Notes on Adjustment Computations

Based on Former Geodetic Science Courses
GS650 and GS651

Taught at
The Ohio State University

by Prof. Burkhard Schaffrin

Compiled by Kyle Snow

September 6, 2017

Contents

Acknowledgments

The compiler of this document would like to thank Dr. Shengjie Ge for his assistance with typesetting an earlier version with LaTeX. Also, Mr. Xiankun Wang, MSc., is acknowledged for his careful reading of the text and for pointing out several typographical errors that were corrected in the edition distributed in January, 2010.

Introduction

This document is based on notes taken by Kyle Snow in Geodetic Science adjustments courses GS650 and GS651 taught by Burkhard Schaffrin at The Ohio State University in 1997 and 1998. The appendix contains several matrix properties and identities used throughout the text. A bibliography at the end includes referenced material and material for suggested reading.

Notation

A few comments about the notation used in this document may be helpful. Matrices are displayed in uppercase. Vectors are lowercase and are set in bold-face type (bold face is not used for any other symbols). Scalar variables are generally lower-case. Greek letters are used for unknown, non-random parameters, while Latin letters are used for unknown, random variables. Symbols denoting estimates of non-random variables use Greek letters with a hat on top, while predictions of random variables are shown as Latin letters with tildes on top. Tables 1 and 2 list variables, mathematical operators, and abbreviations used herein.

Table 1: Variables and mathematical operators

Symbol	Description
A	coefficient (design) matrix in the Gauss-Markov Model
B	coefficient matrix in the Condition Equations Model
$\boldsymbol{c}$	right-side vector in the system of normal equations $N\hat{\boldsymbol{\xi}} = \boldsymbol{c}$
$C\{\cdot\}$	covariance operator
$D\{\cdot\}$	dispersion operator
diag$[\cdot]$	denotes a diagonal matrix with diagonal elements comprised of $[\cdot]$
dim	denotes the dimension of a matrix
$\boldsymbol{e}$	unknown random error vector for the observations
$\tilde{\boldsymbol{e}}$	predicted random error (residual) vector for the observations
$\boldsymbol{e}_0$	unknown random error vector associated with stochastic constraints
$\tilde{\boldsymbol{e}}_0$	predicted random error (residual) vector for $\boldsymbol{e}_0$
$E\{\cdot\}$	expectation operator

Continued on next page

Symbol	Description
H_0	null hypothesis
H_A	alternative hypothesis
K	constraint matrix
m	number of unknown parameters
$\mathrm{MSE}\{\cdot\}$	mean squared error operator
n	number of observations
N	normal-equations matrix in the system of normal equations $N\hat{\boldsymbol{\xi}} = \boldsymbol{c}$
$\mathcal{N}(\cdot)$	stands for the nullspace (kernel) of a matrix or the normal distribution, depending on the context
P	weight matrix for the observations
P_0	weight matrix for stochastic constraints
q	rank of the coefficient (design) matrix A
Q	cofactor matrix for the observations
$Q_{\tilde{\boldsymbol{e}}}$	cofactor matrix for the predicted random errors (residuals)
r	redundancy of data model
$\mathcal{R}(\cdot)$	stands for the range (column) space of a matrix
rk	stands for the rank of a matrix
tr	stands for the trace of a matrix
U	matrix of eigenvectors
$\boldsymbol{w}$	constant vector in the Condition Equations Model
$\boldsymbol{y}$	vector of observations (possibly in linearized form)
$\tilde{\boldsymbol{y}}$	vector of adjusted observations
α	significance level for statistical tests
$\boldsymbol{\alpha}$	observation coefficient vector in the Model of Direct Observations
β	a quantity associated with the power of a statistical test
χ^2	chi-square statistical distribution
δ	denotes small deviation or non-random error, as in δP meaning a non-random error in P
Φ	Lagrange target function
$\boldsymbol{\eta}$	unit vector used in the Outlier Detection Model
$\boldsymbol{\kappa}_0$	vector of constraints
$\boldsymbol{\lambda}$	vector of Lagrange multipliers
$\hat{\boldsymbol{\lambda}}$	estimated vector of Lagrange multipliers
μ, $\boldsymbol{\mu}$	the expected value of a random variable, could be a scalar or vector
$\hat{\mu}$, $\hat{\boldsymbol{\mu}}$	the estimate of a random variable
ν	statistical degrees of freedom
θ	the orientation of a confidence ellipse
σ_0^2	variance component
$\hat{\sigma}_0^2$	estimated variance component
Σ	dispersion (or covariance) matrix for the observations
$\boldsymbol{\tau}$	vector of ones (summation vector)
Ω	(weighted) sum of squared residuals (unconstrained case)
$\boldsymbol{\xi}$	vector of unknown parameters
$\hat{\boldsymbol{\xi}}$	estimated parameter vector

Table 2: List of abbreviations

BLUUE	Best Linear Uniformly Unbiased Estimate
BLIP	Best LInear Prediction
cdf	cumulative distribution function
GMM	Gauss-Markov Model
LESS	LEast-Squares Solution
MSE	Mean Squared Error
pdf	probability density function

Chapter 1

Observations and Random Errors

1.1 Univariate Case

Let us introduce the univariate random variable e with a given probability density function (pdf) $f(t)$ (see Chapter 10 for more about pdf's). The probabilistic mean (or average) is the value that we expect e to take on. We denote the *expectation* of e as $E\{e\}$ and define it as follows:

$$E\{e\} = \mu_e = \int_{t=-\infty}^{\infty} e(t)f(t)dt. \tag{1.1}$$

Equation (1.1) is called the first moment of e. If the random variable e represents measurement error, then, ideally, $E\{e\} = 0$. If $E\{e\} \neq 0$, we say that the measurement error is biased.

The *dispersion*, or *variance*, of e is denoted by σ_e^2 and is defined by

$$\sigma_e^2 = E\{(e - \mu_e)^2\} = \int_{t=-\infty}^{\infty} (e - \mu_e)^2 f(t)dt. \tag{1.2}$$

Equation (1.2) is called the second centralized moment of e. The symbol $D\{e\}$ is also used to denote the dispersion (variance) of e, but often this notation is reserved for the multivariate case. The terms dispersion and variance are used interchangeably throughout these notes. The square root of the variance is called *standard deviation*.

Variance is an indicator of how closely the values taken on by the random variable e are to the expected value of e. It is reflective of measurement precision; a small variance indicates high precision. A succinct expression for the expectation and variance of the random variable e is

$$e \sim (0, \sigma_e^2). \tag{1.3}$$

The expression (1.3) is read as "*e is distributed with zero mean and σ_e^2 variance.*"

1.1.1 Expectation and Variance Propagation

Consider the observation equation

$$y = \mu + e, \quad e \sim (0, \sigma_e^2), \tag{1.4}$$

where y is an observation, μ is an unknown *observable*, and e accounts for the random error inherent in the observation y.

We want to find the expectation and variance of y. In other words, we want to know how the expectation and variance propagate from the random variable e to the random variable y. Note that μ is a constant, or *nonrandom*, variable. The expectation of a constant is the constant itself; for example, $E\{\mu\} = \mu$.

For the expectation of y, we have

$$E\{y\} = \int_{t=-\infty}^{\infty} (\mu + e(t)) f(t) dt.$$

The expectation operator is linear, so the expectation of a sum is the sum of expectations. And, as noted already, μ is a constant variable. Therefore

$$E\{y\} = \mu \cancelto{1}{\int_{t=-\infty}^{\infty} f(t) dt} + \int_{t=-\infty}^{\infty} e(t) f(t) dt = \mu + E\{e\} = \mu + 0 = \mu. \tag{1.5}$$

For the dispersion, or variance, of y we have

$$D\{y\} = \int_{t=-\infty}^{\infty} (y - E\{y\})^2 f(t) dt = \int_{t=-\infty}^{\infty} (\mu + e(t) - \mu)^2 f(t) dt = \int_{t=-\infty}^{\infty} e(t)^2 f(t) dt = \sigma_e^2. \tag{1.6}$$

Given constants α and γ, the above formulas for expectation and dispersion can be summarized in the following properties:

$$E\{\alpha y + \gamma\} = \alpha E\{y\} + \gamma, \tag{1.7a}$$

$$D\{\alpha y + \gamma\} = \alpha^2 D\{y\}. \tag{1.7b}$$

Equation (1.7b) is *the law of error propagation* in its simplest form. It shows that, in contrast to the expectation, the dispersion operator is not linear. Also, it shows that dispersion is not affected by a constant offset.

Another useful formula for the dispersion is derived as follows:

$$
\begin{aligned}
D\{y\} &= E\{(y - E\{y\})^2\} = && (1.8a)\\
&= E\{y^2 - 2yE\{y\} + E\{y\}^2\} = \\
&= E\{y^2 - 2y\mu + \mu^2\} = \\
&= E\{y^2\} - 2\mu E\{y\} + E\{\mu^2\} = \\
&= E\{y^2\} - 2\mu^2 + \mu^2 = \\
&= E\{y^2\} - \mu^2 = \sigma_y^2 && (1.8b)
\end{aligned}
$$

1.1.2 Mean Squared Error

The mean squared error, or MSE, of y is the expectation of the square of the difference of y and its *true value* μ (compare to (1.8a)).

$$\mathrm{MSE}\{y\} = E\{(y - \mu)^2\} \tag{1.9}$$

It is useful to express the MSE as a combination of the dispersion and a (squared) *bias* term. This is done as follows:

$$
\begin{aligned}
\mathrm{MSE}\{y\} &= E\{(y - \mu)^2\} = E\{[(y - E\{y\}) - (\mu - E\{y\})]^2\} = \\
&= E\{(y - E\{y\})^2 - 2(y - E\{y\})(\mu - E\{y\}) + (\mu - E\{y\})^2\} = \\
&= E\{(y - E\{y\})^2\} - 2E\{(y - E\{y\})(\mu - E\{y\})\} + E\{(\mu - E\{y\})^2\}.
\end{aligned}
$$

Note that while y is a random variable, $E\{y\}$ is not. So in the middle term, the expectation operator only applies to y. Thus, we may continue with

$$\mathrm{MSE}\{y\} = D\{y\} - 2(E\{y\} - E\{y\})(\mu - E\{y\}) + (\mu - E\{y\})^2 = D\{y\} + \beta^2, \tag{1.10}$$

where bias is defined formally as

$$\beta \coloneqq E\{\mu - y\} = \mu - E\{y\}. \tag{1.11}$$

Thus, we see that the dispersion of y and the MSE of y are only equal in the absence of bias. Or in other words, only if indeed $\mu = E\{y\}$.

We noted previously that dispersion (variance) is an indicator of precision. In contrast, MSE is a measure of accuracy; it includes both dispersion and bias terms. In general, it is harder to meet accuracy standards than precision standards. We can typically increase our precision by making more observations (though this may come with additional costs in time and money); however it may be very difficult to determine the origin of the bias.

Finally, we note that the square root of MSE is commonly called rms (root mean square). Thus, standard deviation and rms are only equivalent in the absence of bias.

1.2 Multivariate Case

The multivariate case deals with multiple random variables. These variables are typically collected in a column vector. For example, multiple observations of the observable μ in (1.4) can be expressed in the following system of equations:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{\tau}\mu + \boldsymbol{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}, \tag{1.12}$$

where $\boldsymbol{\tau}$ is the summation vector defined as $\boldsymbol{\tau} \coloneqq [1, \ldots, 1]^T$. In the case of unbiased observations, the expectation of the random error vector $\boldsymbol{e}$ is written as

$$E\{\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}\} = \begin{bmatrix} E\{e_1\} \\ \vdots \\ E\{e_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{1.13}$$

And, for the dispersion of each element e_j of $\boldsymbol{e}$ we have

$$D\{e_j\} = E\{\big(e_j - E\{e_j\}\big)^2\} = E\{e_j^2\} = \sigma_j^2. \tag{1.14}$$

For the multivariate case, we introduce the concept of *covariance*, which is a measure of similar behavior between random variables, e.g., between elements e_j and e_k of $\boldsymbol{e}$. Formally, the definition of covariance is

$$C\{e_j, e_k\} = \sigma_{jk} \coloneqq E\{\big(e_j - E\{e_j\}\big)\big(e_k - E\{e_k\}\big)\}. \tag{1.15a}$$

Obviously,

$$C\{e_j, e_k\} = C\{e_k, e_j\}. \tag{1.15b}$$

For a random error vector $\boldsymbol{e}$ having zero expectation, the covariance reduces to

$$C\{e_j, e_k\} = E\{e_j e_k\}, \tag{1.15c}$$

since $E\{e_j\} = E\{e_k\} = 0$. Even though we see from the definition of the covariance (1.15a) that it does not depend on bias, in practice we often find that bias appears as positive correlation.

Two random variables are said to be independent if their joint probability distribution is equal to the product of their individual probability distributions. Mathematically, this is written as

$$f\{e_j, e_k\} = f(e_j)f(e_k) \Leftrightarrow e_j \text{ and } e_k \text{ are independent.} \tag{1.16}$$

If two random variables are independent, their covariance is zero. The converse is not true unless the random variables are normally distributed.

In light of the concept of covariance, the dispersion of a vector of random variables is represented by a matrix. The jth diagonal element of the matrix is denoted by σ_j^2 and the j, k off-diagonal term is written as σ_{jk}. The matrix is called a *covariance matrix* and is represented by Σ. Due to (1.15b),

the covariance matrix is symmetrical. An explicit representation of the covariance matrix Σ is given by

$$D\{\underset{n\times 1}{\boldsymbol{y}}\} = \underset{n\times n}{\Sigma} := \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix} = \Sigma^T. \tag{1.17}$$

Obviously, if the random variables are uncorrelated, the covariance matrix is diagonal.

An important property of a covariance matrix is that it must be at least positive semidefinite. A matrix is positive semidefinite if, and only if, all of its eigenvalues are non-negative. For many applications in geodetic science, the covariance matrix is positive definite, which means that all its eigenvalues are greater than zero. The following statements hold for any positive definite matrix Σ:

- $\boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha} = 0 \Rightarrow \boldsymbol{\alpha} = \mathbf{0}$.
- Σ is a non-singular matrix (also called a *regular* matrix).
- All eigenvalues of Σ are positive and non-zero.
- All principle submatrices of Σ are also positive definite.

In the following chapters, we usually factor out of the covariance matrix Σ a scalar term denoted by σ_0^2, called a variance component, with the resulting matrix called the *cofactor* matrix. We label the cofactor matrix as Q; its inverse is labeled P and is called the *weight matrix*. The relations between these terms are written mathematically as

$$\Sigma = \sigma_0^2 Q = \sigma_0^2 P^{-1}. \tag{1.18}$$

For the remainder of this chapter, we only concern ourselves with the covariance matrix Σ.

1.2.1 The Cauchy-Schwartz Inequality and the Correlation Matrix

We omit the independent-variable argument t from the random variables in the following Cauchy-Schwartz inequality:

$$C\{e_j, e_k\} = \iint e_j e_k f(e_j, e_k) dt_j dt_k = \sigma_{jk} \leq \sqrt{\int e_j^2 f(e_j) dt_j \int e_k^2 f(e_k) dt_k} = \sqrt{\sigma_j^2 \sigma_k^2}. \tag{1.19}$$

The above inequality leads to the notion of a *correlation coefficient*, defined as

$$\rho_{jk} := \frac{\sigma_{jk}}{\sqrt{\sigma_j^2 \sigma_k^2}} \quad \text{for all } j \neq k, \quad \text{with } -1 \leq \rho_{jk} \leq 1. \tag{1.20}$$

Analogous to the covariance matrix, we may form a matrix of correlation coefficients. Such a matrix is called a *correlation matrix* and is defined as

$$\underset{n\times n}{R} := \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix} = R^T. \tag{1.21}$$

Given a covariance matrix Σ, the correlation matrix can easily be generated by

$$R = \operatorname{diag}\big(\big[1/\sigma_1, \ldots, 1/\sigma_n\big]\big) \cdot \Sigma \cdot \operatorname{diag}\big(\big[1/\sigma_1, \ldots, 1/\sigma_n\big]\big). \tag{1.22}$$

Chapter 2

The Model of Direct Observations

When an unknown parameter can be observed directly, the Model of Direct Observations can be formed by

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mu + e_1 \\ \vdots \\ \mu + e_n \end{bmatrix} = \boldsymbol{\tau}\mu + \boldsymbol{e}, \tag{2.1a}$$

$$\boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 Q\right), \;\; Q \coloneqq P^{-1}. \tag{2.1b}$$

The terms in the observational model are defined as follows:

$\boldsymbol{y}$ is an $n \times 1$ vector of observations.

μ is an unknown, non-random parameter to estimate.

$\boldsymbol{\tau}$ is an $n \times 1$ vector of ones, i.e., $\boldsymbol{\tau} = [1, \ldots, 1]^T$.

$\boldsymbol{e}$ is an $n \times 1$ vector of unknown, random errors.

Q is an $n \times n$ cofactor matrix associated with $\boldsymbol{e}$.

P is an $n \times n$ positive-definite weight matrix.

σ_0^2 is an unknown variance component.

2.1 The Least-Squares Solution

We wish to minimize the quadratic form $\boldsymbol{e}^T P \boldsymbol{e}$ (weighted sum of squared errors) subject to the observation equation $\boldsymbol{y} = \boldsymbol{\tau}\mu + \boldsymbol{e}$. This minimization leads to a *LEast-Squares Solution* (LESS) for the unknown parameter μ.

The Lagrange target function (or Lagrangian function)

$$\Phi(\boldsymbol{e}, \boldsymbol{\lambda}, \mu) := \boldsymbol{e}^T P \boldsymbol{e} + 2\boldsymbol{\lambda}^T(\boldsymbol{y} - \boldsymbol{\tau}\mu - \boldsymbol{e}) = \text{ stationary} \tag{2.2}$$

is formed to minimize $\boldsymbol{e}^T P \boldsymbol{e}$ and is said to be stationary with respect to $\boldsymbol{e}$, $\boldsymbol{\lambda}$, and μ. Thus its first partial derivatives are set equivalent to zero, leading to the following Euler-Lagrange necessary conditions (also called first-order conditions):

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{e}} = \frac{1}{2}\left[\frac{\partial \Phi}{\partial e_j}\right]_{n\times 1} = P\tilde{\boldsymbol{e}} - \hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{2.3a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \frac{1}{2}\left[\frac{\partial \Phi}{\partial \lambda_j}\right]_{n\times 1} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu} - \tilde{\boldsymbol{e}} \doteq \boldsymbol{0}, \tag{2.3b}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \mu} = \boldsymbol{\tau}^T\hat{\boldsymbol{\lambda}} \doteq 0. \tag{2.3c}$$

The sufficient condition for minimization is satisfied by the fact that the second partial derivative of Φ is $\partial\Phi^2/\partial\boldsymbol{e}\partial\boldsymbol{e}^T = 2P$, where the weight matrix P is positive definite according to (2.1). Therefore, the solution to the system of equations (2.3) yields the minimum of Φ, and thus the weighted sum of squared residuals $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$ (weighted SSR) is also minimized. See Appendix A for comments on derivatives of quadratic functions with respect to column vectors.

Note that hat and tilde marks are used to denote that variables are associated with a particular solution, i.e., the solution to the first-order condition equations (2.3), which is a least-squares solution (LESS). Throughout these notes, we use a hat for an *estimate* of a non-random variable, whereas a tilde denotes a *prediction* of a random variable. Note that for the vector $\tilde{\boldsymbol{e}}$ we use synonymously the terms *residual* and *predicted error*.

Now we must solve the system of equations (2.3) to obtain a least-squares solution.

$$\hat{\boldsymbol{\lambda}} = P\tilde{\boldsymbol{e}} = P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) \qquad \text{using (2.3a) and (2.3b)} \tag{2.4a}$$

$$\boldsymbol{\tau}^T\hat{\boldsymbol{\lambda}} = \boldsymbol{\tau}^T P\boldsymbol{y} - \big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)\hat{\mu} = 0 \qquad \text{using the previous result and (2.3c)} \tag{2.4b}$$

Equations (2.4a) and (2.4b) lead to

$$\boxed{\hat{\mu} = \frac{\boldsymbol{\tau}^T P\boldsymbol{y}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}} \tag{2.5}$$

for the estimate of the unknown parameter μ. And, from (2.3b), we have

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}} \tag{2.6}$$

for the prediction of the random error vector $\boldsymbol{e}$. The prediction $\tilde{\boldsymbol{e}}$ is also called residual vector.

We say that the quantities $\hat{\mu}$, $\tilde{\boldsymbol{e}}$, and $\hat{\lambda}$ belong to a LEast-Squares Solution (LESS) within the Model of Direct Observations (2.1).

The vectors $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{e}}$ are said to be *P-orthogonal* since

$$\boldsymbol{\tau}^T P\tilde{\boldsymbol{e}} = \boldsymbol{\tau}^T P\big[I_n - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\big]\boldsymbol{y} = \boldsymbol{\tau}^T P\boldsymbol{y} - \boldsymbol{\tau}^T P\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\boldsymbol{y} = 0. \tag{2.7}$$

The vector $\boldsymbol{\tau}\hat{\mu}$ on the right side of (2.6) is called the vector of *adjusted observations*. Obviously, since $\boldsymbol{\tau}^T P\tilde{\boldsymbol{e}} = 0$, we also have

$$\big(\boldsymbol{\tau}\hat{\mu}\big)^T P\tilde{\boldsymbol{e}} = 0. \tag{2.8}$$

Equation (2.8) is an important characteristic of LESS; viz., the vectors of adjusted observations and P-weighted residuals are orthogonal to one another.

In addition to solving for the estimated parameter $\hat{\mu}$ and the predicted random error vector $\tilde{\boldsymbol{e}}$, we are typically interested in their dispersions (variances), which is an indicator of their precision. To compute their dispersions, we apply the law of covariance propagation.

First, for the dispersion of the estimated parameter $\hat{\mu}$ we have

$$D\{\hat{\mu}\} = \frac{\boldsymbol{\tau}^T P}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} D\{\boldsymbol{y}\} \frac{P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} = \frac{\boldsymbol{\tau}^T P\big(\sigma_0^2 P^{-1}\big)P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}\,\boldsymbol{\tau}^T P\boldsymbol{\tau}} = \frac{\sigma_0^2}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}. \tag{2.9}$$

The $n \times n$ dispersion matrix for the residual vector $\tilde{\boldsymbol{e}}$ is derived by

$$\begin{aligned}
D\{\tilde{\boldsymbol{e}}\} &= D\{\big[I_n - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\big]\boldsymbol{y}\} = \\
&= \big[I_n - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\big] D\{\boldsymbol{y}\} \big[I_n - P\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T\big] = \\
&= \sigma_0^2\big[P^{-1} - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T\big]\big[I_n - P\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T\big] = \\
&= \sigma_0^2\big[P^{-1} - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T\big] - \sigma_0^2\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T + \sigma_0^2\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T = \\
&= \sigma_0^2\big[P^{-1} - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T\big].
\end{aligned} \tag{2.10}$$

Formally, neither (2.9) nor (2.10) can be computed, since the variance component σ_0^2 is unknown, though it can be replaced by its estimate shown in (2.16). From (2.10) we see that the dispersion (variance) of the jth element of $\tilde{\boldsymbol{e}}$ is

$$\sigma_{\tilde{\boldsymbol{e}}_j}^2 = \sigma_0^2\Big(\sigma_{jj}^2 - \frac{1}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}\Big), \tag{2.11}$$

where σ_{jj}^2 is the jth diagonal element of P^{-1}, and σ_0^2 is the variance component from the model (2.1). Thus it is apparent that the variance of the jth element of the residual vector $\tilde{\boldsymbol{e}}$ is smaller than the variance of the corresponding jth element of the true, but unknown, random error vector $\boldsymbol{e}$.

2.2 Best Linear Uniformly Unbiased Estimate

Here we take a statistical approach to estimating the unknown parameter μ. We want to find an estimate for μ, expressed as a linear combination of the observations $\boldsymbol{y}$, that extracts the "best" information from the data. The estimate is denoted by $\hat{\mu}$ and is characterized as the *Best Linear Uniformly Unbiased Estimate* (BLUUE) of μ.

1. Linear criterion:

$$\hat{\mu} = \boldsymbol{\alpha}^T\boldsymbol{y}, \quad \text{with } \boldsymbol{\alpha} \text{ to be determined.}$$

2. Uniformly Unbiased criterion:

$$\mu = E\{\hat{\mu}\} = E\{\boldsymbol{\alpha}^T\boldsymbol{y}\} = \boldsymbol{\alpha}^T E\{\boldsymbol{y}\} = \boldsymbol{\alpha}^T E\{\boldsymbol{\tau}\mu + \boldsymbol{e}\} = \boldsymbol{\alpha}^T\boldsymbol{\tau}\mu, \text{ for } \textit{any } \mu \in \mathbb{R} \Leftrightarrow \boldsymbol{\alpha}^T\boldsymbol{\tau} = 1.$$

   Requiring this condition to hold for *any* $\mu \in \mathbb{R}$ satisfies the "uniform" criterion, whereas the requirement that $\boldsymbol{\alpha}^T\boldsymbol{\tau} = 1$ satisfies the "unbiased" criterion.

3. Best criterion: The best criterion requires minimum MSE($\hat{\mu}$), or, equivalently, minimum dispersion, *since $\hat{\mu}$ is unbiased*. Mathematically, the criterion reads

$$\min D\{\hat{\mu}\} = D\{\boldsymbol{\alpha}^T\boldsymbol{y}\} = \boldsymbol{\alpha}^T D\{\boldsymbol{y}\}\boldsymbol{\alpha} = \sigma_0^2\boldsymbol{\alpha}^T Q\boldsymbol{\alpha} \text{ subject to } \boldsymbol{\tau}^T\boldsymbol{\alpha} = 1.$$

A Lagrange target function is formed by

$$\Phi(\boldsymbol{\alpha}, \lambda) := \boldsymbol{\alpha}^T Q\boldsymbol{\alpha} + 2\lambda\big(\boldsymbol{\tau}^T\boldsymbol{\alpha} - 1\big) = \text{stationary}. \tag{2.12}$$

And the Euler-Lagrange necessary conditions result in

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\alpha}} = Q\hat{\boldsymbol{\alpha}} + \boldsymbol{\tau}\hat{\lambda} \doteq 0, \tag{2.13a}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\lambda} = \boldsymbol{\tau}^T\hat{\boldsymbol{\alpha}} - 1 \doteq 0. \tag{2.13b}$$

The sufficient condition for minimization is satisfied by $\partial\Phi^2/(\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T) = 2Q$, which is a positive definite matrix according to (2.1).

Solving (2.13a) and (2.13b) simultaneously yields

$$\hat{\boldsymbol{\alpha}} = -Q^{-1}\boldsymbol{\tau}\hat{\lambda} = -P\boldsymbol{\tau}\hat{\lambda} \text{ using (2.13a)}, \tag{2.14a}$$

$$1 = \boldsymbol{\tau}^T\hat{\boldsymbol{\alpha}} = -\boldsymbol{\tau}^T P\boldsymbol{\tau}\hat{\lambda} \Rightarrow \hat{\lambda} = \frac{-1}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} \text{ using (2.13b) and (2.14a)}. \tag{2.14b}$$

Substituting (2.14b) into (2.14a) we get

$$\hat{\boldsymbol{\alpha}} = \frac{P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}. \tag{2.14c}$$

Finally, substituting the transpose of (2.14c) into the linear requirement $\hat{\mu} = \boldsymbol{\alpha}^T\boldsymbol{y}$ yields the BLUUE of μ as

$$\boxed{\hat{\mu} = \frac{\boldsymbol{\tau}^T P}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}\boldsymbol{y}.} \tag{2.15}$$

Equation (2.15) agrees with (2.5) derived for LESS. Thus we see that the LESS and the BLUUE are equivalent within the Model of Direct Observations.

We may also prove mathematically that (2.15) fulfills the weighted LESS principle by showing that the P-weighted residual norm for any other solution is larger than that obtained via BLUUE, which we do in the following:

Suppose $\hat{\hat{\mu}}$ is any other estimate for μ, then

$$\begin{aligned}
\tilde{\hat{\boldsymbol{e}}}^T P \tilde{\hat{\boldsymbol{e}}} &= \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\hat{\mu}}\big)^T P \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\hat{\mu}}\big) = \\
&= \big[\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) - \boldsymbol{\tau}\big(\hat{\hat{\mu}} - \hat{\mu}\big)\big]^T P \big[\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) - \boldsymbol{\tau}\big(\hat{\hat{\mu}} - \hat{\mu}\big)\big] = \\
&= \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)^T P \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) - 2\big(\hat{\hat{\mu}} - \hat{\mu}\big)\boldsymbol{\tau}^T P \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) + \big(\boldsymbol{\tau}^T P \boldsymbol{\tau}\big)\big(\hat{\hat{\mu}} - \hat{\mu}\big)^2 = \\
&= \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)^T P \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) + \big(\boldsymbol{\tau}^T P \boldsymbol{\tau}\big)\big(\hat{\hat{\mu}} - \hat{\mu}\big)^2 \geq \\
&\geq \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)^T P \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) \\
&\text{Q.E.D}
\end{aligned}$$

We have used the P-orthogonality relation (2.7) in the third line of the proof.

