

ADVANCED ADJUSTMENT COMPUTATIONS

ADVANCED ADJUSTMENT COMPUTATIONS

Part 2 of Notes by
Kyle Snow

February 4, 2021

Based on former Geodetic Science Course GS 762
taught at The Ohio State University
by Prof. Burkhard Schaffrin

Contents

Preface	vii
Introduction	1
1 Review of the Gauss-Markov Model	5
2 Introducing the vec Operator and the Kronecker Product	11
3 Variance Component Estimation	15
4 Expectation-Dispersion Correspondence	23
5 The Rank-Deficient Gauss-Markov Model	29
5.1 Generalized Inverses	31
5.2 Finding a Generalized Inverse	32
5.3 The Singular Value Decomposition	35
5.4 Minimum Norm Least-Squares Solution	38
5.5 Partial Minimum Norm Least-Squares Solution	39
5.6 Best Least-Squares Solution	42
5.7 Best Linear Uniformly Minimum Bias Estimate	46
5.8 Minimum and Inner Constraints	49
5.8.1 Restricted LEast-Squares Solution (RLESS)	49
5.8.2 Reflexive Symmetric G-Inverse	51
5.8.3 (Partial) MINimum NOrm LEast-Squares Solution (MINO- LESS)	51
5.8.4 Summary Formulas for Minimally Constrained LESS	55
5.9 <i>S</i> -Transformations	55
5.9.1 Example <i>S</i> -Transformation	56
5.10 Concluding Remarks about the Restricted Least-Squares Solution . .	57
6 The Variance Component Model	59
7 Prior Information	67
7.1 Pseudo-observations	68
7.2 Alternative Normal Equations	73

7.3	Mixed Linear Model (Helmert's Knack)	76
7.4	Solutions for the Mixed Linear Model	78
7.5	Best Inhomogeneous Linear Predictor	81
7.6	Alternative Normal Equations for the Mixed Linear Model	83
8	The Dynamic Linear Model	87
A	Useful Matrix Relations and Identities	95
	Bibliography	101

Preface

In its earlier versions of many years ago, this document was primarily based on notes taken by Kyle Snow in Geodetic Science adjustment computations course GS 762 taught by Burkhard Schaffrin at The Ohio State University in 1999. Gradually, Kyle added content based on his own wishes and requests made by several readers of the notes. In many places, the notes reflect just what Prof. Schaffrin derived on the chalkboard during his lectures. In many other places, they have been extended beyond that somewhat by Kyle.

The author would like to thank Shengjie Ge for his assistance in typesetting early chapters in the original version of the document with L^AT_EX. Also, Xiankun Wang 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. Many thanks to Dru Smith for his numerous recommendations to improve chapter 6 in a late 2020 version. Thanks to Xing Fang for pointing out an error before the edition distributed in January 2021.

Of course, without the extraordinary teaching of Prof. Burkhard Schaffrin, these notes would not even exist.

Columbus, Ohio, December 2020

Kyle Snow

Introduction

These notes are Part 2 of a two-part series on adjustment computations, Part 1 being introductory and Part 2 being advanced. Together, both parts should provide the reader with a solid foundation in the theory and application of adjustment computations, especially as they are used in the disciplines of geodetic science and engineering and surveying engineering. However, it is expected that researchers, data analysts, and practitioners from other science and engineering fields can benefit from these notes as well.

The subject of adjustment computations is a rich topic spanning many science and engineering disciplines. The need to adjust observations in some meaningful, or perhaps better yet, *optimal*, way is an old one. The need becomes obvious as soon as one realizes that repeated observations of the same phenomenon or physical quantity usually do not yield the same numerical values. And of course, this realization ought to occur as soon as one begins taking notice of the things happening around them, especially if that notice involves the use of instruments and devices that allow measurement or quantification of some physical, abstract, or social phenomenon, whether arising in nature or manufactured by humans.

As a young researcher, Carl Gauss was faced with the problem of how to best use redundant data to predict the trajectory of the asteroid Ceres. Apparently, Gauss settled on a method later known as “least squares,” some 15 years before it was made known to the public by Legendre, in 1805, who claimed to be its original discoverer. The debate over who first discovered least squares, Gauss or Legendre, is discussed by Stigler (1981), who provides evidence from important geodetic surveys that favor Gauss’ claim, though Stigler admits that there still is not conclusive evidence to be absolutely certain about who was first. Perhaps only Gauss will ever know.

The term least squares is often used adjectivally as in least-squares adjustment, least-squares solution, the method of least squares, etc. These terms are all more or less synonymous. The term comes from the mathematical technique of minimizing the *sum of squares of residuals* (or sum of squared residuals), where residual means the difference between an observation and its adjusted value. When observational weights are involved, the descriptive phrase should be modified to “the sum of squares of weighted residuals.” Furthermore, the use of weight matrices in a linear algebra formulation adds another level of detail (correlations between random observation errors) that is often omitted from the descriptive phrase to avoid overburdening it. But that is jumping too far ahead for this introduction.

In the chapters that follow, the term *adjustment computations* will apply to the adjustment of observational data by the method of least squares, unless otherwise noted.

Calculus is used to setup the minimization problem for minimizing the sum of squared residuals, but the resulting solution can be found equivalently through both geometrical (projection of vector spaces) and statistical methods. While most derivations in these notes involve minimization of an objective (target) function that includes Lagrange multipliers (the Lagrangian approach using calculus), equivalent statistical derivations and geometric relationships are also given in some places.

The objectives of the former course these notes are based on (GS 762) were stated by Burkhard Schaffrin as follows:

Objectives: The course makes students aware of various special adjustment techniques. Relations between the Gauss-Markov model and traditional least-squares solutions are explored and compared to the collocation technique. Ranks of matrices are discussed, and they are derived for matrices usually encountered in adjustment computations. The introduction to generalized matrices will give the possibility to solve rank-deficient systems. Estimable and non-estimable quantities in adjustment are defined and discussed, as well as the estimation of variance components. The role of prior information is clarified, and it is shown how the least-squares adjustment in a Dynamic Linear Model leads to Kalman filtering. As a result, students should be able to make a prudent choice of a proper model and the corresponding adjustment techniques for a host of overdetermined problems in geodetic science, no matter how complicated.

Chapter 1 begins with a review of the non-linear Gauss-Markov model, showing how the least-squares solution for the unknown parameters of the model can be arrived at equivalently via both algebraic-geometric and statistical approaches, resulting in an equivalency between LESS (Least Squares-Solution) and BLUE (Best Linearly Uniformly Unbiased Estimate).

Chapter 2 introduces the linear algebra concepts of *vec operator* and *Kronecker product*, which provide powerful tools for the derivations in chapters that follow. These concepts deserve attention here, because they typically are not covered in first courses in linear algebra, which may be the extent of background in linear algebra for many students using these notes.

In Chapter 3, the estimation of the unknown variance component appearing in the Gauss-Markov Model is derived from the statistical concept of Best Invariant Quadratic Uniformly Unbiased Estimate, or BIQUUE.

Chapter 4 treats the concept of *expectation-dispersion correspondence*, which allows a quadratic model for the unknown variance component to be transformed into a new linear model and shows how BIQUUE for the unknown variance component in the Gauss-Markov model is equivalent to LESS within the new model.

Chapter 5 introduces the *rank-deficient Gauss-Markov Model*, providing a useful way to treat many adjustment problems in geodetic science where the observations do not provide enough information to estimate all the unknown parameters of the model. The concept of generalized inverse is presented, which allows for characterizing the solution space for the unknown model parameters, among which the estimators MINOLESS (and its equivalent BLUMBE) and partial MINOLESS are

perhaps most useful.

Chapter 6 introduces *variance component estimation*, a topic that most researchers and engineers in the geodetic sciences and surveying will likely encounter in their work sooner or later. It provides a powerful method for dealing with multiple observational weight matrices for which the relative accuracies among them may be unknown or uncertain, in which case, estimating a variance component for each one of them might be desired. This chapter shows how to do that.

Chapter 7 introduces the notion of *prior information* for the unknown model parameters. Here, the concept of random, rather, than fixed model parameters is introduced, and a model is presented that contains both types of parameters, leading to least-squares estimates of the fixed parameters and predictions for the random parameters.

Chapter 8 presents the *dynamic linear model* and derives the least-squares solution within it, which is also known as a Kalman filter.

An 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.

Chapter 1

Review of the Gauss-Markov Model

The non-linear form of the Gauss-Markov Model (GMM) is written as

$$\mathbf{Y} = a(\boldsymbol{\Xi}) + \mathbf{e}, \quad (1.1a)$$

$$\mathbf{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}). \quad (1.1b)$$

The symbols in equation (1.1) are defined as follows:

\mathbf{Y} is a given $n \times 1$ vector of observations.

a is a (known) non-linear function such that $a : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $m < n$.

$\boldsymbol{\Xi}$ is an $m \times 1$ vector of unknown, non-random parameters.

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

σ_0^2 is an unknown variance component.

P is a given $n \times n$ matrix of observation weights.

