16.000000,      8.6608745,      6.2508929;
        17.000000,      8.9364192,      6.6549629;
        18.000000,      9.2037089,      7.0590192;
        19.000000,      9.4634436,      7.4630635;
        20.000000,      9.7162348,      7.8670992;
        21.000000,      9.9626073,      8.2711260;
        22.000000,      10.203026,      8.6751445;
        23.000000,      10.437908,      9.0791581;
        24.000000,      10.667613,      9.4831645;
        25.000000,      10.892477,      9.8871675;
        26.000000,      11.112784,      10.291164;
        27.000000,      11.328810,      10.695159;
        28.000000,      11.540789,      11.099149;
        29.000000,      11.748943,      11.503136;
```

```
       30.000000,        11.953472,        11.907121;
       31.000000,        12.154556,        12.311102;
       32.000000,        12.352368,        12.715081;
       33.000000,        12.547061,        13.119059;
       34.000000,        12.738776,        13.523034;
       35.000000,        12.927649,        13.927007;
       36.000000,        13.113801,        14.330979;
       37.000000,        13.297347,        14.734950;
       38.000000,        13.478391,        15.138918;
       39.000000,        13.657036,        15.542886;
       40.000000,        13.833375,        15.946852;
       41.000000,        14.007493,        16.350818;
       42.000000,        14.179471,        16.754782;
       43.000000,        14.349388,        17.158745;
       44.000000,        14.517318,        17.562708;
       45.000000,        14.683326,        17.966669;
       46.000000,        14.847478,        18.370630;
       47.000000,        15.009834,        18.774590;
       48.000000,        15.170451,        19.178549;
       49.000000,        15.329387,        19.582508;
       50.000000,        15.486692,        19.986466;
       51.000000,        15.642414,        20.390424;];
if bdp==0.15
  c=x1(p,2);
  k=x1(p,3);
elseif bdp==0.25
  c=x2(p,2);
  k=x2(p,3);
elseif bdp==0.5
  c=x3(p,2);
  k=x3(p,3);
else
  c=0;
  k=0;
end
```

---

solve_LS.m.

```
%  solve_LS.m
%
%   This file uses the Least Square estimate for S and B as the
%   initial guess for S and B in the function solve_sest.m.
%
```

```
% Generate initial guesses for Sigma and beta
S = Y'*(eye(n)-ppo(X))*Y/(n-rank(X));
beta = vec((X'*X)\(X'*Y));

% initial guesses for G0, lam, phi.
[G0,L0,dumm] = svd(S);         %Spectral decomp (symmetric matrix)
lam = M*avg*diag(sparse(L0));  %Avg multiple eigenvalues.
phi = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam); % (Under Model 2)

[S_LS,Gam_LS,Lam_LS,B_LS,throwout,normdw,beta_LS,mu_LS,psi_LS] = ...
   solve_sest(T1,T2,A1,A2,Dp,Dg,nu,b0,c0,X,Y,lam,G0,beta,phi,output);
```

---

solve_c0.m.

```
function [b,c,iter] = solve_c0(p,r);
%=====================================================================
%
% solve_c0    Solve for the constants b0 and c0.
%
%   [b,c,iter] = solve_c0(p,r) solves for the constants b0 and c0 in
%   S-estimation under the assumption that the underlying distribution
%   is normal, and using the integral of Tukey's bivariate function as
%   the rho function
%
%   Inputs
%     p = dimension of data
%     r = asymptotic breakdown point
%
%   Outputs
%     b    = the constant b_0
%     c    = the constant c_0
%     iter = the number of iterations to convergence
%
%=====================================================================

i=1:3;
tol = 1e-13;
pvec = exp(i*log(2)+gammaln(p/2+i)-gammaln(p/2));

% Bound the root between lp and rp.  Since the smallest c0 can be is
% when the bdp is 50%, then any number for lp less than 1.547 will do.
%
lp = 1.5;  %The initial left point
rp = lp;
```

```
lsign = 6*b_0(lp,p,pvec)/lp^2 - r;
sign = lsign;

while sign*lsign > 0,
  rp = 2*rp;
  [b,db] = b_0(rp,p,pvec);
  sign = 6*b/rp^2 - r;
end

%
% Now begin the routine to find the root.  Use either newton or
%   bisection
%
c = rp;
iter = 0;                %count the iterations
stop = 1;                %norm(step) > tol (initially)
while norm(stop) > tol,
  iter = iter + 1;
  c_new = c;

  top = 6*b-r*c^2;
  bot = 6*c*db-12*b;

  % If bot is not zero, then use a newton-step.  Otherwise, force
  % c_new to be outside the interval.
  if abs(bot) > 1e-2,
    stop = c*top/bot;
    c_new = c_new - stop;
  else
    c_new = lp - .2;
  end

  %
  % Update c
  %    if it's in the interval, use the newton-step,
  %    otherwise use the midpoint of the interval.
  in_interval = (lp <= c_new) & (c_new <= rp);
  if in_interval;
    c = c_new;
  else
    c = (lp+rp)/2;
    stop = (rp-lp)/2;  %stop criterion for bisection
  end