2.3 Estimated Variance Component

The variance component σ_0^2 is an unknown quantity in model (2.1), though in practice its expected value is often one. The variance component can be estimated as a function of the P-weighted norm of the residual vector $\tilde{\boldsymbol{e}}$ and the degrees of freedom $n-1$. Its formula is given here without derivation. Later, in Section 3.2, it is derived within the Gauss-Markov Model.

$$\boxed{\hat{\sigma}_0^2 := \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n-1}} \tag{2.16}$$

2.4 Effect of Wrongly Chosen Weight Matrix in the Model of Direct Observations

Assume that the weight matrix P has been wrongly chosen by δP, where δP is assumed to be small, positive semi-definite, and *uncorrelated* with P. Consequently, we have

$$P \to (P + \delta) \Rightarrow \hat{\mu} \to (\hat{\mu} + \delta\hat{\mu}) \;\text{ and }\; D\{\hat{\mu}\} \to D\{\hat{\mu} + \delta\hat{\mu}\} \;\text{ and }\; \hat{\sigma}_0^2 \to \hat{\sigma}_0^2 + \delta\hat{\sigma}_0^2.$$

2.4.1 Effect on the Parameter Estimate

The following shows the effect of a wrongly chosen weight matrix on the estimated parameter $\hat{\mu}$:

$$
\begin{aligned}
\big(\hat{\mu} + \delta\hat{\mu}\big) &= \frac{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} \Rightarrow \\
\delta\hat{\mu} &= \frac{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} - \hat{\mu} = \frac{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} \cdot \frac{\boldsymbol{\tau}^T P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} - \big(\frac{\boldsymbol{\tau}^T P\boldsymbol{y}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}\big) \cdot \frac{\boldsymbol{\tau}^T\big(P+\delta P\big)\boldsymbol{\tau}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} = \\
&= \frac{\cancel{\boldsymbol{\tau}^T P\boldsymbol{y}\boldsymbol{\tau}^T P\boldsymbol{\tau}} + \boldsymbol{\tau}^T\delta P\boldsymbol{y}\boldsymbol{\tau}^T P\boldsymbol{\tau} - \cancel{\boldsymbol{\tau}^T P\boldsymbol{y}\boldsymbol{\tau}^T P\boldsymbol{\tau}} - \boldsymbol{\tau}^T P\boldsymbol{y}\boldsymbol{\tau}^T\delta P\boldsymbol{\tau}}{(\boldsymbol{\tau}^T P\boldsymbol{\tau})\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} = \\
&= \frac{\boldsymbol{\tau}^T\delta P\boldsymbol{y}}{\boldsymbol{\tau}^T\big(P+\delta P\big)\boldsymbol{\tau}} - \frac{\boldsymbol{\tau}^T\delta P\boldsymbol{\tau}\hat{\mu}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} = \frac{\boldsymbol{\tau}^T\delta P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}}
\end{aligned}
$$

Finally, we arrive at

$$
\delta\hat{\mu} = \frac{\boldsymbol{\tau}^T\delta P}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}}\tilde{\boldsymbol{e}}. \tag{2.17}
$$

2.4.2 Effect on the Cofactor Matrix for the Estimated Parameter

The following shows the effect of a wrongly chosen weight matrix on the cofactor matrix $Q_{\hat{\mu}}$ for the estimated parameter $\hat{\mu}$, where $D\{\hat{\mu}\} = \sigma_0^2 Q_{\hat{\mu}}$ is the dispersion of $\hat{\mu}$:

$$
\begin{aligned}
\delta Q_{\hat{\mu}} &= \big(Q_{\hat{\mu}} + \delta Q_{\hat{\mu}}\big) - Q_{\hat{\mu}} = \frac{1}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} - \frac{1}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} = \\
&= \frac{\boldsymbol{\tau}^T P\boldsymbol{\tau} - \boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}}{\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} = \frac{-\boldsymbol{\tau}^T\delta P\boldsymbol{\tau}}{\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}}.
\end{aligned}
$$

Thus we have

$$
\delta Q_{\hat{\mu}} = -\frac{\boldsymbol{\tau}^T\delta P\boldsymbol{\tau}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}} Q_{\hat{\mu}}. \tag{2.18}
$$

2.4.3 Effect on the Estimated Variance Component

The following shows the effect of a wrongly chosen weight matrix on the estimated variance component:

First note that

$$
\begin{aligned}
\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} &= \big(\boldsymbol{y}^T - \hat{\mu}\boldsymbol{\tau}^T\big)P\,(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}) = \\
&= \boldsymbol{y}^T P(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}) - \hat{\mu}\big(\cancel{\boldsymbol{\tau}^T P\boldsymbol{y} - \boldsymbol{\tau}^T P\boldsymbol{\tau}\frac{\boldsymbol{\tau}^T P\boldsymbol{y}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}}\big) = \\
&= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T P\boldsymbol{\tau}\hat{\mu} = \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{\tau}^T P\boldsymbol{y}\hat{\mu} = \boldsymbol{y}^T P\boldsymbol{y} - \hat{\mu}^2\boldsymbol{\tau}^T P\boldsymbol{\tau}.
\end{aligned}
$$

Following the above logic, we have

$$(n-1)\big(\hat{\sigma}_0^2 + \delta\hat{\sigma}_0^2\big) = \boldsymbol{y}^T(P+\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}\big(\hat{\mu}+\delta\hat{\mu}\big) \Rightarrow$$
$$\Rightarrow (n-1)\delta\hat{\sigma}_0^2 = \boldsymbol{y}^T(\cancel{P}+\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}(\hat{\mu}+\delta\hat{\mu}) - \cancel{\boldsymbol{y}^T P\boldsymbol{y}} + (\boldsymbol{\tau}^T P\boldsymbol{y})\hat{\mu} =$$

(Note: the last term will cancel one of the four terms in the binomial product.)

$$= \boldsymbol{y}^T\big(\delta P\big)\boldsymbol{y} - \boldsymbol{\tau}^T\big)\delta P\big)\boldsymbol{y}\big(\hat{\mu}+\delta\hat{\mu}\big) - \big(\boldsymbol{\tau}^T P\boldsymbol{y}\big)\delta\hat{\mu} =$$
$$= \boldsymbol{y}^T(\delta P)\boldsymbol{y} - \hat{\mu}\boldsymbol{\tau}^T(\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}\delta\hat{\mu} =$$
$$= \big(\boldsymbol{y}^T - \hat{\mu}\boldsymbol{\tau}^T\big)(\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}\delta\hat{\mu} =$$
$$= \tilde{\boldsymbol{e}}^T(\delta P)\boldsymbol{y} - \frac{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}}\boldsymbol{\tau}^T(\delta P)\tilde{\boldsymbol{e}} =$$

(Where the previous results for $\delta\hat{\mu}$ have been substituted in the line above.)

$$= \boldsymbol{y}^T(\delta P)\tilde{\boldsymbol{e}} - \big(\hat{\mu}+\delta\hat{\mu}\big)\boldsymbol{\tau}^T(\delta P)\tilde{\boldsymbol{e}} =$$
$$\text{(Using } \boldsymbol{y}^T(\delta P)\tilde{\boldsymbol{e}} = \big(\hat{\mu}\boldsymbol{\tau}^T + \tilde{\boldsymbol{e}}^T\big)\delta P\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{e}}^T\delta P\tilde{\boldsymbol{e}} + \hat{\mu}\boldsymbol{\tau}^T\delta P\tilde{\boldsymbol{e}}\text{)}$$
$$= \tilde{\boldsymbol{e}}^T(\delta P)\tilde{\boldsymbol{e}} - \delta\hat{\mu}\boldsymbol{\tau}^T(\delta P)\tilde{\boldsymbol{e}} =$$
$$= \tilde{\boldsymbol{e}}^T(\delta P)\tilde{\boldsymbol{e}} - \big(\delta\hat{\mu}\big)^2\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}$$

Finally, we arrive at

$$\delta\hat{\sigma}_0^2 = \frac{1}{n-1}\big[\tilde{\boldsymbol{e}}^T(\delta P)\tilde{\boldsymbol{e}} - \big(\delta\hat{\mu}\big)^2\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}\big]. \tag{2.19}$$

2.4.4 Effect on the Estimated Dispersion

The the effect of a wrongly chosen weight matrix on the estimated dispersion of $\hat{\mu}$ is obviously given by

$$\hat{D}\{\hat{\mu}+\delta\hat{\mu}\} = \big(\hat{\sigma}_0^2 + \delta\hat{\sigma}_0^2\big)D\{\hat{\mu}+\delta\hat{\mu}\} = \big(\hat{\sigma}_0^2 + \delta\hat{\sigma}_0^2\big)\big(Q_{\hat{\mu}} + \delta Q_{\hat{\mu}}\big). \tag{2.20}$$

Chapter 3

The Gauss-Markov Model

The Gauss-Markov Model (GMM) is the underlying observational model for many of the topics that follow. In presentation of the model, it is assumed that the observation equations have been linearized, if necessary. The model is written as follows:

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \quad \operatorname{rk} A =: q \leq \{m, n\}, \tag{3.1a}$$

$$\boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 P^{-1}\right). \tag{3.1b}$$

Equation (3.1a) shows the general case, where the *coefficient matrix* A may or may not be of full column rank. Because of linearization, $\boldsymbol{y}$ is a vector of observations minus "zeroth-order" terms; A is the (known) $n \times m$ coefficient matrix (also called *design* or *information* matrix, or *Jacobian* matrix if partial derivatives are involved) containing first-order derivatives of the observations with respect to the m unknown parameters; $\boldsymbol{\xi}$ is a vector of unknown parameters to estimate (corrections to initial values), and $\boldsymbol{e}$ is a vector of random observation errors, having zero expectation.

The $n \times n$ matrix P is symmetric. It contains weights of the observations, which may be correlated. The inverse of P shown in (3.1) implies that P is a positive-definite matrix; this inverse matrix is called the cofactor matrix and is denoted by Q. The symbol σ_0^2 represents a variance component, which is considered unknown but can be estimated. The dispersion matrix $D\{\boldsymbol{e}\} = \sigma_0^2 P^{-1}$ is called the variance-covariance matrix, or simply the covariance matrix, and is denoted as Σ. In summary, we have the following relation between the dispersion, weight, and cofactor matrices of the unknown, random error vector $\boldsymbol{e}$:

$$D\{\boldsymbol{e}\} = \Sigma = \sigma_0^2 Q = \sigma_0^2 P^{-1}. \tag{3.2}$$

The letter q denotes the rank of matrix A. The redundancy r of the system of equations in (3.1a) is defined as

$$r := n - \operatorname{rk} A = n - q. \tag{3.3}$$

Redundancy is also called the *degrees of freedom* in the context of statistical testing discussed in Chapter 10.

The observational model (3.1) has two main components. The first component, (3.1a), shows the functional relation between the observations, and their random errors, to the unknown parameters that are to be estimated. The second component, (3.1b) expresses the expectation and dispersion of the observations due to their random errors, which are called the first and second moments, respectively, of the random error vector $\boldsymbol{e}$.

If the rank of matrix A, namely q, is less than the number of unknown parameters to estimate, m, we say that the problem is rank deficient. It cannot be solved based on the observations alone; additional information about the unknown parameters must be provided. The problem of rank deficiency is covered in Chapter 6 and more thoroughly in the notes for the advanced adjustments course.

3.1 The Least-Squares Solution Within the Gauss-Markov Model

We now derive the LEast-Squares Solution (LESS) for the parameter estimate $\hat{\boldsymbol{\xi}}$ and the predicted random error (residual) vector $\tilde{\boldsymbol{e}}$ with their associated dispersion matrices under the (more restrictive) assumption that $\operatorname{rk} A = m$, i.e., the coefficient matrix A has full column rank. For convenience, we define the $m \times m$ matrix N and the $m \times 1$ vector $\boldsymbol{c}$ as

$$\big[N,\, \boldsymbol{c}\big] \coloneqq A^T P\big[A,\, \boldsymbol{y}\big]. \tag{3.4}$$

To obtain a least-squares solution for $\boldsymbol{\xi}$, the Lagrange target function

$$\Phi(\boldsymbol{\xi}) \coloneqq (\boldsymbol{y} - A\boldsymbol{\xi})^T P(\boldsymbol{y} - A\boldsymbol{\xi}) = \text{stationary} \tag{3.5}$$

must to be minimized.

Forming the The Euler-Lagrange necessary conditions (or first-order conditions) leads directly to the least-squares *normal equations*

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\xi}} = \big(A^T P A\big)\hat{\boldsymbol{\xi}} - A^T P \boldsymbol{y} = N\hat{\boldsymbol{\xi}} - \boldsymbol{c} \doteq \boldsymbol{0}. \tag{3.6}$$

The sufficient condition is satisfied by $(1/2)\cdot(\partial^2\Phi/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T) = N$, which is positive definite since matrix A has full column rank. Equation (3.6) leads to the least-squares solution (LESS)

$$\boxed{\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c},} \tag{3.7}$$

for the unknown parameter vector $\boldsymbol{\xi}$, with its expectation computed by

$$E\{\hat{\boldsymbol{\xi}}\} = N^{-1}E\{\boldsymbol{c}\} = N^{-1}A^T P E\{\boldsymbol{y}\} = N^{-1}A^T P A\boldsymbol{\xi} = \boldsymbol{\xi}. \tag{3.8}$$

The predicted random error vector (also called *residual* vector) is then given by

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{y},} \tag{3.9}$$

with expectation

$$E\{\tilde{\boldsymbol{e}}\} = \big(I_n - AN^{-1}A^TP\big)E\{\boldsymbol{y}\} = \big(I_n - AN^{-1}A^TP\big)A\boldsymbol{\xi} = A\boldsymbol{\xi} - A\boldsymbol{\xi} = \mathbf{0}. \tag{3.10}$$

The adjusted observation vector (also called vector of *predicted* observations) is given by

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - \tilde{\boldsymbol{e}} = A\hat{\boldsymbol{\xi}}, \tag{3.11}$$

with expectation

$$E\{\tilde{\boldsymbol{y}}\} = AE\{\hat{\boldsymbol{\xi}}\} = A\boldsymbol{\xi}. \tag{3.12}$$

Equations (3.8), (3.10) and (3.12) show that the estimated parameters, the residuals, and the adjusted observations, respectively, are unbiased.

The corresponding dispersion matrices are computed by using the law of variance propagation. The dispersion of the estimated parameters is computed by

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}\} = D\{N^{-1}A^TP\boldsymbol{y}\} = \big(N^{-1}A^TP\big)D\{\boldsymbol{y}\}\big(PAN^{-1}\big) = \\ = N^{-1}A^TP\big(\sigma_0^2P^{-1}\big)PAN^{-1} = \sigma_0^2N^{-1}. \end{aligned} \tag{3.13}$$

And, the dispersion of the residual vector $\tilde{\boldsymbol{e}}$ is

$$\begin{aligned} D\{\tilde{\boldsymbol{e}}\} &= \big(I_n - AN^{-1}A^TP\big)D\{\boldsymbol{y}\}\big(I_n - PAN^{-1}A^T\big) = \\ &= \big(I_n - AN^{-1}A^TP\big)\big(\sigma_0^2P^{-1}\big)\big(I_n - PAN^{-1}A^T\big) = \\ &= \sigma_0^2\big(I_n - AN^{-1}A^TP\big)\big(P^{-1} - AN^{-1}A^T\big) = \\ &= \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big) = \\ &= D\{\boldsymbol{y}\} - D\{A\hat{\boldsymbol{\xi}}\} = \sigma_0^2Q_{\tilde{\boldsymbol{e}}}, \end{aligned} \tag{3.14}$$

where the matrix $Q_{\tilde{\boldsymbol{e}}} \coloneqq P^{-1} - AN^{-1}A^T$ is the cofactor matrix of the residual vector $\tilde{\boldsymbol{e}}$. Equation (3.14) reveals that the variances of the residuals are smaller than the corresponding variances of the observations.

Finally, the dispersion of the residual vector is computed by

$$D\{\tilde{\boldsymbol{y}}\} = AD\{\hat{\boldsymbol{\xi}}\}A^T = \sigma_0^2AN^{-1}A^T. \tag{3.15}$$

Summarizing the above equations, the respective distributions for the estimated parameter vector, the residual vector, and the vector of adjusted observations are succinctly expressed by

$$\hat{\boldsymbol{\xi}} \sim \big(\boldsymbol{\xi}, \sigma_0^2N^{-1}\big), \tag{3.16a}$$

$$\tilde{\boldsymbol{e}} \sim \big(\mathbf{0}, \sigma_0^2\big[P^{-1} - AN^{-1}A^T\big]\big), \tag{3.16b}$$

$$\tilde{\boldsymbol{y}} \sim \big(A\boldsymbol{\xi}, \sigma_0^2AN^{-1}A^T\big). \tag{3.16c}$$

Since the variance component σ_0^2 is an unknown quantity, the dispersions shown in (3.16) cannot be computed unless either σ_0^2 is estimated or a value is assigned to it. See Section 3.2 for the derivation of the variance component estimate $\hat{\sigma}_0^2$.

3.1.1 Correlation of Adjusted Observations and Predicted Residuals

Equation (3.14) implies that the covariance between the vector of *adjusted observations* $A\hat{\boldsymbol{\xi}}$ and the vector of residuals $\tilde{\boldsymbol{e}}$ is zero. Considering that both vectors are a function of the random vector $\boldsymbol{y}$, this can also be shown by the following:

$$\begin{aligned} C\{A\hat{\boldsymbol{\xi}}, \tilde{\boldsymbol{e}}\} &= AN^{-1}A^T P \cdot D\{\boldsymbol{y}\} \cdot \big(I_n - AN^{-1}A^T P\big)^T = \\ &= \sigma_0^2 \big[AN^{-1}A^T - AN^{-1}\big(A^T PA\big)N^{-1}A^T\big] = \\ &= \sigma_0^2 \big[AN^{-1}A^T - AN^{-1}A^T\big] = 0. \end{aligned} \tag{3.17}$$

Also, we have the following covariance between the adjusted and original observations:

$$C\{A\hat{\boldsymbol{\xi}}, \boldsymbol{y}\} = AN^{-1}A^T PD\{\boldsymbol{y}\} = \sigma_0^2 AN^{-1}A^T PP^{-1} = \sigma_0^2 AN^{-1}A^T = D\{A\hat{\boldsymbol{\xi}}\}. \tag{3.18}$$

Zero correlation does not necessarily imply statistical independence, though the converse does hold. Analogous to (10.9a), the adjusted observations and predicted residuals are not statistically independent unless the expectation of their product is equal to the product of their expectations. The following shows that this property is not satisfied: Since the trace of a scalar product is the scalar product itself, we start with

$$E\{(A\hat{\boldsymbol{\xi}})^T \tilde{\boldsymbol{e}}\} = E\{\operatorname{tr} \hat{\boldsymbol{\xi}}^T A^T \big(I_n - AN^{-1}A^T P\big)\boldsymbol{y}\}.$$

But the trace is invariant to a cyclic transformation. Thus,

$$\begin{aligned} E\{(A\hat{\boldsymbol{\xi}})^T \tilde{\boldsymbol{e}}\} &= E\{\operatorname{tr}\big(A^T - A^T AN^{-1}A^T P\big)\boldsymbol{y}\hat{\boldsymbol{\xi}}^T\} = \\ = \operatorname{tr}\big(A^T - A^T AN^{-1}A^T P\big)E\{\boldsymbol{y}\hat{\boldsymbol{\xi}}^T\} &\neq 0 = E\{(A\hat{\boldsymbol{\xi}})^T\}E\{\tilde{\boldsymbol{e}}\}, \text{ since } E\{\tilde{\boldsymbol{e}}\} = \boldsymbol{0}. \end{aligned}$$

3.1.2 P-Weighted Norm of the Residual Vector

The P-weighted norm of the residual vector $\tilde{\boldsymbol{e}}$ is an important quantity that can be used to check the overall ("global") fit of the adjustment. The norm is defined as

$$\Omega := \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}. \tag{3.19}$$

In the special case where $P = I_n$, the quadratic form Ω is often called the *sum of squared residuals*, or SSR, in the statistical literature.

Substituting (3.9) into (3.19) leads to some commonly used alternative forms for Ω.

$$\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} = (\boldsymbol{y} - A\hat{\boldsymbol{\xi}})^T P(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \tag{3.20a}$$

$$= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}}^T A^T P\boldsymbol{y} + \hat{\boldsymbol{\xi}}^T A^T PA\hat{\boldsymbol{\xi}} =$$

$$= \boldsymbol{y}^T P\boldsymbol{y} - 2\boldsymbol{c}^T\hat{\boldsymbol{\xi}} + \boldsymbol{c}^T\hat{\boldsymbol{\xi}} =$$

$$= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\hat{\boldsymbol{\xi}} = \tag{3.20b}$$

$$= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T N^{-1}\boldsymbol{c} = \tag{3.20c}$$

$$= \boldsymbol{y}^T P\boldsymbol{y} - (N\hat{\boldsymbol{\xi}})^T N^{-1}N\hat{\boldsymbol{\xi}} =$$

$$= \boldsymbol{y}^T P\boldsymbol{y} - \hat{\boldsymbol{\xi}}^T N\hat{\boldsymbol{\xi}} = \tag{3.20d}$$

$$= \boldsymbol{y}^T (P - PAN^{-1}A^T P)\boldsymbol{y} \tag{3.20e}$$

Note that the target function (3.5) could have been written explicitly as a function of the random error vector $\boldsymbol{e}$ with the introduction of a vector of Lagrange multipliers $\boldsymbol{\lambda}$ as follows:

$$\Phi(\boldsymbol{e}, \boldsymbol{\xi}, \boldsymbol{\lambda}) = \boldsymbol{e}^T P \boldsymbol{e} - 2\boldsymbol{\lambda}^T(\boldsymbol{y} - A\boldsymbol{\xi} - \boldsymbol{e}) = \text{stationary}. \tag{3.21}$$

This approach leads to the estimate of Lagrange multipliers as $-\hat{\boldsymbol{\lambda}} = P\tilde{\boldsymbol{e}}$ and thus leads to yet another expression for the P-weighted norm

$$\Omega = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = -\tilde{\boldsymbol{e}}^T \hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}^T P^{-1} \hat{\boldsymbol{\lambda}}. \tag{3.22}$$

3.2 Estimated Variance Component within the Gauss-Markov Model

As stated in Section 2.3, the variance component σ_0^2 is an unknown quantity in the GMM. We now present the derivation of the estimated variance component $\hat{\sigma}_0^2$. As defined in (3.1), the dispersion matrix for the random error vector $\boldsymbol{e}$ is $D\{\boldsymbol{e}\} = \sigma_0^2 Q$. Also, by definition of dispersion we have $D\{\boldsymbol{e}\} = E\{(\boldsymbol{e} - E\{\boldsymbol{e}\})(\boldsymbol{e} - E\{\boldsymbol{e}\})^T\}$, but for the error vector $E\{\boldsymbol{e}\} = \boldsymbol{0}$. Therefore $D\{\boldsymbol{e}\} = E\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 Q = \sigma_0^2 P^{-1}$.

The following steps lead to an expression for the variance component σ_0^2 in terms of the quadratic product $\boldsymbol{e}^T P \boldsymbol{e}$.

$$\begin{aligned}
&E\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 Q && \text{(by definition)}\\
&PE\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 I_n && \text{(multiply both sides by } P\text{)}\\
&\operatorname{tr} PE\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 \operatorname{tr} I_n = n\sigma_0^2 && \text{(apply the trace operator)}\\
&\operatorname{tr} E\{P\boldsymbol{e}\boldsymbol{e}^T\} = n\sigma_0^2 && \text{(move the constant matrix } P \text{ into the expectation)}\\
&E\{\operatorname{tr} P\boldsymbol{e}\boldsymbol{e}^T\} = n\sigma_0^2 && \text{(interchange the trace and expectation operators—both linear)}\\
&E\{\operatorname{tr} \boldsymbol{e}^T P\boldsymbol{e}\} = n\sigma_0^2 && \text{(the trace is invariant to a cyclic transformation)}\\
&E\{\boldsymbol{e}^T P\boldsymbol{e}\} = n\sigma_0^2 && \text{(a quadratic product is a scalar; trace of scalar is scalar itself)}\\
&\sigma_0^2 = E\{\frac{\boldsymbol{e}^T P\boldsymbol{e}}{n}\} && \text{(dividing through by } n \text{ and placing } n \text{ inside } E\{\cdot\}\text{)}\\
&\bar{\sigma}_0^2 := \frac{\boldsymbol{e}^T P\boldsymbol{e}}{n} && \text{(define a symbol for the term inside } E\{\cdot\}\text{)}\\
&E\{\bar{\sigma}_0^2\} = \sigma_0^2 && \text{(by substitution)}
\end{aligned}$$

Thus we can say that $\big(\boldsymbol{e}^T P\boldsymbol{e}\big)/n$ is an unbiased estimate of σ_0^2. However, we do not actually know the true random error vector $\boldsymbol{e}$, but we do know its predicted value $\tilde{\boldsymbol{e}}$.

We now work with the residual vector $\tilde{\boldsymbol{e}}$ to find an unbiased estimate of σ_0^2. Combining steps similar to those explained above, we can write

$$E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} = \operatorname{tr} E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} = \operatorname{tr} E\{\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}^T\}P = \operatorname{tr} D\{\tilde{\boldsymbol{e}}\}P. \tag{3.23}$$

According to (3.14), the dispersion of the residual vector is $D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big)$, since N is full rank. Substituting this result into (3.23) gives

$$\begin{aligned} E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} &= \operatorname{tr}\sigma_0^2\big(P^{-1} - AN^{-1}A^T\big)P = \\ &= \sigma_0^2\big(\operatorname{tr} I_n - \operatorname{tr} AN^{-1}A^T P\big) = \\ &= \sigma_0^2\big(\operatorname{tr} I_n - \operatorname{tr} N^{-1}A^T PA\big) = && \text{(using (A.5))} \\ &= \sigma_0^2(n - \operatorname{rk} N) = \sigma_0^2(n - \operatorname{rk} A) \Rightarrow && \text{(using (A.4))} \end{aligned}$$

Finally, we arrive at

$$E\{\frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{n - \operatorname{rk} A}\} = \sigma_0^2. \tag{3.24}$$

Now, we simply label the argument of the expectation operator on the lest side of (3.24) as $\hat{\sigma}_0^2$, which allows us to write the expression for the estimated variance component as

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{n - \operatorname{rk} A}.} \tag{3.25}$$

Obviously, $\hat{\sigma}_0^2$ is an uniformly unbiased estimate of σ_0^2, since $E\{\hat{\sigma}_0^2\} = \sigma_0^2$. In the case of the model of direct observations, we replace A with τ, which has rank of 1, and thus we have $\hat{\sigma}_0^2 := \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}/(n-1)$, which verifies (2.16). Alternative expressions for $\hat{\sigma}_0^2$ can be reached by use of (3.20) and (3.22).

The above derivations imply the following relationship between $E\{\boldsymbol{e}^T P\boldsymbol{e}\}$ and $E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\}$:

$$\frac{E\{\boldsymbol{e}^T P\boldsymbol{e}\}}{n} = \frac{E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\}}{n - \operatorname{rk} A} = \sigma_0^2 \Rightarrow \tag{3.26a}$$

$$\Rightarrow E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} < E\{\boldsymbol{e}^T P\boldsymbol{e}\} \tag{3.26b}$$

According Grafarend and Schaffrin (1993), pg. 103, and Schaffrin (1997b), the dispersion, and estimated dispersion, respectively, of $\hat{\sigma}_0^2$ are given by

$$D\{\hat{\sigma}_0^2\} = (n-m)^{-1} \cdot 2\big(\sigma_0^2\big)^2 \tag{3.27}$$

and

$$\hat{D}\{\hat{\sigma}_0^2\} = (n-m)^{-1} \cdot 2\big(\hat{\sigma}_0^2\big)^2, \tag{3.28}$$

where it is assumed that $m = \operatorname{rk} A$.

Chapter 4

Adjustment by Condition Equations

In the adjustment by condition equations, the unknown parameters $\boldsymbol{\xi}$ are not estimated directly, rather the residual vector $\tilde{\boldsymbol{e}}$ is predicted. This approach might be taken if the parameters are of no particular interest, or it might be done to make the problem easier to formulate. An example of the latter is the adjustment of leveling networks, where the parameters (heights of the stations) are of primary interest, but because closed "level loops" within the network sum to zero (a necessary condition), it is convenient to difference the observations along these loops before performing the adjustment (see level loop example in Chapter 9). Another motivation for forming condition equations is that the size of the matrix to invert in the least-squares solution may become smaller.

Let the $r \times n$ matrix B represent a difference operator such that when it is applied to the $n \times 1$ observation equations $\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}$, the parameters are eliminated. More specifically, we require that $BA = 0$, which implies that $B\boldsymbol{y} = B(A\boldsymbol{\xi} + \boldsymbol{e}) = B\boldsymbol{e}$. Therefore, by applying the difference operator B, the Gauss-Markov Model (GMM) is transformed to the following model of condition equations:

$$\boldsymbol{w} := B\boldsymbol{y} = B\boldsymbol{e}, \tag{4.1a}$$

$$\boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 P^{-1}), \tag{4.1b}$$

$$\operatorname{rk} A = q \leq m < n, \tag{4.1c}$$

$$r := n - q = \operatorname{rk} B, \tag{4.1d}$$

where the variable r denotes the redundancy of the model.

The least-squares criteria is then written as

$$\min \boldsymbol{e}^T P \boldsymbol{e} \ \text{ subject to } \boldsymbol{w} = B\boldsymbol{e}, \tag{4.2}$$

from which the Lagrange target function

$$\Phi(\boldsymbol{e}, \boldsymbol{\lambda}) := \boldsymbol{e}^T P \boldsymbol{e} + 2\boldsymbol{\lambda}^T(\boldsymbol{w} - B\boldsymbol{e}) \tag{4.3}$$

can be written, which is stationary with respect to $\boldsymbol{e}$ and $\boldsymbol{\lambda}$. Taking the first partial derivatives of

(4.3) leads to the Euler-Lagrange necessary conditions

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{e}} = P\tilde{\boldsymbol{e}} - B^T\hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{4.4a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \boldsymbol{w} - B\tilde{\boldsymbol{e}} \doteq \boldsymbol{0}. \tag{4.4b}$$

The sufficient condition, required to ensure a minimum is reached, is satisfied by $\partial\Phi^2/\partial\boldsymbol{e}\partial\boldsymbol{e}^T = 2P$, which is positive definite since the weight matrix P is invertible. The simultaneous solution of (4.4a) and (4.4b) leads to the *Best LInear Prediction* (BLIP) of $\boldsymbol{e}$ as derived in the following: Equation (4.4a) leads to

$$\tilde{\boldsymbol{e}} = P^{-1}B^T\hat{\boldsymbol{\lambda}}. \tag{4.5a}$$

Then, (4.4b) and (4.5a) allows

$$\boldsymbol{w} = B\tilde{\boldsymbol{e}} = \left(BP^{-1}B^T\right)\hat{\boldsymbol{\lambda}} \Rightarrow \tag{4.5b}$$

$$\hat{\boldsymbol{\lambda}} = \left(BP^{-1}B^T\right)^{-1}\boldsymbol{w} \Rightarrow \tag{4.5c}$$

$$\tilde{\boldsymbol{e}} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}\boldsymbol{w}, \tag{4.5d}$$

finally leading to the predicted random error vector

$$\boxed{\tilde{\boldsymbol{e}} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}B\boldsymbol{y}.} \tag{4.5e}$$

Note that $BP^{-1}B^T$ is a symmetric, positive definite matrix of size $r \times r$. The predicted random error vector $\tilde{\boldsymbol{e}}$ is also called the *residual* vector. The vector of adjusted observations follows as

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - \tilde{\boldsymbol{e}}. \tag{4.6}$$

The estimated variance component is given by

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{\operatorname{rk} B} = \frac{\tilde{\boldsymbol{e}}^T B^T\hat{\boldsymbol{\lambda}}}{r} = \frac{\boldsymbol{w}^T\hat{\boldsymbol{\lambda}}}{r} = \frac{\boldsymbol{w}^T(BP^{-1}B^T)^{-1}\boldsymbol{w}}{r}.} \tag{4.7}$$

In words it is described as the P-weighted residual norm divided by the degrees of freedom of the model.

Note that B is not a unique matrix, but regardless of how B is chosen the results will be the same, assuming the following necessary conditions for B are satisfied:

(i) *Dimensionality:* $\operatorname{rk} B = n - \operatorname{rk} A = n - q = r$, which means that $\operatorname{rk} B + \operatorname{rk} A = (n-q) + q = n$.

(ii) *Orthogonality:* $BA = 0$.

As we did earlier within the GMM (Section 3.1.1), we compute the covariance between the residual

vector $\tilde{\boldsymbol{e}}$ and the vector adjusted observations $\boldsymbol{y} - \tilde{\boldsymbol{e}}$ as follows:

$$\begin{aligned} C\{\boldsymbol{y} - \tilde{\boldsymbol{e}}, \tilde{\boldsymbol{e}}\} &= C\{\big[I - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{y},\, P^{-1}B^T\big(BP^{-1}B^T\big)B\boldsymbol{y}\} = \\ &= \big[I - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] \cdot D\{\boldsymbol{y}\} \cdot \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]^T = \\ &= \big[I - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] \cdot \sigma_0^2 P^{-1} \cdot B^T\big(BP^{-1}B^T\big)^{-1}BP^{-1} = \\ &= \sigma_0^2\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}BP^{-1} - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}BP^{-1}B^T\big(BP^{-1}B^T\big)^{-1}BP^{-1}\big] = 0 \end{aligned} \tag{4.8}$$

Thus it has been shown that the residuals and adjusted observations are uncorrelated.

4.1 Equivalence Between Least-Solutions Within the GMM and the Model of Condition Equations

To show the equivalence between the least-squares adjustments within the GMM and the model of condition equations, it must be shown that the predicted random error vectors (residuals) from both adjustments are equivalent. The residual vector $\tilde{\boldsymbol{e}}$ from each adjustment can be expressed as a projection matrix times the true random error vector $\boldsymbol{e}$ (or equivalently, times the observation vector $\boldsymbol{y}$) as shown below.

Recall that the residual vector within the GMM can be written as

$$\tilde{\boldsymbol{e}} = \big[I_n - AN^{-1}A^TP\big]\boldsymbol{e}. \tag{4.9}$$

And the residual vector from within the model of condition equations can be written as

$$\tilde{\boldsymbol{e}} = \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{e}. \tag{4.10}$$

Note that the right sides of (4.9) and (4.10) cannot be computed since $\boldsymbol{e}$ is unknown, but the equations do hold since, for the GMM,

$$\begin{aligned} \tilde{\boldsymbol{e}} &= \big[I_n - AN^{-1}A^TP\big]\boldsymbol{y} = \\ &= \big[I_n - AN^{-1}A^TP\big](A\boldsymbol{\xi} + \boldsymbol{e}) = \\ &= \big[A\boldsymbol{\xi} - AN^{-1}(A^TPA)\boldsymbol{\xi}\big] + \big[I_n - AN^{-1}A^TP\big]\boldsymbol{e} = \\ &= \big[I_n - AN^{-1}A^TP\big]\boldsymbol{e}, \end{aligned}$$

and, for the model of condition equations,

$$\begin{aligned} \tilde{\boldsymbol{e}} &= P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\boldsymbol{y} = \\ &= P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B(A\boldsymbol{\xi} + \boldsymbol{e}) = \\ &= \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{e}, \end{aligned}$$

using the fact that $BA = 0$.