We linearize (1.1a) by Taylor-series expansion with respect to a fixed (i.e., non-random) approximation $\boldsymbol{\Xi}_0$ (called expansion point), which leads to the following *linearized GMM*:

$$\mathbf{y} := \mathbf{Y} - a(\boldsymbol{\Xi}_0) = \left. \frac{\partial a(\boldsymbol{\Xi})}{\partial \boldsymbol{\Xi}^T} \right|_{\boldsymbol{\Xi}=\boldsymbol{\Xi}_0} (\boldsymbol{\Xi} - \boldsymbol{\Xi}_0) + (\text{higher order terms}) + \mathbf{e}. \quad (1.2)$$

Neglecting the higher order terms and defining a coefficient matrix $A := \partial a / \partial \boldsymbol{\Xi}^T : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $m < n$, and an incremental parameter vector $\boldsymbol{\xi} := \boldsymbol{\Xi} - \boldsymbol{\Xi}_0$, we can then write the linearized GMM as

$$\mathbf{y} = A\boldsymbol{\xi} + \mathbf{e}, \quad (1.3a)$$

$$\mathbf{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}). \quad (1.3b)$$

Note that $\text{rk } A = \dim \mathcal{R}(A) = m$ if, and only if, there is no rank deficiency in the model. For the linearized case this relation holds everywhere in the m -dimensional space, but it may not hold in the non-linear case.

We review two approaches to estimating the parameter vector $\boldsymbol{\xi}$:

- I. Algebraic-geometric approach (e.g., least-squares adjustment).
- II. Statistical approach (e.g., minimum variance, unbiased estimators).

I. Algebraic-geometric approach: This approach uses a weighted least-squares adjustment, which is derived from the minimization of the weighted L_2 -norm

$$\|\mathbf{y} - A\boldsymbol{\xi}\|_P^2 = \min_{\boldsymbol{\xi}} \quad (1.4)$$

and leads to an estimate $\hat{\boldsymbol{\xi}}$ for the unknown parameters. Or, being more general, we can express the problem as a minimization of the random error vector \mathbf{e} , which leads to both the parameter estimate $\hat{\boldsymbol{\xi}}$ and the predicted random error (or residual) vector $\tilde{\mathbf{e}}$.

$$\|\mathbf{e}\|_P^2 = \min_{\mathbf{e}, \boldsymbol{\xi}} \{\mathbf{e} = \mathbf{y} - A\boldsymbol{\xi}\} \quad (1.5)$$

The Lagrange target function (or Lagrangian function) to minimize is a scalar-valued function that is quadratic in the unknown random error vector \mathbf{e} . It is written as

$$\Phi(\mathbf{e}, \boldsymbol{\xi}, \boldsymbol{\lambda}) = \mathbf{e}^T P \mathbf{e} + 2\boldsymbol{\lambda}^T (\mathbf{y} - A\boldsymbol{\xi} - \mathbf{e}), \quad (1.6)$$

which must be made stationary with respect to the unknown variables \mathbf{e} , $\boldsymbol{\xi}$, and $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is an $n \times 1$ vector of Lagrange multipliers. Accordingly, the Euler-Lagrange necessary conditions (or first-order conditions) lead to a minimization of (1.6). In forming the Euler-Lagrange necessary conditions, we take the first partial derivatives of the target function (1.6), set them to be zero, and use hat and tilde symbols to denote the particular solutions to these condition equations.

$$\frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{e}} = P\tilde{\mathbf{e}} - \hat{\boldsymbol{\lambda}} \doteq \mathbf{0} \quad (1.7a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\xi}} = -A^T \hat{\boldsymbol{\lambda}} \doteq \mathbf{0} \quad (1.7b)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \mathbf{y} - A\hat{\boldsymbol{\xi}} - \tilde{\mathbf{e}} \doteq \mathbf{0} \quad (1.7c)$$

For convenience we define *normal equation* variables N and \mathbf{c} as

$$[N, \mathbf{c}] := A^T P [A, \mathbf{y}]. \quad (1.8)$$

In the following, we often refer to matrix N as the *normal equations matrix*.

Solving the three equations (1.7) in the three unknowns $\tilde{\mathbf{e}}$, $\hat{\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\lambda}}$ leads to

$$\hat{\boldsymbol{\lambda}} = P\tilde{\mathbf{e}}, \quad (1.9a)$$

$$A^T\hat{\boldsymbol{\lambda}} = A^TP\tilde{\mathbf{e}} = \mathbf{0}, \quad (1.9b)$$

implying that

$$A^T\hat{\boldsymbol{\lambda}} = A^TP(\mathbf{y} - A\hat{\boldsymbol{\xi}}) = \mathbf{0}, \quad (1.9c)$$

which leads to the LEast-Squares Solution (LESS) for $\boldsymbol{\xi}$ as

$$\hat{\boldsymbol{\xi}} = (A^T P A)^{-1} A^T P \mathbf{y} = N^{-1} \mathbf{c} \quad (1.9d)$$

and the predicted residual vector

$$\tilde{\mathbf{e}} = (I - AN^{-1}A^T P)\mathbf{y}. \quad (1.9e)$$

Note the following relation between the predicted residual vector $\tilde{\mathbf{e}}$ and the unknown random error vector \mathbf{e} :

$$\tilde{\mathbf{e}} = (I - AN^{-1}A^T P)\mathbf{y} = (I - AN^{-1}A^T P)(\mathbf{y} - A\boldsymbol{\xi}) = (I - AN^{-1}A^T P)\mathbf{e}. \quad (1.10)$$

Equation (1.10) shows that, in general, the predicted random error vector $\tilde{\mathbf{e}}$ is not the same as the true (unknown) random error vector \mathbf{e} . They would only be the same if \mathbf{e} belonged to the null space of $A^T P$, which is hardly possible since random measurement errors are involved.

The expectations of the estimated parameter vector $\hat{\boldsymbol{\xi}}$ and the predicted random error vector $\tilde{\mathbf{e}}$ are given as follows:

$$E\{\hat{\boldsymbol{\xi}}\} = N^{-1}A^T P \cdot E\{\mathbf{y}\} = N^{-1}A^T P A \boldsymbol{\xi} = \boldsymbol{\xi}. \quad (1.11)$$

Equation (1.11) holds for all $\boldsymbol{\xi} \in \mathbb{R}^m$. Therefore, $\hat{\boldsymbol{\xi}}$ is said to be a *uniformly unbiased estimate* of $\boldsymbol{\xi}$.

$$E\{\tilde{\mathbf{e}}\} = (I - AN^{-1}A^T P) \cdot E\{\mathbf{y}\} = (I - AN^{-1}A^T P)A\boldsymbol{\xi} = A\boldsymbol{\xi} - A\boldsymbol{\xi} = \mathbf{0} = E\{\mathbf{e}\} \quad (1.12)$$

Because the $n \times 1$ vector $\mathbf{0}$ is only one element of \mathbb{R}^n , $\tilde{\mathbf{e}}$ is considered to be a *weakly unbiased prediction* of \mathbf{e} .

The associated dispersion and covariance matrices are derived as follows:

$$D\{\hat{\boldsymbol{\xi}}\} = N^{-1}A^T P \cdot D\{\mathbf{y}\} \cdot P A N^{-1} = N^{-1}A^T P (\sigma_0^2 P^{-1}) P A N^{-1} \Rightarrow \quad (1.13)$$

$$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 N^{-1},$$

$$\begin{aligned} D\{\tilde{\mathbf{e}}\} &= (I - AN^{-1}A^T P) \cdot D\{\mathbf{y}\} \cdot (I - P A N^{-1}A^T) = \\ &= (I - AN^{-1}A^T P) (\sigma_0^2 P^{-1}) (I - P A N^{-1}A^T) = \\ &= \sigma_0^2 (P^{-1} - AN^{-1}A^T) (I - P A N^{-1}A^T) = \\ &= \sigma_0^2 (P^{-1} - AN^{-1}A^T) - P^{-1} P A N^{-1}A^T + AN^{-1}A^T P A N^{-1}A^T \Rightarrow \\ &D\{\tilde{\mathbf{e}}\} = \sigma_0^2 (P^{-1} - AN^{-1}A^T), \end{aligned} \quad (1.14)$$

$$\begin{aligned}
C\{\hat{\boldsymbol{\xi}}, \tilde{\mathbf{e}}\} &= N^{-1}A^T P \cdot D\{\mathbf{y}\} \cdot (I - PAN^{-1}A^T) = \\
&= N^{-1}A^T P(\sigma_0^2 P^{-1})(I - PAN^{-1}A^T) = \\
&= \sigma_0^2 N^{-1}A^T - \sigma_0^2 N^{-1}A^T P A N^{-1}A^T = \sigma_0^2(N^{-1}A^T - N^{-1}A^T) \Rightarrow \\
&\boxed{C\{\hat{\boldsymbol{\xi}}, \tilde{\mathbf{e}}\} = 0.} \tag{1.15}
\end{aligned}$$

Equation (1.15) shows the ‘‘covariance orthogonality.’’ This is opposed to the algebraic orthogonality depicted in Figure 1.1, where it is shown that the residual vector $\tilde{\mathbf{e}}$ is added to \mathbf{y} to make equation (1.3) consistent. Through the least-squares principle, we have found a particular residual vector $\tilde{\mathbf{e}}$ that is closest to (geometrically orthogonal to) the column space of matrix A .