  %
  % Make a new lp or rp.
```

```
  %
  [b,db] = b_0(c,p,pvec);
  sign = 6*b/c^2 - r;
  if sign*lsign > 0,
    lp = c;
    lsign = sign;
  else
    rp = c;
  end

%  string = '[%6.4e %6.4e %6.4e] stop=%6.4e\n';
%  fprintf(string,lp,c,rp,stop);
%  pause

end;

%======================================================================
%======================================================================

function [b,db] = b_0(c,p,pvec)
%
% Evaluative the function b and the derivative of b for a given value
% of c.
%

i=1:3;
cs = c^2;
con = pvec./([2 -2*cs 6*cs^2]);

ccdf0 = chi2cdf(cs,p);
cpdf0 = chi2pdf(cs,p);
cpdf = chi2pdf(cs,p+2*i);
ccdf = chi2cdf(cs,p+2*i);

b = con*ccdf' + cs/6*(1-ccdf0);
db = 2*c*con*cpdf' + con(2:3)*diag([-2 -4]/c)*ccdf(2:3)' + ...
      c/3*(1-ccdf0) - c^3/3*cpdf0;
return;
```

---

solve_dq_betamu.m.

```
function [Gam,beta,q,stop,mu] = ...
   solve_dq_betamu(T1,T2,A1,A2,Dp,tNp,Dg,nu,b0,c0,X,Y,lam,Gam,...
                   beta,varargin);
```

```
%=====================================================================
%
% solve_dq_betamu   Solve dq_betamu = 0
%
%  [Gam,beta,q,stop,mu] =
%    solve_dq_betamu(T1,T2,A1,A2,Dp,tNp,Dg,nu,b0,c0,X,Y,lam,Gam,beta);
%
%   This will solve part of the S estimation problem, with first
%   minimizing the function q = avg(rho(dk)) - b0 with respect to beta
%   and mu, keeping the eigenvalue parameters (psi) constant.
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     A1   = Matrix associated with mu
%     A2   = Matrix associated with eta
%     Dp   = Duplication matrix (dup.m)
%     tNp  = 2*Np (ppo of space of symmetric matrices)
%     Dg   = Derivative of vecG wrt mu
%     nu   = vector containing sizes of beta, mu, phi
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     X    = Design Matrix (location parameters)
%     Y    = Data Matrix Y
%     lam  = Current guess eigenvalues
%     Gam  = Current guess for Eigenvector matrix
%     beta = Current guess for the location parameters
%     varargin = output indicator variable
%                1 = ouput iterations
%                0 or blank = don't output iterations
%
%   Outputs
%     Gam_est  = new estimate of the Eigenvector matrix
%     beta     = new estimate of beta
%     q        = q evaluated at psi
%     stop     = convergence criterion
%     mu       = new estimate of mu
%
%=====================================================================

if nargin == 16,
  output = varargin{1};
else
  output = 0;
end
```

```
%
% Constants
%

tol = 1e-10;
iter = 0;
p = size(T1,1);
d = size(X,2);
n = size(Y,1);

%
% Constant Matrices
%
Ip = speye(p);
Inu2 = speye(nu(2));
Lam = spdiag(lam);
Laminv = spdiag(1./lam);
kLamIp = kron(Lam,Ip);
tNpkLamIpDg = tNp*kLamIp*Dg;
KL = kron3(Ip,vec(Lam)',Ip);
Dg2 = A2*Dp'*kron3(Ip,vec(Ip)',Ip);
second = tNp*(kLamIp*Dg2-KL)*kron(Dg,Dg);

stop = 1;

if output == 1,
  str1 = strcat('Stage 1 - Solve dq/dbetamu = 0\n',dashes(30),'\n');
  fprintf(str1);
end

while (stop > tol) & (iter < 100),

  iter = iter + 1;
  %
  % Next guess for theta (Always start with a column of zeros for
  %   mu)
  %
  theta1 = [ beta; zeros(nu(2),1) ];
  Siginv = Gam*Laminv*Gam';
  %--------------------------------------------------
  %
  % Construct F11_mumu and F1_mumu
  %

  % The next line is the same as (but cheaper)
  %   F1_mu = tNp*kron(Gam,Gam)*kron(Lam,Ip)*Dg;
```

```
kGamGam = sparse(kron(Gam,Gam));
F1{2} = kGamGam*tNpkLamIpDg;

% The next line is the same as (but cheaper)
%   F2_mumu = tNp*kron(Gam,Gam)*(kron(Lam,Ip)*Dg2-KL)*kron(Dg,Dg);
F2{2,2} = kGamGam*second;
F11{2,2} = dvec(F2{2,2},p^2*nu(2),nu(2));

%----------------------------------------------------
%
% Construct dk, W1, and wk
%
Z = Y-X*dvec(beta,d,p);
[q,dk] = qeval(Z,Siginv,b0,c0);
W1 = (1/n)*spdiag(.5*rho(1.1,dk,c0));              %diag(rho(dk)/dk)
W2 = spdiag((rho(2,dk,c0)-rho(1.1,dk,c0))./(4*n*dk.^2)); %in the V's

%Construct W and V matrices
Wxx = X'*W1*X;
Wxz = X'*W1*Z;
Wzz = Z'*W1*Z;