To show that (4.9) and (4.10) are equivalent, it must be shown that the range spaces and the nullspaces are equivalent for their respective projection matrices $[I_n - AN^{-1}A^TP]$ and $[P^{-1}B^T(BP^{-1}B^T)^{-1}B]$.

(i) Equivalent range spaces: Show that $\mathcal{R}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]$.
Proof: Since $A^TPP^{-1}B^T = A^TB^T = 0$, then

$$\begin{aligned}&\big[I_n - AN^{-1}A^TP\big]\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{z} = \\ &= \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{z} - \mathbf{0} \ \text{ for any } \boldsymbol{z} \in \mathbb{R}^n,\end{aligned}$$

which, with the help of (A.6), implies that

$$\mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] \subset \mathcal{R}\big[I_n - AN^{-1}A^TP\big].$$

Also:

$$\begin{aligned}&\dim\mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] = \\ &= \operatorname{rk}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] = \\ &= \operatorname{tr}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] = && \text{using (A.12)} \\ &= \operatorname{tr}\big[BP^{-1}B^T\big(BP^{-1}B^T\big)^{-1}\big] = && \text{using (A.5)} \\ &= \operatorname{tr}\big(I_r\big) = r.\end{aligned}$$

Furthermore:

$$\begin{aligned}&\dim\mathcal{R}\big[I_n - AN^{-1}A^TP\big] = \\ &= \operatorname{rk}\big(I_n - AN^{-1}A^TP\big) = \\ &= \operatorname{tr}\big(I_n - AN^{-1}A^TP\big) = && \text{using (A.12)} \\ &= \operatorname{tr}\big(I_n\big) - \operatorname{tr}\big(N^{-1}A^TPA\big) = \\ &= n - \operatorname{rk}N = n - \operatorname{rk}A = \\ &= n - q = r,\end{aligned}$$

which implies that

$$\mathcal{R}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]. \tag{4.11}$$

(ii) Equivalent Nullspaces: Show that

$$\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{N}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big].$$

Proof:
First show that $\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}(A)$.

$$\begin{aligned}&\big[I_n - AN^{-1}A^TP\big]A\boldsymbol{\alpha} = \mathbf{0} \ \text{ for all } \boldsymbol{\alpha} \\ &\Rightarrow \mathcal{R}(A) \subset \mathcal{N}\big[I_n - AN^{-1}A^TP\big]\end{aligned}$$

Also:

$$\dim\mathcal{R}(A) = \operatorname{rk}A = q$$

and

$$\begin{aligned}&\dim\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \\ &= n - \dim\mathcal{R}\big[I_n - AN^{-1}A^TP\big] = n - r = q\end{aligned}$$

(See dimension of nullspace in Appendix.)

$$\Rightarrow \mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}(A)$$
$$\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]A = 0,$$

since $BA = 0$. The preceding development implies that

$$\mathcal{R}(A) = \mathcal{N}\big[I_n - AN^{-1}A^TP\big] \subset \mathcal{N}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big].$$

We showed in part (i) that the dimensions of the range spaces of the respective projection matrices are equivalent. And, since $\dim \mathcal{N}(\cdot) = n - \dim \mathcal{R}(\cdot)$, it follows that the dimension of the nullspaces of the respective projection matrices are also equivalent. If one space is a subset of another and both spaces have the same dimension, the subspaces are equivalent.

We have showed that the range spaces and nullspaces of the projections matrices in (4.9) and (4.10) are the same. Therefore, the projections are the same, and thus the adjustments are equivalent.

Chapter 5

The Gauss-Markov Model with Constraints

The Gauss-Markov Model (GMM) with constraints is written as:

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\boldsymbol{\xi} + \boldsymbol{e}, \quad \boldsymbol{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}), \quad \operatorname{rk} A =: q \leq \{m, n\}, \tag{5.1a}$$

$$\boldsymbol{\kappa}_0 = \underset{l\times m}{K}\boldsymbol{\xi}, \quad \operatorname{rk} K =: l \geq m - q, \quad \operatorname{rk}\big[A^T,\, K^T\big] = m. \tag{5.1b}$$

The variables are as defined on page 17, but now with the addition of the $l \times m$ constraint matrix K and the $l \times 1$ vector of constrains $\boldsymbol{\kappa}_0$. Symbols for the normal equations were introduced in (3.4) and are repeated here:

$$[N,\, \boldsymbol{c}] = A^T P[A,\, \boldsymbol{y}]. \tag{5.2}$$

The given rank conditions imply that $(N+K^TK)^{-1}$ exists, and, if N^{-1} exists, so does $(KN^{-1}K^T)^{-1}$. The range space of $[A^T,\, K^T]$ spans $\mathbb{R}^m$ as is evident from (5.1b). The redundancy of the system is computed by

$$r := n - \operatorname{rk} A + \operatorname{rk} K = n - q + l. \tag{5.3}$$

The Lagrange target function to minimize is

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}) := (\boldsymbol{y} - A\boldsymbol{\xi})^T P(\boldsymbol{y} - A\boldsymbol{\xi}) - 2\boldsymbol{\lambda}^T\big(\boldsymbol{\kappa}_0 - K\boldsymbol{\xi}\big) = \text{stationary}. \tag{5.4}$$

Its first partial derivates are taken to form the following Euler-Lagrange necessary conditions:

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\xi}} = N\hat{\boldsymbol{\xi}} - \boldsymbol{c} + K^T\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{5.5a}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\lambda}} = -\boldsymbol{\kappa}_0 + K\hat{\boldsymbol{\xi}} \doteq \mathbf{0}. \tag{5.5b}$$

In matrix form (5.5a) and (5.5b) are expressed as

$$\begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\kappa}_0 \end{bmatrix}. \tag{5.6}$$

The sufficient condition, required for minimization, is satisfied by $(1/2)\left(\partial^2\Phi/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T\right) = N$, which is positive (semi) definite. We refer to the matrix on the left side of (5.6) as the least-squares *normal equation matrix*. It is invertible if, and only if, $\mathrm{rk}\left[A^T,\, K^T\right] = m$. This rank condition means that for the normal equation matrix:

- among the first m columns, $m - l$ must be linearly independent, and
- the additional l columns are complementary.

We consider two cases: (1) N is invertible, and (2) N is singular.

Case 1: N is invertible (also said to be *regular*)
Equations (5.5a) and (5.5b) imply

$$\hat{\boldsymbol{\xi}} = N^{-1}\left(\boldsymbol{c} - K^T\hat{\boldsymbol{\lambda}}\right) \tag{5.7a}$$

$$\boldsymbol{\kappa}_0 = K\hat{\boldsymbol{\xi}} = KN^{-1}\boldsymbol{c} - KN^{-1}K^T\hat{\boldsymbol{\lambda}} \tag{5.7b}$$

$$\Rightarrow \hat{\boldsymbol{\lambda}} = -\left(KN^{-1}K^T\right)^{-1}\left(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\right) \Rightarrow \tag{5.7c}$$

$$\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} + N^{-1}K^T\left(KN^{-1}K^T\right)^{-1}\left(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\right). \tag{5.7d}$$

The vector difference $\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}$ in (5.7d) is *a vector of discrepancies*. It shows the mismatch between the vector of constraint values $\boldsymbol{\kappa}_0$ and a linear combination (as generated by the matrix K) of the solution without constraints (i.e., $N^{-1}\boldsymbol{c}$).

Case 2: N is singular (i.e., not invertible)
Multiply equation (5.5b) by K^T and add the result to (5.5a), leading to

$$\begin{aligned}\left(N + K^TK\right)\hat{\boldsymbol{\xi}} &= \boldsymbol{c} + K^T\left(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\right) \Rightarrow \\ \hat{\boldsymbol{\xi}} &= \left(N + K^TK\right)^{-1}\boldsymbol{c} + \left(N + K^TK\right)^{-1}K^T\left(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\right).\end{aligned} \tag{5.8}$$

Then from (5.5b) and (5.8) we have

$$\begin{aligned}\boldsymbol{\kappa}_0 = K\hat{\boldsymbol{\xi}} &= K\left(N + K^TK\right)^{-1}\boldsymbol{c} + K\left(N + K^TK\right)^{-1}K^T\left(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\right) \Rightarrow \\ \left(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\right) &= \left[K\left(N + K^TK\right)^{-1}K^T\right]^{-1}\left[\boldsymbol{\kappa}_0 - K\left(N + K^TK\right)^{-1}\boldsymbol{c}\right].\end{aligned} \tag{5.9}$$

Substituting (5.9) into (5.8) leads to the solution

$$\begin{aligned}\hat{\boldsymbol{\xi}} = &\left(N + K^TK\right)^{-1}\boldsymbol{c} \\ &+ \left(N + K^TK\right)^{-1}K^T\left[K\left(N + K^TK\right)^{-1}K^T\right]^{-1}\left[\boldsymbol{\kappa}_0 - K\left(N + K^TK\right)^{-1}\boldsymbol{c}\right].\end{aligned} \tag{5.10}$$

The form of (5.10) is identical to (5.7d) except that all occurrences of matrix N in (5.7d) have been replaced by $N + K^TK$ in (5.10).

We now compute the formal dispersion of $\hat{\boldsymbol{\xi}}$ for both cases. From (5.6) we have

$$\begin{bmatrix}\hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}}\end{bmatrix} = \begin{bmatrix}N & K^T \\ K & 0\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{c} \\ \boldsymbol{\kappa}_0\end{bmatrix}, \tag{5.11}$$

which, from the law of covariance propagation, implies that

$$D\{\begin{bmatrix}\hat{\boldsymbol{\xi}}\\ \hat{\boldsymbol{\lambda}}\end{bmatrix}\} = \begin{bmatrix}N & K^T\\ K & 0\end{bmatrix}^{-1} D\{\begin{bmatrix}\boldsymbol{c}\\ \boldsymbol{\kappa}_0\end{bmatrix}\} \begin{bmatrix}N & K^T\\ K & 0\end{bmatrix}^{-1} =$$

$$= \sigma_0^2 \begin{bmatrix}N & K^T\\ K & 0\end{bmatrix}^{-1} \begin{bmatrix}N & 0\\ 0 & 0\end{bmatrix} \begin{bmatrix}N & K^T\\ K & 0\end{bmatrix}^{-1}. \tag{5.12}$$

Here, the symmetry of the normal-equation matrix and the fact that $\boldsymbol{\kappa}_0$ is a non-random vector have been applied. Upon algebraic reduction of (5.12), we find that

$$\begin{bmatrix}D\{\hat{\boldsymbol{\xi}}\} & X\\ X & -D\{\hat{\boldsymbol{\lambda}}\}\end{bmatrix} = \sigma_0^2 \begin{bmatrix}N & K^T\\ K & 0\end{bmatrix}^{-1}. \tag{5.13}$$

Here, the symbol X represents a term of no particular interest. Note that $X \neq C\{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}\} = 0$. The inverse on the right side of (5.13) reveals the following dispersion matrices for cases 1 and 2, respectively:

Case 1:

$$\boxed{D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\big[N^{-1} - N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}KN^{-1}\big]} \tag{5.14}$$

Case 2:

$$\boxed{D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\big(N + K^TK\big)^{-1} - \sigma_0^2\big(N + K^TK\big)^{-1}K^T\big[K\left(N + K^TK\right)^{-1}K^T\big]^{-1}K\big(N + K^TK\big)^{-1}} \tag{5.15}$$

As with the parameter estimates, the dispersion matrices for both cases have a similar form, with every occurrence of N in Case 1 being replaced by $N+K^TK$ in Case 2. Also note that the dispersions in (5.14) and (5.15) are nothing more than the coefficient matrices multiplying the vector $\boldsymbol{c}$ in (5.7d) and (5.10), respectively, multiplied by the (unknown) variance component σ_0^2. Finally, it is clear from the above that the constraints reduce the dispersion matrix of $\hat{\boldsymbol{\xi}}$ compared to the corresponding dispersion matrix of the solution within the GMM (without constraints) derived in Chapter 3.

5.1 Estimated Variance Component

The estimated variance component for the GMM with constraints is derived similar to that for the GMM without constraints as shown in Section 3.2. The estimation is based on the principle

$$\frac{\hat{\sigma}_0^2}{\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}}} = \frac{\sigma_0^2}{E\{\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}}\}}. \tag{5.16}$$

Furthermore, for the purpose of validating the constraints, we wish to decompose the quadratic form $\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}}$ into the sum $\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} = \Omega + R$. In the following, we derive these components for both cases I and II.

5.1.1 Case I - Matrix N is invertible

$$
\begin{aligned}
\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} &= \big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\big)^T P \big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\big) = \\
&= \big[\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) - AN^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)\big]^T P \\
&\quad \times \big[\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) - AN^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)\big] = \\
&= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) \\
&\quad - \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T PAN^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big) \\
&\quad - \big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)^T\big(KN^{-1}K^T\big)^{-1}KN^{-1}A^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) \\
&\quad + \big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)^T\big(KN^{-1}K^T\big)^{-1}KN^{-1}\big(A^T PA\big)N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big) = \\
&\qquad\qquad \text{(Note that } A^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) = 0.\text{)} \\
&= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big) = \\
&= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \hat{\boldsymbol{\lambda}}^T\big(KN^{-1}K^T\big)\hat{\boldsymbol{\lambda}} = \Omega + R
\end{aligned}
\tag{5.17}
$$

The scalars Ω and R defined as

$$\Omega := \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) \tag{5.18a}$$

$$R := \big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big). \tag{5.18b}$$

Thus we have decomposed the quadratic form $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ into components Ω and R. Obviously, both Ω and R are random numbers since the vector $\boldsymbol{c}$ is random. It turns out that they are also uncorrelated. The variable Ω is associated with the LESS with the GMM without constraints, and R is due to the constraints $\boldsymbol{\kappa}_0 = K\boldsymbol{\xi}$. From (5.18b) we see that R is always positive, and thus the constraints will increase the value of $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$. The variables Ω and R are used for hypothesis testing as discussed in Chapter 10.

We now derive the expectation of $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$.

$$
\begin{aligned}
E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\} &= E\{\Omega\} + E\{R\} = \\
&= (n-m)\sigma_0^2 + E\{\hat{\boldsymbol{\lambda}}^T\big(KN^{-1}K^T\big)\hat{\boldsymbol{\lambda}}\} = E\{\Omega\} \qquad \text{using (3.24) for} \\
&= (n-m)\sigma_0^2 + \operatorname{tr}\big[\big(KN^{-1}K^T\big)E\{\hat{\boldsymbol{\lambda}}\,\hat{\boldsymbol{\lambda}}^T\}\big] = \\
&= (n-m)\sigma_0^2 + \operatorname{tr}\big[\big(KN^{-1}K^T\big)\big(D\{\hat{\boldsymbol{\lambda}}\} + E\{\hat{\boldsymbol{\lambda}}\}E\{\hat{\boldsymbol{\lambda}}\}^T\big)\big] = \\
&\qquad \text{(with } E\{\hat{\boldsymbol{\lambda}}\} = \boldsymbol{0} \text{ and } D\{\hat{\boldsymbol{\lambda}}\} = \sigma_0^2\big(KN^{-1}K^T\big)^{-1}\text{)} \\
&= (n-m)\sigma_0^2 + \operatorname{tr}\big[\big(KN^{-1}K^T\big)\sigma_0^2\big(KN^{-1}K^T\big)^{-1}\big] = \\
&= (n-m+l)\sigma_0^2
\end{aligned}
\tag{5.19}
$$

Substitution of (5.17) and (5.19) into (5.16) yields the following formula for the estimated variance component:

$$\boxed{\hat{\sigma}_0^2 = \frac{\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)}{n-m+l}.} \tag{5.20}$$

Other useful forms of $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ are derived below starting with the line above (5.17).

$$
\begin{aligned}
\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} &= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \hat{\boldsymbol{\lambda}}^T \big(KN^{-1}K^T\big)\hat{\boldsymbol{\lambda}} = \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T N^{-1} \boldsymbol{c} - \big(\boldsymbol{\kappa}_0^T - \boldsymbol{c}^T N^{-1} K^T\big)\hat{\boldsymbol{\lambda}} = && \text{using (5.7c)} \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T N^{-1} \big(\boldsymbol{c} - K^T \hat{\boldsymbol{\lambda}}\big) - \boldsymbol{\kappa}_0^T \hat{\boldsymbol{\lambda}} = && \text{using (5.7a)} \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T \hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0^T \hat{\boldsymbol{\lambda}} = \\
&= \boldsymbol{y}^T P \big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\big) - \boldsymbol{\kappa}_0^T \hat{\boldsymbol{\lambda}} = \\
&= \boldsymbol{y}^T P \tilde{\boldsymbol{e}} - \boldsymbol{\kappa}_0^T \hat{\boldsymbol{\lambda}} && (5.21)
\end{aligned}
$$

5.1.2 Case II - Matrix N is singular

$$
\begin{aligned}
&\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \\
&= \big\{\boldsymbol{y} - A\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]\big\}^T P \big\{\boldsymbol{y} - A\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]\big\} = \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{y}^T P A \big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] - \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T \big(N + K^T K\big)^{-1} A^T P \boldsymbol{y} + \\
&\quad + \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T \big(N + K^T K\big)^{-1}\big(A^T P A + K^T K - K^T K\big)\big(N + K^T K\big)^{-1} \times \\
&\quad \times \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] = \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T \big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] - \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T \big(N + K^T K\big)^{-1} \boldsymbol{c} + \\
&\quad + \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T \big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] - \\
&\quad - \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T \big(N + K^T K\big)^{-1} K^T K \big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] = \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \cancel{\boldsymbol{c}^T \big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]} - \\
&\quad - \boldsymbol{c}^T \big(N + K^T K\big)^{-1} \boldsymbol{c} - \cancel{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T K \big(N + K^T K\big)^{-1} \boldsymbol{c}} + \\
&\quad + \cancel{\boldsymbol{c}^T \big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]} + \cancel{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T K \big(N + K^T K\big)^{-1} \boldsymbol{c}} + \\
&\quad + \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T K \big(N + K^T K\big)^{-1} K^T \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \hat{\boldsymbol{\xi}}^T K^T K \hat{\boldsymbol{\xi}} = \\
&= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T \big(N + K^T K\big)^{-1} \boldsymbol{c} + \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T \big[K\big(N + K^T K\big)^{-1} K^T\big]\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \boldsymbol{\kappa}_0^T \boldsymbol{\kappa}_0 && (5.22)
\end{aligned}
$$

Now we compute the expectation for $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$.

$$
\begin{aligned}
&E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\} = \\
&= E\{\boldsymbol{y}^T P\big[\boldsymbol{y} - A\big(N + K^T K\big)^{-1} c\big] - \boldsymbol{\kappa}_0^T \boldsymbol{\kappa}_0\} + E\{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T \big[K\big(N + K^T K\big)^{-1} K^T\big]\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\} = \\
&= \operatorname{tr} P E\{\big[I_n - A\big(N + K^T K\big)^{-1} A^T P\big]\boldsymbol{y}\boldsymbol{y}^T\} + \operatorname{tr} K\big(N + K^T K\big)^{-1} K^T \cdot E\{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\} =
\end{aligned}
$$

(Note that $E\{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\} = D\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} + E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}^T$ and $D\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} = D\{\hat{\boldsymbol{\lambda}}\}$)

$$
\begin{aligned}
&= \operatorname{tr}\big[P - PA\big(N + K^T K\big)^{-1} A^T P\big]\big(\sigma_0^2 P^{-1} + A\boldsymbol{\xi}^T \boldsymbol{\xi} A^T\big) + \\
&\quad + \operatorname{tr}\big[K\big(N + K^T K\big)^{-1} K^T \big[K\big(N + K^T K\big)^{-1} - I_l\big]\sigma_0^2 + E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}^T\big] = \\
&= \sigma_0^2\big[\operatorname{tr} I_n - \operatorname{tr}\big(N + K^T K\big)N + \operatorname{tr} I_l - \operatorname{tr}\big(N + K^T K\big)N\big] + 0 =
\end{aligned}
$$

(using $E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}} \ = 0)$

$$E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\} = \sigma_0^2 (n - m + l) \tag{5.23}$$

Finally, substituting (5.22) and (5.23) into (5.16) yields

$$\hat{\sigma}_0^2 = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T \big(N + K^T K\big)^{-1} \boldsymbol{c} + \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T \big[K \big(N + K^T K\big)^{-1} K^T\big] \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \boldsymbol{\kappa}_0^T \boldsymbol{\kappa}_0}{(n - m + l)}. \tag{5.24}$$

We cannot directly identify Ω and R in (5.22) as we could in case I. Therefore, we define Ω as

$$\Omega = \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big), \tag{5.25}$$

and R as

$$R = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} - \Omega, \tag{5.26}$$

where $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ is given in (5.22).

The following ratio is formed (for both cases I and II) for the purposes of hypothesis testing (see Chapter 10 for more details on hypothesis testing):

$$T := \frac{R/(l - m + q)}{\Omega/(n - q)} \sim F(l - m + q, n - q), \ \text{ with } q := \mathrm{rk}(A). \tag{5.27}$$

The hypothesis test is then

$$H_0 : K\boldsymbol{\xi} = \boldsymbol{\kappa}_0 \ \text{ versus } H_A : K\boldsymbol{\xi} \neq \boldsymbol{\kappa}_0. \tag{5.28}$$

For some chosen significance level α,

$$\begin{aligned} &\text{Accept } H_0 : \ \text{if } T \leq F_{\alpha, l-m+q, n-q} \\ &\text{Reject } H_0 : \ \text{if } T > F_{\alpha, l-m+q, n-q}, \end{aligned}$$

where $F_{\alpha, l-m+q, n-q}$ is the critical value from the F-distribution table. Note that the redundancy $r_2 := n - q$ represents the degrees of freedom for the system of equations if no constraints were applied, whereas the redundancy $r_1 := l - m + q$ represents the *increase* in degrees of freedom due to the constraints. In the case that A has full column rank (i.e., $\mathrm{rk}\, A = q = m$), then the redundancies reduce to $r_1 := l$ and $r_2 := n - m$, respectively.

Chapter 6

Introduction of a Datum to Treat the Rank-Deficient Gauss-Markov Model

Consider the following linearized Gauss-Markov Model (GMM) with rank-deficient matrix A:

$$\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}, \quad \boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 P^{-1}\right), \quad \operatorname{rk} A =: q < m. \tag{6.1}$$

We can partition the matrix A as

$$\underset{n\times m}{A} = \left[\underset{n\times q}{A_1} \;\middle|\; \underset{n\times(m-q)}{A_2}\right], \quad \text{with } \operatorname{rk} A_1 = q = \operatorname{rk} A. \tag{6.2}$$

A similar partitioning of the parameter vector $\hat{\boldsymbol{\xi}}$ leads to the following system of partitioned normal equations:

$$\begin{aligned}\begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} P \left[A_1,\ A_2\right] \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} &= \begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} P\boldsymbol{y} = \left[\begin{array}{c|c} A_1^T P A_1 & A_1^T P A_2 \\ \hline A_2^T P A_1 & A_2^T P A_2 \end{array}\right] \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} = \begin{bmatrix} A_1^T P\boldsymbol{y} \\ A_2^T P\boldsymbol{y} \end{bmatrix} = \\ &= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}.\end{aligned} \tag{6.3}$$

The subscripted terms in (6.3) may be defined more succinctly as

$$\left[N_{ij},\, \boldsymbol{c}_i\right] := A_i^T P \left[A_j,\, \boldsymbol{y}\right]. \tag{6.4}$$

Defining a datum for $m-q$ parameters means that $\hat{\boldsymbol{\xi}}_2 \to \boldsymbol{\xi}_2^0$, where $\boldsymbol{\xi}_2^0$ is known. The rank of A_1 given in (6.2) implies that the inverse of the $q \times q$ matrix N_{11} exists. Therefore, from the top row of (6.3) and with a given *datum* $\boldsymbol{\xi}_2^0$ substituted for $\hat{\boldsymbol{\xi}}_2$, we can write

$$N_{11}\hat{\boldsymbol{\xi}}_1 = \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \Rightarrow \tag{6.5a}$$

$$\hat{\boldsymbol{\xi}}_1 = N_{11}^{-1}\left(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\right). \tag{6.5b}$$

Equation (6.5b) shows that the datum can be chosen after adjustment of the observations. Moreover, since the only random component in (6.5b) is $\boldsymbol{c}_1$, we have

$$D\{\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2 N_{11}^{-1} \tag{6.6}$$

for the dispersion of the vector of estimated parameters $\hat{\boldsymbol{\xi}}_1$.

The predicted random error (residual) vector and its dispersion are then defined as follows:

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \boldsymbol{y} - \begin{bmatrix} A_1 & \mid & A_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \boldsymbol{\xi}_2^0 \end{bmatrix} = \boldsymbol{y} - A_1\hat{\boldsymbol{\xi}}_1 - A_2\boldsymbol{\xi}_2^0, \tag{6.7a}$$

$$D\{\tilde{\boldsymbol{e}}\} = D\{\boldsymbol{y}\} - D\{A_1\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2\big(P^{-1} - A_1 N_{11}^{-1} A_1^T\big). \tag{6.7b}$$

Note that $C\{\boldsymbol{y}, \hat{\boldsymbol{\xi}}_1\} = 0$, which is implied by (6.7b).

Substituting (6.7a) into the quadratic product $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$, and considering (6.5a), leads to

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - \operatorname{rk} A} = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}_1^T \hat{\boldsymbol{\xi}}_1 - \boldsymbol{c}_2^T \boldsymbol{\xi}_2^0}{n - q} \tag{6.8}$$

as an estimate for the unknown variance component σ_0^2. Here, the relation $\hat{\boldsymbol{\xi}}_1^T N_{11} \hat{\boldsymbol{\xi}}_1 + \hat{\boldsymbol{\xi}}_1^T N_{12} \hat{\boldsymbol{\xi}}_2 = \hat{\boldsymbol{\xi}}_1^T \boldsymbol{c}_1$ has been used. However, since $\operatorname{rk} A_1 = \operatorname{rk} A = q$, the $n \times (m-q)$ submatrix A_2 must be in the column space of the $n \times q$ matrix A_1 so that $A_2 = A_1 L$ for some $q \times (m-q)$ matrix L. Therefore,

$$N_{12} = A_1^T P A_2 = A_1^T P A_1 L = N_{11} L \Rightarrow \tag{6.9a}$$

$$N_{11}^{-1} N_{12} = L. \tag{6.9b}$$

With this result, and using (6.5a), we have

$$\begin{aligned} \boldsymbol{c}_1^T \hat{\boldsymbol{\xi}}_1 + \boldsymbol{c}_2^T \boldsymbol{\xi}_2^0 &= \boldsymbol{y}^T P A_1 \big(N_{11}^{-1} \boldsymbol{c}_1 - N_{11}^{-1} N_{12} \boldsymbol{\xi}_2^0\big) + \boldsymbol{y}^T P A_2 \boldsymbol{\xi}_2^0 = \\ &= \boldsymbol{y}^T P A_1 \big(N_{11}^{-1} \boldsymbol{c}_1 - L \boldsymbol{\xi}_2^0\big) + \boldsymbol{y}^T P A_2 \boldsymbol{\xi}_2^0 = \\ &= \boldsymbol{y}^T P A_1 N_{11}^{-1} \boldsymbol{c}_1 - \boldsymbol{y}^T P \big(A_1 L\big) \boldsymbol{\xi}_2^0 + \boldsymbol{y}^T P A_2 \boldsymbol{\xi}_2^0 = \\ &= \boldsymbol{y}^T P A_1 N_{11}^{-1} \boldsymbol{c}_1 = \boldsymbol{c}_1^T N_{11}^{-1} \boldsymbol{c}_1, \end{aligned} \tag{6.10}$$

which upon substitution into (6.8) yields

$$\boxed{\hat{\sigma}_0^2 = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}_1^T N_{11}^{-1} \boldsymbol{c}_1}{n - q}} \tag{6.11}$$

as an alternative form for the estimated variance component.

It is instructive to compare the dispersion of $\hat{\boldsymbol{\xi}}_1$ shown in (6.6) with the corresponding dispersion in the case that matrix A has full row rank, i.e., $\operatorname{rk} A = m$. In this full-rank case, we could invert the coefficient matrix of (6.3) and find the upper $q \times q$ block of the inverse, scaled by σ_0^2, to be the dispersion of $\hat{\boldsymbol{\xi}}_1$. Referring to (A.14) for the inverse of the partitioned matrix N, we find

$$\begin{aligned} \underbrace{D\{\hat{\boldsymbol{\xi}}_1\}}_{\text{no datum}} &= \sigma_0^2 \big[N_{11}^{-1} + N_{11}^{-1} N_{12} \big(N_{22} - N_{21} N_{11}^{-1} N_{12}\big)^{-1} N_{21} N_{11}^{-1}\big] = \\ &= \sigma_0^2 \big(N_{11} - N_{12} N_{22}^{-1} N_{21}\big)^{-1} > \sigma_0^2 N_{11}^{-1} = \underbrace{D\{\hat{\boldsymbol{\xi}}_1\}}_{\text{datum supplied}} . \end{aligned} \tag{6.12}$$

The smaller dispersion in the last line of (6.12) shows that if a datum is introduced (increase in information), the unknown parameters $\boldsymbol{\xi}$ are estimated with smaller variance.

6.1 Generation of Equivalent Condition Equations

We may also wish to transform the rank-deficient model of (6.1) to a model of condition equations. To do so, consider the further splitting of the rank-deficient matrix A defined in (6.2) as follows:

$$\underset{n\times m}{A} = \left[A_1 \mid A_2\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{6.13a}$$

$$\text{with } \dim(A_{11}) = q \times q \text{ and } \dim(A_{22}) = (n-q) \times (m-q). \tag{6.13b}$$

Also, we have $\operatorname{rk} A_{11} = q = \operatorname{rk} A$. And, based on the definition of L in the preceding section, we may write

$$A_2 = \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} = A_1 L = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} L. \tag{6.14}$$

The matrix B within the model of condition equations could be defined as

$$\underset{r\times n}{B} := \left[A_{21}A_{11}^{-1} \mid -I_{n-q}\right], \tag{6.15}$$

with

$$r := n - q \tag{6.16}$$

as the system redundancy.

As discussed in Chapter 4, two conditions must be satisfied in order to reach an equivalent model of condition equations:

i. dimensionality condition,

ii. orthogonality condition.

The first condition requires that the dimensions of the column spaces of A and B sum to n. The second condition requires that the rows of matrix B are orthogonal to the columns of A, i.e., $BA = 0$. Taken together, these conditions mean that A and B^T are orthogonal complements in n-dimensional space, or, stated more succinctly,

$$\mathcal{R}(A) \overset{\perp}{\oplus} \mathcal{R}(B^T) = \mathbb{R}^N. \tag{6.17}$$

Both conditions i and ii are satisfied for (6.15) as shown below.

i. Dimensionality condition:

$$\operatorname{rk} B = n - q = n - \operatorname{rk} A \Rightarrow \operatorname{rk} A + \operatorname{rk} B = n \tag{6.18a}$$

ii. Orthogonality condition:

$$BA = B\left[A_1 \mid A_2\right] = BA_1\left[I_q \mid L\right], \tag{6.18b}$$

$$\text{but } BA_1 = \left[A_{21}A_{11}^{-1} \mid -I_{n-q}\right]\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} = A_{21}A_{11}^{-1}A_{11} - A_{21} = 0 \Rightarrow \tag{6.18c}$$

$$BA = 0 \tag{6.18d}$$

Note that as long as the rank of matrix A is known, we can always generate a splitting of A as shown in (6.13a); however, we may need to reorder the columns of A (tantamount to reordering the elements of the parameter vector) to ensure that A_{11} has full column rank.