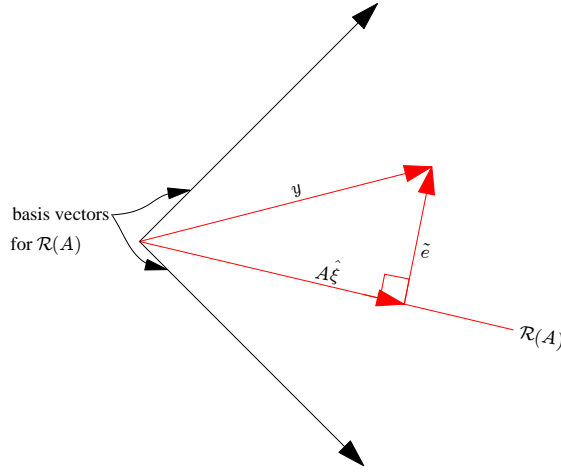


Figure 1.1: Orthogonality between $\mathcal{R}(A)$ and $\tilde{\mathbf{e}}$

Let us consider the P -weighted L_2 -norm of the residual vector $\tilde{\mathbf{e}}$, which can be expressed in the following forms:

$$\|\tilde{\mathbf{e}}\|_P^2 = \tilde{\mathbf{e}}^T P \tilde{\mathbf{e}} = (\mathbf{y} - A\hat{\boldsymbol{\xi}})^T P (\mathbf{y} - A\hat{\boldsymbol{\xi}}) = \tag{1.16a}$$

$$= \mathbf{y}^T P (\mathbf{y} - A\hat{\boldsymbol{\xi}}) - \hat{\boldsymbol{\xi}}^T A^T P \tilde{\mathbf{e}} = \tag{1.16b}$$

$$= \mathbf{y}^T P (\mathbf{y} - A\hat{\boldsymbol{\xi}}) = \mathbf{y}^T P \mathbf{y} - \mathbf{c}^T \hat{\boldsymbol{\xi}} = \mathbf{y}^T P \mathbf{y} - \hat{\boldsymbol{\xi}}^T N \hat{\boldsymbol{\xi}}. \tag{1.16c}$$

In (1.16) we have used the orthogonality property $A^T P \tilde{\mathbf{e}} = \mathbf{0}$ shown in (1.9b). However, we have not yet shown that equations (1.16) are connected to the ‘‘best’’ estimate for the variance component σ_0^2 . That is, the connection between $\tilde{\mathbf{e}}$ and $\hat{\sigma}_0^2$ is unknown at this point. Furthermore, there is no algebraic principle that allows us to make this connection; we must use statistics.

II) Statistical approach: This approach gives a more indirect estimate of the unknown parameters $\boldsymbol{\xi}$. Our principal unknown is an $m \times n$ matrix L that the

parameter estimate $\hat{\boldsymbol{\xi}}$ can be derived from. We seek a linear estimator. That is, the parameter estimates $\hat{\boldsymbol{\xi}}$ must depend linearly on the observation vector \mathbf{y} . Furthermore, we require the estimate to be *uniformly unbiased* and to be *best* in terms of minimum variance. Together these requirements comprise the *Best Linear Uniformly Unbiased Estimate*, or BLUUE, of $\boldsymbol{\xi}$. The components of BLUUE are outlined in the following:

(i) Linear requirement:

$$\hat{\boldsymbol{\xi}} = L\mathbf{y} + \boldsymbol{\gamma}, \text{ where } L \in \mathbb{R}^{m \times n} \text{ and } \boldsymbol{\gamma} \in \mathbb{R}^m. \quad (1.17a)$$

Equation (1.17a) is an inhomogeneous linear form due the $m \times 1$ vector $\boldsymbol{\gamma}$. It requires that both L and $\boldsymbol{\gamma}$ be determined. Therefore, there are $m(n+1)$ unknowns, a relatively large number.

(ii) Uniformly unbiased requirement:

$$\begin{aligned} \boldsymbol{\xi} &= E\{\hat{\boldsymbol{\xi}}\} \text{ for all } \boldsymbol{\xi} \in \mathbb{R}^m \Rightarrow \\ E\{L\mathbf{y} + \boldsymbol{\gamma}\} &= LE\{\mathbf{y}\} + \boldsymbol{\gamma} = LA\boldsymbol{\xi} + \boldsymbol{\gamma}, \end{aligned} \quad (1.17b)$$

leading to the two requirements

$$LA = I_m \text{ and } \boldsymbol{\gamma} = \mathbf{0}. \quad (1.17c)$$

Equation (1.17c) specifies $m \times m$ constraints in the first equation and m constraints in the second equation. Thus the number of unknowns minus the number of constraints is $m(n+1) - m(m+1) = m(n-m)$. Therefore, we have reduced our search space somewhat from $m(n+1)$.

(iii) Best requirement: By “best” we mean minimum average-variance. An average variance can be computed by dividing the trace of the $m \times m$ parameter dispersion matrix by m . However, division by m only scales the quantity to be minimized, so we can just as well minimize the trace itself.

$$\text{tr } D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \text{tr}(LP^{-1}L^T) = \min_L \{LA = I_m\} \quad (1.17d)$$

The quadratic form $LP^{-1}L^T$ in (1.17d) is the term to minimize. The term $LA = I_m$ imposes $m \times m$ constraints.

Because unbiasedness is required, the dispersion of $\hat{\boldsymbol{\xi}}$ is the same as the MSE of $\hat{\boldsymbol{\xi}}$. Thus we can write

$$\text{tr MSE}\{\hat{\boldsymbol{\xi}}\} = \text{tr } E\{(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^T\} = E\{(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^T(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\} = E\{\|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2\}. \quad (1.18)$$

The result is an expectation of a vector norm, which is a scalar. Note that the property of the trace being invariant with respect to a cyclic transformation was used in (1.18).

The Lagrange target function associated with (1.17d) is

$$\Phi(L, \boldsymbol{\lambda}) = \text{tr } LP^{-1}L^T + 2 \text{tr } \boldsymbol{\lambda}(LA - I_m), \quad (1.19)$$

which must be made stationary with respect to L and Λ . Here, Λ is an $m \times m$ symmetric matrix of Lagrange multipliers. Accordingly, the Euler-Lagrange necessary conditions are formed by

$$\frac{1}{2} \frac{\partial \Phi}{\partial L} = \hat{L}P^{-1} + \hat{\Lambda}A^T \doteq 0 \quad (1.20a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \Lambda} = \hat{L}A - I_m \doteq 0. \quad (1.20b)$$

In equation (1.20a) we have used rules (12) and (4) from section 10.3.2 of Lütkepohl (1996) for derivatives of the trace. Likewise, in equation (1.20b) we have used rule (5) from the same section of Lütkepohl (see equations (A.20) herein). The two condition equations are solved simultaneously as follows:

$$\hat{L} = -\hat{\Lambda}^T A^T P$$

implies that

$$-\hat{\Lambda}^T A^T P A = I_m \Rightarrow \hat{\Lambda} = -(A^T P A)^{-1},$$

finally leading to

$$\hat{L} = (A^T P A)^{-1} A^T P. \quad (1.21)$$

Substituting the solution for \hat{L} into (1.17a) and using the condition $\gamma = \mathbf{0}$ in (1.17c) leads to the BLUE for the parameters ξ as

$$\hat{\xi} = \hat{L}y + \gamma = (A^T P A)^{-1} A^T P y. \quad (1.22)$$

Comparing (1.22) to (1.9d), reveals that the BLUE of ξ is equivalent to the BLUE of ξ within the (full-rank) Gauss-Markov model.

Chapter 2

Introducing the vec Operator and the Kronecker-Zehfuss Product

The vec operator forms a column vector from the columns of the matrix that it operates on by stacking one column on top of the next, from first to last.

$$G \underset{p \times q}{:=} [\mathbf{g}_1, \dots, \mathbf{g}_q] \Rightarrow \text{vec } G \underset{pq \times 1}{:=} [\mathbf{g}_1^T, \dots, \mathbf{g}_q^T]^T \quad (2.1)$$

Here $\mathbf{g}_1, \dots, \mathbf{g}_q$ are the $p \times 1$ column vectors of G . Note that the reverse operation *is not unique*.

Given two $p \times q$ matrices A , B such that $A \underset{p \times q}{:=} [a_{ij}]$ and $B \underset{p \times q}{:=} [b_{ij}]$, the following relationship between the trace and the vec operator holds:

$$\begin{aligned} \text{tr}(A^T B) &= \text{tr}(BA^T) = \quad (\text{trace invariant with respect to a cyclic transf.}) \quad (2.2) \\ &= \sum_{i=1}^p \sum_{j=1}^q a_{ij} b_{ij} = \quad (\text{first sum for trace, second for matrix product}) \\ &= (\text{vec } A)^T \text{vec } B = \\ &= \sum_{ij} a_{ij} b_{ij}, \quad (\text{multiplies corresponding elements}) \end{aligned}$$

which finally allows us to write

$$\text{tr}(A^T B) = (\text{vec } A)^T \text{vec } B. \quad (2.3)$$

Given matrices A of size $p \times q$, B of size $q \times r$, and C of size $s \times r$, we have the following important relationship, which connects the vec operator and the Kronecker-Zehfuss product (or Kronecker product).

$$\text{vec}(ABC^T) \underset{ps \times 1}{=} (C \otimes A) \underset{ps \times qr}{\text{vec } B}, \underset{qr \times 1}{\quad} \quad (2.4)$$

where the Kronecker product is defined by

$$(C \otimes A) := [c_{ij}A] = \begin{bmatrix} c_{11}A & c_{12}A & \dots & c_{1r}A \\ c_{21}A & \ddots & & \\ \vdots & & & \\ c_{s1}A & c_{s2}A & \dots & c_{sr}A \end{bmatrix}. \quad (2.5)$$

The definition of the Kronecker product, as well as many of its properties, is given in the appendix and is used in several of the following sections.