% This code takes an average of .1432 seconds per iteration,
% whereas the code below it computing Vzz,Vxz,Vxx takes .0071
% seconds per iteration, a much improved time
%
%  Vxx = sparse(p*d,p*d);
%  Vxz = sparse(p*d,p^2);
%  Vzz = sparse(p^2,p^2);
%  for k=1:n
%    x = X(k,:)';
%    zpiS = (Z(k,:)*Siginv)';
%    xk_star = vec(x*zpiS');
%    zk_star = vec(zpiS*zpiS');
%    Vxx = Vxx + xk_star*wk(k)*xk_star'/n;
%    Vxz = Vxz + xk_star*wk(k)*zk_star'/n;
%    Vzz = Vzz + zk_star*wk(k)*zk_star'/n;
%  end
%
ZSiginv = Z*Siginv;

Kzz = khatrirao(ZSiginv',ZSiginv');
Kzx = khatrirao(ZSiginv',X');
W2Kzz = W2*Kzz';

Vzz = Kzz*W2Kzz;
```

```
Vxz = Kzx*W2Kzz;
Vxx = Kzx*W2*Kzx';

%---------------------------------------------------
%
% Compute derivative dq1_theta1
%
WxzS = Wxz*Siginv;
SWzzS = Siginv*Wzz*Siginv;

dq1_1 = -2*vec(WxzS);
dq1_2 = -F1{2}'*vec(SWzzS);
dq1 = [ dq1_1 ; dq1_2 ];

%
% Compute second derivate d2q11_(dtheta1 x dtheta1')
%

d2q11_11 = 2*(kron(Siginv,Wxx) + 2*Vxx);
d2q11_21 = 2*(kron(Siginv,WxzS) + Vxz)*F1{2};
d2q11_22 = -F11{2,2}'*kron(Inu2,vec(SWzzS)) + 2*F1{2}'* ...
    (kron(Siginv,SWzzS) + .5*Vzz)*F1{2};
d2q11 = [ d2q11_11 d2q11_21; d2q11_21' d2q11_22 ];

%---------------------------------------------------
%
% Make update to theta1
%
step = -d2q11\dq1;
% If it's not a descent direction, then use
% steepest descent step.
if (dq1'*step > 0) | (abs(dq1'*step)>1e6),
  step = -dq1*(norm(step)/norm(dq1));
end
theta1_new = theta1 + step;

%
% If max(abs(mu)) < 0.2 continue normally, otherwise force max of
% mu to be 0.2.  (Change step also also).)
%
mu = theta1_new(nu(1)+1:end);
maxmu = max(abs(mu));
if maxmu > .05
  alpha = .05/maxmu;
  step_new = alpha*step;
  theta1_new = theta1 + step_new;
```

```
    mu = theta1_new(nu(1)+1:end);
  end;
  stop = sqrt(abs(dq1'*step));

  %
  %Form current estimates for beta and G
  %
  beta_new = theta1_new(1:nu(1));
  eta = solve_eta(A1,A2,mu);
  G = dvec(A1*mu+A2*eta,p,p);

  %
  % Estimate matrices
  %
  Gam_new = Gam*G;
  Siginv = Gam_new*Laminv*Gam_new';

  %
  % Iteration output
  %
  if output == 1,
    str = '    It#%3.0f, max(mu)=%6.4e, stop=%6.4e, q=%6.4g\n';
    fprintf(str,iter,maxmu,stop,q);
  end

  % Update Gam and beta
  Gam = Gam_new;
  beta = beta_new;

end

B_est = dvec(beta,d,p);

%======================================================================

function dash = dashes(n);
%
% returns a string with n dashes
%
dash = char(45*ones(1,n));
return;
```

---

solve_dw_psi.m.

```
function [lam,phi_new,q,stop,throwout,psi] = ...
    solve_dw_psi(T1,T2,nu,b0,c0,X,Y,Gam,beta,phi,varargin);
%======================================================================
%
%  solve_dw_psi   Solve dw/dpsi = 0
%
%  [lam,phi_new,q,stop,throwout,psi]
%    = solve_dw_psi(T1,T2,nu,b0,c0,X,Y,Gam,beta,phi)
%
%   This function gives the s-estimators for the location and the
%   eigenvectors and eigenvalues for the model Y = XB + E, where
%   Disp(Y) = I x alpha*Sig
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     V1   = Implicit model matrix
%     V2   = Implicit model matrix
%     Z    = Centered Data Matrix (Y-XB)
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     Gam  = Current guess for Eigenvector matrix
%     beta = Current guess for the location parameters
%     phi  = Current value for phi.
%     psi  = Current value for psi.
%     varargin = output indicator variable
%                1 = ouput iterations
%                0 or blank = don't output iterations
%
%   Outputs
%     lam      = updated estimate of lam
%     phi_new  = updated estimate of phi
%     q        = q evaluated at psi
%     stop     = convergence criterion
%     throwout = Was convergence reached?
%     psi      = updated estimate of psi
%
%   For Model found in make_F11_phiphi.m  only!
%
%======================================================================

if nargin == 11,
  output = varargin{1};
else
  output = 0;
end
```

```
%
% Constants
%
tol = 1e-11;
tol1 = 1e-5;
p = size(T1,1);
d = size(X,2);
n = size(Y,1);
stop = 1;