Chapter 7

The Gauss-Markov Model with Stochastic Constraints

The Gauss-Markov model (GMM) with stochastic constraints is similar in form to the GMM with fixed constraints, with one important difference. The constraints in the stochastic case are specified with some level of uncertainty, expressed in the form of a given weight matrix P_0, or an associated cofactor matrix $Q_0 \coloneqq P_0^{-1}$. The model reads

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \tag{7.1a}$$

$$\boldsymbol{z}_0 = \underset{l\times m}{K}\,\boldsymbol{\xi} + \boldsymbol{e}_0, \tag{7.1b}$$

$$\begin{bmatrix}\boldsymbol{e}\\ \boldsymbol{e}_0\end{bmatrix} \sim \left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2 \begin{bmatrix}P^{-1} & 0\\ 0 & P_0^{-1}\end{bmatrix}\right). \tag{7.1c}$$

Note that in this model there is no correlation between the random error vectors $\boldsymbol{e}$ and $\boldsymbol{e}_0$. Also, the unknown variance component σ_0^2 is common to both cofactor matrices P^{-1} and P_0^{-1}. However, there may be correlations within one or both of the cofactor matrices, just not between them. Depending on the application, the data in the vector $\boldsymbol{y}$ can be thought of as new information, while the constraints in the vector $\boldsymbol{\kappa}_0$ can be thought of as prior information (for example, coordinates estimated from a previous adjustment as prior information).

The ranks of the coefficient matrices A and K are expressed as

$$\operatorname{rk} A =: q \leq \{m, n\}, \;\; \operatorname{rk} K =: l \geq m - q, \;\; \operatorname{rk}\big[A^T,\, K^T\big] = m. \tag{7.2}$$

The least-squares solution (LESS) for $\boldsymbol{\xi}$ within model (7.1) may be derived by minimizing the Lagrange target function

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}) = \boldsymbol{e}^T P \boldsymbol{e} + 2\boldsymbol{\lambda}^T\big(K\boldsymbol{\xi} - \boldsymbol{z}_0\big) - \boldsymbol{\lambda}^T P_0^{-1}\boldsymbol{\lambda} = \text{stationary}. \tag{7.3}$$

Here we simply consider (7.1) as an extended GMM and apply *the addition theory of normal equations* as follows:

$$\begin{aligned}\begin{bmatrix}A^T & K^T\end{bmatrix}\begin{bmatrix}P & 0\\ 0 & P_0\end{bmatrix}\begin{bmatrix}A\\ K\end{bmatrix}\cdot\hat{\boldsymbol{\xi}} = \begin{bmatrix}A^T & K^T\end{bmatrix}\begin{bmatrix}P & 0\\ 0 & P_0\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} = \\ = \big(N + K^T P_0 K\big)\hat{\boldsymbol{\xi}} = \boldsymbol{c} + K^T P_0 \boldsymbol{z}_0,\end{aligned} \tag{7.4}$$

where

$$[N, \boldsymbol{c}] \coloneqq A^T P [A, \boldsymbol{y}]. \tag{7.5}$$

In the case where the matrix N is invertible, the Sherman-Morrison-Woodbury-Schur formula (A.7) may be used to invert the matrix on the left side of (7.4) as in the following:

$$\hat{\boldsymbol{\xi}} = \big(N + K^T P_0 K\big)^{-1}\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big) = \tag{7.6a}$$

$$\begin{aligned}&= \big[N^{-1} - N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}\big]\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big) = \\ &= N^{-1}\boldsymbol{c} + N^{-1}K^T P_0 \boldsymbol{z}_0 + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(-KN^{-1}\boldsymbol{c} - KN^{-1}K^T P_0 \boldsymbol{z}_0\big) = \\ &= N^{-1}\boldsymbol{c} + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\times \\ &\quad\times\big[\big(P_0^{-1} + KN^{-1}K^T\big)P_0\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c} - KN^{-1}K^T P_0 \boldsymbol{z}_0\big] = \\ &= N^{-1}\boldsymbol{c} + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big).\end{aligned} \tag{7.6b}$$

Thus, the LESS (7.6b) can be viewed as a weighted average between the prior and the new information. The vector $\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}$ is referred to as the vector of discrepancies. The solution can also be recognized as an update to the solution $\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c}$ within the GMM (3.1). It is also interesting to express it as an update to the LESS within the GMM with fixed constraints (5.1). This can be done by changing the symbols $\hat{\boldsymbol{\xi}}$ and $\boldsymbol{\kappa}_0$ in (5.7d) to $\hat{\boldsymbol{\xi}}_K$ and $\boldsymbol{z}_0$, respectively, solving for $N^{-1}\boldsymbol{c}$ in terms of these renamed variables, and substituting into (7.6b), which yields the following:

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_K + N^{-1}K^T\big[\big(P_0^{-1} + KN^{-1}K^T\big)^{-1} - \big(KN^{-1}K^T\big)^{-1}\big]\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big). \tag{7.7}$$

Applying the laws of variance propagation to (7.6a), the dispersion of the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ is computed as follows:

$$\begin{aligned}D\{\hat{\boldsymbol{\xi}}\} &= \big(N + K^T P_0 K\big)^{-1} D\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\}\big(N + K^T P_0 K\big)^{-1} = \\ &= \sigma_0^2\big(N + K^T P_0 K\big)^{-1}\big(N + K^T P_0 K\big)\big(N + K^T P_0 K\big)^{-1} = \\ &= \sigma_0^2\big(N + K^T P_0 K\big)^{-1} = \sigma_0^2\big[N^{-1} - N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}\big].\end{aligned} \tag{7.8}$$

The subtraction in (7.8) implies that our knowledge of the parameters has improved by supplying the additional prior information, provided the estimated variance component $\hat{\sigma}_0^2$ does not change much. Indeed, if the new data is consistent with the old, $\hat{\sigma}_0^2$ is not expected to change very much. In contrast, $\hat{\sigma}_0^2$ is expected to increase if there is inconsistency between the old and new information.

Let us now determine the residual vectors $\tilde{\boldsymbol{e}}$ and $\tilde{\boldsymbol{e}}_0$ (also called predicted random error vectors). The residual vector $\tilde{\boldsymbol{e}}$ for the observations $\boldsymbol{y}$ is

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}. \tag{7.9}$$

The residual vector $\tilde{\boldsymbol{e}}_0$ associated with the prior information $\boldsymbol{z}_0$ is

$$\begin{aligned}
\tilde{\boldsymbol{e}}_0 &= \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}} = && (7.10a)\\
&= \big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) - \big(KN^{-1}K^T + P_0^{-1} - P_0^{-1}\big)\times \\
&\quad \times \big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \\
&= \big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) - \Big[\big(KN^{-1}K^T + P_0^{-1}\big)\big(P_0^{-1} + KN^{-1}K^T\big)^{-1} - \\
&\quad - P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\Big]\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \\
&= \big\{I - \big[I - P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big]\big\}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \\
&= P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \\
&= \big(I_l + KN^{-1}K^T P_0\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big). && (7.10b)
\end{aligned}$$

The dispersion matrix of the residual vectors is derived as follows:

$$\begin{aligned}
D\{\begin{bmatrix}\tilde{\boldsymbol{e}}\\ \tilde{\boldsymbol{e}}_0\end{bmatrix}\} &= D\{\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}\} - D\{\begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\} = \\
&= \sigma_0^2\begin{bmatrix}P^{-1} & 0\\ 0 & P_0^{-1}\end{bmatrix} - \sigma_0^2\begin{bmatrix}A\\ K\end{bmatrix}\Big[N^{-1} - N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}\Big]\begin{bmatrix}A^T & K^T\end{bmatrix} = \\
&= \sigma_0^2\begin{bmatrix}P^{-1} - AN^{-1}A^T & -AN^{-1}K^T\\ -KN^{-1}A^T & P_0^{-1} - KN^{-1}K^T\end{bmatrix} + \\
&\quad + \sigma_0^2\begin{bmatrix}AN^{-1}K^T\\ KN^{-1}K^T\end{bmatrix}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\begin{bmatrix}KN^{-1}A^T & KN^{-1}K^T\end{bmatrix}. && (7.11)
\end{aligned}$$

From (7.11) we can write the dispersion matrices for the residual vectors individually as

$$\begin{aligned}
D\{\tilde{\boldsymbol{e}}\} &= \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big) + \sigma_0^2 AN^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}A^T = && (7.12a)\\
&= \sigma_0^2\big[P^{-1} - A\big(N + K^T P_0 K\big)^{-1}A^T\big], && (7.12b)
\end{aligned}$$

and

$$\begin{aligned}
D\{\tilde{\boldsymbol{e}}_0\} &= \sigma_0^2 P_0^{-1} - \sigma_0^2 KN^{-1}K^T + \sigma_0^2 KN^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}K^T = \\
&= \sigma_0^2 P_0^{-1} - \sigma_0^2 KN^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(P_0^{-1} + KN^{-1}K^T - KN^{-1}K^T\big) = \\
&= \sigma_0^2 P_0^{-1} - \sigma_0^2 KN^{-1}K^T\big(I_l + P_0KN^{-1}K^T\big)^{-1} = \\
&= \sigma_0^2 P_0^{-1}\big(I_l + P_0KN^{-1}K^T\big)\big(I_l + P_0KN^{-1}K^T\big)^{-1} - \sigma_0^2 KN^{-1}K^T\big(I_l + P_0KN^{-1}K^T\big)^{-1} = \\
&= \sigma_0^2 P_0^{-1}\big(I_l + P_0KN^{-1}K^T\big)^{-1} + \sigma_0^2 KN^{-1}K^T\big(I_l + P_0KN^{-1}K^T\big)^{-1} - \\
&\quad - \sigma_0^2 KN^{-1}K^T\big(I_l + P_0KN^{-1}K^T\big)^{-1} = \\
&= \sigma_0^2 P_0^{-1}\big(I_l + P_0KN^{-1}K^T\big)^{-1}. && (7.13)
\end{aligned}$$

We summarize by listing a few equivalent formulas for $D\{\tilde{\boldsymbol{e}}_0\}$.

$$D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 P_0^{-1}\big(I_l + P_0 K N^{-1} K^T\big)^{-1} = \tag{7.14a}$$

$$= \sigma_0^2 \big(I_l + K N^{-1} K^T P_0\big)^{-1} P_0^{-1} = \tag{7.14b}$$

$$= \sigma_0^2 P_0^{-1}\big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} = \tag{7.14c}$$

$$= \sigma_0^2 \big(P_0 + P_0 K N^{-1} K^T P_0\big)^{-1} = \tag{7.14d}$$

$$= \sigma_0^2 \big[P_0^{-1} - K\big(N + K^T P_0 K\big)^{-1} K^T\big] \tag{7.14e}$$

The symmetry of $D\{\tilde{\boldsymbol{e}}_0\}$ has been exploited to get from (7.14a) to (7.14b), using the rule for the transpose of a matrix product (A.1) and the rule for the transpose of an inverse (A.2). Also (A.3) has been used in the above.

Now it remains to write a succinct form for the covariance matrix $C\{\tilde{\boldsymbol{e}}, \tilde{\boldsymbol{e}}_0\}$, beginning with the off-diagonal element of (7.11).

$$C\{\tilde{\boldsymbol{e}}, \tilde{\boldsymbol{e}}_0\} = -\sigma_0^2 A N^{-1} K^T + \sigma_0^2 A N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1} K^T = \tag{7.15a}$$

$$= -\sigma_0^2 A N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} \big(P_0^{-1} + K N^{-1} K^T - K N^{-1} K^T\big) = \tag{7.15b}$$

$$= -\sigma_0^2 A N^{-1} K^T \big(I_l + P_0 K N^{-1} K^T\big)^{-1} = \tag{7.15c}$$

$$= -\sigma_0^2 A \big(I_m + N^{-1} K^T P_0 K\big)^{-1} N^{-1} K^T = \tag{7.15d}$$

$$= -\sigma_0^2 A \big(N + K^T P_0 K\big)^{-1} K^T \tag{7.15e}$$

The line following (7.15c) is based on relations shown in equations (A.9). To see how these equations are used, compare what follows the term $-\sigma_0^2 A$ in (7.15c) and (7.15d), with the first and last lines in (A.9).

Further insight may be gained by minimizing a Lagrange target function, as in the MS thesis by K. Snow (OSU Report 465). This leads to the following system of normal equations, which includes an estimated vector of Lagrange multipliers $\hat{\boldsymbol{\lambda}}$:

$$\begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{z}_0 \end{bmatrix}. \tag{7.16}$$

Using (7.1b) and (7.16), we can express the predicted residual vector $\tilde{\boldsymbol{e}}_0$ as a function of the vector of Lagrange multipliers $\hat{\boldsymbol{\lambda}}$ as follows:

$$\boldsymbol{z}_0 = K\hat{\boldsymbol{\xi}} + \tilde{\boldsymbol{e}}_0 = K\hat{\boldsymbol{\xi}} - P_0^{-1}\hat{\boldsymbol{\lambda}} \Rightarrow \tilde{\boldsymbol{e}}_0 = -P_0^{-1}\hat{\boldsymbol{\lambda}}. \tag{7.17}$$

Therefore, the dispersion of $\tilde{\boldsymbol{e}}_0$ is given by

$$D\{\tilde{\boldsymbol{e}}_0\} = P_0^{-1} D\{\hat{\boldsymbol{\lambda}}\} P_0^{-1}. \tag{7.18}$$

Assuming matrix N is invertible, from (7.16) we see that the dispersion of $\hat{\boldsymbol{\lambda}}$ can be found from

$$
\begin{aligned}
D\{\begin{bmatrix}\hat{\boldsymbol{\xi}}\\ \hat{\boldsymbol{\lambda}}\end{bmatrix}\} &= \begin{bmatrix}N & K^T\\ K & -P_0^{-1}\end{bmatrix}^{-1} D\{\begin{bmatrix}\boldsymbol{c}\\ \boldsymbol{z}_0\end{bmatrix}\} \begin{bmatrix}N & K^T\\ K & -P_0^{-1}\end{bmatrix}^{-1} = \\
&= \sigma_0^2 \begin{bmatrix}N & K^T\\ K & -P_0^{-1}\end{bmatrix}^{-1} \begin{bmatrix}N & 0\\ 0 & P_0^{-1}\end{bmatrix} \begin{bmatrix}N & K^T\\ K & -P_0^{-1}\end{bmatrix}^{-1} = \\
&= \sigma_0^2 \begin{bmatrix}N & K^T\\ K & -P_0^{-1}\end{bmatrix}^{-1} \begin{bmatrix}N^{-1} & 0\\ 0 & P_0\end{bmatrix}^{-1} \begin{bmatrix}N & K^T\\ K & -P_0^{-1}\end{bmatrix}^{-1} = \\
&= \sigma_0^2 \left[\begin{array}{c|c} N + K^T P_0 K & 0 \\ \hline 0 & P_0^{-1} + K N^{-1} K^T \end{array}\right]^{-1}.
\end{aligned}
\tag{7.19}
$$

The last line was reached by successively applying the rule for the product of two inverses (A.3). From (7.19) we see that

$$
D\{\hat{\boldsymbol{\lambda}}\} = \sigma_0^2\big(P_0^{-1} + KN^{-1}K^T\big)^{-1} = \sigma_0^2\big[P_0 - P_0K\big(N + K^TP_0K\big)^{-1}K^TP_0\big]. \tag{7.20}
$$

Finally, applying the product-of-inverses rule to (7.18), we can write

$$
\boxed{D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}P_0^{-1} = \sigma_0^2\big(P_0 + P_0KN^{-1}K^TP_0\big)^{-1}.} \tag{7.21}
$$

Also, we see from (7.19) that

$$
C(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}) = 0. \tag{7.22}
$$

We also note that in the GMM with stochastic constraints, the predicted residual vector $\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}$ by itself is no longer a projection of $\boldsymbol{y}$. However, the vector $\big[\tilde{\boldsymbol{e}}^T,\, \tilde{\boldsymbol{e}}_0^T\big]^T$ does represent a projection of $\big[\boldsymbol{y}^T,\, \boldsymbol{z}_0^T\big]^T$ since

$$
\begin{bmatrix}\tilde{\boldsymbol{e}}\\ \tilde{\boldsymbol{e}}_0\end{bmatrix} = \begin{bmatrix}\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\\ \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}\end{bmatrix} = \left(\begin{bmatrix}I_n & 0\\ 0 & I_l\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\big(N + K^TP_0K\big)^{-1}\begin{bmatrix}A^TP & K^TP_0\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}, \tag{7.23}
$$

and the matrix in parenthesis is idempotent, which can be verified by application of (A.11).

7.1 Variance Component Estimate

The derivation of the variance component estimate is shown here in detail. The trace operator is employed analogously to what was done in Section 3.2. We also make use of the following expectation and dispersion relationships:

$$
\begin{aligned}
E\{\boldsymbol{c} + K^TP_0\boldsymbol{z}_0\} &= \begin{bmatrix}A^TP & K^TP_0\end{bmatrix} E\{\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}\} = \\
&= \begin{bmatrix}A^TP & K^TP_0\end{bmatrix}\begin{bmatrix}A\\ K\end{bmatrix}\boldsymbol{\xi} = \big(N + K^TP_0K\big)\boldsymbol{\xi},
\end{aligned}
\tag{7.24a}
$$

$$D\{\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\} = D\{\begin{bmatrix}A^TP & K^TP_0\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}\} =$$
$$= \sigma_0^2\begin{bmatrix}A^TP & K^TP_0\end{bmatrix}\begin{bmatrix}P^{-1} & 0\\ 0 & P_0^{-1}\end{bmatrix}\begin{bmatrix}PA\\ P_0K\end{bmatrix} = \sigma_0^2\big(N+K^TP_0K\big). \tag{7.25a}$$

The following expectation and dispersion relationships are employed as well:

$$E\{\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)^T\} = D\{\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\} + E\{\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\}E\{\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\}^T, \tag{7.26a}$$
$$E\{\boldsymbol{y}\boldsymbol{y}^T\} = D\{\boldsymbol{y}\} + E\{\boldsymbol{y}\}E\{\boldsymbol{y}\}^T = \sigma_0^2P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^TA^T, \tag{7.26b}$$
$$E\{\boldsymbol{z}_0\boldsymbol{z}_0^T\} = D\{\boldsymbol{z}_0\} + E\{\boldsymbol{z}_0\}E\{\boldsymbol{z}_0\}^T = \sigma_0^2P_0^{-1} + K\boldsymbol{\xi}\boldsymbol{\xi}^TK^T. \tag{7.26c}$$

Then, the estimated variance component is computed from the expectation of the combined quadratic forms of the residual vectors: $\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^TP_0\tilde{\boldsymbol{e}}_0$.

$$\begin{aligned}
&E\{\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^TP_0\tilde{\boldsymbol{e}}_0\} =\\
&= E\{\Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\Big)^T\begin{bmatrix}P & 0\\ 0 & P_0\end{bmatrix}\Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\Big)\} =\\
&= E\{\boldsymbol{y}^TP\boldsymbol{y} + \boldsymbol{z}_0^TP_0\boldsymbol{z}_0 - 2\hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big) + \hat{\boldsymbol{\xi}}^T\big(N+K^TP_0K\big)\hat{\boldsymbol{\xi}}\} =\\
&= E\{\boldsymbol{y}^TP\boldsymbol{y} + \boldsymbol{z}_0^TP_0\boldsymbol{z}_0 - 2\hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big) + \hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\} =\\
&= E\{\boldsymbol{y}^TP\boldsymbol{y} + \boldsymbol{z}_0^TP_0\boldsymbol{z}_0 - \hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\} =\\
&= E\{\boldsymbol{y}^TP\boldsymbol{y} + \boldsymbol{z}_0^TP_0\boldsymbol{z}_0 - \big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)^T\big(N+K^TP_0K\big)^{-1}\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\} =\\
&= E\{\mathrm{tr}\big(\boldsymbol{y}^TP\boldsymbol{y}\big) + \mathrm{tr}\big(\boldsymbol{z}_0^TP_0\boldsymbol{z}_0\big) - \mathrm{tr}\big[\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)^T\big(N+K^TP_0K\big)^{-1}\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\big]\} =\\
&= E\{\mathrm{tr}\big(P\boldsymbol{y}\boldsymbol{y}^T\big) + \mathrm{tr}\big(P_0\boldsymbol{z}_0\boldsymbol{z}_0^T\big) - \mathrm{tr}\big[\big(N+K^TP_0K\big)^{-1}\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)^T\big]\} =\\
&= \mathrm{tr}\big(PE\{\boldsymbol{y}\boldsymbol{y}^T\}\big) + \mathrm{tr}\big(P_0E\{\boldsymbol{z}_0\boldsymbol{z}_0^T\}\big) - \mathrm{tr}\big[\big(N+K^TP_0K\big)^{-1}E\{\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)\big(\boldsymbol{c}+K^TP_0\boldsymbol{z}_0\big)^T\}\big] =\\
&= \mathrm{tr}\big[P\big(\sigma_0^2P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^TA^T\big)\big] + \mathrm{tr}\big[P_0\big(\sigma_0^2P_0^{-1} + K\boldsymbol{\xi}\boldsymbol{\xi}^TK^T\big)\big] -\\
&\quad - \sigma_0^2\,\mathrm{tr}\big[\big(N+K^TP_0K\big)^{-1}\big(N+K^TP_0K\big)\big] -\\
&\quad - \mathrm{tr}\big[\big(N+K^TP_0K\big)^{-1}\big(N+K^TP_0K\big)\boldsymbol{\xi}\boldsymbol{\xi}^T\big(N+K^TP_0K\big)\big] =\\
&= \sigma_0^2\,\mathrm{tr}\big(PP^{-1}\big) + \mathrm{tr}\big(PA\boldsymbol{\xi}\boldsymbol{\xi}^TA^T\big) + \sigma_0^2\,\mathrm{tr}\big(P_0P_0^{-1}\big) + \mathrm{tr}\big(P_0K\boldsymbol{\xi}\boldsymbol{\xi}^TK^T\big) -\\
&\quad - \sigma_0^2\,\mathrm{tr}\big(I_m\big) - \mathrm{tr}\big(\boldsymbol{\xi}\boldsymbol{\xi}^TN + \boldsymbol{\xi}\boldsymbol{\xi}^TK^TP_0K\big) =\\
&= \sigma_0^2\,\mathrm{tr}\big(I_n\big) + \mathrm{tr}\big(\boldsymbol{\xi}^TN\boldsymbol{\xi}\big) + \sigma_0^2\,\mathrm{tr}\big(I_l\big) + \mathrm{tr}\big(\boldsymbol{\xi}^TK^TP_0K\boldsymbol{\xi}\big) -\\
&\quad - \sigma_0^2\,\mathrm{tr}\big(I_m\big) - \mathrm{tr}\big(\boldsymbol{\xi}^TN\boldsymbol{\xi}\big) - \mathrm{tr}\big(\boldsymbol{\xi}^TK^TP_0K\boldsymbol{\xi}\big) =\\
&= \sigma_0^2(n+l-m) = E\{\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^TP_0\tilde{\boldsymbol{e}}_0\}
\end{aligned}$$

From the preceding derivation, it follows that

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^TP_0\tilde{\boldsymbol{e}}_0}{(n-m+l)}} \tag{7.27}$$

represents the estimated variance component.

7.2 Hypothesis Test for the Estimated Variance Component

Hypothesis testing can be used to validate that the least-squares solution satisfies the stochastic constraints in the model (7.1). The test statistic to be computed is comprised of a ratio of two estimated, and therefore random, variances and thus has an F-distribution (see Section 10.4). The idea is to extract from the sum of the quadratic products in (7.27) the associated sum of squared residuals that would have been computed for the LESS within the unconstrained GMM solution, i.e., $\hat{\boldsymbol{\xi}}_u = N^{-1}\boldsymbol{c}$, had it been estimated. We label this quantity Ω. What remains after extracting Ω from the numerator of (7.27) is a quantity that depends on the weight matrix P_0. We denote this quantity as $R(P_0)$ to indicate that it is a function of P_0. Both Ω and $R(P_0)$ are scalars, and both have random properties. These two variables, which are used to form the test statistic, are defined as follows:

$$\Omega \coloneqq \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) = \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T N^{-1}\boldsymbol{c}, \tag{7.28a}$$

$$R(P_0) \coloneqq \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0\tilde{\boldsymbol{e}}_0 - \Omega. \tag{7.28b}$$

Again we note that $\hat{\boldsymbol{\xi}}_u = N^{-1}\boldsymbol{c}$ represents the least-squares solution within model (7.1) had the stochastic constraints been omitted. In the following derivations, we also make use of (7.6b), (7.10a), (7.10b), (7.17), and (7.18) to write formulas for $\tilde{\boldsymbol{e}}_0$ and $\hat{\boldsymbol{\xi}}$ in terms of $\hat{\boldsymbol{\xi}}_u$ as follows:

$$\tilde{\boldsymbol{e}}_0 = \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}} = -P_0^{-1}\hat{\boldsymbol{\lambda}} = \big(I_l + KN^{-1}K^T P_0\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big), \tag{7.29}$$

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_u + N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0 = \hat{\boldsymbol{\xi}}_u - N^{-1}K^T\hat{\boldsymbol{\lambda}}. \tag{7.30}$$

We now begin with the quadratic form for the full predicted residual vector appearing in (7.27) (also called sum of squared residuals, SSR) and decompose it into Ω and $R(P_0)$. The crossed-out vector in the first line below is neglected since its contribution vanishes in the quadratic product.

$$\begin{aligned}
&\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0\tilde{\boldsymbol{e}}_0 = \Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \cancel{\begin{bmatrix}A\\ K\end{bmatrix}}\hat{\boldsymbol{\xi}}\Big)^T \begin{bmatrix}P & 0\\ 0 & P_0\end{bmatrix}\Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\Big) = \\
&= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\hat{\boldsymbol{\xi}} + \boldsymbol{z}_0^T P_0\boldsymbol{z}_0 - \boldsymbol{z}_0^T P_0 K\hat{\boldsymbol{\xi}} = \\
&= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\big(\hat{\boldsymbol{\xi}}_u + N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0\big) + \boldsymbol{z}_0^T P_0\boldsymbol{z}_0 - \boldsymbol{z}_0^T P_0 K\big(\hat{\boldsymbol{\xi}}_u + N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0\big) = \\
&= \underbrace{\big(\boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\hat{\boldsymbol{\xi}}_u\big)}_{\Omega} + \boldsymbol{z}_0^T P_0 \underbrace{\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)}_{\left(I_l + KN^{-1}K^T P_0\right)\tilde{\boldsymbol{e}}_0} - \underbrace{\big(\boldsymbol{c} + K^T P_0\boldsymbol{z}_0\big)^T}_{\left(N + K^T P_0 K\right)\hat{\boldsymbol{\xi}}} N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \boldsymbol{z}_0^T P_0\big(I_l + KN^{-1}K^T P_0\big)\tilde{\boldsymbol{e}}_0 - \hat{\boldsymbol{\xi}}^T\big(N + K^T P_0 K\big)N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \boldsymbol{z}_0^T\big(I_l + P_0KN^{-1}K^T\big)P_0\tilde{\boldsymbol{e}}_0 - \big(K\hat{\boldsymbol{\xi}}\big)^T\big(I_l + P_0KN^{-1}K^T\big)P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}\big)^T\big(I_l + P_0KN^{-1}K^T\big)P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)^T\big(I_l + P_0KN^{-1}K^T\big)^{-1}\big(I_l + P_0KN^{-1}K^T\big)\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big) = \\
&= \Omega + \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big) = \\
&= \Omega + R(P_0)
\end{aligned}$$

Thus, $R(P_0)$ is defined as

$$R(P_0) \coloneqq \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big). \tag{7.31}$$

Finally, the test statistic t can be expressed as a ratio of $R(P_0)$ to Ω, viz.

$$\boxed{t = \frac{(\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0 - \Omega)/(l-m+q)}{\Omega/(n-q)} = \frac{R(P_0)/(l-m+q)}{\Omega/(n-q)} \sim F(l-m+q, n-q).} \tag{7.32}$$

Recall from (7.2) that $l := \mathrm{rk}(K)$ and $q := \mathrm{rk}(A)$.

The following hypothesis test can now be performed, where $\mathcal{N}$ stands for the normal distribution and $\boldsymbol{\kappa}_0$ is an unknown quantity:

$$H_0 : \boldsymbol{z}_0 \sim \mathcal{N}\big(K\boldsymbol{\xi}, \sigma_0^2 P_0^{-1}\big) \text{ against } H_a : \boldsymbol{z}_0 \sim \mathcal{N}\big(\boldsymbol{\kappa}_0 \neq K\boldsymbol{\xi}, \sigma_0^2 P_0^{-1}\big). \tag{7.33}$$

The term H_0 is called the null hypothesis, and H_a is the alternative hypothesis.

After taking $F_{\alpha,l-m+q,n-q}$ from a table of critical values for the F-distribution, and choosing a level of significance α, the following logic can be applied:

$$\text{If } t \leq F_{\alpha,l-m+q,n-q} \text{ accept } H_0\text{; else reject } H_0. \tag{7.34}$$

7.3 Some Comments on Reproducing Estimators

In this section we briefly discuss two estimators within the Gauss-Markov model with stochastic constraints (7.1) that leave the constrained parameters unchanged. Such estimators are called *reproducing estimators.* For simplicity, we restrict the discussion to full rank models, i.e., $\mathrm{rk}\, A = m$, where m is the number of columns of matrix A and also the number of parameters to estimate.

One possible choice for a reproducing estimator is the estimator within the Gauss-Markov model with fixed constraints shown in (5.7d), which is optimal for that model. Two points should be made regarding the use of this estimator within the model (7.1). First, it is not an optimal estimator within model (7.1), and, second, its dispersion matrix shown in (5.14) and (5.15) is not correct within model (7.1). In the following, we show the proper dispersion matrix for the reproducing estimator within model (7.1). First, we introduce different subscripts to denote various linear estimators for $\boldsymbol{\xi}$.

$\hat{\boldsymbol{\xi}}_u$ denotes the unconstrained estimator $\hat{\boldsymbol{\xi}}_u = N^{-1}\boldsymbol{c}$, which is not optimal within model (7.1).

$\hat{\boldsymbol{\xi}}_K$ denotes the reproducing estimator from equation (5.7d), which is not optimal within model (7.1).

$\hat{\boldsymbol{\xi}}_S$ denotes the optimal estimator within the Gauss-Markov model with stochastic constraints shown in (7.1).