Now we generalize formula (2.4) using a quadruple product of matrices, which is commonly found in the variance-component estimation problem.

$$\text{tr}(ABC^T D^T) = \text{tr}(D^T ABC^T) = \quad (\text{trace invariant w.r.t. a cyclic transf.}) \quad (2.6a)$$

$$= (\text{vec } D)^T \text{vec}(ABC^T) = \quad (\text{using equation (2.3)}) \quad (2.6b)$$

$$= (\text{vec } D)^T (C \otimes A) \text{vec } B. \quad (\text{using equation (2.4)}) \quad (2.6c)$$

It is also required at times to apply the vec operator to a vector outer-product. Given a vector \mathbf{a} , substitute into equation (2.4) $A = \mathbf{a}$ and $C^T = \mathbf{a}^T$; also let $B = I_1 = 1$. Then we have

$$\text{vec}(\mathbf{a}\mathbf{a}^T) = \mathbf{a} \otimes \mathbf{a}. \quad (2.7)$$

Commutation matrices appear in the rules for the Kronecker product in the appendix. Here we comment that a commutation matrix K is square and has only ones and zeros for its elements. Each row has exactly a single one, and likewise for each column. Thus the identity matrix is one example of a commutation matrix. The commutation matrix is not symmetric (except for the identity matrix), but it is orthogonal, meaning that $K^{-1} = K^T$. A commutation matrix is also a vec-permutation matrix. We illustrate this property by the following example:

$$K := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad A := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \Rightarrow \text{vec } A = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \end{bmatrix},$$

$$\text{vec } A^T = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \end{bmatrix} = K \text{vec } A.$$

To demonstrate the usefulness of the Kronecker product, we now show certain applications of it to the Gauss-Markov Model (GMM) and the associated Best Linear Uniformly Unbiased Estimate (BLUUE). We begin by deriving an alternative form for the target function (1.19) in order to exploit the Kronecker product.

$$\Phi(L, \Lambda) = \text{tr} LP^{-1}L^T + 2 \text{tr} \Lambda^T(LA - I_m) = \quad (2.8a)$$

$$= \text{tr} LP^{-1}L^T I_m + 2 \text{tr}(LA - I_m)\Lambda I_m = \quad (\text{noting the symmetry of } \Lambda)$$

$$= \text{tr} LP^{-1}L^T I_m + 2 \text{tr} LA\Lambda I_m - 2 \text{tr} I_m\Lambda I_m =$$

$$= (\text{vec } L^T)^T (I_m \times P^{-1}) \text{vec } L^T +$$

$$+ 2(\text{vec } L^T)^T (I_m \times A) \text{vec } \Lambda - 2(\text{vec } I_m)^T \text{vec } \Lambda \Rightarrow$$

$$\Phi(\mathbf{l}, \boldsymbol{\lambda}) = \mathbf{l}^T (I_m \times P^{-1}) \mathbf{l} + 2[\mathbf{l}^T (I_m \times A) - (\text{vec } I_m)^T] \boldsymbol{\lambda} \quad (2.8b)$$

Here, \mathbf{l} and $\boldsymbol{\lambda}$ are variables for the vectorized forms of the unknown matrices L and Λ and are define as follows:

$$\mathbf{l} := \text{vec}(L^T) \text{ is an } nm \times 1 \text{ vector containing the rows of } L \text{ in vector form.} \quad (2.8c)$$

$$\boldsymbol{\lambda} := \text{vec } \Lambda \text{ is an } m^2 \times 1 \text{ vector comprised of the columns of } \Lambda. \quad (2.8d)$$

Using vectors \mathbf{l} and $\boldsymbol{\lambda}$, the following Lagrange target function can be written as an alternative to (1.19):

$$\Phi(\mathbf{l}, \boldsymbol{\lambda}) = \mathbf{l}^T (I_m \otimes P^{-1}) \mathbf{l} - 2[\mathbf{l}^T (I_m \otimes A) - (\text{vec } I_m)^T] \boldsymbol{\lambda}, \quad (2.9)$$

which must be made stationary with respect to \mathbf{l} and $\boldsymbol{\lambda}$. Accordingly, the Euler-Lagrange necessary conditions are written as

$$\frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{l}} = (I_m \otimes P^{-1}) \hat{\mathbf{l}} - (I_m \otimes A) \hat{\boldsymbol{\lambda}} \doteq 0, \quad (2.10a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = -(I_m \otimes A^T) \hat{\mathbf{l}} + \text{vec } I_m \doteq 0. \quad (2.10b)$$

The normal equations are then solved for $\hat{\mathbf{l}}$ and $\hat{\boldsymbol{\lambda}}$ as follows: Equation (2.10a) implies that

$$\hat{\mathbf{l}} = (I_m \otimes P^{-1})^{-1} (I_m \otimes A) \hat{\boldsymbol{\lambda}} = (I_m \otimes PA) \hat{\boldsymbol{\lambda}}, \quad (2.11a)$$

which, together with (2.10b), further implies

$$(I_m \otimes A^T PA) \hat{\boldsymbol{\lambda}} = \text{vec } I_m, \quad (2.11b)$$

leading to the estimates

$$\hat{\boldsymbol{\lambda}} = (I_m \otimes A^T PA)^{-1} \text{vec } I_m = \text{vec}(A^T PA)^{-1}, \quad (2.11c)$$

$$\hat{\mathbf{l}} = \text{vec}(\hat{L}^T) = (I_m \otimes PA) \text{vec}(A^T PA)^{-1} = \text{vec}[PA(A^T PA)^{-1}], \quad (2.11d)$$

and finally to

$$\hat{L} = (A^T PA)^{-1} A^T P. \quad (2.11e)$$

The sufficient condition for minimization is satisfied by

$$\frac{1}{2} \frac{\partial^2 \Phi}{\partial \mathbf{l} \partial \mathbf{l}^T} = \frac{\partial (I_m \otimes P^{-1}) \hat{\mathbf{l}}}{\partial \mathbf{l}^T} = I_m \otimes P^{-1}, \quad (2.12)$$

which is positive definite.

The Best Linear Uniformly Unbiased Estimation (BLUUE) of $\boldsymbol{\xi}$ and its dispersion are, respectively,

$$\hat{\boldsymbol{\xi}} = \hat{L} \mathbf{y} = \hat{\Lambda}^T A^T P \mathbf{y} = (A^T P A)^{-1} A^T P \mathbf{y}, \quad (2.13a)$$

and

$$D\{\hat{\boldsymbol{\xi}}\} = \hat{\Lambda}^T A^T P (\sigma_0^2 P^{-1}) P A \hat{\Lambda} = \sigma_0^2 (A^T P A)^{-1} = \sigma_0^2 N^{-1}. \quad (2.13b)$$

Likewise, the predicted residual vector and its dispersion are, respectively,

$$\tilde{\mathbf{e}} := \mathbf{y} - A \hat{\boldsymbol{\xi}} = [I_n - A(A^T P A)^{-1} A^T P] \mathbf{y}, \quad (2.14a)$$

and

$$D\{\tilde{\mathbf{e}}\} = \sigma_0^2 [P^{-1} - A(A^T P A)^{-1} A^T]. \quad (2.14b)$$

Corollary: In the Gauss-Markov Model (GMM) with full-rank matrices A and P , the BLUUE of $\boldsymbol{\xi}$ is automatically generated by the LESS $\hat{\boldsymbol{\xi}}$ with the associated dispersion matrix $D\{\hat{\boldsymbol{\xi}}\}$ and residual vector $\tilde{\mathbf{e}}$. This fact is called “*Gauss’ second argument in favor of the least-squares adjustment.*”

Chapter 3

Variance Component Estimation

In this chapter we develop estimators for the unknown variance component σ_0^2 from the Gauss-Markov model (GMM) (1.3b). We begin by restating the full-rank GMM from chapter Chapter 1.

$$\mathbf{y} = A\boldsymbol{\xi} + \mathbf{e}, \quad \text{rk } A = m < n, \quad \mathbf{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}) \quad (3.1)$$

Our goal is to determine the estimated variance component $\hat{\sigma}_0^2$ in such a way that it is independent of the estimated parameter vector $\hat{\boldsymbol{\xi}}$.