%
% Constant Matrices
%
T1 = sparse(T1);
T2 = sparse(T2);
G = sparse(p,p);
Ip = speye(p);
Inu3m1 = speye(nu(3)-1);
Inu3 = speye(nu(3));
B_est = dvec(beta,d,p);
Z = Y - X*B_est;

%----------------------------------------------------------------------
%  Make initial guesses for phi, psi, V1, V2
%----------------------------------------------------------------------
[V1,V2,psi]=initial_guess(T1,T2,Z,b0,c0,Gam,beta,phi);
[a,V1,V2,phi,psi,eta_psi,lam,D1_lam,dk,W1,Wzz]= ...
    satisfy_constraint(T1,T2,V1,V2,Z,b0,c0,Gam,beta,phi,psi);
step = 1;
iter = 0;
throwout = 0;
Siginv = Gam*spdiag(1./lam)*Gam';

%----------------------------------------------------------------------
%
%    Main branch for minimizing w = ln(det(Sigma)) wrt psi while
%       beta and mu are constant
%
%----------------------------------------------------------------------

if output == 1,
  str2 = strcat('Stage 2 - solve dw/dpsi = 0\n',dashes(27),'\n');
  fprintf(str2);
end
```

```
while (stop > tol) & (iter < 500) & (throwout == 0);

  %--------------------------------
  % Organize some preliminary vars
  %--------------------------------

  F1{3} = khatrirao(Gam,Gam)*D1_lam;        %Same as kGamGam*L*D1_lam;
  P1{3} = V1;                               %V1 - V2*inv(a'*V2)*a'*V1;
  F11{3,3} = make_F11_phiphi(T1,T2,Gam,phi);%Depends on MODEL!!!!

  W2 = spdiag((rho(2,dk,c0)-rho(1.1,dk,c0))./(4*n*dk.^2));

  % This code takes an average of .0987 seconds per iteration,
  % whereas the code below it computing Vzz takes .0077 seconds per
  % iteration, a much improved time
  %
  %  Vzz = sparse(p^2,p^2);
  %  for k=1:n
  %    zpiS = (Z(k,:)*Siginv)';
  %    vzz = vec(zpiS*zpiS');
  %    Vzz = Vzz + vzz*ww(k)*vzz'/n;
  %  end;
  ZSiginv = Z*Siginv;
  Kzz = khatrirao(ZSiginv',ZSiginv');
  Vzz = Kzz*W2*Kzz';  % perform the sum using matrix multp.

  %--------------------------------
  % Compute dw/dpsi
  %--------------------------------
  dw{3} = P1{3}'*(F1{3}'*vec(Siginv));

  %--------------------------------
  % Compute d2w/(dpsi x dpsi)
  %--------------------------------

  temp1 = (inv(V2'*a)*V2')*(F1{3}'*vec(Siginv));
  SWzzS = Siginv*Wzz*Siginv;
  SS = kron(Siginv,Siginv);
  Mzz = kron(Siginv,SWzzS) + .5*Vzz;

  d2w11{3,3} =  F11{3,3}'*kron(Inu3,vec(Siginv)) - F1{3}'*SS*F1{3};
  d2q11{3,3} = -F11{3,3}'*kron(Inu3,vec(SWzzS)) + 2*F1{3}'*Mzz*F1{3};

  d2w{3,3} = -P1{3}'*(d2q11{3,3}*P1{3}*temp1 - d2w11{3,3}*P1{3});

  %----------------------------------------
```

```
  % Update psi first, then everything else
  %----------------------------------------

  % We're approaching a boundary on the parameter space.  Throw this
  % out!
  if cond(d2w{3,3}) > 1e15
    d2w{3,3} = eye(size(d2w{3,3}));
    throwout == 1;
  end;

  step = - d2w{3,3}\dw{3};
%  psi_new = psi + step;
%  if (abs(dw{3}'*step) > tol1) & (dw{3}'*step > tol1),
    %If step is not a descent direction, use
    %steepest descent direction.
%    step = -dw{3};
%    disp('Wrong way');
%  end

%  psi

  normcond = norm(exp(T2*V1*step))/norm(lam);    %Specific to Model 2
  if normcond > .5,
     step_new = .5*step/abs(norm(step))
     disp('Step is too big!');
     psi_new = psi + step_new;
  else
     psi_new = psi + step;
  end;

  normT2phi = norm(T2*phi);
  if normT2phi > 30,
    throwout = 1;
  end;

  stop = sqrt(abs(dw{3}'*step));

  %------------------------------------------------------
  % update all variables involved to satisfy constraint
  %------------------------------------------------------
  [a,V1,V2,phi,psi,eta_psi,lam,D1_lam,dk,W1,Wzz]= ...
      satisfy_constraint(T1,T2,V1,V2,Z,b0,c0,Gam,beta,phi,psi_new);

  Siginv = model_lam(T1,T2,V1,V2,Gam,phi);
  q = qeval(Z,Siginv,b0,c0);
```

```
  iter = iter + 1;
  if output == 1,
    string = '   It#%3.0f, step=%6.4e stop=%6.4e, det=%6.4f q=%6.4e\n';
    fprintf(string,iter,norm(step),stop,prod(lam),q);
  end;
end;

if output == 1,
  fprintf('\n');
end

phi_new = phi;
Siginv = model_lam(T1,T2,V1,V2,Gam,phi_new);
q = qeval(Z,Siginv,b0,c0);

%==================================================================

function dash = dashes(n);
%
% returns a string with n dashes
%
dash = char(45*ones(1,n));
return;
```

---

solve_eta_psi.m.