First we express the estimator $\hat{\boldsymbol{\xi}}_K$ as a function of the optimal estimator $\hat{\boldsymbol{\xi}}_S$. Using (7.4), we can write

$$\big(N + K^T P_0 K\big)^{-1} \boldsymbol{c} = \hat{\boldsymbol{\xi}}_S - \big(N + K^T P_0 K\big)^{-1} K^T P_0 \boldsymbol{z}_0. \tag{7.35}$$

We then repeat (5.7d) for the estimator $\hat{\boldsymbol{\xi}}_K$ with N replaced by $\big(N + K^T P_0 K\big)$ and $\boldsymbol{\kappa}_0$ replaced by $\boldsymbol{z}_0$ according to the model (7.1). This is our starting point.

$$\begin{aligned}\hat{\boldsymbol{\xi}}_K = \big(N + K^T P_0 K\big)^{-1}\boldsymbol{c} + \\ + \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}\big[\boldsymbol{z}_0 - K\big(N + K^T P_0 K\big)^{-1}\boldsymbol{c}\big]\end{aligned} \tag{7.36}$$

Now using (7.35) in (7.36):

$$\begin{aligned}\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S - \big(N + K^T P_0 K\big)^{-1} K^T P_0 \boldsymbol{z}_0 + \\ + \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}\big[\boldsymbol{z}_0 - K\big(N + K^T P_0 K\big)^{-1}\boldsymbol{c}\big].\end{aligned}$$

Factoring out $-\big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}$ yields

$$\begin{aligned}\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S - \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1} \times \\ \times \big\{\big[K\big(N + K^T P_0 K\big)^{-1} K^T\big] P_0 \boldsymbol{z}_0 - \boldsymbol{z}_0 + K\big(N + K^T P_0 K\big)^{-1}\boldsymbol{c}\big\}.\end{aligned}$$

Now, from (7.6a) we recognize $K\hat{\boldsymbol{\xi}}_S$ in the above line; thus we write:

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S + \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big). \tag{7.37}$$

We now have the fixed-constraint estimator $\hat{\boldsymbol{\xi}}_K$ expressed as a function of the optimal estimator for model (7.1), namely $\hat{\boldsymbol{\xi}}_S$. Using a familiar formula for $\big(N + K^T P_0 K\big)^{-1}$ and noting that

$$\big(N + K^T P_0 K\big)^{-1} K^T P_0 = N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1},$$

we can rewrite (7.37) as:

$$\begin{aligned}\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S + \big[N^{-1} - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1}\big] K^T \times \\ \times \big[K N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1}\big]^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big).\end{aligned} \tag{7.38}$$

Note the following useful relations:

$$\big[N^{-1} - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1}\big] K^T = N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} \tag{7.39}$$

and

$$\big(K N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1}\big)^{-1} = P_0 \big(P_0^{-1} + K N^{-1} K^T\big)\big(K N^{-1} K^T\big)^{-1}. \tag{7.40}$$

Equation (7.39) is derived as follows:

$$\begin{aligned}&\big[N^{-1} - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1}\big] K^T = \\ &= N^{-1} K^T - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1}\big(P_0^{-1} + K N^{-1} K^T - P_0^{-1}\big) = \\ &= N^{-1} K^T - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1}\big(P_0^{-1} + K N^{-1} K^T\big) - \\ &\quad - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1}\big(-P_0^{-1}\big) = \\ &= N^{-1} K^T - N^{-1} K^T + N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} = \\ &= N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1}.\end{aligned}$$

Successive application of the rule for the product of inverted matrices was used in equation (7.40). Substituting (7.39) and (7.40) into (7.38) yields:

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}P_0^{-1}P_0\big(P_0^{-1} + KN^{-1}K^T\big)\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big) = \\ = \hat{\boldsymbol{\xi}}_S + N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big). \tag{7.41}$$

Equation (7.41) gives an elegant expression of the fixed-constraint estimator $\hat{\boldsymbol{\xi}}_K$ in terms of the optimal estimator $\hat{\boldsymbol{\xi}}_S$. Realizing that the model with stochastic constraints (7.1) becomes the model with fixed constraints (5.1) when P_0^{-1} is replaced by zero, we can replace (7.41) with (7.42) below, which is also obvious from our starting equation (7.36). This also makes the appropriate dispersion matrix $D\{\hat{\boldsymbol{\xi}}_K\}$ under model (7.1) easier to compute.

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_u + N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big) \tag{7.42}$$

Note that $C\{\boldsymbol{z}_0, \boldsymbol{y}\} = 0$, which allows us to apply the dispersion operator to (7.42) as follows:

$$D\{\hat{\boldsymbol{\xi}}_K\} = D\{\hat{\boldsymbol{\xi}}_u - N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}K\hat{\boldsymbol{\xi}}_u\} + D\{N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\boldsymbol{z}_0\} = \\ D\{\hat{\boldsymbol{\xi}}_S \to \hat{\boldsymbol{\xi}}_K\} = \sigma_0^2N^{-1} - \sigma_0^2N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}KN^{-1} + \\ + \sigma_0^2N^{-1}K^T\big(KN^{-1}K^TP_0KN^{-1}K^T\big)^{-1}KN^{-1}. \tag{7.43}$$

Compare (7.43) to (5.12) to see that $D\{\hat{\boldsymbol{\xi}}_K\}$ increases by $\big[\sigma_0^2N^{-1}K^T\big(KN^{-1}K^TP_0KN^{-1}K^T\big)^{-1}\cdot \cdot KN^{-1}\big]$ when the estimator $\hat{\boldsymbol{\xi}}_K$ is used for the model with stochastic constraints (7.1).

We already noted that $\hat{\boldsymbol{\xi}}_K$ is a sub-optimal (reproducing) estimator within model (7.1). We now give the optimal reproducing estimator without derivation (for details see Schaffrin (1997a)).

$$\boxed{\hat{\boldsymbol{\xi}}_{opt/rep} = \hat{\boldsymbol{\xi}}_S + K^T\big(KK^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big)} \tag{7.44}$$

The symbol $\hat{\boldsymbol{\xi}}_S$ on the right side of (7.44) represents the optimal ("non-reproducing") estimator. Equation (7.44) is identical to (7.41) when N^{-1} is replaced by I.

The dispersion is given by

$$D\{\hat{\boldsymbol{\xi}}_{opt/rep}\} = D\{\hat{\boldsymbol{\xi}}_S\} + D\{K^T\big(KK^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big)\} = \\ = \sigma_0^2N^{-1} - \sigma_0^2N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1} + \\ + \sigma_0^2K^T\big(KK^T\big)^{-1}P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}P_0^{-1}\big(KK^T\big)^{-1}K. \tag{7.45}$$

Also note that

$$E\{\hat{\boldsymbol{\xi}}_{opt/rep}\} = \boldsymbol{\xi}, \tag{7.46a}$$

$$\boldsymbol{z}_0 - K\boldsymbol{\xi}_{opt/rep} = \mathbf{0}, \tag{7.46b}$$

$$D\{K\hat{\boldsymbol{\xi}}_{opt/rep}\} = D\{\boldsymbol{z}_0\} = \sigma_0^2P_0^{-1}. \tag{7.46c}$$

Chapter 8

Sequential Adjustments

The data model for sequential adjustments is based on two data sets, denoted by subscripts 1 and 2. The first data set is comprised of n_1 observations, and the second is comprised of n_2. It is assumed that the observations from the first data set, $\boldsymbol{y}_1$, are uncorrelated with those from the second, $\boldsymbol{y}_2$, i.e., $C\{\boldsymbol{y}_1, \boldsymbol{y}_2\} = 0$. The data model is written as

$$\underset{n_1\times 1}{\boldsymbol{y}_1} = \underset{n_1\times m}{A_1}\ \boldsymbol{\xi} + \boldsymbol{e}_1, \tag{8.1a}$$

$$\underset{n_2\times 1}{\boldsymbol{y}_2} = \underset{n_2\times m}{A_2}\ \boldsymbol{\xi} + \boldsymbol{e}_2, \tag{8.1b}$$

$$\begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \sim \Big(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \sigma_0^2 \begin{bmatrix}P_1^{-1} & 0\\ 0 & P_2^{-1}\end{bmatrix}\Big). \tag{8.1c}$$

The ranks of the coefficient (design) matrices A_1 and A_2 are

$$\operatorname{rk} A_1 = \operatorname{rk}\begin{bmatrix}A_1\\ A_2\end{bmatrix} = m. \tag{8.2}$$

Note that the coefficient matrix A_1 has full column rank, that there is no correlation between the random error vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, and that both data sets share a common variance component σ_0^2. Also, the total number of observations from both data sets is defined as $n := n_1 + n_2$.

The following notation is adopted for variables used in the subsequent least-squares normal equations:

$$[N_{ii},\, \boldsymbol{c}_i] = A_i^T P_i\, [A_i,\, \boldsymbol{y}_i] \ \text{ and } N_{ij} = A_i^T P_i A_i + A_j^T P_j A_j \ \text{ for } i \neq j. \tag{8.3}$$

We use a single hat to denote that estimates are based only on the first data set and a double hat to denote that they are based on both data sets. This makes it convenient to show estimates based on both data sets as an update to estimates based on only the first data set. For example, the estimate $\hat{\boldsymbol{\xi}}$ is based only on the first data set, while the estimate $\hat{\hat{\boldsymbol{\xi}}}$ is based on both data sets.

We recognize a structural similarity between the data model shown in (8.1) and the Gauss-Markov Model with stochastic constraints shown in (7.1). Given this similarity, we may immediately write down a least-squares solution for $\boldsymbol{\xi}$, and its dispersion matrix, in the form of (7.6b) and (7.8), respectively, viewing the second data set as "stochastic constraints."

$$\hat{\hat{\boldsymbol{\xi}}} = \hat{\boldsymbol{\xi}} + N_{11}^{-1} A_2^T \big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)^{-1} \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big) = \tag{8.4a}$$

$$= \hat{\boldsymbol{\xi}} + \big(N_{11} + A_2^T P_2 A_2\big)^{-1} A_2^T P_2 \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big) \tag{8.4b}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}\} = D\{\hat{\boldsymbol{\xi}}\} - \sigma_0^2 N_{11}^{-1} A_2^T \big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)^{-1} A_2 N_{11}^{-1} \tag{8.4c}$$

Equation (A.9a) was used in going from (8.4a) to (8.4b). It is important to note that the matrix $\big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)$ is of size $n_2 \times n_2$; whereas the size of matrix $\big(N_{11} + A_2^T P_2 A_2\big)$ is $m \times m$. Therefore, if the second data set has only one observation, then $n_2 = 1$, and the update is very fast! This may be the case in a real-time application where one new observation is added at each epoch in time.

8.1 Verification of the Sequential Adjustment

In this section we discuss verification of the sequential adjustment, the aim of which is to confirm that the adjustment based on both data sets is consistent with an adjustment based only on the first data set. We can make use of the work done in Chapter 7 to write the estimated variance component in a form composed of the sum of squared residuals Ω, based on adjustment of the first data set only, and an update $R(P_2)$ for the inclusion of the second data set, analogous to the derivation of (7.31). This facilitates hypothesis testing for the purpose of determining if the combined adjustment is consistent with an adjustment based only on the first data set.

$$\hat{\hat{\sigma}}_0^2 (n - m) = \Omega + R(P_2) \;\text{ with } \Omega = \hat{\sigma}_0^2 (n_1 - m) \tag{8.5a}$$

$$R(P_2) = -\big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big)^T \hat{\hat{\boldsymbol{\lambda}}} \;\text{ with } \hat{\hat{\boldsymbol{\lambda}}} = -\big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)^{-1} \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big) \Rightarrow \tag{8.5b}$$

$$\hat{\hat{\sigma}}_0^2 (n - m) = \Omega + \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big)^T \big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)^{-1} \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big) \tag{8.5c}$$

Then, the test statistic

$$t = \frac{R/n_2}{\Omega/(n_1 - m)} \sim F(n_2, n_1 - m) \tag{8.6}$$

can be computed to verify the sequential adjustment. It has an F-distribution with n_2 and $n_1 - m$ degrees for freedom. For some specified significance level α, we claim that the observations from the second data set are consistent with those from the first if $t \leq F_{\alpha, n_2, n_1 - m}$. See Chapter 10 for more on hypothesis testing.

8.2 Alternative Solution for the Normal Equations

Using the *addition theory of normal equations*, we may find a matrix representation of the normal equations as follows, where again the double hats refer to a solution based on both data sets:

$$\big(N_{11} + N_{22}\big) \hat{\hat{\boldsymbol{\xi}}} = \big(\boldsymbol{c}_1 + \boldsymbol{c}_2\big) \tag{8.7}$$

These normal equations lead to

$$N_{11}\hat{\hat{\boldsymbol{\xi}}} + N_{22}\hat{\hat{\boldsymbol{\xi}}} - \boldsymbol{c}_2 = \boldsymbol{c}_1 \Rightarrow \tag{8.8a}$$

$$N_{11}\hat{\hat{\boldsymbol{\xi}}} + A_2^T\hat{\hat{\boldsymbol{\lambda}}}_2 = \boldsymbol{c}_1, \text{ with } \hat{\hat{\boldsymbol{\lambda}}} = P_2\big(A_2\hat{\hat{\boldsymbol{\xi}}} - \boldsymbol{y}_2\big) \Rightarrow \tag{8.8b}$$

$$\boldsymbol{y}_2 = A_2\hat{\hat{\boldsymbol{\xi}}} - P_2^{-1}\hat{\hat{\boldsymbol{\lambda}}}. \tag{8.8c}$$

Then, from (8.8b) and (8.8c), we can write the following system of least-squares normal equations:

$$\begin{bmatrix} N_{11} & A_2^T \\ A_2 & -P_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}} \\ \hat{\hat{\boldsymbol{\lambda}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{y}_2 \end{bmatrix}. \tag{8.9}$$

From the first row of (8.9) we get

$$\hat{\hat{\boldsymbol{\xi}}} = N_{11}^{-1}\boldsymbol{c}_1 - N_{11}^{-1}A_2^T\hat{\hat{\boldsymbol{\lambda}}} = \tag{8.10a}$$

$$= \hat{\boldsymbol{\xi}} - N_{11}^{-1}A_2^T\hat{\hat{\boldsymbol{\lambda}}}. \tag{8.10b}$$

Equation (8.10b) is the update formula as a function of the vector of estimated Lagrange multipliers $\hat{\hat{\boldsymbol{\lambda}}}$. Without further derivation, we can compare (8.10b) to (8.4a) to get an expression for the estimated vector of Lagrange-multiplier as

$$\hat{\hat{\boldsymbol{\lambda}}} = -\big(P_2^{-1} + A_2N_{11}^{-1}A_2^T\big)^{-1}\big(\boldsymbol{y}_2 - A_2\hat{\boldsymbol{\xi}}\big). \tag{8.11}$$

Applying covariance propagation to (8.10b), we find the dispersion matrix of $\hat{\hat{\boldsymbol{\xi}}}$ to be

$$D\{\hat{\hat{\boldsymbol{\xi}}}\} = D\{\hat{\boldsymbol{\xi}}\} - \sigma_0^2N_{11}^{-1}A_2^T\big(P_2^{-1} + A_2N_{11}^{-1}A_2^T\big)^{-1}A_2N_{11}^{-1}, \tag{8.12}$$

where we used the fact that $C\{\boldsymbol{y}_2, \hat{\boldsymbol{\xi}}\} = 0$, which indicates that the observations from the second data set are uncorrelated with the estimated parameters based on the first data set only.

8.3 Sequential Adjustment, Rank-Deficient Case

Given $\operatorname{rk} A_1 =: q_1 < m$, we may introduce a datum by splitting the system as was done in Section 6.1. We split A_1 into an $n_1 \times q_1$ part denoted A_{11} and an $n_1 \times (m - q_1)$ part denoted A_{12}. We also split the parameter vector $\boldsymbol{\xi}$ into $q_1 \times 1$ part $\boldsymbol{\xi}_1$ and a $(m - q_1) \times 1$ part $\boldsymbol{\xi}_2$. Thus, we have

$$A_1 = [A_{11},\, A_{12}]\,, \;\; \operatorname{rk} A_{11} = q_1 \;\text{ and } \boldsymbol{\xi} = \big[\boldsymbol{\xi}_1^T,\, \boldsymbol{\xi}_2^T\big]^T. \tag{8.13}$$

Next we introduce a datum $\boldsymbol{\xi}_2^0$ such that $\boldsymbol{\xi}_2 \to \boldsymbol{\xi}_2^0$, where the subscript 2 now refers to the datum, rather than a second data set, for this symbol. The solution and dispersion formulas based on the first data set only can be copied from (6.5b) and (6.6), respectively. The estimated variance

component is nearly identical to (6.8) and (6.11), except for the splitting of the observation vector $\boldsymbol{y}$.

$$\hat{\boldsymbol{\xi}}_1 = N_{11}^{-1}\big(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\big) \tag{8.14a}$$

$$D\{\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2 N_{11}^{-1} \tag{8.14b}$$

$$\hat{\sigma}_0^2 = \frac{\boldsymbol{y}_1^T P\big(\boldsymbol{y}_1 - A_{11}\hat{\boldsymbol{\xi}}_1 - A_{12}\boldsymbol{\xi}_2^0\big)}{\big(n_1 - q_1\big)} = \tag{8.14c}$$

$$= \frac{\big(\boldsymbol{y}_1^T P\boldsymbol{y}_1 - \boldsymbol{c}_1^T N_{11}^{-1}\boldsymbol{c}_1\big)}{\big(n_1 - q_1\big)} \tag{8.14d}$$

Note that the steps from (6.8) to (6.11) can be used to go from (8.14c) to (8.14d). Now we introduce the second data set with a splitting analogous to the first.

$$\boldsymbol{y}_2 = A_{21}\boldsymbol{\xi}_1 + A_{22}\boldsymbol{\xi}_2 + \boldsymbol{e}_2, \;\; \boldsymbol{e}_2 \sim \big(\boldsymbol{0}, \sigma_0^2 P_2^{-1}\big) \tag{8.15}$$

The matrix A_{21} is of size $n_2 \times q_1$, and A_{22} is of size $n_2 \times (m - q_1)$. No information in the second data set refers to the datum choice; it only adds to the redundancy provided by the first data set. The rank of the normal equation matrix is unchanged and is expressed as

$$\operatorname{rk}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} =: q = q_1 \tag{8.16}$$

The full least-squares normal equations are then written as

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{11}^T P_1 A_{12} + A_{21}^T P_2 A_{22} \\ \hline A_{12}^T P_1 A_{11} + A_{22}^T P_2 A_{21} & A_{12}^T P_1 A_{12} + A_{22}^T P_2 A_{22} \end{array}\right] \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \boldsymbol{\xi}_2^0 \end{bmatrix} = \left[\begin{array}{c|c} A_{11}^T P_1 & A_{21}^T P_2 \\ \hline A_{12}^T P_1 & A_{22}^T P_2 \end{array}\right] \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}. \tag{8.17}$$

From the first row of (8.17), we may write the least-squares solution for $\hat{\hat{\boldsymbol{\xi}}}_1$ directly, followed by its dispersion matrix, as

$$\hat{\hat{\boldsymbol{\xi}}}_1 = \big(A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21}\big)^{-1}\big[\big(A_{11}^T P_1 \boldsymbol{y}_1 + A_{21}^T P_2 \boldsymbol{y}_2\big) - \big(A_{11}^T P_1 A_{12} + A_{21}^T P_2 A_{22}\big)\boldsymbol{\xi}_2^0\big], \tag{8.18a}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = \sigma_0^2\big(A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21}\big)^{-1}. \tag{8.19}$$

In order to derive update formulas, it is helpful to introduce an alternative expression for the normal equations analogous to what was done in (8.8a) through (8.9).

$$\big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 = \big(A_{11}^T P_1 \boldsymbol{y}_1\big) - \big(A_{11}^T P_1 A_{12}\big)\boldsymbol{\xi}_2^0 \Rightarrow \tag{8.20a}$$

$$N_{11}\hat{\boldsymbol{\xi}}_1 = \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \tag{8.20b}$$

$$\big(A_{21}^T P_2 A_{21}\big)\hat{\boldsymbol{\xi}}_1 = \big(A_{21}^T P_2 \boldsymbol{y}_2\big) - \big(A_{21}^T P_2 A_{22}\big)\boldsymbol{\xi}_2^0 \Rightarrow \tag{8.20c}$$

$$N_{21}\hat{\boldsymbol{\xi}}_1 = \boldsymbol{c}_2 - N_{22}\boldsymbol{\xi}_2^0 \tag{8.20d}$$

Here we have used the symbols N_{12} and N_{21} differently than defined in (8.3). Together, (8.20b) and (8.20d) comprise the first row of (8.17). Recombining (8.20b) and (8.20d) gives

$$\big(N_{11} + N_{21}\big)\hat{\hat{\boldsymbol{\xi}}}_1 = \boldsymbol{c}_1 + \boldsymbol{c}_2 - \big(N_{12} + N_{22}\big)\boldsymbol{\xi}_2^0 \Rightarrow \tag{8.21a}$$

$$N_{11}\hat{\hat{\boldsymbol{\xi}}}_1 + A_{21}^T\hat{\hat{\boldsymbol{\lambda}}} = \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0, \;\; \text{with } \hat{\hat{\boldsymbol{\lambda}}} = P_2\big(A_{21}\hat{\hat{\boldsymbol{\xi}}}_1 - \boldsymbol{y}_2 + A_{22}\boldsymbol{\xi}_2^0\big). \tag{8.21b}$$

Note that in (8.20b) and (8.20d) a single hat was used for the estimate of $\boldsymbol{\xi}_1$ since each respective equation represents only one set of data. The double hat in (8.21a) denotes the estimate of $\boldsymbol{\xi}_1$ based on both data sets. From (8.21b) we can write the system of normal equations in matrix form as follows:

$$\begin{bmatrix} N_{11} & A_{21}^T \\ A_{21} & -P_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\lambda}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \\ \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0 \end{bmatrix}. \tag{8.22}$$

The solution of (8.22) can be obtained by applying the inversion formula for a partitioned matrix as shown in (A.14).

$$\begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\lambda}}} \end{bmatrix} = \begin{bmatrix} N_{11} & A_{21}^T \\ A_{21} & -P_2^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \\ \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0 \end{bmatrix} = \left[\begin{array}{c|c} N_{11}^{-1} - N_{11}^{-1}A_{21}^T S_2 A_{21} N_{11}^{-1} & N_{11}^{-1}A_{21}^T S_2 \\ \hline S_2 A_{21} N_{11}^{-1} & -S_2 \end{array}\right] \begin{bmatrix} \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \\ \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0 \end{bmatrix}, \tag{8.23}$$

with

$$S_2 \coloneqq \left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}. \tag{8.24}$$

Finally, the estimated parameters and Lagrange multipliers are expressed as

$$\begin{aligned} \hat{\hat{\boldsymbol{\xi}}}_1 &= N_{11}^{-1}\left(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\right) + \\ &\quad + N_{11}^{-1}A_{21}^T\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}\left[A_{21}N_{11}^{-1}\left(-\boldsymbol{c}_1 + N_{12}\boldsymbol{\xi}_2^0\right) + \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0\right] = \end{aligned} \tag{8.25a}$$

$$= \hat{\boldsymbol{\xi}}_1 + N_{11}^{-1}A_{21}^T\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}\left(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\right), \tag{8.25b}$$

$$\hat{\hat{\boldsymbol{\lambda}}} = -\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}\left(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\right). \tag{8.25c}$$

The dispersion matrix of the estimated vector of Lagrange multipliers is

$$D\{\hat{\hat{\boldsymbol{\lambda}}}\} = \left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1} D\{\boldsymbol{y} - A_{21}\hat{\boldsymbol{\xi}}_1\}\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}, \tag{8.26}$$

since $D\{\boldsymbol{\xi}_2^0\} = 0$. The following relations also hold:

$$C\{\boldsymbol{y}_2, \hat{\boldsymbol{\xi}}_1\} = 0, \tag{8.27a}$$

$$D\{\boldsymbol{y} - A_{21}\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right), \tag{8.27b}$$

$$D\{\hat{\hat{\boldsymbol{\lambda}}}\} = \sigma_0^2\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}, \tag{8.27c}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = D\{\hat{\boldsymbol{\xi}}_1\} - \sigma_0^2 N_{11}^{-1}A_{21}^T\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}A_{21}N_{11}^{-1}. \tag{8.27d}$$

The estimated variance component is expressed as follows:

$$\hat{\hat{\sigma}}_0^2(n-q) = \hat{\sigma}_0^2\left(n_1 - q_1\right) + \left(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\right)^T\left(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\right)^{-1}\left(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\right) = \tag{8.28a}$$

$$= \hat{\sigma}_0^2\left(n_1 - q_1\right) - \hat{\hat{\boldsymbol{\lambda}}}^T\left(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\right). \tag{8.28b}$$

Once again, from (8.20b) to (8.28b), we have used $N_{11} = A_{11}^T P_1 A_{11}$.

8.4 Sequential Adjustment with New Parameters

In this section we consider the case where the second data set refers to all the parameters of the first data set, plus some additional new parameters. Thus we speak of m_1 parameters associated with the first data set and m_2 with the second, with $m_2 > m_1$. In the double subscripts used below, the first one refers to the data set, and the second subscript refers to the matrix splitting. For example, A_{21} is that part of the design matrix from the second data set that refers to the original parameters, whereas A_{22} is associated with the new parameters observed in the second data set. The original lecture used $\bar{\boldsymbol{y}}$ to denote a preprocessed observation vector that included the datum choice. However, we leave the bar off here and simply note that $\boldsymbol{y}$ could include both datum information as well as observations. The observation equations that follow imply that we have assumed that there are no correlations between the observations of data set one and those of data set two; they also imply that both sets of observations share a common variance component.

$$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix} = \begin{bmatrix}A_{11} & 0\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix}, \quad \begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \sim \left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2 \begin{bmatrix}P_1^{-1} & 0\\ 0 & P_2^{-1}\end{bmatrix}\right) \tag{8.29}$$

The size of the system of equations is described as follows:

$$\boldsymbol{y}_1 \in \mathbb{R}^{n_1}, \;\; \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1}, \;\; \boldsymbol{y}_2 \in \mathbb{R}^{n_2}, \;\; \boldsymbol{\xi}_2 \in \mathbb{R}^{m_2-m_1}, \;\; n = n_1 + n_2, \;\; m_2 > m_1. \tag{8.30}$$

Now, using the addition theory of normal equations, we can write

$$\begin{bmatrix}A_{11}^T & A_{21}^T\\ 0 & A_{22}^T\end{bmatrix}\begin{bmatrix}P_1 & 0\\ 0 & P_2\end{bmatrix}\begin{bmatrix}A_{11} & 0\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\hat{\hat{\boldsymbol{\xi}}}_1\\ \hat{\hat{\boldsymbol{\xi}}}_2\end{bmatrix} = \begin{bmatrix}A_{11}^T P_1 & A_{21}^T P_2\\ 0 & A_{22}^T P_2\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix} \Rightarrow \tag{8.31a}$$

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{21}^T P_2 A_{22}\\ \hline A_{22}^T P_2 A_{21} & A_{22}^T P_2 A_{22}\end{array}\right]\begin{bmatrix}\hat{\hat{\boldsymbol{\xi}}}_1\\ \hat{\hat{\boldsymbol{\xi}}}_2\end{bmatrix} = \begin{bmatrix}A_{11}^T P_1 \boldsymbol{y}_1 + A_{21}^T P_2 \boldsymbol{y}_2\\ A_{22}^T P_2 \boldsymbol{y}_2\end{bmatrix}. \tag{8.31b}$$

Here again, the double-hats refer to estimates based on both data sets.

Now, the first data set may no longer be available, rather we may be given only the estimates from the first adjustment. In this case we can use the bottom row of (8.31b) to solve for the estimates of the new parameters in terms of only the second data set, leading to

$$\hat{\hat{\boldsymbol{\xi}}}_2 = \left(A_{22}^T P_2 A_{22}\right)^{-1} A_{22}^T P_2 \left(\boldsymbol{y}_2 - A_{21}\hat{\hat{\boldsymbol{\xi}}}_1\right) \tag{8.32}$$

Then, from the normal equations based solely on the first data set, we may substitute

$$A_{11}^T P_1 \boldsymbol{y}_1 = \left(A_{11}^T P_1 A_{11}\right)\hat{\boldsymbol{\xi}}_1 \tag{8.33}$$

into the top row of the right side of (8.31b) and invert the normal-equation matrix on the left to solve for the parameter estimates. For convenience, we introduce the following symbols to use in the inverted matrix:

$$S_1 \coloneqq A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} - A_{21}^T P_2 A_{22}\left(A_{22}^T P_2 A_{22}\right)^{-1} A_{22}^T P_2 A_{21} = \tag{8.34a}$$

$$= A_{11}^T P_1 A_{11} + A_{21}^T \bar{P}_2 A_{21}, \tag{8.34b}$$

$$\bar{P}_2 \coloneqq P_2 - P_2 A_{22}\left(A_{22}^T P_2 A_{22}\right)^{-1} A_{22}^T P_2, \tag{8.34c}$$

$$N_{22} = A_{22}^T P_2 A_{22}. \tag{8.34d}$$

We refer to $\bar{P}_2$ as a reduced weight matrix. Upon inverting the normal-equations matrix (see (A.14) for the inverse of a partitioned matrix), we find the following solution for $\hat{\hat{\boldsymbol{\xi}}}_1$ and $\hat{\hat{\boldsymbol{\xi}}}_2$:

$$\begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\xi}}}_2 \end{bmatrix} = \left[\begin{array}{c|c} S_1^{-1} & -S_1^{-1}\left(A_{21}^T P_2 A_{22}\right) N_{22}^{-1} \\ \hline -N_{22}^{-1}\left(A_{22}^T P_2 A_{21}\right) S_1^{-1} & N_{22}^{-1} + N_{22}^{-1}\left(A_{22}^T P_2 A_{21}\right) S_1^{-1}\left(A_{21}^T P_2 A_{22}\right) N_{22}^{-1} \end{array}\right] \times$$
$$\times \begin{bmatrix} \left(A_{11}^T P_1 A_{11}\right)\hat{\boldsymbol{\xi}}_1 + A_{21}^T P_2 \boldsymbol{y}_2 \\ A_{22}^T P_2 \boldsymbol{y}_2 \end{bmatrix}. \quad (8.35)$$

We can continue by using (8.33), (8.34b) and (8.34c) with the first row of (8.35) to arrive at

$$\begin{aligned}
\hat{\hat{\boldsymbol{\xi}}}_1 &= S_1^{-1}\big[\big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T P_2 \boldsymbol{y}_2 - \big(A_{21}^T P_2 A_{22}\big) N_{22}^{-1} A_{22}^T P_2 \boldsymbol{y}_2\big] = && (8.36a)\\
&= S_1^{-1}\big\{\big[\big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T \bar{P}_2 \boldsymbol{y}_2\big] + \big[\big(A_{21}^T \bar{P}_2 A_{21}\big) - \big(A_{21}^T \bar{P}_2 A_{21}\big)\big]\hat{\boldsymbol{\xi}}_1\big\} = && (8.36b)\\
&= S_1^{-1} A_{21}^T \bar{P}_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) + S_1^{-1}\big(A_{11}^T P_1 A_{11} + A_{21}^T \bar{P}_2 A_{21}\big)\hat{\boldsymbol{\xi}}_1 = && (8.36c)\\
&= S_1^{-1} A_{21}^T \bar{P}_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) + \hat{\boldsymbol{\xi}}_1 \Rightarrow && (8.36d)\\
\hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 &= S_1^{-1} A_{21}^T \bar{P}_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big), && (8.36e)
\end{aligned}$$

where (8.36e) is in the form of an update formula.

We assume that P_2 is invertible, as implied in the given model (8.29). We now wish to check the rank of the reduced weight matrix $\bar{P}_2$. It is easy to check that the product $P_2^{-1}\bar{P}_2$ is idempotent. Then using (A.4) and (A.12) we find

$$\begin{aligned}
\operatorname{rk}\bar{P}_2 &= \operatorname{rk}\big(P_2^{-1}\bar{P}_2\big) = \operatorname{tr}\big(P_2^{-1}\bar{P}_2\big) = \operatorname{tr}\big(I_{n_2} - A_{22}\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2\big) = && (8.37a)\\
&= n_2 - \operatorname{tr}\big[A_{22}\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2\big] = n_2 - \operatorname{tr}\big[\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 A_{22}\big] = && (8.37b)\\
&= n_2 - m_2 < n_2. && (8.37c)
\end{aligned}$$

Thus we reduce the rank by modifying the original weight matrix P_2 to obtain $\bar{P}_2$. Moreover, we find that matrix $\bar{P}_2$ is singular.

The parameter dispersion matrices, $D\{\hat{\hat{\boldsymbol{\xi}}}_1\}$ and $D\{\hat{\hat{\boldsymbol{\xi}}}_2\}$, are shown at the end of the next section.

8.5 Sequential Adjustment with New Parameters and Small Second Data Set

In (8.36e) we must invert the $m_1 \times m_1$ matrix S_1 to solve the system of equations. However, in some applications, the number of observations n_2 in the second data set may be significantly less than m_1. In this case we would like to reformulate the solution in (8.36e) so that only a matrix of size $n_2 \times n_2$ needs to be inverted.