First note that

$$E\{\mathbf{e}\mathbf{e}^T\} = \sigma_0^2 P^{-1} = D\{\mathbf{e}\}, \quad (3.2a)$$

and

$$\begin{aligned} E\{\mathbf{e}^T \mathbf{e}\} &= E\{\text{tr } \mathbf{e}^T \mathbf{e}\} = E\{\text{tr } \mathbf{e}\mathbf{e}^T\} = \text{tr } E\{\mathbf{e}\mathbf{e}^T\} = \sigma_0^2 \text{tr } P^{-1} = \\ &= E\{(\mathbf{y} - A\boldsymbol{\xi})^T (\mathbf{y} - A\boldsymbol{\xi})\} = E\{\mathbf{y}^T \mathbf{y}\} - E\{\mathbf{y}\}^T E\{\mathbf{y}\} = E\{\mathbf{y}^T \mathbf{y}\} - \boldsymbol{\xi}^T A^T A \boldsymbol{\xi}. \end{aligned} \quad (3.2b)$$

We see from (3.2b) that σ_0^2 and $\boldsymbol{\xi}$ are not decoupled. But we want the estimates $\hat{\sigma}_0^2$ and $\hat{\boldsymbol{\xi}}$ to be decoupled so that estimating the variance component σ_0^2 has nothing to do with estimating the parameter vector $\boldsymbol{\xi}$. To this end we seek the *Best Invariant Quadratic Uniformly Unbiased Estimate*, or BIQUUE. Each term in the acronym BIQUUE is explained below.

BIQUUE

(i) Quadratic requirement: $\hat{\sigma}_0^2$ is quadratic in the observation vector \mathbf{y} , such that

$$\hat{\sigma}_0^2 = \mathbf{y}^T M \mathbf{y}, \quad (3.3a)$$

where M is an unknown $n \times n$ matrix to be determined. Since $\hat{\sigma}_0^2$ is a scalar, it is equal to its transpose; therefore

$$\hat{\sigma}_0^2 = (\mathbf{y}^T M \mathbf{y})^T = \mathbf{y}^T M^T \mathbf{y} = \mathbf{y}^T \left(\frac{M + M^T}{2} \right) \mathbf{y}. \quad (3.3b)$$

Thus we can use M or M^T ; it does not matter. So, without loss of generality, we require the matrix M to be symmetric. This reduces the number of unknowns from n^2 to $n(n+1)/2$.

- (ii) Invariant requirement: We require the estimate to be invariant with respect to translation, i.e., invariant with respect to a shift of \mathbf{y} along the range space of A . The motivation for this requirement is to ensure that $\hat{\sigma}_0^2$ is independent of the estimated parameter vector $\hat{\boldsymbol{\xi}}$, an objective already stated above.

$$\hat{\sigma}_0^2 = (\mathbf{y} - A\boldsymbol{\xi})^T M (\mathbf{y} - A\boldsymbol{\xi}) \text{ for all } \boldsymbol{\xi} \in \mathbb{R}^m. \quad (3.4a)$$

Obviously, (3.4a) includes the estimate $\hat{\boldsymbol{\xi}}$ since it also belongs to \mathbb{R}^m . Due to invariance we can write

$$\begin{aligned} \hat{\sigma}_0^2 &= \mathbf{y}^T M \mathbf{y} = (\mathbf{y} - A\boldsymbol{\xi})^T M (\mathbf{y} - A\boldsymbol{\xi}) = \\ &= \mathbf{y}^T M \mathbf{y} - \mathbf{y}^T M A \boldsymbol{\xi} - \boldsymbol{\xi}^T A^T M \mathbf{y} + \boldsymbol{\xi}^T A^T M A \boldsymbol{\xi}, \end{aligned} \quad (3.4b)$$

implying that

$$\boldsymbol{\xi}^T A^T M A \boldsymbol{\xi} = 2\mathbf{y}^T M A \boldsymbol{\xi} \text{ for any } \mathbf{y} \text{ and any } \boldsymbol{\xi}. \quad (3.4c)$$

“For any $\boldsymbol{\xi}$ ” means that $\boldsymbol{\xi}$ could be positive or negative. The left side of (3.4c) would not change if $\boldsymbol{\xi}$ is replaced by $-\boldsymbol{\xi}$, but the right side would change in sign. The only quantity that remains equal when we change the sign of $\boldsymbol{\xi}$ in (3.4c) is zero. Therefore, the condition becomes

$$\mathbf{y}^T M A \boldsymbol{\xi} = \mathbf{0} \text{ for any } \mathbf{y} \text{ and any } \boldsymbol{\xi}, \text{ which is true if, and only if, } M A = \mathbf{0}. \quad (3.4d)$$

This matrix constraint satisfies the invariant condition; that is, the “decoupling” between $\hat{\boldsymbol{\xi}}$ and $\hat{\sigma}_0^2$ guarantees invariance.

- (iii) Uniformly Unbiased requirement: The equality

$$\sigma_0^2 = E\{\hat{\sigma}_0^2\} = E\{\mathbf{y}^T M \mathbf{y}\} = E\{\mathbf{e}^T M \mathbf{e}\} \quad (3.5a)$$

holds since $\hat{\sigma}_0^2 = \mathbf{y}^T M \mathbf{y}$ according to (3.3a). Also, due to the invariant principle, $\mathbf{e} = \mathbf{y} - A\boldsymbol{\xi}$ holds for all $\boldsymbol{\xi}$. Thus, we may continue with

$$\sigma_0^2 = \text{tr} E\{M \mathbf{e} \mathbf{e}^T\} = \text{tr}(M D\{\mathbf{e}\}) = \sigma_0^2 \text{tr}(M P^{-1}) \text{ for all } \sigma_0^2 \in \mathbb{R}^+ \Leftrightarrow (3.5b)$$

$$1 = \text{tr}(M P^{-1}). \quad (3.5c)$$

Here, \mathbb{R}^+ denotes the field of positive real numbers. Equation (3.5c) provides the uniformly-unbiased condition; it holds due to the invariance principle.

- (iv) Best requirement: “Best” means that the dispersion of the estimated variance component $\hat{\sigma}_0^2$ must be minimized.

$$D\{\hat{\sigma}_0^2\} = \min_M \begin{cases} MA = 0 \\ \text{tr}(MP^{-1}) = 1 \end{cases} \quad (3.6)$$

We are dealing with the dispersion of a quadratic form, which is a fourth moment. Therefore we need to make an additional assumption. We must assume quasi-normality, which says that the fourth moment behaves as if the random errors \mathbf{e} are normally distributed.

Aside: For the normal distribution, all moments can be written as a function of the first and second moments. Therefore, for the i th random error e_i , we make the following assumptions:

$$E\{e_i^4\} = 3(\sigma_i^2)^2 \quad \text{and} \quad E\{e_{2i+1}\} = 0. \quad (3.7)$$

The left side of (3.6) can also be expressed as follows:

$$D\{\hat{\sigma}_0^2\} = D\{\mathbf{e}^T M \mathbf{e}\} = E\{\mathbf{e}^T M \mathbf{e} \mathbf{e}^T M \mathbf{e}\} - E\{\mathbf{e}^T M \mathbf{e}\}^2 = \quad (3.8a)$$

$$= E\{\mathbf{e}^T M \mathbf{e} \mathbf{e}^T M \mathbf{e}\} - [\sigma_0^2 \text{tr}(MP^{-1})]^2 = E\{\mathbf{e}^T M \mathbf{e} \mathbf{e}^T M \mathbf{e}\} - (\sigma_0^2)^2. \quad (3.8b)$$

Note that in numerical computations we would replace \mathbf{e} with \mathbf{y} in (3.8) since \mathbf{e} is unknown. However, analytically the results are the same due to the invariance property.

Now, the expectation term in (3.8b) consists of products in the random variable \mathbf{e} ; therefore, it can be expressed as a sum of the expectations of all combinations of the products. We illustrate this with symbols under the respective occurrences of \mathbf{e} as follows (obviously each of these accented vectors \mathbf{e} are actually equivalent to one another):

$$\begin{aligned} E\{\mathbf{e}^T M \underline{\mathbf{e}} \underline{\mathbf{e}}^T M \underline{\mathbf{e}}\} &= E\{\underline{\mathbf{e}}^T M \underline{\mathbf{e}}\} E\{\underline{\mathbf{e}}^T M \underline{\mathbf{e}}\} + E\{M \underline{\mathbf{e}} \underline{\mathbf{e}}^T\} E\{M \underline{\mathbf{e}} \underline{\mathbf{e}}^T\} + \\ &\quad + E\{\underline{\mathbf{e}} \underline{\mathbf{e}}^T\} E\{M \underline{\mathbf{e}} \underline{\mathbf{e}}^T M^T\} = \\ &= (\sigma_0^2)^2 + ME\{\mathbf{e} \mathbf{e}^T\} ME\{\mathbf{e} \mathbf{e}^T\} + E\{\mathbf{e} \mathbf{e}^T\} ME\{\mathbf{e} \mathbf{e}^T\} M^T = \\ &= (\sigma_0^2)^2 + (\sigma_0^2)^2 MP^{-1} MP^{-1} + (\sigma_0^2)^2 P^{-1} MP^{-1} M^T. \end{aligned} \quad (3.9)$$

Noting that M is symmetrical and substituting (3.9) into (3.8b) and then applying the trace operator yields

$$D\{\hat{\sigma}_0^2\} = 2(\sigma_0^2)^2 \text{tr}(P^{-1} MP^{-1} M^T) = 2(\sigma_0^2)^2 (\text{vec } M)^T (P^{-1} \otimes P^{-1}) \text{vec } M, \quad (3.10)$$

leading to the following expression that must be minimized:

$$D\{\hat{\sigma}_0^2\} = \min_M \{MA = 0, \text{tr}(MP^{-1}) = 1\}. \quad (3.11)$$

Introducing an $m \times n$ matrix of Lagrange multipliers Λ and a scalar Lagrange multiplier λ_0 leads to the following Lagrange target function:

$$\Phi(M, \Lambda, \lambda_0) = \text{tr}(MP^{-1}MP^{-1}) - 2 \text{tr}(M^T A \Lambda) - 2[\text{tr}(MP^{-1}) - 1]\lambda_0, \quad (3.12)$$

which must be made stationary with respect to the unknown variables M , Λ , and λ_0 .