```
function [ep,iter,qfunc] = solve_eta_psi(T1,T2,V1,V2,Gam,b0,c0,...
                                         Z,phi,psi);
%==================================================================
%
%  [ep,iter,qfunc] = solve_eta_psi(T1,T2,V1,V2,Gam,b0,c0,Z,phi,psi)
%
%   Solves q = 1/n * sum(rho(dk)) - b0 = 0 for eta_psi given psi (Note
%   that dk is a function of eta_psi and psi.)
%
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     V1   = Implicit model matrix
%     V2   = Implicit model matrix
%     Gam  = Current guess for Eigenvector matrix
```

```
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     Z    = Centered Data Matrix (Y-XB)
%     phi  = current guess for phi (probably doesn't satisfy the
%            constraint)
%     psi  = current estimate of psi.
%
%   Outputs
%     ep    = solution for eta_psi as a function of psi
%     iter  = number of iterations to convergence
%     qfunc = q evaluated at eta_psi and psi (should be zero).
%
%     (Solves for Model found in model_lam.m)
%
%======================================================================

tol = 1e-12;
step = 1;
iter = 0;
n = size(Z,1);

%
% Starting point for Newton's method (and one side of bisection).
%
ep = (V2'*V2)\(V2'*phi);

%
% Find starting point for the other side of bisection method.
%
ep_a = ep;

Siginv = model_lam(T1,T2,V1,V2,Gam,psi,ep_a);
qa = qeval(Z,Siginv,b0,c0);

if abs(qa) < tol,
  ep_b = ep_a;     %The constraint is already satisfied!
  qb = qa;
else
  % Since q is increasing as a function of eta_psi, then the sign for
  % ep_b is opposite the sign of qa.  Then update ep_b until a root is
  % bound between ep_a and ep_b
  %
  if qa > 0,
    ep_b = -2;       %choose ep_b < 0
  else
    ep_b =  2;       %choose ep_b > 0
```

```
  end
  %
  % Update ep_b so root is bounded between ep_a and ep_b
  %
  Siginv = model_lam(T1,T2,V1,V2,Gam,psi,ep_b);
  qb = qeval(Z,Siginv,b0,c0);
  while qa*qb > 0,
    ep_b = 2*ep_b;
    Siginv = model_lam(T1,T2,V1,V2,Gam,psi,ep_b);
    qb = qeval(Z,Siginv,b0,c0);
  end;
end;

%
% Choose next newton iterate based on smallest q evalution
%
if abs(qa) > abs(qb),
  ep_c = ep_b;
else
  ep_c = ep_a;
end

%
%Make sure ep_a is on the left and ep_b is on the right
%
if ep_b < ep_a,
  temp = ep_b; ep_b = ep_a; ep_a = temp;
  temp = qb;   qb = qa;     qa = temp;
end

%---------------
% main branch
%---------------

while abs(step) > tol & (iter < 300),
  iter = iter + 1;

  [ep_c_new,stepn,chg_slp] = newton(T1,T2,V1,V2,b0,c0,Z,Gam,psi,ep_c);
  in_interval = (ep_a <= ep_c_new) & (ep_c_new <= ep_b);

  if  ( in_interval & chg_slp == 0 ) | (stepn < tol)
    %
    % Use Newton's method if the new estimate is in the interval
    %   (assuming the slope was not bounded away from zero)
    ep_c = ep_c_new;
    Siginv = model_lam(T1,T2,V1,V2,Gam,psi,ep_c);
```

```
    qc = qeval(Z,Siginv,b0,c0);
    [qa,ep_a,ep_b] = switchq(qa,qc,ep_a,ep_b,ep_c);
    step = stepn;
  else
    %
    % Bisection is needed if Newton's method decides to blow up.
    %
    ep_c = (ep_a + ep_b)/2;
    Siginv = model_lam(T1,T2,V1,V2,Gam,psi,ep_c);
    qc = qeval(Z,Siginv,b0,c0);
    [qa,ep_a,ep_b] = switchq(qa,qc,ep_a,ep_b,ep_c);
    step = ep_a - ep_b;
  end;

  % Uncomment the following lines to watch the convergence
  %
  %st = 'It#%3.0f, [%6.4f < %6.4f < %6.4f ] stop=%6.4e, cstr=%6.4e\n';
  %fprintf(st,iter,ep_a,ep_c,ep_b,step,min(abs(qa),min(qb)));
  %pause;
end;
%fprintf('----------------------\n')

ep = ep_c;
qfunc = qc;