We have an alternative expression for matrix S_1 in (8.34b), the inverse of which can also be derived as follows:

$$S_1^{-1} = \big[\big(A_{11}^T P_1 A_{11}\big) + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big]^{-1} = \tag{8.38a}$$

$$= \big\{\big[I_{m_1} + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big(A_{11}^T P_1 A_{11}\big)^{-1}\big]\big(A_{11}^T P_1 A_{11}\big)\big\}^{-1} = \tag{8.38b}$$

$$= \big(A_{11}^T P_1 A_{11}\big)^{-1}\big[I_{m_1} + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big(A_{11}^T P_1 A_{11}\big)^{-1}\big]^{-1}. \tag{8.38c}$$

Using (8.38c) we may rewrite (8.36e) as

$$\hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 = \big(A_{11}^T P_1 A_{11}\big)^{-1}\big[I_{m_1} + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big(A_{11}^T P_1 A_{11}\big)^{-1}\big]^{-1} A_{21}^T \bar{P}_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) = \tag{8.39a}$$

$$= \big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big). \tag{8.39b}$$

Here, we have made use of (A.9) in the step from (8.39a) to (8.39b), with two of the matrices in (A.9) set to identity. Note that the matrix to invert inside the square brackets is of size $m_1 \times m_1$ in (8.39a) but is size $n_2 \times n_2$ in (8.39b). The choice of which equation to use will usually be determined by the smaller of m_1 and n_2. Also, we have the relation

$$-\hat{\hat{\boldsymbol{\lambda}}} = \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big), \tag{8.40}$$

which means that the solution for the first subset of parameters may also be expressed as

$$\hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 = -\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2 \hat{\hat{\boldsymbol{\lambda}}}. \tag{8.41}$$

Now we begin with (8.32) to find a solution for the parameters $\hat{\hat{\boldsymbol{\xi}}}_2$ in terms of the Lagrange multipliers $\hat{\hat{\boldsymbol{\lambda}}}$:

$$\hat{\hat{\boldsymbol{\xi}}}_2 = \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2\big(\boldsymbol{y}_2 - A_{21}\hat{\hat{\boldsymbol{\xi}}}_1\big) = \tag{8.42a}$$

$$= \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \cdot \big\{\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) - A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2 \times$$

$$\times \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)\big\} = \tag{8.42b}$$

$$= \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2\big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) = \tag{8.42c}$$

$$= -\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \hat{\hat{\boldsymbol{\lambda}}}. \tag{8.42d}$$

The inverse formula of (A.7) was used in the last step to reach (8.42c), with matrices T, W, and V in (A.7) set to identity.

To facilitate computing the parameter dispersion matrix we write the following system of normal equations, noting that the second line, (8.43b), is in the form of an update solution:

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{21}^T P_2 A_{22} \\ \hline A_{22}^T P_2 A_{21} & A_{22}^T P_2 A_{22} \end{array}\right] \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\xi}}}_2 \end{bmatrix} = \begin{bmatrix} \big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T P_2 \boldsymbol{y}_2 \\ A_{22}^T P_2 \boldsymbol{y}_2 \end{bmatrix} \Rightarrow \tag{8.43a}$$

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{21}^T P_2 A_{22} \\ \hline A_{22}^T P_2 A_{21} & A_{22}^T P_2 A_{22} \end{array}\right] \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 \\ \hat{\hat{\boldsymbol{\xi}}}_2 \end{bmatrix} = \begin{bmatrix} A_{21}^T P_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) \\ A_{22}^T P_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) \end{bmatrix} \tag{8.43b}$$

Note that (8.43a) is equivalent to (8.31b) shown earlier.

We have already inverted the normal-equation matrix in (8.35). Taking elements from (8.35), we may write the parameter dispersion and covariance matrices as follows:

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = \sigma_0^2 S_1^{-1} = \sigma_0^2 \big(A_{11}^T P_1 A_{11} + A_{21}^T \bar{P}_2 A_{21}\big)^{-1}, \tag{8.44a}$$

$$C\{\hat{\hat{\boldsymbol{\xi}}}_1, \hat{\hat{\boldsymbol{\xi}}}_2\} = -D\{\hat{\hat{\boldsymbol{\xi}}}_1\}\big(A_{21}^T P_2 A_{22}\big)\big(A_{22}^T P_2 A_{22}\big)^{-1}, \tag{8.44b}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}_2\} = \sigma_0^2 \big(A_{22}^T P_2 A_{22}\big)^{-1} - \big(A_{22}^T P_2 A_{22}\big)^{-1}\big(A_{22}^T P_2 A_{21}\big) C\{\hat{\hat{\boldsymbol{\xi}}}_1, \hat{\hat{\boldsymbol{\xi}}}_2\}. \tag{8.44c}$$

Each of the above covariance matrices (8.44a) through (8.44c) include the matrix S_1^{-1}, which implies that a matrix of size $m_1 \times m_1$ must be inverted. However, with the insertion of I_{n_2} into (8.44a), and with appropriate matrix groupings, we may apply the inversion formula (A.7) to find an inverse of smaller dimension as shown in the following:

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = \sigma_0^2 \big[\big(A_{11}^T P_1 A_{11}\big) + \big(A_{21}^T \bar{P}_2\big) I_{n_2} A_{21}\big]^{-1} = \tag{8.45a}$$

$$= \sigma_0^2 N_{11}^{-1} - \sigma_0^2 N_{11}^{-1} A_{21}^T \bar{P}_2 \big(I_{n_2} + A_{21} N_{11}^{-1} A_{21}^T \bar{P}_2\big)^{-1} A_{21} N_{11}^{-1} = \tag{8.45b}$$

$$= \sigma_0^2 N_{11}^{-1} - \sigma_0^2 N_{11}^{-1} A_{21}^T \bar{P}_2 \hat{\hat{\boldsymbol{\lambda}}}. \tag{8.45c}$$

Here, we have used $N_{11} \coloneqq A_{11}^T P_1 A_{11}$ for compactness. The parenthetical term that must be inverted in equation (8.45b) is an $n_2 \times n_2$ matrix, which, again, may be much smaller than an $m_1 \times m_1$ matrix, depending on the application. Of course, the matrix $(A_{11}^T P_1 A_{11})^{-1}$ is also size $m_1 \times m_1$, but it is assumed that this inverse has been performed and saved in the first adjustment.

The estimated variance component is expressed as

$$\hat{\hat{\sigma}}_0^2 (n - m_2) = \hat{\sigma}_0^2 (n_1 - m_1) - \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)^T \bar{P}_2 \hat{\hat{\boldsymbol{\lambda}}}. \tag{8.46a}$$

Then, substituting (8.40) leads to:

$$\begin{aligned} \hat{\hat{\sigma}}_0^2 (n - m_2) &= \hat{\sigma}_0^2 (n_1 - m_1) + \\ &\quad + \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)^T \bar{P}_2 \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1} \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) = \qquad (8.46b) \\ &= \hat{\sigma}_0^2 (n_1 - m_1) + \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)^T \bar{P}_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) \qquad (8.46c) \end{aligned}$$

Chapter 9

Condition Equations with Parameters — the Gauss-Helmert Model

Data models introduced prior to this chapter have treated parameters and condition equations separately. The Gauss-Helmert model (GHM) combines condition equations and parameters into a single model. In some cases the GHM might be useful for dealing with complicated observation equations.

We begin our discussion with a leveling-network example in order to contrast the Gauss-Markov model (GMM) with the model of condition equations presented in Chapter 4 and to show how the Gauss-Helmert model combines the information used in these two models. The diagram in Figure 9.1 shows a leveling network with four points (P_1, P_2, P_3, P_4) that has been observed in two closed loops comprised of a total of five observations $(y_1, y_2, y_3, y_4, y_5)$. The general form of the observation equations is

$$\boldsymbol{y} = A_1\boldsymbol{\xi}_1 + A_2\boldsymbol{\xi}_2 + \boldsymbol{e}, \tag{9.1a}$$

$$\operatorname{rk} A_1 = \operatorname{rk}\big[A_1,\ A_2\big] =: q < m, \tag{9.1b}$$

$$\boldsymbol{e} \sim \big(\boldsymbol{0}, \sigma_0^2 P^{-1}\big), \tag{9.1c}$$

where the coefficient matrix A and the vector of unknown parameters $\boldsymbol{\xi}$ have been partitioned, respectively, as

$$A = \left[\begin{array}{c|c} \underset{n\times q}{A_1} & \underset{n\times(m-q)}{A_2} \end{array}\right] \text{ and } \boldsymbol{\xi} = \left[\begin{array}{c|c} \underset{1\times q}{\boldsymbol{\xi}_1^T} & \underset{1\times(m-q)}{\boldsymbol{\xi}_2^T} \end{array}\right]^T. \tag{9.2}$$

In this example, $m = 4$ (four heights). Since leveled height-differences supply no information about the height datum, we can only estimate the heights of three of the points $(m - 1 = q = 3)$ with respect to the remaining one $(m - q = 1)$. Thus, the model has been partitioned so that $\boldsymbol{\xi}_1$ contains three estimable heights, and $\boldsymbol{\xi}_2$ contains a non-estimable height, which could be assigned a datum value. In this example, we arbitrarily chose point P_4 for the non-estimable height. As was stated in Chapter 6, we have the relationship $A_2 = A_1 L$, for some $q \times (m - q)$ matrix L, which means

that matrix A_2 is a linear combination of the columns of matrix A_1, reflecting the rank deficiency of matrix A.

The problem could also be solved within the model of condition equations introduced in Chapter 4, which reads

$$B\boldsymbol{y} = B\boldsymbol{e}, \tag{9.3a}$$

with conditions

$$\text{i. } B\begin{bmatrix} A_1 & A_2 \end{bmatrix} = 0, \tag{9.3b}$$

and

$$\text{ii. } \operatorname{rk} B = r = n - \operatorname{rk} A_1. \tag{9.3c}$$

These two conditions ensure equivalent solutions within the models of (9.1) and (9.3a) as discussed in Section 4.1.

We have the following design matrices and parameter vector for the example leveling network, for which it is easy to verify that conditions i and ii are satisfied:

$$A_1 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}, \quad \boldsymbol{\xi}_1 = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}, \quad \boldsymbol{\xi}_2 = \begin{bmatrix} h_4 \end{bmatrix}, \tag{9.4}$$

where h_i is the height of point P_i.

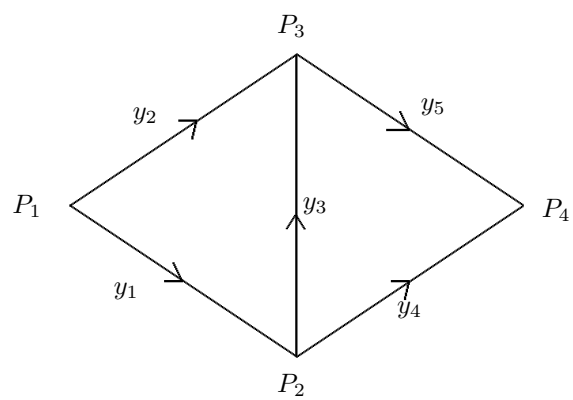

Figure 9.1: Leveling network. Arrows point in the direction of the level runs.

Now we wish to introduce a new coefficient matrix B that does not contain matrix A in its nullspace, so that we are left with a model of condition equations with parameters. For now we use the symbol $\bar{B}$ in order to distinguish from the coefficient matrix used in the model of condition equations (which does contain matrix A in its nullspace). Similarly, we introduce other bar-terms as follows:

$$\bar{\boldsymbol{y}} = \bar{B}\boldsymbol{y} = \bar{\boldsymbol{w}} = \bar{B}A_1\boldsymbol{\xi}_1 + \bar{B}A_2\boldsymbol{\xi}_2 + \bar{B}\boldsymbol{e}, \tag{9.5a}$$

$$\operatorname{rk}\big(\bar{B}\big) =: \bar{r}, \tag{9.5b}$$

$$\bar{B}\boldsymbol{e} \sim \big(\boldsymbol{0}, \sigma_0^2 \bar{B} P^{-1} \bar{B}^T\big). \tag{9.5c}$$

The size of $\bar{B}$ is $\bar{r} \times n$. The model in (9.5) is equivalent to (9.1), if and only if,

iii) $\bar{B}A_1$ has $n - \bar{r}$ columns of zeros, and

iv) $\mathrm{rk}(\bar{B}A_1) + r = \bar{r} \Leftrightarrow n = \bar{r} + q - \mathrm{rk}(\bar{B}A_1) = \mathrm{rk}\,\bar{B} + \mathrm{rk}\,A - \mathrm{rk}(\bar{B}A_1)$

Note that, through the matrix $\bar{B}$, one observation is eliminated for each eliminated parameter. Referring to the level network example, we may wish to eliminate the height of point P_3 from the parameter list (perhaps it is a temporary benchmark of no particular interest). This can be done by introducing the following example matrix $\bar{B}$:

$$\bar{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \Rightarrow \bar{B}A_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \bar{B}A_1 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

With these example matrices we have $n = 5$, $\bar{r} = \mathrm{rk}\,\bar{B} = 4$, $q = \mathrm{rk}\,A_1 = 3$, and $\mathrm{rk}(\bar{B}A_1) = 2$. Since $n - \bar{r} = 1$, the single column of zeros in $\bar{B}A_1$ satisfies condition iii. Also, condition iv is satisfied since $5 = 4 + 3 - 2$.

As an aside, we note that it is also possible to remove l estimable parameters via the splitting of the constraint equation introduced in (5.1).

$$\boldsymbol{\kappa}_0 = \underset{l \times m}{K}\, \boldsymbol{\xi} = \begin{bmatrix} K_1, & K_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \Rightarrow \tag{9.6a}$$

$$\boldsymbol{\xi}_1 = K_1^{-1}\boldsymbol{\kappa}_0 - K_1^{-1}K_2\boldsymbol{\xi}_2 \tag{9.6b}$$

Here K_1 is a $l \times l$ invertible matrix, and K_2 is of size $l \times (m - l)$. Upon substitution for $\boldsymbol{\xi}_1$ of (9.6b) into (9.1) we find the following modified system of observation equations with l parameters eliminated:

$$\boldsymbol{y} = A_1\boldsymbol{\xi}_1 + A_2\boldsymbol{\xi}_2 + \boldsymbol{e} = A_1K_1^{-1}\boldsymbol{\kappa}_0 + (A_2 - A_1K_1^{-1}K_2)\boldsymbol{\xi}_2 + \boldsymbol{e}. \tag{9.7}$$

The $l \times 1$ vector $\boldsymbol{\xi}_1$ has vanished on the right side of (9.7). While this technique is possible, it might not be used frequently in practice.

We could derive the solution for $\boldsymbol{\xi}$ within (9.5) from statistical principles via BLUUE (Best Linear Uniformly Unbiased Estimate), but here we use the equivalent principle of LESS (LEast-Squares Solution) shown in Section 4.1. In the following, we recombine coefficient matrices A_1 and A_2 back into the single matrix A and recombine the partitioned parameter vector back into a single vector $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^T, \boldsymbol{\xi}_2^T]^T$. Therefore, we can rewrite (9.5) as

$$\bar{\boldsymbol{w}} = \bar{B}A_1\boldsymbol{\xi}_1 + \bar{B}A_2\boldsymbol{\xi}_2 + \bar{B}\boldsymbol{e} = \bar{A}\boldsymbol{\xi} + \bar{B}\boldsymbol{e}. \tag{9.8}$$

Our target function should minimize a quadratic form in the random error vector $\boldsymbol{e}$. That is, we minimize $\boldsymbol{e}^T P\boldsymbol{e}$ rather than $(\bar{B}\boldsymbol{e})^T(\bar{B}P^{-1}\bar{B}^T)^{-1}(\bar{B}\boldsymbol{e})$.

For convenience, another bar-symbol is introduced at this point: $\bar{A} := \bar{B}A$. Then, the Lagrange Target function is written as

$$\Phi(\boldsymbol{e}, \boldsymbol{\xi}, \boldsymbol{\lambda}) =: \boldsymbol{e}^T P\boldsymbol{e} + 2\boldsymbol{\lambda}^T\big(\bar{B}\boldsymbol{e} + \bar{A}\boldsymbol{\xi} - \bar{\boldsymbol{w}}\big), \tag{9.9}$$

which is stationary with respect to the unknown vectors $\boldsymbol{e}$, $\boldsymbol{\xi}$, and $\boldsymbol{\lambda}$.

The Euler-Lagrange necessary conditions result in the following system of equations:

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{e}} = P\tilde{\boldsymbol{e}} + \bar{B}^T\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{9.10a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\xi}} = \bar{A}^T\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{9.10b}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \bar{B}\tilde{\boldsymbol{e}} + \bar{A}\hat{\boldsymbol{\xi}} - \bar{\boldsymbol{w}} \doteq \mathbf{0}. \tag{9.10c}$$

The predicted error and estimated parameter vectors are then solved for as follows:

$$\begin{aligned}
&\tilde{\boldsymbol{e}} = -\big(P^{-1}\bar{B}^T\big)\hat{\boldsymbol{\lambda}} \Rightarrow && \text{from equation (9.10a)}\\
&-\big(\bar{B}P^{-1}\bar{B}^T\big)\hat{\boldsymbol{\lambda}} = \bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}} \Rightarrow && \text{multiplying by } \bar{B} \text{ and using (9.10c)}\\
&-\hat{\boldsymbol{\lambda}} = \big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big(\bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}}\big) \Rightarrow && \big(\bar{B}P^{-1}\bar{B}^T\big) \text{ is invertible}\\
&-\bar{A}^T\hat{\boldsymbol{\lambda}} = \bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big(\bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}}\big) = \mathbf{0} \Rightarrow && \text{multiplying by } \bar{A}^T \text{ and using (9.10b)}\\
&\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{A}\hat{\boldsymbol{\xi}} = \bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{\boldsymbol{w}}
\end{aligned} \tag{9.11}$$

Finally, we arrive at

$$\hat{\boldsymbol{\xi}} = \big[\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{A}\big]^{-1}\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{\boldsymbol{w}}, \tag{9.12a}$$

$$\tilde{\boldsymbol{e}} = \big(P^{-1}\bar{B}^T\big)\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big(\bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}}\big). \tag{9.12b}$$

for the estimated parameters and predicted residuals, respectively. Equation (9.12a) has the same form as the normal equations derived within the GMM. The dispersion of the estimate $\hat{\boldsymbol{\xi}}$ is expressed by

$$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\big[\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{A}\big]^{-1}. \tag{9.13}$$

Notation change: For the remainder of the discussion we drop the bars from the symbols as a matter of convenience. Recall that the bars were introduced in the first place to distinguish between the matrix B introduced in (9.5) and that used in Chapter 4 for the model of condition equations. Dropping the bars means that $B \coloneqq \bar{B}$, $\boldsymbol{w} \coloneqq \bar{\boldsymbol{w}}$, $BA \coloneqq \bar{A}$. With this simplified notation, we rewrite the solution (9.12a) as follows:

$$\boxed{\hat{\boldsymbol{\xi}} = \big[(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big]^{-1}(BA)^T\big(BP^{-1}B^T\big)^{-1}\boldsymbol{w}.} \tag{9.14}$$

The dispersion of $\boldsymbol{\xi}$ is derived in parts as follows:

$$\begin{aligned}
D\{(BA)^T\big(BP^{-1}B^T\big)^{-1}\boldsymbol{w}\} &= (BA)^T\big(BP^{-1}B^T\big)^{-1}D\{\boldsymbol{w}\}\big(BP^{-1}B^T\big)^{-1}BA = \\
&= \sigma_0^2(BA)^T\big(BP^{-1}B^T\big)^{-1}\big(BP^{-1}B^T\big)\big(BP^{-1}B^T\big)^{-1}BA = \sigma_0^2(BA)^T\big(BP^{-1}B^T\big)^{-1}BA,
\end{aligned}$$

therefore

$$\begin{aligned}
D\{\hat{\boldsymbol{\xi}}\} &= \big[(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big]^{-1}D\{(BA)^T\big(BP^{-1}B^T\big)^{-1}\boldsymbol{w}\}\big[(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big]^{-1} = \\
&= \big[(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big]^{-1}\sigma_0^2(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big[(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big]^{-1} = \\
&\boxed{D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\big[(BA)^T\big(BP^{-1}B^T\big)^{-1}BA\big]^{-1}.}
\end{aligned} \tag{9.15}$$

9.1 Estimated Variance Component

The P-weighted norm of the residual vector $\tilde{\boldsymbol{e}}$ is defined as

$$\Omega \coloneqq \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \tag{9.16a}$$

$$= \big(\hat{\boldsymbol{\lambda}}^T B P^{-1}\big) P \big(P^{-1} B^T \hat{\boldsymbol{\lambda}}\big) = \tag{9.16b}$$

$$= \big[\big(\boldsymbol{w} - BA\hat{\boldsymbol{\xi}}\big)^T \big(BP^{-1}B^T\big)^{-1}\big]\big(BP^{-1}B^T\big)\hat{\boldsymbol{\lambda}} = \tag{9.16c}$$

$$= \big(\boldsymbol{w} - BA\hat{\boldsymbol{\xi}}\big)^T \big(BP^{-1}B^T\big)^{-1}\big(\boldsymbol{w} - BA\hat{\boldsymbol{\xi}}\big) = \tag{9.16d}$$

$$= \big(B\tilde{\boldsymbol{e}}\big)^T \big(BP^{-1}B^T\big)^{-1}\big(B\tilde{\boldsymbol{e}}\big). \tag{9.16e}$$

Thus is follows that, the *uniformly unbiased estimate* of the variance component σ_0^2 is given by

$$\boxed{\hat{\sigma}_0^2 = \frac{\big(B\tilde{\boldsymbol{e}}\big)^T \big(BP^{-1}B^T\big)^{-1}\big(B\tilde{\boldsymbol{e}}\big)}{r} = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{r} = \frac{-\boldsymbol{w}^T \hat{\boldsymbol{\lambda}}}{r},} \tag{9.17}$$

where, using the recombined matrix A, the redundancy r is defined as $r \coloneqq \operatorname{rk} B - \operatorname{rk}(BA)$.

9.2 Equivalent Normal Equations

From (9.10b) and the second equation following (9.10c), we can recognize the following system of normal equations:

$$\begin{bmatrix} BP^{-1}B^T & -BA \\ -(BA)^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \left[\begin{array}{c|c} BP^{-1}B^T & -BA \\ \hline -(BA)^T & 0 \end{array}\right]^{-1} \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \end{bmatrix}. \tag{9.18}$$

We want to show that the solution to this system yields the same $\hat{\boldsymbol{\xi}}$ as that of (9.14). The formula for the inverse of a partitioned matrix (see (A.14)) leads to the following solution:

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} =$$

$$\left[\begin{array}{c|c} X_1 & X_2 \\ \hline -\big[(BA)^T(BP^{-1}B^T)^{-1}BA\big]^{-1}(BA)^T(BP^{-1}B^T)^{-1} & \big[0 - (BA)^T(BP^{-1}B^T)^{-1}BA\big]^{-1} \end{array}\right] \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \end{bmatrix},$$

and finally to

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} -X_1\boldsymbol{w} \\ \big[(BA)^T(BP^{-1}B^T)^{-1}BA\big]^{-1}(BA)^T(BP^{-1}B^T)^{-1}\boldsymbol{w} \end{bmatrix}. \tag{9.19}$$

Here the symbols X_1 and X_2 represent quantities of no interest. We see that the solution for the parameters $\hat{\boldsymbol{\xi}}$ is the same in (9.14).

Chapter 10

Statistical Testing and Confidence Ellipses

Consider a normally distributed random (univariate) variable y with the following first through fourth moments:

$$E\{y\} = \mu, \tag{10.1a}$$
$$E\{(y-\mu)^2\} = D\{y\} = \sigma^2, \tag{10.1b}$$
$$E\{(y-\mu)^3\} = 0, \tag{10.1c}$$
$$E\{(y-\mu)^4\} = 3(\sigma^2)^2. \tag{10.1d}$$

The third moment being zero in (10.1c) means there is no skewness in the distribution of the random variable. The right side of (10.1d) indicates that there is no kurtosis (peak) in the distribution.

If (10.1c) or (10.1d) are not satisfied, the variable is not normally distributed and can be characterized as follows:

$$E\{(y-\mu)^3\} > 0 \Leftrightarrow \text{the distribution is skewed to the positive side.} \tag{10.2a}$$
$$E\{(y-\mu)^3\} < 0 \Leftrightarrow \text{the distribution is skewed to the negative side.} \tag{10.2b}$$
$$E\{(y-\mu)^4\} - 3(\sigma^2)^2 > 0 \Leftrightarrow \text{the distribution has positive kurtosis.} \tag{10.2c}$$
$$E\{(y-\mu)^4\} - 3(\sigma^2)^2 < 0 \Leftrightarrow \text{the distribution has negative kurtosis.} \tag{10.2d}$$

Skewness appears in a graph of a sample of the random variable (e.g., a histogram) as a shift in the peak value from center. Positive kurtosis shows higher probability near the expected value μ, which results in a taller, narrower graph. Negative kurtosis shows higher probability in the tails of the graph; thus the graph appears flatter than that of a normally distributed variable.

The pdf (probability density function, or density function) of a normally distributed random (univariate) variable y is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(y-\mu)^2/2\sigma^2}, \tag{10.3}$$

where μ is the expectation of the distribution, σ is standard deviation, and σ^2 is variance. Note that the term $1/\sqrt{2\pi\sigma^2} \approx 0.4/\sigma$ denotes the amplitude of the graph of the curve, μ shows the offset of the peak from center, and σ is the distance from the center to the inflection points of the curve.

The cdf (cumulative distribution function, or distribution function) of a *normally distributed* random variable is expressed as:

$$F(y) = \int_{-\infty}^{y} f(t)dt = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(t-\mu)^2/2\sigma^2} dt. \tag{10.4}$$

Figure 10.1 shows pdf and cdf plots for the normal distribution using various values for μ and σ_0^2. Line colors and types match between the pdf and cdf plots. The solid, green line represents the respective standard-normal pdf and cdf curves.

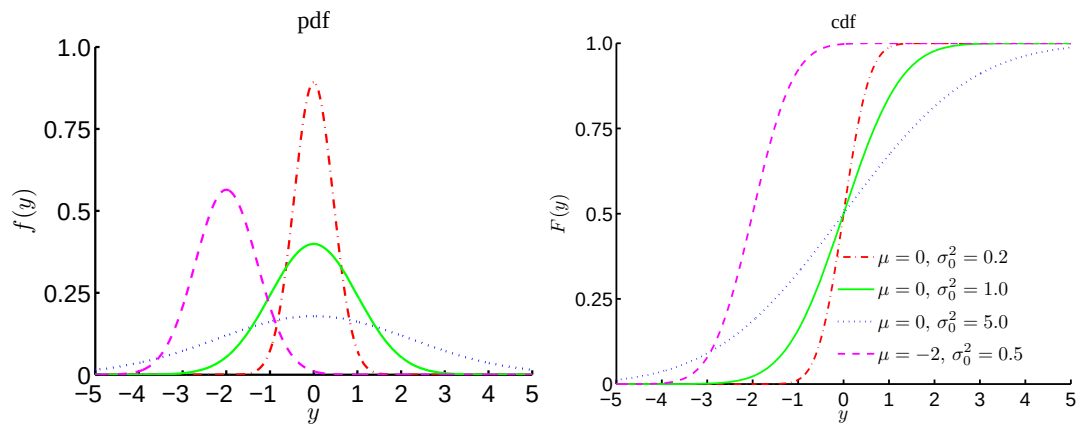

Figure 10.1: pdf and cdf for the normal distribution with matching line types and colors

Note that, in geodetic-science applications, the random variable y might be an observation, an adjusted observation, a predicted residual, etc. We can standardize the random variable y with the following transformation, which subtracts out the mean and divides by the standard deviation:

$$z = \frac{y-\mu}{\sigma}. \tag{10.5}$$

The standardized random variable z has the following moments and probability functions:

$$E\{z\} = 0, \tag{10.6a}$$

$$D\{z\} = 1, \tag{10.6b}$$

$$\text{pdf}: \ f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}, \tag{10.6c}$$

$$\text{cdf}: \ F(z) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt. \tag{10.6d}$$

A plot of the pdf of z is shown in Figure 10.2, along with example Student's-t distribution curves (discussed below).

In the multivariate case, the random variable $\boldsymbol{y}$ is an $n \times 1$ vector, with dispersion (covariance) matrix Σ and expectation $\boldsymbol{\mu}$, which is also size $n \times 1$. The pdf is then written as

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}}\, e^{-(\boldsymbol{y}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})/2}. \tag{10.7}$$

And the cdf is written as

$$F(y_1, \ldots, y_n) = \int_{-\infty}^{y_n} \ldots \int_{-\infty}^{y_1} f(t_1, \ldots, t_n) dt_1, \ldots, dt_n. \tag{10.8}$$

The elements of $\boldsymbol{y}$, i.e. $y_1, \ldots, y_n$, are statistically independent if, and only if,

$$f(t_1, \ldots, t_n) = f(t_1) \cdot f(t_2) \ldots f(t_n), \tag{10.9a}$$

which implies

$$C\{y_i, y_j\} = 0 \;\text{ for } i \neq j. \tag{10.9b}$$

Equation (10.9b) states that there is no covariance between the elements of random vector $\boldsymbol{y}$.

The third and fourth moments for the multivariate case are given in (10.10a) and (10.10b), respectively.

$$E\{(y_i - \mu_i)(y_j - \mu_j)(y_k - \mu_k)\} = 0 \;\text{ for } i, j, k = \{1, \ldots, n\} \tag{10.10a}$$

$$E\{(y_i - \mu_i)(y_j - \mu_j)(y_k - \mu_k)(y_l - \mu_l)\} = 3(\sigma_i^2)\delta_{ijkl} \;\text{ for } i, j, k, l = \{1, \ldots, n\} \tag{10.10b}$$

In the following, we discuss *studentized residuals*, which have a t-distribution (or Student's t-distribution). The pdf for a (univariate) variable having a t-distribution and having $\nu = n - 1$ degrees of freedom is defined as follows:

$$f(t) = \frac{1}{\sqrt{(n-1)\pi}} \cdot \frac{\Gamma(n/2)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}}, \tag{10.11}$$

where the gamma function is defined by

$$\Gamma(n) := (n-1)\Gamma(n-1) = \int_0^\infty e^{-t} t^{n-1} dt = (n-1)! \;\text{ for } n \in \mathbb{N}. \tag{10.12}$$

As is known from introductory statistics texts, the pdf for the Student's t-distribution converges to the pdf of the normal distribution as n approaches 30. A plot of the pdf for the Student's t-distribution, with $\nu = 2, 4$, together with the pdf for the normal distribution, is shown in Figure 10.2.

10.1 Standardized and Studentized Residuals

We begin this section by restating the (full rank) Gauss-Markov model and writing the predicted vector of random errors within the model.

$$\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}, \;\; \boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 P^{-1}\right), \;\; \operatorname{rk} A = m \tag{10.13a}$$

$$\tilde{\boldsymbol{e}} = \left(I_n - AN^{-1}A^T P\right)\boldsymbol{y} = \left(I_n - AN^{-1}A^T P\right)\boldsymbol{e} \tag{10.13b}$$

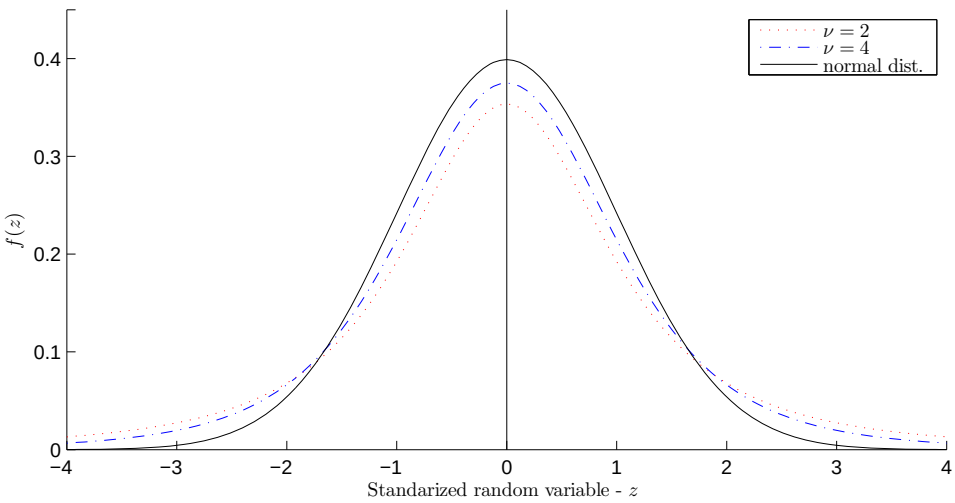

Figure 10.2: Student's t and normal distributions for a standardized random variable z

As usual, the observation vector $\boldsymbol{y}$ is of size $n \times 1$, and the coefficient matrix A is of size $n \times m$. Obviously, the far-right side of (10.13b) cannot be computed since $\boldsymbol{e}$ is an unknown variable. However, the expression is useful for analytical purposes.