An alternative form of the Lagrangian function can be written as

$$\begin{aligned} \Phi(\text{vec } M, \boldsymbol{\lambda}, \lambda_0) &= 2(\text{vec } M)^T (P^{-1} \otimes P^{-1}) \text{vec } M + \\ &+ 4(\text{vec } M)^T (I_n \otimes A) \boldsymbol{\lambda} + 4[(\text{vec } M)^T \text{vec } P^{-1} - 1] \lambda_0, \end{aligned} \quad (3.13)$$

which must be made stationary with respect to $\text{vec } M$, $\boldsymbol{\lambda}$, and λ_0 .

Note that $\text{vec } M$ is an $n^2 \times 1$ vector; $\boldsymbol{\lambda}$ is an $nm \times 1$ vector, and λ_0 is a scalar. The equivalence between (3.12) and (3.13) is seen by noting that $\text{vec } \Lambda = \boldsymbol{\lambda}$ and by use of (2.2) and (2.5).

The Euler-Lagrange necessary conditions result in

$$\frac{1}{4} \frac{\partial \Phi}{\partial \text{vec } M} = (P^{-1} \otimes P^{-1}) \text{vec } M + (I_n \otimes A) \hat{\boldsymbol{\lambda}} + \text{vec } P^{-1} \hat{\lambda}_0 \doteq \mathbf{0}, \quad (3.14a)$$

$$\frac{1}{4} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = (I_n \otimes A^T) \text{vec } M \doteq \mathbf{0}, \quad (3.14b)$$

$$\frac{1}{4} \frac{\partial \Phi}{\partial \lambda_0} = (\text{vec } P^{-1})^T \text{vec } M - 1 \doteq 0. \quad (3.14c)$$

For convenience, a hat symbol is not used over the particular vector $\text{vec } M$ that we solve for in the minimization, as has been done for $\hat{\boldsymbol{\lambda}}$ and $\hat{\lambda}_0$. From (3.14a) we get the following expression for $\text{vec } M$:

$$\begin{aligned} \text{vec } M &= -(P^{-1} \otimes P^{-1})^{-1} (I_n \otimes A) \hat{\boldsymbol{\lambda}} - (P^{-1} \otimes P^{-1})^{-1} \text{vec } (P^{-1}) \hat{\lambda}_0 = \\ &= -(P \otimes P) (I_n \otimes A) \hat{\boldsymbol{\lambda}} - (P \otimes P) \text{vec } (P^{-1}) \hat{\lambda}_0 = \\ &= -(P I_n \otimes P A) \hat{\boldsymbol{\lambda}} - \text{vec } (P P^{-1} P) \hat{\lambda}_0 \Rightarrow \\ \text{vec } M &= -(P \otimes P A) \hat{\boldsymbol{\lambda}} - \text{vec } (P) \hat{\lambda}_0. \end{aligned} \quad (3.15a)$$

Now, substituting (3.15a) into (3.14b) and dropping the negative sign results in

$$\begin{aligned} (I_n \otimes A^T) [(P \otimes P A) \hat{\boldsymbol{\lambda}} + \text{vec } (P) \hat{\lambda}_0] &= 0 \\ (I_n \otimes A^T) (P \otimes P A) \hat{\boldsymbol{\lambda}} + (I_n \otimes A^T) \text{vec } (P) \hat{\lambda}_0 &= 0 \\ (I_n P \otimes A^T P A) \hat{\boldsymbol{\lambda}} + \text{vec } (A^T P I_n^T) \hat{\lambda}_0 &= 0 \\ (P \otimes N) \hat{\boldsymbol{\lambda}} + \text{vec } (A^T P) \hat{\lambda}_0 &= 0. \end{aligned} \quad (3.15b)$$

Finally, we can write the estimated vector of Lagrange multipliers as

$$\hat{\boldsymbol{\lambda}} = -(P^{-1} \otimes N^{-1}) \text{vec } (A^T P) \hat{\lambda}_0 = -\text{vec } (N^{-1} A^T P P^{-1}) \hat{\lambda}_0 = -\text{vec } (N^{-1} A^T) \hat{\lambda}_0. \quad (3.15c)$$

Substituting (3.15c) into (3.15a) yields

$$\begin{aligned} \text{vec } M &= (P \otimes PA) \text{vec}(N^{-1}A^T)\hat{\lambda}_0 - \text{vec}(P)\hat{\lambda}_0 = \\ &= \text{vec}(PAN^{-1}A^TP)\hat{\lambda}_0 - \text{vec}(P)\hat{\lambda}_0 \Rightarrow \\ \text{vec } M &= -[\text{vec}(P) - \text{vec}(PAN^{-1}A^TP)]\hat{\lambda}_0. \end{aligned} \quad (3.15d)$$

Then we substitute (3.15d) into (3.14c) to obtain

$$-(\text{vec } P^{-1})^T [\text{vec}(P) - \text{vec}(PAN^{-1}A^TP)]\hat{\lambda}_0 = 1. \quad (3.15e)$$

Using (2.2) allows us to rewrite (3.15e) and solve for $\hat{\lambda}_0$ as follows:

$$\begin{aligned} -\text{tr}(P^{-1}[P - PAN^{-1}A^TP])\hat{\lambda}_0 &= 1 \Rightarrow \\ -\text{tr}(I_n - AN^{-1}A^TP)\hat{\lambda}_0 &= 1 \Rightarrow \\ -[\text{tr } I_n - \text{tr}(N^{-1}A^TPA)]\hat{\lambda}_0 &= 1 \Rightarrow \\ (n - m)\hat{\lambda}_0 &= -1 \Rightarrow \\ \hat{\lambda}_0 &= -1/(n - m). \end{aligned} \quad (3.15f)$$

Now we substitute (3.15f) into (3.15d) and obtain

$$\text{vec } M = \text{vec} \{(n - m)^{-1}[P - PAN^{-1}A^TP]\}. \quad (3.15g)$$

Because the matrices within the vec operator in (3.15g) are of the same size, we can write

$$M = (n - m)^{-1}[P - PAN^{-1}A^TP]. \quad (3.15h)$$

Now we substitute (3.15h) into (3.3a) to obtain an expression for the estimated variance component as

$$\hat{\sigma}_0^2 = \mathbf{y}^T M \mathbf{y} = \mathbf{y}^T [(n - m)^{-1}(P - PAN^{-1}A^TP)] \mathbf{y} = \quad (3.16a)$$

$$= (n - m)^{-1}(\mathbf{y}^T P \mathbf{y} - \mathbf{y}^T PAN^{-1}A^TP \mathbf{y}). \quad (3.16b)$$

Equation (3.16b) is the BIQUUE for the unknown variance component σ_0^2 . The estimated variance component $\hat{\sigma}_0^2$ has been determined independently from the parameter estimate $\hat{\boldsymbol{\xi}}$, which was our objective.

Let us verify that the two conditions stated in (3.6) are satisfied for matrix M .

First condition: $MA = 0$

$$\begin{aligned} MA &= \{(n - m)^{-1}[P - PAN^{-1}A^TP]\}A = (n - m)^{-1}[PA - PAN^{-1}A^TPA] \Rightarrow \\ MA &= (n - m)^{-1}[PA - PA] = 0 \end{aligned} \quad (3.17a)$$

Second condition: $\text{tr}(MP^{-1}) = 1$

$$\begin{aligned} \text{tr}(MP^{-1}) &= \text{tr}(\{(n-m)^{-1}[P - PAN^{-1}A^T P]\})P^{-1} \Rightarrow \\ \text{tr}(MP^{-1}) &= (n-m)^{-1}[\text{tr} I_n - \text{tr} PAN^{-1}A^T] = (n-m)^{-1}(\text{tr} I_n - \text{tr} I_m) = 1 \end{aligned} \quad (3.17b)$$

Now using the symbols N and \mathbf{c} introduced in (1.8), we can rewrite BIQUUE (3.16b) as follows:

$$\hat{\sigma}_0^2 = \frac{\mathbf{y}^T P \mathbf{y} - \mathbf{c}^T \hat{\boldsymbol{\xi}}}{n-m} = \frac{\mathbf{y}^T P \mathbf{y} - \hat{\boldsymbol{\xi}}^T N \hat{\boldsymbol{\xi}}}{n-m} = \frac{\tilde{\mathbf{e}}^T P \tilde{\mathbf{e}}}{n-m}. \quad (3.18)$$

The vector $\tilde{\mathbf{e}}$ in (3.18) is the same as the predicted residual vector associated with the BLUE of $\boldsymbol{\xi}$. Thus the BIQUUE variance component $\hat{\sigma}_0^2$ agrees with that associated with BLUE for $\boldsymbol{\xi}$. Also note that the use of the symbol $\hat{\boldsymbol{\xi}}$ in (3.18) is only done for convenience and does not mean that BIQUUE depends on the estimate for the parameter vector.