%======================================================================
%======================================================================

function [qa,ep_a,ep_b] = switchq(qa,qc,ep_a,ep_b,ep_c);
%
% This bounds the root between ep_a and ep_b (The essential step of
%   the bisection method.)
%

if qc*qa > 0,
  ep_a = ep_c;
  qa = qc;
else
  ep_b = ep_c;
end

%======================================================================

function [ep_n,step,chg_slp] = newton(T1,T2,V1,V2,b0,c0,Z,Gam,psi,ep);
%
% Compute one step of Newton's Method (modified Newton's when the
```

```
% slope is too close to zero).
%
%--------------------
% Find q and dq
%--------------------
[Siginv,Laminv,D1_lam] = model_lam(T1,T2,V1,V2,Gam,psi,ep);
[q,dk] = qeval(Z,Siginv,b0,c0);
a = compute_a(b0,c0,Z,Gam,dk,Laminv,D1_lam);
dq = a'*V2;

%----------------------------------
% Make sure slope is not near zero
%----------------------------------
chg_slp = 0;            %Initialize to 0
if abs(dq) < 1e-5,
  % Change slope to .5 or -.5 if slope is too small
  dq = .5*sign(dq)*(dq ~= 0) + .5*(dq == 0);
  chg_slp = 1;          %Report this is a false slope
end;

%----------------------
% Update psi to psi_new
%----------------------
step = -q/dq;
ep_n = ep + step;

%=================================================================
```

---

solve_newguess.m.

```
%  solve_newguess.m
%
%  Find the S-estimator using an initial guess from the current
%  subsample of the data.  (Generated from newguess.m)
%

% Generate initial guesses for Sigma and beta
Xsub = X(J,:);
Is = eye(size(subsample,1));
S = subsample'*(Is-ppo(Xsub))*subsample/(n-rank(Xsub));
beta = vec((Xsub'*Xsub)\(Xsub'*subsample));

% initial guesses for G0, lam, phi.
[G0,L0,dumm] = svd(S);         %Spectral decomp (symmetric matrix)
```

```
lam = M*avg*diag(sparse(L0));  %Avg multiple eigenvalues.
phi = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam); % (Under Model 2)

[S_mine,Gam_mine,Lam_mine,B_mine,throwout,normdw,beta_mine,...
 mu_mine,psi_mine] = ...
   solve_sest(T1,T2,A1,A2,Dp,Dg,nu,b0,c0,X,Y,lam,G0,beta,phi,output);
```

---

solve_s_estimate_sim.m.

```
%function [S_min,Gam_min,Lam_min,beta_min,mu_min,psi_min] =
%solve_s_estimate();
%
%
% Solve for the S-estimate using various starting points.  If an
% error is found, throw that guess out and start from a different
% spot in the space
%
%
str1 = '\n\n It#%4.0f Starting pt=%3.0f  normdw='
str2 = '%9.4e   detS=%12.6e\n\n\n';
string = strcat(str1,str2);

startpts = 80;
if surreal==1, startpts=1;end;
normdet = [];

for iii=1:startpts
  fprintf('\n It#%4.0f Starting pt=%3.0f\n',ii,iii)
  throwout = 1;

  iters=2000;
  while throwout == 1,
    try
      if surreal==1,
        res=Smultivar(Y,iters,bdp);
        S = res.cov;        %Use initial guess of cov from Smultivar.
        beta = res.mean';   %Use initial guess of mean from Smultivar.
      else
        newguess
      end
      solve_newguess
      if throwout == 1, iters = iters + 500;end;
    catch
      throwout = 1;
```

```
    end;
  end;

  if iii == 1,
    S_min = S_mine;
    Gam_min = Gam_mine;
    lam_min = diag(Lam_mine);
    beta_min = beta_mine;
    mu_min = mu_mine;
    psi_min = psi_mine;
  else
    if (sum(log(diag(Lam_mine))) < sum(log(lam_min))) & ...
  (normdw/n < 1e-6)
      S_min = S_mine;
      Gam_min = Gam_mine;
      Lam_min = Lam_mine;
      lam_min = diag(Lam_mine);
      beta_min = beta_mine;
      mu_min = mu_mine;
      psi_min = psi_mine;
    end;
  end;
  fprintf(string,ii,iii,normdw,det(S_mine));
  normdet = [ normdet; normdw det(S_mine) ];
%  pause
end;

%
%This counts the number of local minimizers
%
eee = ones(startpts,1);
diff = spdiags([eee -eee], 0:1, startpts-1,startpts);

sum(round(diff*sort(normdet(:,2))*100)/100~=0)+1
```

---

solve_sest.m.