The so-called standardized residual is a function of the residual vector $\tilde{\boldsymbol{e}}$ and its dispersion matrix $D\{\tilde{\boldsymbol{e}}\}$ as shown below.

$$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big), \tag{10.14a}$$

$$\boldsymbol{\eta}_j \coloneqq \big[0, \ldots, 0, \underset{j\text{th}}{1}, 0, \ldots, 0\big]^T, \tag{10.14b}$$

$$\sigma_{\tilde{e}_j}^2 = \boldsymbol{\eta}_j^T D\{\tilde{\boldsymbol{e}}\}\boldsymbol{\eta}_j = E\{\tilde{e}_j^2\}, \tag{10.14c}$$

Then, the jth standardized residual is defined as

$$\tilde{e}_j / \sigma_{\tilde{e}_j}. \tag{10.15}$$

Since the variance component σ_0^2 is considered unknown in the model (10.13a), we typically replace it with its estimate $\hat{\sigma}_0^2$. This replacement leads to the following analogous set of equations for the *studentized residual*:

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - \mathrm{rk}(A)} = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T N^{-1} \boldsymbol{c}}{n - m}, \tag{10.16a}$$

$$\hat{D}\{\tilde{\boldsymbol{e}}\} = \hat{\sigma}_0^2\big(P^{-1} - AN^{-1}A^T\big), \tag{10.16b}$$

$$\hat{\sigma}_{\tilde{e}_j}^2 = \boldsymbol{\eta}_j^T \hat{D}\{\tilde{\boldsymbol{e}}\}\boldsymbol{\eta}_j = \hat{E}\{\tilde{e}_j^2\}. \tag{10.16c}$$

Then the studentized residual is defined as

$$\tilde{e}_j / \hat{\sigma}_{\tilde{e}_j}. \tag{10.17}$$

Note that the denominator in (10.15) is constant (due to the unknown but constant variance component σ_0^2), whereas the denominator of (10.17) is random due to the introduction of the estimate $\hat{\sigma}_0^2$,

which is random. Of course the numerator is random in both cases. Using Q to represent cofactor matrices, we can rewrite the standardized/studentized residuals in the following alternative form:

$$\text{standardized residual:} \quad \tilde{e}_j \Big/ \sqrt{\sigma_0^2 \big(Q_{\tilde{\boldsymbol{e}}}\big)_{jj}} \sim \mathcal{N}(0,1), \tag{10.18a}$$

$$\text{studentized residual:} \quad \tilde{e}_j \Big/ \sqrt{\hat{\sigma}_0^2 \big(Q_{\tilde{\boldsymbol{e}}}\big)_{jj}} \sim t(n-1). \tag{10.18b}$$

Here $D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 Q_{\tilde{\boldsymbol{e}}}$, and $(Q_{\tilde{\boldsymbol{e}}})_{jj}$ denotes the jth diagonal element of the residual cofactor matrix $Q_{\tilde{\boldsymbol{e}}}$. Standardized residuals are normally distributed, whereas Studentized residuals follow the student t-distribution. Again, it is noted that (10.18a) cannot be computed unless the variance component σ_0^2 is known.

Example: Direct observations of a single parameter μ with weight matrix $P = I_n$.

$$\boldsymbol{y} = \boldsymbol{\tau}\mu + \boldsymbol{e}, \quad \boldsymbol{e} \sim \mathcal{N}\big(\boldsymbol{0}, \sigma_0^2 I_n\big), \quad \text{with } \boldsymbol{\tau} = [1, \ldots, 1]^T$$

$$\hat{\mu} = \frac{\boldsymbol{\tau}^T \boldsymbol{y}}{\boldsymbol{\tau}^T \boldsymbol{\tau}} = \frac{1}{n}\big(y_1 + \ldots + y_n\big) \sim \mathcal{N}\big(\mu, \sigma_0^2/n\big)$$

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu} \sim \mathcal{N}(\boldsymbol{0}, \sigma_0^2 \left[I_n - \frac{\boldsymbol{\tau}\boldsymbol{\tau}^T}{n} \right])$$

$$Q_{\tilde{\boldsymbol{e}}} = I_n - \frac{\boldsymbol{\tau}\boldsymbol{\tau}^T}{n}$$

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T \tilde{\boldsymbol{e}}}{(n-1)}$$

The formula for $Q_{\tilde{\boldsymbol{e}}}$ in the above example means that $(Q_{\tilde{\boldsymbol{e}}})_{jj} = (n-1)/n$, which shows that the more observations we have (i.e., the larger n is), the more the dispersion of the *predicted* random error $D\{\tilde{\boldsymbol{e}}\}$ approaches the dispersion of the *true* random error $D\{\boldsymbol{e}\}$. In this example the standardized and studentized residuals are written as follows:

$$\text{standardized:} \quad \frac{\tilde{e}_j}{\sqrt{\sigma_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}}} = \frac{\tilde{e}_j \sqrt{n}}{\sigma_0 \sqrt{n-1}} \sim \mathcal{N}(0,1) \tag{10.19a}$$

$$\text{or alternatively:} \quad \frac{\tilde{e}_j}{\sqrt{(Q_{\tilde{\boldsymbol{e}}})_{jj}}} = \frac{\tilde{e}_j \sqrt{n}}{\sqrt{n-1}} \sim \mathcal{N}(0, \sigma_0^2) \tag{10.19b}$$

$$\text{studentized:} \quad \frac{\tilde{e}_j}{\sqrt{\hat{\sigma}_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}}} = \frac{\tilde{e}_j}{\sqrt{\tilde{\boldsymbol{e}}^T \tilde{\boldsymbol{e}}}} \sqrt{n} \sim t(n-1) \tag{10.19c}$$

We extend the example by including a hypothesis test for the parameter estimate $\hat{\mu}$ against a specified value μ_0 at a significance level α (see Section 10.2 for a more complete discussion of hypothesis testing).

$$\text{Hypothesis test:} \quad H_0 : E\{\hat{\mu}\} = \mu_0 \ \text{ against } H_A : E\{\hat{\mu}\} \neq \mu_0.$$

$$\text{Test statistic:} \quad t = \frac{\hat{\mu} - \mu_0}{\sqrt{\hat{\sigma}_0^2}} \sqrt{n} \sim t(n-1).$$

We accept the null hypothesis H_0 if $t_{-\alpha/2} \leq t \leq t_{\alpha/2}$; otherwise we reject H_0. We may perform a similar test $H_0 : E\{\tilde{e}_j\} = 0$ for the jth residual. In this case the test statistic is the studentized residual computed by (10.19c).

10.2 Hypothesis Testing Within the Gauss-Markov Model

The hypothesis test introduced in Section 10.1 for direct observations of a single parameter is now extended to the Gauss-Markov Model (GMM). In introducing the GMM in Chapter 3, a probability density function was not given for the random observation errors; only the first and second moments of the errors were specified. This is indeed all that is necessary to formulate and solve the least-squares estimation problem within the GMM. However, in order to perform hypothesis testing after the least-squares estimate has been computed, the probability distribution must be specified. Typically, we assume that the observation errors have a normal distribution. Then, the (full rank) GMM is succinctly written as

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \;\; \mathrm{rk}(A) = m, \;\; \boldsymbol{e} \sim \mathcal{N}\big(\boldsymbol{0}, \sigma_0^2 P^{-1}\big). \tag{10.20}$$

Minimization of the observation errors via a least-squares adjustment leads to the following parameter estimate and predicted-error vectors, shown with their corresponding (assumed) normal distributions:

$$\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} \sim \mathcal{N}\big(\boldsymbol{\xi}, \sigma_0^2 N^{-1}\big), \tag{10.21a}$$

$$\tilde{\boldsymbol{e}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{y} \sim \mathcal{N}\big(\boldsymbol{0}, \sigma_0^2\big[P^{-1} - AN^{-1}A^T\big]\big). \tag{10.21b}$$

Or equivalently, we could write for the predicted residual vector

$$\tilde{\boldsymbol{e}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{e} = Q_{\tilde{\boldsymbol{e}}} P\boldsymbol{y} \sim \mathcal{N}\big(\boldsymbol{0}, \sigma_0^2 Q_{\tilde{\boldsymbol{e}}}\big), \tag{10.22a}$$

with

$$Q_{\tilde{\boldsymbol{e}}} := P^{-1} - AN^{-1}A^T. \tag{10.22b}$$

The jth standardized and studentized residuals are then written as follows:

$$j\text{th standardized residual: } \tilde{e}_j \Big/ \sqrt{\sigma_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}} \sim \mathcal{N}(0,1), \tag{10.23}$$

$$j\text{th studentized residual: } \tilde{e}_j \Big/ \sqrt{\hat{\sigma}_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}} \sim t(n - \mathrm{rk}\,A). \tag{10.24}$$

As shown in Chapter 3, we compute the estimated reference variance within this model by

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - \mathrm{rk}\,A}. \tag{10.25}$$

The hypothesis test for the jth studentized residual then becomes

$$H_0 : E\{\tilde{e}_j\} = 0 \;\text{ versus }\; H_A : E\{\tilde{e}_j\} \neq 0. \tag{10.26}$$

Likewise, we may test individual elements of the estimated parameter vector $\hat{\boldsymbol{\xi}}$. For example, we may want to compare the jth element of the parameter vector, $\hat{\xi}_j$, against some specified value $\xi_j^{(0)}$. In this case, the null hypothesis and computed test statistic are defined as follows:

$$H_0 : E\{\hat{\xi}_j\} = \xi_j^{(0)} \;\text{ versus }\; H_A : E\{\hat{\xi}_j\} \neq \xi_j^{(0)}, \tag{10.27a}$$

$$t_j = \frac{\hat{\xi}_j - \xi_j^{(0)}}{\sqrt{\hat{\sigma}_0^2 \big(N^{-1}\big)_{jj}}} \sim t(n - \mathrm{rk}\,A) \;\text{ or }\; t_j^2 \sim F(1, n - \mathrm{rk}\,A). \tag{10.27b}$$

Note that from (10.27b) we see that the square of a test statistic having a Student's t-distribution has a Fisher distribution (also called F-distribution). For a given significance level α, we accept H_0 if $t_{-\alpha/2} \leq t_j \leq t_{\alpha/2}$; otherwise we reject H_0. We can use a cdf table for the t-distribution to find the value of $t_{\alpha/2}(n - \operatorname{rk} A)$. Note that α is the probability of making a Type I error (also called the significance level of the test), and $n - \operatorname{rk} A$ is the degrees of freedom, often denoted by ν in the statistical literature.

10.3 Confidence Intervals for Ellipses, Ellipsoids, and Hyperellipsoids

Confidence intervals specify the statistical probability that a univariate random variable X lies within a certain range of values. Confidence ellipses, ellipsoids, and hyperellipsoids are the respective 2-D, 3-D, and n-D analog to confidence intervals.

10.3.1 Confidence Intervals (univariate case)

By definition, the cdf (cumulative distribution function) of X is

$$F_X(x) = P(X \leq x), \quad -\infty < x < \infty. \tag{10.28}$$

It follows from (10.28) that the probability the random variable X lies within a given range is computed by

$$P(a < X \leq b) = F_X(b) - F_X(a). \tag{10.29}$$

Applying (10.29) to the normalized random variable z of (10.5), we can write the following probabilities for confidence intervals bounded by $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$, respectively, from the mean, where $\sigma = 1$ since $z \sim \mathcal{N}(0, 1)$ according to (10.6a) and (10.6b).

$$P(-1 < z \leq 1) = P(\mu - \sigma < y \leq \mu + \sigma) \quad = 68.3\% \tag{10.30a}$$
$$P(-2 < z \leq 2) = P(\mu - 2\sigma < y \leq \mu + 2\sigma) = 95.5\% \tag{10.30b}$$
$$P(-3 < z \leq 3) = P(\mu - 3\sigma < y \leq \mu + 3\sigma) = 99.7\% \tag{10.30c}$$

Often we speak of a confidence interval in terms of percent and wish to find the range of values the random variable can take on in this interval. Some common intervals for a normally distributed random (univariate) variable z are

$$90\% = P(-1.645 < z \leq 1.645), \tag{10.31a}$$
$$95\% = P(-1.960 < z \leq 1.960), \tag{10.31b}$$
$$99\% = P(-2.576 < z \leq 2.576). \tag{10.31c}$$

10.3.2 Confidence Ellipses — Bivariate Case

Analogous to the univariate case, we write the probability for the standard ("1-sigma") confidence ellipse as follows:

$$P\big((\boldsymbol{y}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})<1\big)=19.9\%. \tag{10.32}$$

We are using vector notation again: $\boldsymbol{y}$ is a random 2-D vector and $\boldsymbol{\mu}$ is the expected value of $\boldsymbol{y}$. Also, we have a 2×2 dispersion matrix for $\boldsymbol{y}$, namely Σ. More specifically we have

$$\boldsymbol{y}=\begin{bmatrix}y_1\\y_2\end{bmatrix},\quad \boldsymbol{\mu}=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix},\quad \Sigma\coloneqq D\{\boldsymbol{y}\}=\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{bmatrix},\quad \sigma_{12}=\sigma_{21}. \tag{10.33}$$

When speaking of the elements of the vectors and matrix in (10.33), we say that μ_1 is the expected value of y_1; σ_1^2 is the variance of y_1 (with σ_1 called standard deviation), and σ_{12} is the covariance between y_1 and y_2.

Using the above definitions, together with equation (10.7), we can write the pdf of $\boldsymbol{y}$ explicitly as

$$\begin{aligned}f(\boldsymbol{y})=&\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2-\sigma_{12}^2}}\times\\ &\times\exp\left\{-\frac{\sigma_1^2\sigma_2^2}{2(\sigma_1^2\sigma_2^2-\sigma_{12}^2)}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2}-\left(2\sigma_{12}\frac{(y_1-\mu_1)}{\sigma_1^2}\frac{(y_2-\mu_2)}{\sigma_2^2}\right)+\frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]\right\},\end{aligned} \tag{10.34}$$

where exp stands for the exponential function, i.e., $\exp\{x\}=e^x$.

Each element of the vector $\boldsymbol{y}$ may be normalized according to (10.5), so that the jth element of the normalized vector $\boldsymbol{z}$ is expressed in terms of the corresponding jth element of $\boldsymbol{y}$; that is $z_j=(y_j-\mu_j)/\sigma_j$, $j=1,2$. Also, we define the *correlation coefficient* as

$$\rho_{12}=\frac{\sigma_{12}}{\sigma_1\sigma_2}. \tag{10.35}$$

Substituting z_j and ρ_{12} into (10.34) we can write the following pdf for the vector $\boldsymbol{z}$:

$$f(z_1,z_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}}\cdot\exp\left\{-\frac{1}{2\big(1-\rho_{12}^2\big)}\big(z_1^2-2\rho_{12}z_1z_2+z_2^2\big)\right\}. \tag{10.36}$$

Setting the quadratic form in (10.32) equal to 1 gives the equation for the standard confidence ellipse (also called standard error ellipse by Mikhail). From (10.36) we see that the standard confidence ellipse can also be described by

$$\big(z_1^2-2\rho_{12}z_1z_2+z_2^2\big)=1-\rho_{12}^2. \tag{10.37}$$

The size, shape, and orientation of the confidence ellipse is described by the eigenvalues and eigenvectors of Σ.

The eigenvector-eigenvalue decomposition of the 2×2 matrix Σ is described as follows: Denote the eigenvectors of Σ as $\boldsymbol{u}_j$ and the eigenvalues as λ_j, $j=1,2$. Then we have the relation

$$\Sigma\boldsymbol{u}_j=\lambda\boldsymbol{u}_j, \tag{10.38}$$

for which we write the following characteristic equation:

$$\det\big(\Sigma - \lambda I_2\big) = \big(\sigma_1^2 - \lambda\big)\big(\sigma_2^2 - \lambda\big) - \sigma_{12}^2 = \lambda^2 - \big(\sigma_1^2 + \sigma_2^2\big)\lambda + \big(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\big) = 0. \tag{10.39}$$

In (10.39) λ has been used in general to represent either eigenvalue λ_1 or λ_2. We require $\lambda_1 \geq \lambda_2 > 0$ and write the following solution for the roots of the characteristic equation (10.39):

$$\lambda_{1 \text{ or } 2} = \frac{\sigma_1^2 + \sigma_2^2}{2} \pm \sqrt{\left(\frac{\sigma_1^2 + \sigma_2^2}{2}\right)^2 - \frac{1}{4}4\sigma_1^2\sigma_2^2 + \frac{4\sigma_{12}^2}{4}} = \tag{10.40a}$$

$$= \lambda_{1 \text{ or } 2} = \frac{\sigma_1^2 + \sigma_2^2}{2} \pm \frac{1}{2}\sqrt{\big(\sigma_1^2 - \sigma_2^2\big)^2 + 4\sigma_{12}^2} > 0, \tag{10.40b}$$

which shows that the eigenvalues must be greater than zero, since Σ is positive definite.

Now we must find the two corresponding eigenvectors. Let the matrix U be comprised of the two eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ such that $U \coloneqq [\boldsymbol{u}_1,\ \boldsymbol{u}_2]$. Also define a diagonal matrix comprised of the corresponding eigenvalues $\Lambda \coloneqq \operatorname{diag}(\lambda_1,\ \lambda_2)$. Then according to (10.38) we have

$$\Sigma U = U\Lambda = \tag{10.41a}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \tag{10.41b}$$

$$= \left[\begin{array}{c|c} \sigma_1^2 u_{11} + \sigma_{12}u_{21} & \sigma_1^2 u_{12} + \sigma_{12}u_{22} \\ \hline \sigma_{12}u_{11} + \sigma_2^2 u_{21} & \sigma_{12}u_{12} + \sigma_2^2 u_{22} \end{array}\right] = \begin{bmatrix} \lambda_1 \cdot u_{11} & \lambda_2 \cdot u_{12} \\ \lambda_1 \cdot u_{21} & \lambda_2 \cdot u_{22} \end{bmatrix}. \tag{10.41c}$$

From (10.41c) we can write the following four equations in the four unknowns u_{11}, u_{12}, u_{21}, and u_{22}:

$$u_{21} = \frac{\big(\lambda_1 - \sigma_1^2\big)u_{11}}{\sigma_{12}},\quad u_{21} = \frac{\sigma_{12}u_{11}}{\lambda_1 - \sigma_2^2},\quad u_{12} = \frac{\sigma_{12}u_{22}}{\lambda_2 - \sigma_1^2},\quad u_{12} = \frac{\big(\lambda_2 - \sigma_2^2\big)u_{22}}{\sigma_{12}}. \tag{10.42}$$

The eigenvector $\boldsymbol{u}_1$ defines the direction of the *semimajor axis* of the confidence ellipse, while the eigenvector $\boldsymbol{u}_2$, orthogonal to $\boldsymbol{u}_1$, defines the *semiminor axis* direction. The square root of the eigenvalue λ_1 gives the semimajor-axis length, and the square root of the eigenvalue λ_2 gives the semiminor-axis length. Also, if θ is the angle measured counter clockwise from the positive z_1-axis to the semimajor axis of the confidence ellipse, then we can write the matrix U as

$$U = [\boldsymbol{u}_1,\ \boldsymbol{u}_2] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{10.43}$$

Using (10.42) and (10.43), the angle θ is derived as follows:

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{u_{21}}{u_{11}} = \frac{\lambda_1 - \sigma_1^2}{\sigma_{12}} = \frac{\sigma_{12}}{\lambda_1 - \sigma_2^2} = -\frac{u_{12}}{u_{22}} = \frac{\sigma_2^2 - \lambda_2}{\sigma_{12}} = \frac{\sigma_{12}}{\sigma_1^2 - \lambda_2} \quad \text{and} \tag{10.44a}$$

$$\tan(2\theta) = \tan(\theta + \theta) = \frac{2\tan\theta}{1 - \tan^2\theta} = \left(\frac{2\sigma_{12}}{\lambda_1 - \sigma_2^2}\right)\frac{1}{1 - \frac{\sigma_{12}^2}{\left(\lambda_1 - \sigma_2^2\right)^2}}\left(\frac{\lambda_1 - \sigma_2^2}{\lambda_1 - \sigma_2^2}\right) = \tag{10.44b}$$

$$= \tan(2\theta) = \frac{2\sigma_{12}\big(\lambda_1 - \sigma_2^2\big)}{\big(\lambda_1 - \sigma_2^2\big)^2 - \sigma_{12}^2} = \frac{2\sigma_{12}\big(\lambda_1 - \sigma_2^2\big)4}{\big[2\big(\lambda_1 - \sigma_2^2\big)\big]^2 - 4\sigma_{12}^2}. \tag{10.44c}$$

By manipulating (10.40b), we have

$$2\big(\lambda_1 - \sigma_2^2\big) = \big(\sigma_1^2 - \sigma_2^2\big) \pm \sqrt{\big(\sigma_1^2 - \sigma_2^2\big)^2 + 4\sigma_{12}^2} \Rightarrow \tag{10.45a}$$

$$\big[2\big(\lambda_1 - \sigma_2^2\big)\big]^2 = 2\big(\sigma_1^2 - \sigma_2^2\big)^2 \pm 2\big(\sigma_1^2 - \sigma_2^2\big)\sqrt{\big(\sigma_1^2 - \sigma_2^2\big)^2 + 4\sigma_{12}^2} + 4\sigma_{12}^2. \tag{10.45b}$$

Substituting (10.45a) and (10.45b) into (10.44c) gives

$$\tan(2\theta) = \frac{4\sigma_{12}\left[\big(\sigma_1^2 - \sigma_2^2\big) \pm \sqrt{\big(\sigma_1^2 - \sigma_2^2\big)^2 + 4\sigma_{12}^2}\right]}{2\big(\sigma_1^2 - \sigma_2^2\big)\left[\big(\sigma_1^2 - \sigma_2^2\big) \pm \sqrt{\big(\sigma_1^2 - \sigma_2^2\big)^2 + 4\sigma_{12}^2}\right]} = \tag{10.46a}$$

$$= \tan(2\theta) = \frac{2\sigma_{12}}{\sigma_1^2 - \sigma_2^2}. \tag{10.46b}$$

The sign of the numerical value of the right side of (10.46b) tells which quadrant the positive axis falls in.

An empirical error ellipse differs from the confidence ellipse in that the matrix Σ is replaced by the estimated matrix $\hat{\Sigma}$ such that $\hat{\Sigma}^{-1} = \hat{\sigma}_0^{-2}P$, where $\hat{\sigma}_0^2$ is the estimated variance component. In this case, the empirical error ellipse is described by

$$\frac{(\boldsymbol{y} - \hat{\boldsymbol{\mu}})^T P(\boldsymbol{y} - \hat{\boldsymbol{\mu}})}{\hat{\sigma}_0^2} = 1. \tag{10.47}$$

If we are evaluating $n/2$ number of 2-D points, so that P is of size $n \times n$, we may simply work with each of the ($n/2$ number of) 2×2 block diagonal matrices of $\hat{\sigma}_0^{-2}P$ independently to form the empirical error ellipses. However, we must bear in mind that these block diagonal matrices do not tell the whole story since the off-block-diagonal elements have been ignored. In any case, it may be prudent to verify that the associated correlation-coefficients of the off-block-diagonal elements are relatively small in magnitude.

The following two examples apply to the Gauss-Markov model (GMM):

1. Consider the GMM (10.20), with an associated least-squares solution and dispersion given in (10.21a). Assume that the parameter vector $\boldsymbol{\xi}$ is comprised of successive 2-D point coordinates such that $(\hat{\xi}_{2i-1},\ \hat{\xi}_{2i})$ represents the coordinate estimates of the ith point. Now, also assume that we wish to compare the estimates with given (fixed) values $(\xi_{2i-1}^0,\ \xi_{2i}^0)$, perhaps from published results of a previous adjustment. Then we may write the following equations for the null hypothesis and the error ellipse (for convenience, let $k = 2i$ and $j = k - 1$):

$$H_0 : E\{\big[\hat{\xi}_j,\ \hat{\xi}_k\big]^T\} = \big[\xi_j^0,\ \xi_k^0\big]^T, \tag{10.48a}$$

$$\frac{1}{\hat{\sigma}_0^2}\begin{bmatrix}\hat{\xi}_j - \xi_j^0 \\ \hat{\xi}_k - \xi_k^0\end{bmatrix}^T \begin{bmatrix} N_{j,j} & N_{j,k} \\ N_{k,j} & N_{k,k}\end{bmatrix}\begin{bmatrix}\hat{\xi}_j - \xi_j^0 \\ \hat{\xi}_k - \xi_k^0\end{bmatrix} = 1. \tag{10.48b}$$

   Here we have formed the sub-blocks of N before inverting N.

2. Suppose that instead of comparing the solution to given, fixed values we want to compare the results (2-D coordinate estimates) of two adjustments. Using the previously defined indices, let the estimates of the ith point of the second adjustment be represented by $(\hat{\hat{\xi}}_j,\ \hat{\hat{\xi}}_k)$. We ask the question: is the outcome of the second adjustment statistically equivalent to the first? Unless there is statistically significant overlap of the respective error ellipses, the answer is no. The null hypothesis H_0 and the test statistic f are defined as follows:

$$H_0 : E\{\left[\hat{\xi}_j,\ \hat{\xi}_k\right]^T\} = E\{\left[\hat{\hat{\xi}}_j,\ \hat{\hat{\xi}}_k\right]^T\} \tag{10.49a}$$

$$f := \frac{1}{2}\frac{1}{\hat{\sigma}_0^2/\sigma_0^2}\begin{bmatrix}\hat{\xi}_j - \hat{\hat{\xi}}_j\\ \hat{\xi}_k - \hat{\hat{\xi}}_k\end{bmatrix}^T D\{\begin{bmatrix}\hat{\xi}_j - \hat{\hat{\xi}}_j\\ \hat{\xi}_k - \hat{\hat{\xi}}_k\end{bmatrix}\}^{-1}\begin{bmatrix}\hat{\xi}_j - \hat{\hat{\xi}}_j\\ \hat{\xi}_k - \hat{\hat{\xi}}_k\end{bmatrix} \sim F(2, n - \operatorname{rk} A) \tag{10.49b}$$

In computing the test statistic f shown in (10.49b), it is assumed that the estimated variance component $\hat{\sigma}_0^2$ is common to both adjustments. This assumption is based on the homogeneity test $H_0 : E\{\hat{\sigma}_0^2\} = E\{\hat{\hat{\sigma}}_0^2\}$, which is discussed in Section 10.4. Note that in the case that the two adjustments are uncorrelated, we could replace the differences of parameters in the inverted dispersion matrix with their sums.

10.3.3 Confidence Ellipsoids and Hyperellipsoids — Multivariate Case

In the 3-D case, confidence ellipses are extended to confidence ellipsoids. But, in our general formulation of the GMM we may be working with any arbitrary higher-dimensional space, and thus we speak of confidence hyperellipsoids. Since 3-D and higher dimensions are natural extensions of the 2-D case, no further discussion is necessary.

10.4 χ^2-distribution, Variance Testing, and F-distribution

This section includes the statistical topics of χ^2- and F-distributions as well as the topic of variance testing.

10.4.1 χ^2-distribution (F.R. Helmert, 1876)

If we claim that the (unknown) random error vector $\boldsymbol{e}$ from the GMM is distributed as $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 P^{-1})$, this leads to a quadratic product in $\boldsymbol{e}$ that has a χ^2-distribution with $\nu := \operatorname{rk} P = n$ *degrees of freedom*, expressed by

$$\frac{\boldsymbol{e}^T P \boldsymbol{e}}{\sigma_0^2} \sim \chi^2(\nu). \tag{10.50}$$

Now, define $x := \boldsymbol{e}^T P \boldsymbol{e}/\sigma_0^2$ (which cannot actually be computed since both $\boldsymbol{e}$ and σ_0^2 are unknown). Nevertheless, the pdf of x is written as

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu-2)/2} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0, \end{cases} \tag{10.51}$$

where e is mathematical constant $2.718\ldots$ The gamma function $\Gamma(\cdot)$ was defined in (10.12). Figure 10.3 shows plots of the χ^2-distribution for $\nu = \{1, 3, 5, 8, 10, 30\}$ with respective colors: black, magenta, cyan, red, green, blue. Note that the peaks of the curves move to the right as ν increases and that the curves appear to approximate the normal-distribution curve as ν grows to 10 and larger. This agrees with our expectation that the χ^2-distribution is *asymptotically normal*, due to the *central limit theorem*.

From the variance component derivations in Section 3.2, we can write

$$E\{\boldsymbol{e}^T P \boldsymbol{e}/\sigma_0^2\} = \operatorname{tr}\big(P \cdot E\{\boldsymbol{e}\boldsymbol{e}^T\}\big) = \operatorname{tr} I_n = n, \tag{10.52a}$$

$$E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}/\sigma_0^2\} = \operatorname{tr}\big(P \cdot E\{\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}^T\}\big) = \operatorname{tr}\big(I_n - AN^{-1}A^T P\big) = n - \operatorname{rk} A = n - m. \tag{10.52b}$$

Equations (10.25) and (10.52b) lead to

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}/\sigma_0^2 = \nu \hat{\sigma}_0^2/\sigma_0^2 \sim \chi^2(\nu), \tag{10.53a}$$

with

$$\nu := n - \operatorname{rk}(A) \tag{10.53b}$$

as the degrees of freedom.

Note that though we have been discussing the random error vector e and the predicted residual $\tilde{\boldsymbol{e}}$, the relations expressed in (10.53a) apply to all quadratic forms in normally distributed variables. Thus, when we have a vector of normally distributed variables, the corresponding quadratic form will have a χ^2-distribution.

10.4.2 Variance Testing

Suppose we want to compare the estimated variance component $\hat{\sigma}_0^2$ to a given quantity σ^2 (here the 0-subscript is not used so as not to confuse the given value with the "true value"). We do so by performing the following hypothesis test at a chosen significance level α:

$$H_0 : E\{\hat{\sigma}_0^2\} \leq \sigma^2 \;\text{ vs. }\; H_A : E\{\hat{\sigma}_0^2\} > \sigma^2 \tag{10.54a}$$

$$t := (n - \operatorname{rk} A) \cdot \big(\hat{\sigma}_0^2/\sigma^2\big) \sim \chi^2(n - \operatorname{rk} A) \tag{10.54b}$$

$$\text{If } t \leq \chi_\alpha^2 \;\text{ accept } H_0, \;\text{ else if } t > \chi_\alpha^2 \;\text{ reject} H_0. \tag{10.54c}$$

Note that we could have constructed a two-tailed hypothesis test, rather than a one-tailed test. In that case we would have used an equality sign in the null hypothesis H_0 and a non-equality sign in the alternative hypothesis. Generally speaking, if the estimate $\hat{\sigma}_0^2$ proves statistically to be less than the given value σ^2, we deem our measurements to be more precise than that reflected in the weight matrix P. On the other hand, if $\hat{\sigma}_0^2$ proves statistically to be greater than the given value, we deem our measurements to be less precise. Usually our main concern is that $\hat{\sigma}_0^2$ reflects that our measurements are at least as precise as what is specified in P, thus the use of a single-tail hypothesis may be more commonly used in practice.