Incidentally, if we omit the uniformly-unbiased condition of (3.6), we arrive at BIQE, which differs from BIQUUE by an addition of 2 in the denominator.

$$\hat{\sigma}_0^2 = \frac{\tilde{\mathbf{e}}^T P \tilde{\mathbf{e}}}{n-m+2} = \text{BIQE}\{\sigma_0^2\} \quad (3.19)$$

The BIQUUE variance component $\hat{\sigma}_0^2$ is a random variable, so we want to find its expectation and dispersion.

First it is useful to compute the expectation $E\{\mathbf{y}\mathbf{y}^T\}$.

$$\begin{aligned} E\{\mathbf{y}\mathbf{y}^T\} &= E\{(A\boldsymbol{\xi} + \mathbf{e})(A\boldsymbol{\xi} + \mathbf{e})^T\} = E\{A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T + A\boldsymbol{\xi}\mathbf{e}^T + \mathbf{e}\boldsymbol{\xi}^T A^T + \mathbf{e}\mathbf{e}^T\} = \\ &= E\{A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T\} + E\{A\boldsymbol{\xi}\mathbf{e}^T\} + E\{\mathbf{e}\boldsymbol{\xi}^T A^T\} + E\{\mathbf{e}\mathbf{e}^T\} = \\ &= A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T + A\boldsymbol{\xi}E\{\mathbf{e}^T\} + E\{\mathbf{e}\}\boldsymbol{\xi}^T A^T + D\{\mathbf{e}\} \Rightarrow \\ E\{\mathbf{y}\mathbf{y}^T\} &= \sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T \end{aligned} \quad (3.20a)$$

Next we compute the expectation of $\hat{\sigma}_0^2$.

$$\begin{aligned} (n-m)E\{\hat{\sigma}_0^2\} &= E\{\mathbf{y}^T P \mathbf{y} - \mathbf{y}^T P A N^{-1} A^T P \mathbf{y}\} = \\ &= \text{tr} E\{\mathbf{y}^T P \mathbf{y} - \mathbf{y}^T P A N^{-1} A^T P \mathbf{y}\} = \text{tr}(PE\{\mathbf{y}\mathbf{y}^T\}) - \text{tr}(P A N^{-1} A^T P E\{\mathbf{y}\mathbf{y}^T\}) = \\ &= \text{tr}[P(\sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T)] - \text{tr}[P A N^{-1} A^T P(\sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T)] = \\ &= \text{tr}(\sigma_0^2 I_n + \boldsymbol{\xi}^T N \boldsymbol{\xi}) - \text{tr}(P A N^{-1} A^T \sigma_0^2 + P A N^{-1} A^T P A \boldsymbol{\xi}\boldsymbol{\xi}^T A^T) = \\ &= \text{tr}(\sigma_0^2 I_n + \boldsymbol{\xi}^T N \boldsymbol{\xi}) - \text{tr}(I_m \sigma_0^2 + \boldsymbol{\xi}^T N \boldsymbol{\xi}) \Rightarrow \\ (n-m)E\{\hat{\sigma}_0^2\} &= \sigma_0^2(n-m) \end{aligned} \quad (3.20b)$$

Finally, we can write the expectation of the BIQUUE variance component as

$$E\{\hat{\sigma}_0^2\} = \sigma_0^2. \quad (3.21)$$

Equation (3.21) shows that BIQUUE $\hat{\sigma}_0^2$ is indeed an unbiased estimate of σ_0^2 .

The dispersion of BIQUUE $\hat{\sigma}_0^2$ is computed as follows: Considering (3.10) and that matrix M is symmetric, we write

$$D\{\hat{\sigma}_0^2\} = 2(\sigma_0^2)^2 \text{tr}(MP^{-1}MP^{-1}).$$

Then, considering (3.15h)

$$\begin{aligned} D\{\hat{\sigma}_0^2\} &= 2(\sigma_0^2)^2 \text{tr}([P - PAN^{-1}A^T P]P^{-1}[P - PAN^{-1}A^T P]P^{-1})(n - m)^{-2} = \\ &= 2(\sigma_0^2)^2 \text{tr}([I_n - PAN^{-1}A^T]^2)(n - m)^{-2} = \end{aligned}$$

(because $[I_n - PAN^{-1}A^T]$ is idempotent)

$$\begin{aligned} &= 2(\sigma_0^2)^2 [\text{tr} I_n - \text{tr}(A^T PAN^{-1})](n - m)^{-2} = \\ &= 2(\sigma_0^2)^2 (n - m)(n - m)^{-2}, \end{aligned}$$

finally resulting in

$$D\{\hat{\sigma}_0^2\} = 2(\sigma_0^2)^2 / (n - m). \quad (3.22)$$

Equation (3.22) shows the true dispersion of the BIQUUE variance component $D\{\hat{\sigma}_0^2\}$ in terms of the true variance component σ_0^2 . Equation (3.22) also implies that the estimated dispersion is provided by

$$\boxed{\hat{D}\{\hat{\sigma}_0^2\} = 2(\hat{\sigma}_0^2)^2 / (n - m)}. \quad (3.23)$$

From equation (3.23) we see that the estimated dispersion of the BIQUUE variance component will turn out to be relatively large unless the model redundancy $n - m$ is large.

We gave the solution to BIQE above; it can be shown that its dispersion is given by

$$D\{\hat{\hat{\sigma}}_0^2\} = 2(\sigma_0^2)^2 / (n - m + 2). \quad (3.24)$$

Chapter 4

Expectation-Dispersion Correspondence

An alternative approach to estimating the variance component σ_0^2 exploits the vec operator to a larger degree by changing the quadratic estimate to a linear estimate. Mathematically, this change is expressed by

$$\mathbf{y}^T M \mathbf{y} \rightarrow \text{vec}(\mathbf{y}^T M \mathbf{y}) = (\text{vec } M)^T (\mathbf{y} \otimes \mathbf{y}). \quad (4.1)$$

The first term in (4.1) is quadratic in \mathbf{y} . The rightmost term is linear in $(\mathbf{y} \otimes \mathbf{y})$. Note that the equation in (4.1) holds since (using (A.2))

$$\mathbf{y}^T M \mathbf{y} = \text{tr}(\mathbf{y}^T M \mathbf{y}) = \text{tr}(M^T \mathbf{y} I_1 \mathbf{y}^T) = (\text{vec } M)^T (\mathbf{y} \otimes \mathbf{y}), \quad (4.2)$$

where M is symmetric by definition.

The key idea is to change our original (quadratic) model so that the Best Linear Uniformly Unbiased Estimate (BLUUE) of the variance component in the revised (linear) model is the same as the BIQUUE for the variance component in the original model. We call this equivalence Expectation-Dispersion (E-D) Correspondence, so named because we rephrase the dispersion $D\{\hat{\sigma}_0^2\}$ as an expectation.

We begin by computing the expectation of the Kronecker product $\mathbf{y} \otimes \mathbf{y}$ in (4.1). Using (2.7) and (3.20a), we can write

$$E\{\mathbf{y} \otimes \mathbf{y}\} = \text{vec } E\{\mathbf{y} \mathbf{y}^T\} = \text{vec}(\sigma_0^2 P^{-1} + A \boldsymbol{\xi} \boldsymbol{\xi}^T A^T) = \quad (4.3a)$$

(applying (2.7) to $A \boldsymbol{\xi}$)

$$= (\text{vec } P^{-1}) \sigma_0^2 + (A \boldsymbol{\xi} \otimes A \boldsymbol{\xi}) \Rightarrow \quad (4.3b)$$

(using (A.8))

$$E\{\mathbf{y} \otimes \mathbf{y}\} = (\text{vec } P^{-1}) \sigma_0^2 + (A \otimes A)(\boldsymbol{\xi} \otimes \boldsymbol{\xi}). \quad (4.3c)$$

Both unknown quantities σ_0^2 and $\boldsymbol{\xi}$ appear in equation (4.3c). Note that we could estimate the term $\boldsymbol{\xi} \otimes \boldsymbol{\xi}$ appearing (4.3c); however, $\boldsymbol{\xi} \otimes \boldsymbol{\xi}$ tells us nothing about $\boldsymbol{\xi}$ itself, and therefore we consider $\boldsymbol{\xi} \otimes \boldsymbol{\xi}$ to be a nuisance parameter. Note that the

size of $\boldsymbol{\xi} \otimes \boldsymbol{\xi}$ is $m^2 \times 1$; however the product contains only $(m+1)/2$ independent elements. We need a matrix B that, when multiplied on the left of (4.3c), will eliminate the nuisance parameters $\boldsymbol{\xi} \otimes \boldsymbol{\xi}$, i.e., $B(A \otimes A) = 0$. Also, the matrix B must satisfy

$$\text{tr } B + m(m+1)/2 = n(m+1)/2. \quad (4.4)$$

For example, we could choose B as

$$B := (I_n - AN^{-1}A^T P) \otimes (I_n - AN^{-1}A^T P). \quad (4.5)$$

It is apparent from (4.5) that $B(A \otimes A) = 0$. Also, using (A.13), we have $\text{tr } B = (n-m)^2$, which is the number of independent equations left in the model. The rank of matrix B is easily computed by noting that the matrix within the parenthetical terms in (4.5) is idempotent and that the rank of an idempotent matrix equals its trace, and by using (A.13).