```
function [S_est,Gam_est,Lam_est,B_est,throwout,nmdw,beta,mu,psi] = ...
 solve_sest(T1,T2,A1,A2,Dp,Dg,nu,b0,c0,X,Y,lam,Gam,beta,phi,varargin);
%=====================================================================
%
%  [S_est,Gam_est,Lam_est,B_est,throwout,nmdw,beta,mu,psi]
%   = solve_sest(T1,T2,A1,A2,Dp,tNp,Dg,nu,b0,c0,X,Y,lam,Gam,beta,phi);
%
```

```
%   This groups the two subproblems together to solve for a local
%   minimum to the S-estimating equation of multivariate regression.
%
%   Inputs
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     A1   = Matrix associated with mu
%     A2   = Matrix associated with eta
%     Dp   = Duplication matrix (dup.m)
%     Dg   = Derivative of vecG wrt mu
%     nu   = vector containing sizes of beta, mu, phi
%     b0   = Constant associated with constraint q
%     c0   = Constant associated with breakdown point
%     X    = Design Matrix (location parameters)
%     Y    = Data Matrix Y
%     lam  = Current guess eigenvalues
%     Gam  = Current guess for Eigenvector matrix
%     beta = Current guess for the location parameters
%     phi  = Current value for phi.
%     varargin= Output stages and/or overall iteration
%            [ 1 1 ] - ouput both stages and overall (pause at end
%                      of iteration.
%            [ 1 0 ] - ouput overall, but no stages
%            blank   - no output
%
%   Outputs
%     S_est    = S-estimator of the scatter matrix
%     Gam_est  = S-estimator of the Eigenvector matrix
%     Lam_est  = S-estimator of the Eigenvalue matrix
%     B_est    = S-estimator of the Location parameters
%     throwout = Was convergence met?
%     nmdw     = normdw should be zero
%     beta     = S-estimator of beta
%     mu       = S-estimator of mu
%     psi      = S-estimator of psi
%
%======================================================================

%
% Constants
%
p = size(T1,1);
d = size(X,2);
tNp = speye(p^2) + commute(p,p);

%
```

```
% Iteration messages.
%
% Stage 1 message
% Divider
str1 = strcat(dashes(55),'\n\n');
% Big loop iteration message
str2 = 'It#%3.0f, step_dw=%6.4e, step_dq=%6.4e q=%4.2e stop=%6.4e\n';

%
% Constants for the conditions of convergence
%
tol = 1e-12;
nmdw_tol = 1e-7;
iter = 30;
throwout = 0;

% initialize convergence variables
stop = 1;
iter = 0;
lam_old = lam;
dw = 1;

% Decide whether to output stages
if nargin == 16,
  output = varargin{1};
  if length(output) == 2,
    output2 = output(2);
  else
    output2 = 0;
  end
else
  output = 0;
  output2 = 0;
end


while ((stop > tol) | (nmdw > nmdw_tol)) & ...
      (iter < 30) & (throwout == 0);
  iter = iter + 1;
  if output2 == 1,
    % clear the screen
    !clear;
   end

  [Gam,beta,q,stepdq,mu] = solve_dq_betamu(T1,T2,A1,A2,Dp,tNp,Dg, ...
                             nu,b0,c0,X,Y,lam,Gam,beta,output2);
```

```
  [lam,phi,q,stepdw,throwout,psi] = solve_dw_psi(T1,T2,nu,b0,c0,X, ...
                              Y,Gam,beta,phi,output2);
  stop = abs(sum(log(lam_old)) - sum(log(lam)));
  dw = deriv1_w(T1,T2,A1,A2,Dg,Y,X,d,b0,c0,nu,Gam,beta,mu,psi,phi);
  nmdw = norm(dw);
  if output2 == 1,
    fprintf(str1);
  end
  if output(1) == 1,
    fprintf(str2,iter,stepdw,stepdq,q,stop);
  end
  if (output(1) == 1) & (output2 == 1),
    pause;
  end

  lam_old = lam;
end

if (throwout == 1)
  disp('throw out!');
end

if nmdw > nmdw_tol
  throwout = 1;
  disp('Solution not found');
end;

Gam_est = Gam;
Lam_est = diag(lam);
S_est = Gam*Lam_est*Gam';
B_est = dvec(beta,d,p);

function dash = dashes(n);
%
% returns a string with n dashes
%
dash = char(45*ones(1,n));
return;
```

---

solve_sestimator.m.

```
function [S_min,Gam_min,Lam_min,B_min,sand,bias] = ...
    solve_sestimator(T1,T2,m,X,Y,bdp,method,output);
%======================================================================
```

```
%
% solve_sestimator       Find the global S-estimates.
%
%   Solve for the S-estimate using various starting points.
%   Specify which starting point to use in the method variable.
%   When using the Least Squares estimate or the Surreal estimate
%   as the initial guess, only a local S-estimate is found.  To get
%   a global S-estimates, random starting points must be used.
%
%   When using random starting points, if an error is found, the
%   guess is thrown out and a new guess is generated from a
%   different spot in the parameter space.
%
%   Inputs:
%     T1   = Design matrix (models linear relations btwn eigvals)
%     T2   = Design matrix (models exponential relations btwn eigvals)
%     m    = multiplicity vector
%     X    = Design matrix for B
%     Y    = Data matrix
%     bdp  = breakdown point
%     method = Designate which initial guess to use
%              1 = random starting points
%              2 = Least Squares estimate
%              3 = Surreal estimate
%     output = Output stages and/or overall iteration
%            [ 1 1 ] - ouput both stages and overall (pause at end
%                      of iteration.
%            [ 1 0 ] - ouput overall, but no stages
%
%   Outputs:
%     S_min    = S-estimate of the scatter matrix
%     Gam_min  = S-estimate of the eigenvector matrix
%     Lam_min  = S-estimate of the eigenvalue matrix
%     beta_min = S-estimate of the location parameters
%     sand     = Sandwich estimator of the asymptotic variance of
%                the estimators
%     bias     = Estimate of the bias of the estimators.
%
%======================================================================