In the case where we need to compare two estimated reference variances $\hat{\sigma}_{0,1}^2$ and $\hat{\sigma}_{0,2}^2$ from independent adjustments, we must compute a ratio of test statistics, which has a Fisher distribution

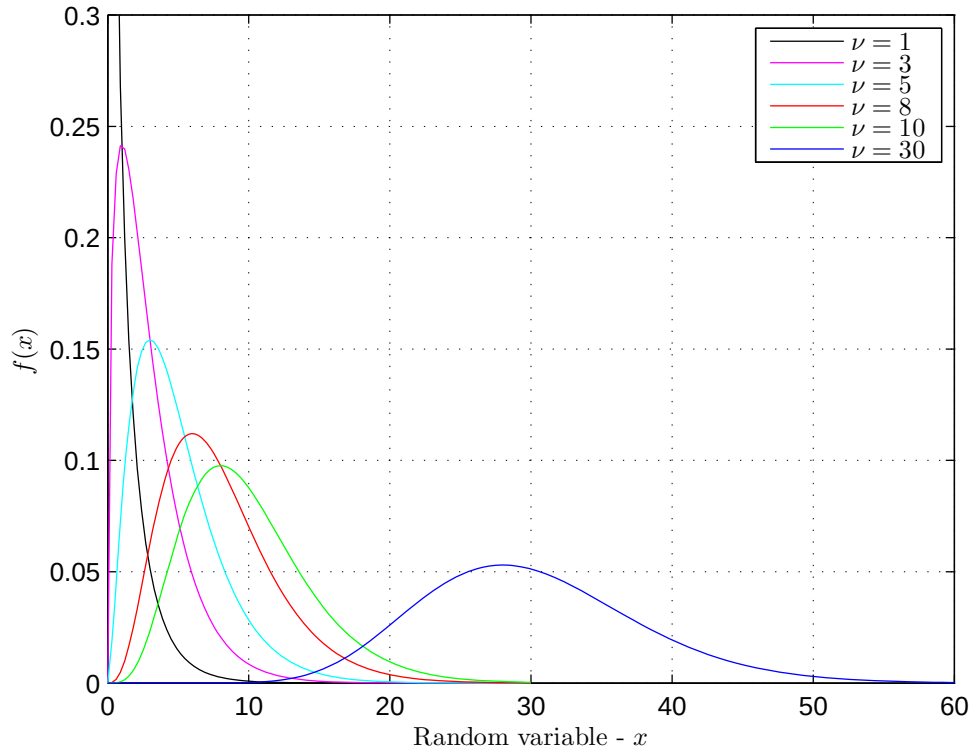

Figure 10.3: χ^2-distribution with various degrees of freedom ν

(assuming both the numerator and denominator have χ^2 distributions). Let t_1 and t_2 be the test statistics from the respective adjustments; then we can write

$$\frac{t_1/(n_1 - m_1)}{t_2/(n_2 - m_2)} = \hat{\sigma}^2_{0,1}/\hat{\sigma}^2_{0,2} \sim F(n_1 - m_1, n_2 - m_2), \tag{10.55}$$

where $n_i - m_i$, $i = 1, 2$, are the respective degrees of freedom of the two independent data sets.

10.4.3 F-distribution (R.A. Fisher and Snedacor, 1925)

The ratio of two mutually independent χ^2-distributed variables has an F-distribution. The pdf for the F-distribution with degrees of freedom $v_1 := m$ and $v_2 := n - m$ is given by

$$f(w) = \frac{\Gamma\big(\frac{m}{2} + \frac{n-m}{2}\big) m^{m/2} (n-m)^{(n-m)/2} w^{(m/2)-1}}{\Gamma(\frac{m}{2})\Gamma\big(\frac{n-m}{2}\big)(n - m + mw)^{(m/2+(n-m)/2)}} = \tag{10.56a}$$

$$= \frac{(v_1/v_2)^{v_1/2}\Gamma\big((v_1+v_2)/2\big) w^{(v_1/2)-1}}{\Gamma(v_1/2)\Gamma(v_2/2)\big(1 + v_1 w/v_2\big)^{(v_1+v_2)/2}}. \tag{10.56b}$$

As n becomes large compared to m, the curve of the F-distribution pdf approaches the normal distribution pdf curve.

10.5 Hypothesis Testing on the Estimated Parameters

In the GMM, we may wish to perform a global model check by comparing a specified parameter vector $\boldsymbol{\xi}^0$ to the estimated vector $\hat{\boldsymbol{\xi}}$. In such a case, we may use as the test statistic the ratio of weighted norms of the difference vector $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0$ and the predicted residual vector $\tilde{\boldsymbol{e}}$ as follows:

$$w := \frac{(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0)^T A^T P A (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0)}{\sigma_0^2 \operatorname{rk} A} \cdot \frac{\sigma_0^2 (n - \operatorname{rk} A)}{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}} \sim F(m, n-m). \tag{10.57}$$

Assuming that matrix A has full rank, i.e., $\operatorname{rk} A = m$, the numerator and denominator both have χ^2-distributions with m and $n-m$ degrees of freedom, respectively. Since the numerator and denominator are statistically independent of one another, the test statistic w has an F-distribution with m and $n-m$ degrees of freedom, as shown in (10.57). Therefore, our global model check is made by the following hypothesis test:

$$H_0 : E\{\hat{\boldsymbol{\xi}}\} = \boldsymbol{\xi}^0 \ \text{ vs. } \ H_A : E\{\hat{\boldsymbol{\xi}}\} \neq \boldsymbol{\xi}^0 \tag{10.58a}$$

$$\text{If } w \leq F_{\alpha,m,n-m} \ \text{ accept } H_0; \ \text{ else } \ w > F_{\alpha,m,n-m} \ \text{ therefore reject } H_0. \tag{10.58b}$$

We now show that the numerator and denominator of w are indeed independent, as required for use of the F-distribution. To do so, we only need to show that

$$C\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}, (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^T (A^T P A)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\} = \mathbf{0}. \tag{10.59}$$

Note that, without loss of generality, we have replaced $\boldsymbol{\xi}^0$ with $\boldsymbol{\xi}$. From (4.5e) we have $\tilde{\boldsymbol{e}} = [I_n - AN^{-1}A^T P]\boldsymbol{e}$. Therefore,

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \boldsymbol{e}^T \big[I_n - PAN^{-1}A^T\big] P \big[I_n - AN^{-1}A^T P\big]\boldsymbol{e} = \boldsymbol{e}^T \big[P - PAN^{-1}A^T P\big]\boldsymbol{e} = \boldsymbol{e}^T M_1 \boldsymbol{e}. \tag{10.60a}$$

Also

$$A(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \boldsymbol{e} - \tilde{\boldsymbol{e}} = \big(AN^{-1}A^T P\big)\boldsymbol{e} \Rightarrow \tag{10.60b}$$

$$(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^T \big(A^T P A\big)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \boldsymbol{e}^T \big(PAN^{-1}A^T\big) P \big(AN^{-1}A^T P\big)\boldsymbol{e} = \tag{10.60c}$$

$$= \boldsymbol{e}^T \big(PAN^{-1}A^T P\big)\boldsymbol{e} = \boldsymbol{e}^T M_2 \boldsymbol{e}. \tag{10.60d}$$

By substitution of (10.60a) and (10.60d), the condition (10.59) is equivalent to the condition that $\boldsymbol{e}^T M_1 \boldsymbol{e}$ and $\boldsymbol{e}^T M_2 \boldsymbol{e}$ are independent, which holds if, and only if,

$$\boldsymbol{e}^T M_1 D\{\boldsymbol{e}\} M_2 \boldsymbol{e} = \boldsymbol{e}^T \big(P - PAN^{-1}A^T P\big)\big(\sigma_0^2 P^{-1}\big)\big(PAN^{-1}A^T P\big)\boldsymbol{e} = 0, \tag{10.60e}$$

which is obviously true.

10.6 Checking an Individual Element (or 2-D or 3-D Point) in the Parameter Vector

We may use the $l \times m$ matrix K, with $\operatorname{rk} K = l$, to select a subset of the parameter vector for hypothesis testing as follows:

$$H_0 : E\{K\hat{\boldsymbol{\xi}}\} = K\boldsymbol{\xi}^0 = \boldsymbol{\kappa}_0, \tag{10.61a}$$

$$H_A : E\{K\hat{\boldsymbol{\xi}}\} = K\boldsymbol{\xi}^0 \neq \boldsymbol{\kappa}_0. \tag{10.61b}$$

If $l = 1$, K is a unit row vector that extracts the relevant element from the parameter vector, in which case κ_0 is simply a scalar quantity. The following examples show the matrix K used for extracting a single element, a 2-D point, and a 3-D point, respectively:

$$K \coloneqq \big[0, \ldots, 0, 1, 0, \ldots, 0\big], \text{ where 1 appears at the } j\text{th element;} \tag{10.62a}$$

$$K \coloneqq \big[0_2, \ldots, 0_2, I_2, 0_2, \ldots, 0_2\big], \text{ where } K \text{ is size } 2 \times m; \tag{10.62b}$$

$$K \coloneqq \big[0_3, \ldots, 0_3, I_3, 0_3, \ldots, 0_3\big], \text{ where } K \text{ is size } 3 \times m. \tag{10.62c}$$

For the 2-D and 3-D points, the subscripts denote the dimension of the square sub-matrices (zero matrix or identity matrix), and I_n ($n =\in \{2, 3\}$) is the jth sub-matrix, which means it "selects" the jth point from $\boldsymbol{\xi}$.

The test statistic is then defined as

$$w := \frac{\big[K\big(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0\big)\big]^T D\{K\big(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0\big)\}^{-1}\big[K\big(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0\big)\big]/\operatorname{rk} K}{1/\sigma_0^2} = \tag{10.63a}$$

$$= \frac{\big[K\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0\big]^T\big[KN^{-1}K^T\big]^{-1}\big[K\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0\big]/l}{\hat{\sigma}_0^2} =: \frac{R/l}{\Omega/(n-m)}, \tag{10.63b}$$

where σ_0^2 is assumed to be 1, and thus omitted from (10.63b). Note that since $\boldsymbol{\xi}^0$ is a chosen (and therefore non-random) quantity to test against, the dispersion is not affected by the constant shift, i.e.,

$$D\{K(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0)\} = D\{K\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 KN^{-1}K^T. \tag{10.64}$$

The symbols R and Ω are used for convenience and are analogous to the symbols introduced in Sections 5.1 and 7.2, respectively. They are statistically independent of one another and have the following distributions:

$$R \sim \chi^2(l), \;\; \Omega \sim \chi^2(n-m). \tag{10.65}$$

The statistical independence means that the joint pdf is equivalent to the product of the individual pdf's: $f(R, \Omega) = f(R) \cdot f(\Omega)$. Independence can be shown by following the same line of thought as that used at the end of the previous section, where M_1 remains unchanged and M_2 is now $PAN^{-1}K^T\big[KN^{-1}K^T\big]^{-1}KN^{-1}A^TP$. Therefore, the test statistic (10.63b) has an F-distribution represented by

$$w \sim F(l, n-m). \tag{10.66}$$

An alternative, more compact, form for w when $l = 1$ is given by

$$w = \frac{(\hat{\xi}_j - (\kappa_0)_j)^2}{\hat{\sigma}_0^2\big(N^{-1}\big)_{jj}} \sim F(1, n-m). \tag{10.67}$$

The decision to accept or reject the null hypothesis is made analogous to (10.58b).

10.6.1 Non-central F-distribution

If the null hypothesis H_0 is false, the test statistic w is said to have a non-central F-distribution (denoted here as F'), which requires a non-centrality parameter θ so that $w \sim F'\big(v_1, v_2, \theta\big)$ under H_A,

where v_1 and v_2 have been used to denote the degrees of freedom, in general. The qualification "under H_A" implies that we must pose a specific alternative hypothesis H_A in this case, rather than just the negation of H_0. For a one-tailed test, the area under the non-central F-distribution curve and to the right of F_α (from the F-distribution table) is denoted as β. The value of β is also the probability of making an *error of the second kind*, namely to accept the null hypothesis H_0 when the specified alternative hypothesis H_A is actually true. The quantity $1-\beta$ is known as the *power of the test*. As the value of θ increases, so does the value $1-\beta$. Below we have rewritten (10.66) for the non-central case, with the theoretical formula for 2θ following.

$$w \sim F'(l, n-m, \theta) \tag{10.68a}$$

$$2\theta = \big(K\boldsymbol{\xi} - \boldsymbol{\kappa}_0\big)^T \big(KN^{-1}K^T\big)^{-1} \big(K\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0\big) \tag{10.68b}$$

Note that the non-centrality property is reflected in (10.68b) by using both the true (unknown) $\boldsymbol{\xi}$ and the estimated $\hat{\boldsymbol{\xi}}$ in a bilinear form.

10.7 Detection of a Single Outlier in the Gauss-Markov Model

An outlier in the jth observation can be modeled by

$$y_j = \boldsymbol{a}_j^T \boldsymbol{\xi}^{(j)} + \xi_0^{(j)} + e_j. \tag{10.69}$$

The symbol j is being used as both a vector index and as an indicator of a particular vector. The following explanation should make this clear. The variable y_j is the jth element of the $n \times 1$ observation vector $\boldsymbol{y}$. The symbol $\boldsymbol{a}_j$ is an $m \times 1$ column vector that is comprised of the m elements of the jth row of matrix A so that $[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]^T := A$. We have used $\boldsymbol{\xi}^{(j)}$ to denote the $m \times 1$ estimated parameter vector associated with that set of observations that contains an outlier in the jth observation, as opposed to using $\boldsymbol{\xi}$, which is associated with the same set of observations except that the jth observation would not contain an outlier. The variable e_j is the jth element of the true (but unknown) random error vector $\boldsymbol{e}$. The symbol $\xi_0^{(j)}$ is a scalar that accounts for the effect of the outlier. The formula for its estimate is developed below.

The following example may be illustrative: The observation y_j should have been 100 m but only a value of 10 m was recorded, then $\xi_0^{(j)}$ accounts for a 90 m blunder.

The full system of equations for the GMM with a single outlier in the jth observation is expressed as

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi}^{(j)} + \underset{n\times 1}{\boldsymbol{\eta}_j}\,\xi_0^{(j)} + \boldsymbol{e}, \quad \boldsymbol{\eta}_j := \big[0, \ldots, 0, 1, 0, \ldots, 0\big]^T, \tag{10.70a}$$

$$\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_0^2 P^{-1}). \tag{10.70b}$$

Note that the number 1 in $\boldsymbol{\eta}_j$ appears at the jth element; all other elements are 0. We must compare the model in (10.70) with the original GMM (3.1) that does not include an outlier. Since the model (10.70) assumes only one outlier in the data set, n comparisons of the two models are necessary in

order to test all y_i observations independently, i.e., $i = 1, \ldots, n$. For each comparison we introduce the following constraint:

$$\xi_0^{(j)} = K \begin{bmatrix} \boldsymbol{\xi}^{(j)} \\ \xi_0^{(j)} \end{bmatrix} = \kappa_0 = 0. \tag{10.71}$$

Here $K \coloneqq [0, 0, \ldots, 1]$ is of size $1 \times (m+1)$. When we impose the constraint (10.71) upon the model (10.70), we obtain a model equivalent to the original GMM (3.1) that does not include a parameter for an outlier.

Note: For the remainder of this section, we will assume that the weight matrix P is diagonal: $P = \operatorname{diag}(p_1, \ldots, p_n)$, where p_i is the weight of the ith observation. See Schaffrin (1997) for a treatment of outlier detection with correlated observations.

Now, we begin with following Lagrange target function to derive a least-squares estimator in the unconstrained model (10.70):

$$\Phi\big(\boldsymbol{\xi}^{(j)}, \xi_0^{(j)}\big) = \big(\boldsymbol{y} - A\boldsymbol{\xi}^{(j)} - \boldsymbol{\eta}_j \xi_0^{(j)}\big)^T P \big(\boldsymbol{y} - A\boldsymbol{\xi}^{(j)} - \boldsymbol{\eta}_j \xi_0^{(j)}\big), \tag{10.72}$$

which is stationary with respect to $\boldsymbol{\xi}^{(j)}$ and $\xi_0^{(j)}$. Setting the first partial derivatives of (10.72) to zero results in the following Euler-Lagrange necessary conditions:

$$\frac{1}{2}\left[\frac{\partial \Phi}{\partial \boldsymbol{\xi}^{(j)}}\right]^T = -A^T P \boldsymbol{y} + A^T P \boldsymbol{\eta}_j \hat{\xi}_0^{(j)} + A^T P A \hat{\boldsymbol{\xi}}^{(j)} \doteq \boldsymbol{0}, \tag{10.73a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \xi_0^{(j)}} = -\boldsymbol{\eta}_j^T P \boldsymbol{y} + \boldsymbol{\eta}_j^T P A \hat{\boldsymbol{\xi}}^{(j)} + \boldsymbol{\eta}_j^T P \boldsymbol{\eta}_j \hat{\xi}_0^{(j)} \doteq \boldsymbol{0}. \tag{10.73b}$$

Of course the second partial derivatives are functions of P, which is positive-definite by definition, thereby satisfying the sufficient condition required for obtaining the minimum of (10.72). In matrix form we have

$$\begin{bmatrix} N & A^T P \boldsymbol{\eta}_j \\ \boldsymbol{\eta}_j^T P A & \boldsymbol{\eta}_j^T P \boldsymbol{\eta}_j \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\eta}_j^T P \boldsymbol{y} \end{bmatrix}, \tag{10.74a}$$

or, because P was assumed to be diagonal,

$$\begin{bmatrix} N & \boldsymbol{a}_j p_j \\ p_j \boldsymbol{a}_j^T & p_j \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix}. \tag{10.74b}$$

Here, as in previous chapters, we have used the definition $[N, \boldsymbol{c}] \coloneqq A^T P [A, \boldsymbol{y}]$.

Using (A.14) for the inverse of a partitioned matrix, and decomposing the resulting inverse into a sum of two matrices, results in

$$\begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} N^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix} + \begin{bmatrix} N^{-1} \boldsymbol{a}_j p_j \\ -1 \end{bmatrix} \big(p_j - p_j \boldsymbol{a}_j^T N^{-1} \boldsymbol{a}_j p_j\big)^{-1} \begin{bmatrix} p_j \boldsymbol{a}_j^T N^{-1} & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix}, \tag{10.75a}$$

or

$$\begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} N^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix} - \begin{bmatrix} N^{-1} \boldsymbol{a}_j p_j \\ -1 \end{bmatrix} \big(p_j - p_j^2 \boldsymbol{a}_j^T N^{-1} \boldsymbol{a}_j\big)^{-1} p_j \big(y_j - \boldsymbol{a}_j^T N^{-1} \boldsymbol{c}\big). \tag{10.75b}$$

From (10.75b), and recalling that $\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c}$ is based on a data set assumed to have no outliers, we can write the following difference between estimations:

$$\hat{\boldsymbol{\xi}}^{(j)} - \hat{\boldsymbol{\xi}} = -N^{-1}\boldsymbol{a}_j\left(\frac{y_j - \boldsymbol{a}_j^T\hat{\boldsymbol{\xi}}}{p_j^{-1} - \boldsymbol{a}_j^T N^{-1}\boldsymbol{a}_j}\right) = -N^{-1}\boldsymbol{a}_j\frac{\tilde{e}_j}{(Q_{\tilde{\boldsymbol{e}}})_{jj}}, \tag{10.76}$$

where $(Q_{\tilde{\boldsymbol{e}}})_{jj}$ is the jth diagonal element of the cofactor matrix for the residual vector $\tilde{\boldsymbol{e}}$. For the estimated outlier itself we have

$$\hat{\xi}_0^{(j)} = \frac{y_j - \boldsymbol{a}_j^T\hat{\boldsymbol{\xi}}}{1 - p_j\boldsymbol{a}_j^T N^{-1}\boldsymbol{a}_j} = \frac{\tilde{e}_j}{(Q_{\tilde{\boldsymbol{e}}}P)_{jj}} = \frac{\tilde{e}_j/p_j}{(Q_{\tilde{\boldsymbol{e}}})_{jj}}. \tag{10.77}$$

The hypothesis test for an outlier in the jth observation is then written as

$$H_0 : E\{\hat{\xi}_0^{(j)}\} = 0 \;\text{ versus }\; H_A : E\{\hat{\xi}_0^{(j)}\} \neq 0. \tag{10.78}$$

The test statistic has an F-distribution and is computed by

$$T_j = \frac{R_j/1}{(\Omega - R_j)/(n-m-1)} \sim F(1, n-m-1). \tag{10.79}$$

The definition of R_j, in terms of $\hat{\xi}_0^{(j)}$, is

$$R_j := \frac{\left(\hat{\xi}_0^{(j)} - 0\right)^2}{KN_1^{-1}K^T} = \frac{\left(\hat{\xi}_0^{(j)}\right)^2}{\left(p_j - p_j^2\boldsymbol{a}_j^T N^{-1}\boldsymbol{a}_j\right)^{-1}} = \frac{\tilde{e}_j^2}{(Q_{\tilde{\boldsymbol{e}}}P)_{jj}^2}p_j(Q_{\tilde{\boldsymbol{e}}}P)_{jj} = \frac{\tilde{e}_j^2}{(Q_{\tilde{\boldsymbol{e}}})_{jj}}. \tag{10.80}$$

It is important to note that the symbols $\tilde{\boldsymbol{e}}$ and $Q_{\tilde{\boldsymbol{e}}}$ represent the residual vector and its cofactor matrix, respectively, as predicted within the GMM model (3.1). As was already mentioned, when we impose the constraint (10.71) on model (10.70b) we reach a solution identical to the LESS within model (3.1). It is also important to to understand the denominator of (10.79). As stated previously, the symbol R is used to account for that portion of the P-weighted residual norm due to the constraints. The parenthetical term in the denominator accounts for that part of the norm coming from the unconstrained solution. Here we have used $\Omega := \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$, with $\tilde{\boldsymbol{e}}$ belonging to the constrained solution (determined within the model (3.1)). Therefore, we must subtract R from Ω, as it is defined here, to arrive at the portion of the norm coming from the unconstrained LESS computed within model (10.70).

We note again that the equations from (10.74b) to (10.80) hold only in the case of a diagonal weight matrix P. Regardless of whether or not P is diagonal, the quantity

$$r_j := (Q_{\tilde{e}}P)_{jj} \tag{10.81}$$

is the jth so-called *redundancy number*, for the unconstrained solution in this case. The following properties hold for r_j:

$$0 < r_j \leq 1 \;\text{ for } i = \{1, \ldots, n\} \;\text{ and }\; \sum_j r_j = n - \operatorname{rk} A. \tag{10.82}$$

Note that $(Q_{\tilde{e}}P)_{jj} = (Q_{\tilde{e}})_{jj}\,p_j$ for the case that matrix P is diagonal.

Finally, the matrix N_1 in (10.80) is defined as

$$N_1 = \begin{bmatrix} N & \boldsymbol{a}_j p_j \\ p_j \boldsymbol{a}_j^T & p_j \end{bmatrix}, \tag{10.83}$$

which appears in (10.74b). Pre- and post-multiplying N_1^{-1} by K extracts only its last diagonal element, which, according to the formula for inverting a partitioned matrix, turns out to be the scalar quantity $(p_j - p_j^2 \boldsymbol{a}_j^T N^{-1} \boldsymbol{a}_j)^{-1}$, also appearing in (10.80).

We comment that outlier detection at the 2-D and 3-D level can also be performed, for example, in testing outliers in observations of 2-D and 3-D points. The 3-D case is also appropriate for GPS baseline adjustments, and its development is shown by Snow (2002); see also Snow and Schaffrin (2003).

Appendix A

Useful Matrix Relations and Identities

Product of transposes:

$$A^T B^T = (BA)^T \tag{A.1}$$

Transpose of inverse:

$$(A^T)^{-1} = (A^{-1})^T \tag{A.2}$$

Product of inverses:

$$A^{-1} B^{-1} = (BA)^{-1} \tag{A.3}$$

Rank of triple product: Given: $A(m \times n)$, $B(m \times m)$, $C(n \times n)$:

$$B, C \text{ nonsingular} \Rightarrow \operatorname{rk}(BAC) = \operatorname{rk}(A) \;\text{ or }\; \operatorname{rk}(BA) = \operatorname{rk}(A) \;\text{ if }\; C = I \tag{A.4}$$

Trace invariant with respect to a cyclic permutation of factors: If the product ABC is square, then the following trace operations are equivalent:

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB). \tag{A.5}$$

Column space and nullspace: The column space of A is denoted by $\mathcal{R}(A)$ and is also called the range of A. Its dimension equals the rank of A. The nullspace of A is denoted by $\mathcal{N}(A)$ and is also called the kernel of A. The dimension of the nullspace is $m - \operatorname{rk} A$, where m is the number of columns of A; the dimension is also called the nullity.

$$\text{The column space of } AB \text{ is contained in the column space of } A. \tag{A.6}$$

Sherman-Morrison-Woodbury-Schur formula:

$$(T - UW^{-1}V)^{-1} = T^{-1} + T^{-1}U(W - VT^{-1}U)^{-1}VT^{-1} \tag{A.7}$$

As a consequence of (A.7), we also have:

$$(I \pm W^{-1}V)^{-1} = I \mp (W \pm V)^{-1}V, \tag{A.8a}$$
$$(I \pm V)^{-1} = I \mp (I \pm V)^{-1}V, \tag{A.8b}$$
$$(I \pm W^{-1})^{-1} = I \mp (W \pm I)^{-1}. \tag{A.8c}$$

Equations (39–43) of "Useful Matrix Equalities" (handout from Prof. Schaffrin, possibly originating from Urho A. Uotila).

$$DC(A + BDC)^{-1} = (D^{-1} + CA^{-1}B)^{-1}CA^{-1} = \tag{A.9a}$$
$$= D(I + CA^{-1}BD)^{-1}CA^{-1} = \tag{A.9b}$$
$$= DC(I + A^{-1}BDC)^{-1}A^{-1} = \tag{A.9c}$$
$$= DCA^{-1}(I + BDCA^{-1})^{-1} = \tag{A.9d}$$
$$= (I + DCA^{-1}B)^{-1}DCA^{-1} \tag{A.9e}$$

Suppose the matrices A and B in (A.9) are identity matrices, then we have

$$DC(I + DC)^{-1} = (D^{-1} + C)^{-1}C = \tag{A.10a}$$
$$= D(I + CD)^{-1}C = \tag{A.10b}$$
$$= (I + DC)^{-1}DC. \tag{A.10c}$$

Definition of idempotent:

$$\text{The matrix } P \text{ is idempotent if } PP = P. \text{ Projection matrices are idempotent.} \tag{A.11}$$
$$\text{If } P \text{ is idempotent, } \operatorname{tr} P = \operatorname{rk} P. \tag{A.12}$$

Inverse of the partitioned normal equation matrix: Assume the matrix N is of full rank and is partitioned as follows:

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}. \tag{A.13}$$

The following steps lead to the inverse of N expressed in terms of the partitioned blocks:

$$\left[\begin{array}{cc|cc} N_{11} & N_{12} & I & 0 \\ N_{21} & N_{22} & 0 & I \end{array}\right] \to \left[\begin{array}{c|c|c|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline N_{21} & N_{22} & 0 & I \end{array}\right] \to$$
$$\left[\begin{array}{c|c|c|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline 0 & N_{22} - N_{21}N_{11}^{-1}N_{12} & -N_{21}N_{11}^{-1} & I \end{array}\right] \to$$
$$\left[\begin{array}{cc|c|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ 0 & I & -\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1}N_{21}N_{11}^{-1} & \left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} \end{array}\right] \to$$
$$\left[\begin{array}{cc|c|c} I & 0 & N_{11}^{-1} + N_{11}^{-1}N_{12}\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1}N_{21}N_{11}^{-1} & -N_{11}^{-1}N_{12}\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} \\ 0 & I & -\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1}N_{21}N_{11}^{-1} & \left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} \end{array}\right].$$

Finally we may write

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \left[\begin{array}{c|c} N_{11}^{-1} + N_{11}^{-1}N_{12}\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} N_{21}N_{11}^{-1} & -N_{11}^{-1}N_{12}\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} \\ \hline -\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} N_{21}N_{11}^{-1} & \left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} \end{array}\right]. \tag{A.14}$$

Note that other equivalent representations of this inverse exist. Taking directly from the Useful Matrix Equalities handout mentioned above, we write some additional expressions for the inverse.

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{A.15}$$

$$Q_{11} = \left(N_{11} - N_{12}N_{22}^{-1}N_{21}\right)^{-1} = \tag{A.16a}$$

$$= N_{11}^{-1} + N_{11}^{-1}N_{12}\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1}N_{21}N_{11}^{-1} = \tag{A.16b}$$

$$= N_{11}^{-1} + N_{11}^{-1}N_{12}Q_{22}N_{21}N_{11}^{-1} \tag{A.16c}$$

$$Q_{22} = \left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} = \tag{A.17a}$$

$$= N_{22}^{-1} + N_{22}^{-1}N_{21}\left(N_{11} - N_{12}N_{22}^{-1}N_{21}\right)^{-1}N_{12}N_{22}^{-1} = \tag{A.17b}$$

$$= N_{22}^{-1} + N_{22}^{-1}N_{21}Q_{11}N_{12}N_{22}^{-1} \tag{A.17c}$$

$$Q_{12} = -\left(N_{11} - N_{12}N_{22}^{-1}N_{21}\right)^{-1} N_{12}N_{22}^{-1} = -Q_{11}N_{12}N_{22}^{-1} = \tag{A.18a}$$

$$= -N_{11}^{-1}N_{12}\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} = -N_{11}^{-1}N_{12}Q_{22} \tag{A.18b}$$

$$Q_{21} = -N_{22}^{-1}N_{21}\left(N_{11} - N_{12}N_{22}^{-1}N_{21}\right)^{-1} = -N_{22}^{-1}N_{21}Q_{11} = \tag{A.19a}$$

$$= -\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)^{-1} N_{21}N_{11}^{-1} = -Q_{22}N_{21}N_{11}^{-1} \tag{A.19b}$$

In the case that $N_{22} = 0$, we have:

$$Q_{22} = -\left(N_{21}N_{11}^{-1}N_{12}\right)^{-1} \tag{A.20a}$$

$$Q_{11} = N_{11}^{-1} + N_{11}^{-1}N_{12}Q_{22}N_{21}N_{11}^{-1} \tag{A.20b}$$

$$Q_{12} = -N_{11}^{-1}N_{12}Q_{22} \tag{A.20c}$$

$$Q_{21} = -Q_{22}N_{21}N_{11}^{-1} \tag{A.20d}$$

Schur Complement: the parenthetical term $\left(N_{22} - N_{21}N_{11}^{-1}N_{12}\right)$ shown above is called the Schur Complement of N_{11}.

Derivative of quadratic form:
While some authors write the derivative of a quadratic form (a scalar-valued vector function) with respect to a column vector as a row vector, we write such a derivative as a column vector. This is in agreement with the following authors: Grafarend and Schaffrin, pg. 443, (1993); Harville (1997), pg. 295; Koch (1999), pg. 69; Lüetkepohl (1996), pg. 175; Strang and Boore (1997), pg. 300. For example, given $\boldsymbol{x} \in \mathbb{R}^n$ and $Q \in \mathbb{R}^{n \times n}$, we have

$$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T Q \boldsymbol{x} \Rightarrow \frac{\partial \Phi}{\partial \boldsymbol{x}} = 2Q\boldsymbol{x}. \tag{A.21}$$

References and Suggested Reading

[1] Björck, Å. (1996), *Numerical Methods for Least-Squares Problems*, SIAM, Philadelphia.

[2] Christensen, R. (1996), *Plane Answers to Complex Questions*, Springer Verlag, New York.

[3] Grafarend, E. and B. Schaffrin (1993), *Ausgleichungsrechnung In Linearen Modellen*, Wissenschaftsverlag, Mannheim.

[4] Harville, D.A. (1997), *Matrix Algebra From A Statistician's Perspective*, Springer, New York.

[5] Koch, K.-R. (1997), *Parameter Estimation and Hypothesis Testing in Linear Models*, 2nd ed., Springer Verlag, Berlin/Heidelberg.

[6] Lütkepohl, H. (1996), *Handbook of Matrices*, John Wiley & Sons, Chichester, England.

[7] Rao, C.R. and H. Toutenburg (1995), *Linear Models, Least-Squares and Alternatives*, Springer, NY.

[8] Schaffrin, B. (1997a), *On suboptimal geodetic network fusion*, IAG General Meeting, Rio de Janeiro, Brazil.

[9] Schaffrin, B. (1997b), *Reliability measures for correlated observations*, J. of Surveying Engrg. 123, No. 3, 126-137.

[10] Snow, K. (2002), *Applications of Parameter Estimation and Hypothesis Testing to GPS Network Adjustments*, OSU Report 465, Geodetic Science and Surveying, The Ohio State University, Columbus, Ohio.

[11] Snow, K. and B. Schaffrin (2003), *Three-dimensional outlier detection for GPS networks and their densification via the BLIMPBE approach*, GPS Solutions (7) 130–139 Springer Verlag, Berlin / Heidelberg.

[12] Strang, G. and K. Borre (1997), *Linear Algebra, Geodesy, and GPS*, Wellesley-Cambridge Press, Wellesley, MA.