[n,p] = size(Y);
d = size(X,2);
nu = [ p*d; (p^2-m'*m)/2; size(T2,2) ];

[Dg,A1,A2,L,Dp]= constantmatrices(m); %Constant matrices.
```

```
% Compute the breakdown point constant (Under a gaussian model)
[b0,c0] = solve_c0(p,bdp);

if sum(m) ~= p,
  error('Multiplicity vector invalid');
end


M=[];
avgi=[];
for i=1:length(m);
  M = [ M repmat(i,1,m(i)) ];
  avgi = [ avgi repmat(1/m(i),1,m(i)) ];
end;
M = sparse(dummyvar(M));%Simple multiplicity structure.
avg = (spdiag(avgi)*M)';%inv(M'*M)*M' (Used to calculate estimates for
                        %  eigenvalues.  It averages eigenvalues)

% Output string
str1 = '\n\n Starting pt=%3.0f  normdw=';
str2 = '%9.4e   detS=%12.6e\n\n\n';
string = strcat(str1,str2);

if method == 1,  % Use random starting points
  startpts = 80;

  for iii=1:startpts
    if output(1) == 1,
      fprintf('\n Starting pt=%3.0f\n',iii)
    end
    throwout = 1;

    while throwout == 1,
      try
        %Compute a random sample
        J = sample_wor(n/2,n);
        subsample = Y(J,:);

        % Generate initial guesses for Sigma and beta
        Xsub = X(J,:);
        Is = eye(size(subsample,1));
        S = subsample'*(Is-ppo(Xsub))*subsample/(n-rank(Xsub));
        beta = vec((Xsub'*Xsub)\(Xsub'*subsample));

        % initial guesses for G0, lam, phi.
        [G0,L0,dumm] = svd(S);
```

```
        lam = M*avg*diag(sparse(L0)); %Avg multiple eigenvalues.
        phi = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam); % (Model 2)

        [S_mine,Gam_mine,Lam_mine,B_mine,throwout,normdw] = ...
        solve_sest(T1,T2,A1,A2,Dp,Dg,nu,b0,c0,X,Y,lam,G0,...
                   beta,phi,output);
      catch
        throwout = 1;
      end;
    end;

    if iii == 1,
      S_min = S_mine;
      Gam_min = Gam_mine;
      Lam_min = Lam_mine;
      detlogSig = sum(log(diag(Lam_mine)));
      B_min = B_mine;
    else
      if (sum(log(diag(Lam_mine))) < detlogSig) ...
           & (normdw/n < 1e-6)
        S_min = S_mine;
        Gam_min = Gam_mine;
        Lam_min = Lam_mine;
        detlogSig = sum(log(diag(Lam_mine)));
        B_min = B_mine;
      end;
    end;

    if output(1) == 1,
      fprintf(string,iii,normdw,det(S_mine));
    end
  end;
else
  if method == 2,
    % Generate initial guesses for Sigma and beta
    S = Y'*(eye(n)-ppo(X))*Y/(n-rank(X));
    beta = vec((X'*X)\(X'*Y));
  elseif method == 3,
    res = Smultivar(Y,2000,bdp);
    S = res.cov;           %Use initial guess of cov from Smultivar.
    beta = res.mean';      %Use initial guess of mean from Smultivar.
  end

  % initial guesses for G0, lam, phi.
  [G0,L0,dumm] = svd(S);
  lam = M*avg*diag(sparse(L0));  %Avg multiple eigenvalues.
```

```
  phi = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam); % (Under Model 2)

  [S_min,Gam_min,Lam_min,B_min,throwout] = ...
    solve_sest(T1,T2,A1,A2,Dp,Dg,nu,b0,c0,X,Y,lam,G0,beta,phi,output);
end

% Model 2 for phi
phi_min = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*diag(Lam_min));
[dw,dw1,d2w,d3w,K2,FF,sand,EZ2D0,ED1] = deriv_w(T1,T2,A1,A2,Dg,...
                      Y,X,b0,c0,nu,Gam_min,vec(B_min),phi_min);

bias = 1/sqrt(n)*ED1;
```

---

solve_surreal.m.

```
%  solve_surreal.m
%
%   This file uses the Smultivar.m estimates for cov and mean as
%   initial guesses for S and B in the solve_sest.m function
%
%   Execute res = Smultivar(Y,2000,bdp) before running.
%
S = res.cov;             %Use initial guess of cov from Smultivar.
beta = res.mean';        %Use initial guess of mean from Smultivar.

% initial guesses for G0, lam, phi.
[G0,L0,dumm] = svd(S);         %Spectral decomp (symmetric matrix)
lam = M*avg*diag(sparse(L0));  %Avg multiple eigenvalues.
phi = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam); % (Under Model 2)

[S_surreal,Gam_surreal,Lam_surreal,B_surreal,throwout] = ...
   solve_sest(T1,T2,A1,A2,Dp,Dg,nu,b0,c0,X,Y,lam,G0,beta,phi,output);
```