After multiplication by matrix B , the resulting model (now linear in $\mathbf{y} \otimes \mathbf{y}$) is not equivalent to the original model (which is linear in \mathbf{y}), but we choose to proceed with this matrix B anyway. Our next step is to find the expectation $E\{\tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}\}$, and to do so we begin with $B(\mathbf{y} \otimes \mathbf{y})$ since

$$B(\mathbf{y} \otimes \mathbf{y}) = [(I_n - AN^{-1}A^T P) \otimes (I_n - AN^{-1}A^T P)](\mathbf{y} \otimes \mathbf{y}) = \quad (4.6a)$$

$$= (I_n - AN^{-1}A^T P)\mathbf{y} \otimes (I_n - AN^{-1}A^T P)\mathbf{y} = \quad (4.6b)$$

$$= (\mathbf{y} - A\hat{\boldsymbol{\xi}}) \otimes (\mathbf{y} - A\hat{\boldsymbol{\xi}}) = \tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}. \quad (4.6c)$$

Here, $\hat{\boldsymbol{\xi}}$ is the BLUE for the parameter vector in the Gauss-Markov model (GMM). Continuing, with the help of (4.3c), we find

$$E\{\tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}\} = B \cdot E\{\mathbf{y} \otimes \mathbf{y}\} = B[(\text{vec } P^{-1})\sigma_0^2 + (A \otimes A)(\boldsymbol{\xi} \otimes \boldsymbol{\xi})] = \quad (4.7a)$$

$$= B(\text{vec } P^{-1})\sigma_0^2 + B(A \otimes A)(\boldsymbol{\xi} \otimes \boldsymbol{\xi}) \Rightarrow \quad (4.7b)$$

$$\boxed{E\{\tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}\} = B(\text{vec } P^{-1})\sigma_0^2.} \quad (4.7c)$$

With equation (4.7c) we have a linear model in σ_0^2 . An alternative expression for $E\{\tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}\}$ is derived as follows:

$$E(\tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}) = B(\text{vec } P^{-1})\sigma_0^2 = \quad (4.8a)$$

$$= [(I_n - AN^{-1}A^T P) \otimes (I_n - AN^{-1}A^T P)](\text{vec } P^{-1}) \cdot \sigma_0^2 =$$

(applying (A.1))

$$= \text{vec}[(I_n - AN^{-1}A^T P)P^{-1}(I_n - PAN^{-1}A^T) \cdot \sigma_0^2] = \quad (4.8b)$$

(transposing the symmetrical part)

$$= \text{vec}[(I_n - AN^{-1}A^T P)(I_n - AN^{-1}A^T P)P^{-1} \cdot \sigma_0^2] \Rightarrow \quad (4.8c)$$

(exploiting the idempotent property)

$$\boxed{E(\tilde{\boldsymbol{\epsilon}} \otimes \tilde{\boldsymbol{\epsilon}}) = \text{vec}(P^{-1} - AN^{-1}A^T) \cdot \sigma_0^2.} \quad (4.8d)$$

Note that the matrix expression $(P^{-1} - AN^{-1}A^T)$ might be singular, but this poses no problem due to use of the vec operator. We now derive the dispersion of $\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}$.

First, note that

$$\tilde{\mathbf{e}} = \mathbf{y} - A\hat{\boldsymbol{\xi}} = (A\boldsymbol{\xi} + \mathbf{e}) - A\hat{\boldsymbol{\xi}} = \mathbf{e} + A(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}). \quad (4.9)$$

This relation, along with the invariance principle, is exploited in the following:

$$D\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\} = D\{B(\mathbf{y} \otimes \mathbf{y})\} = D\{B[(\mathbf{y} - A\boldsymbol{\xi}) \otimes (\mathbf{y} - A\boldsymbol{\xi})]\} = \quad (4.10a)$$

$$= D\{B(\mathbf{e} \otimes \mathbf{e})\} = BD\{\mathbf{e} \otimes \mathbf{e}\}B^T = \quad (4.10b)$$

$$= B(E\{(\mathbf{e} \otimes \mathbf{e})(\mathbf{e} \otimes \mathbf{e})^T\} - E\{\mathbf{e} \otimes \mathbf{e}\}E\{\mathbf{e} \otimes \mathbf{e}\}^T)B^T. \quad (4.10c)$$

Now, considering the first expectation term in (4.10c), and temporarily using various symbols beneath the variables to illustrate how the combinations are formed, we find

$$\begin{aligned} E\{(\underline{\mathbf{e}} \otimes \underline{\mathbf{e}})(\underline{\mathbf{e}} \otimes \underline{\mathbf{e}})^T\} &= E\{\underline{\mathbf{e}} \otimes \underline{\mathbf{e}}\}E\{(\underline{\mathbf{e}} \otimes \underline{\mathbf{e}})^T\} + \\ &+ E\{\underline{\mathbf{e}}\underline{\mathbf{e}}^T\} \otimes E\{\underline{\mathbf{e}}\underline{\mathbf{e}}^T\} + K(E\{\underline{\mathbf{e}}\underline{\mathbf{e}}^T\} \otimes E\{\underline{\mathbf{e}}\underline{\mathbf{e}}^T\}), \end{aligned} \quad (4.11a)$$

or, more simply

$$\begin{aligned} E\{(\mathbf{e} \otimes \mathbf{e})(\mathbf{e} \otimes \mathbf{e})^T\} &= E\{(\mathbf{e} \otimes \mathbf{e})\}E\{(\mathbf{e} \otimes \mathbf{e})^T\} + \\ &+ E\{\mathbf{e}\mathbf{e}^T\} \otimes E\{\mathbf{e}\mathbf{e}^T\} + K(E\{\mathbf{e}\mathbf{e}^T\} \otimes E\{\mathbf{e}\mathbf{e}^T\}). \end{aligned} \quad (4.11b)$$

In equation (4.11b) a commutation matrix K has been introduced by way of (A.10). Inserting (4.11b) into (4.10c) and making use of (3.2a) leads to

$$D\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\} = B(I + K)(P^{-1} \otimes P^{-1}) \cdot (\sigma_0^2)^2 \cdot B^T = \quad (4.12a)$$

$$= (\sigma_0^2)^2 \cdot (I + K)[B(P^{-1} \otimes P^{-1})B^T] \Rightarrow \quad (4.12b)$$

$$D\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\} = (\sigma_0^2)^2 \cdot (I + K)[(P^{-1} - AN^{-1}A^T) \otimes (P^{-1} - AN^{-1}A^T)]. \quad (4.12c)$$

In (4.12b) we have used the fact that B is a Kronecker product of the same matrix (see (4.5)), so that $BK = KB$. We may now combine equations (4.8d) and (4.12c) into one succinct expression describing the distribution of $\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}$ as follows:

$$\boxed{\begin{aligned} \tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}} &\sim (\sigma_0^2 \cdot \text{vec}(P^{-1} - AN^{-1}A^T), \\ &(\sigma_0^2)^2 \cdot (I + K)[(P^{-1} - AN^{-1}A^T) \otimes (P^{-1} - AN^{-1}A^T)]). \end{aligned}} \quad (4.13)$$

Note that both the expectation and dispersion in (4.13) contain the parameter σ_0^2 . Also note that the matrix comprised of the Kronecker product is singular. This equation has some similarities to the GMM, enough to try the least-squares solution (LESS) approach to estimate σ_0^2 . We call this approach *E-D Correspondence*, the concept of which is summarized in the diagram below.

$$\boxed{\mathbf{y} \sim (A\boldsymbol{\xi}, \sigma_0^2 P^{-1}) \xrightarrow[\text{Correspondence}]{\text{E-D}} \tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}} \sim (E\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\}, D\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\})} \quad (4.14)$$

In order to proceed with the estimation of σ_0^2 using LESS, we need to handle the singular dispersion matrix in (4.13), which requires a generalized inverse (g-inverse). The g-inverse of a $p \times q$ matrix G is defined as the $q \times p$ matrix G^- such that

$$GG^-G = G. \quad (4.15)$$

We seek a g-inverse for the matrix $(P^{-1} - AN^{-1}A^T)$, which is provided by $(P - PAN^{-1}A^TP)$, since

$$(P^{-1} - AN^{-1}A^T)(P - PAN^{-1}A^TP)(P^{-1} - AN^{-1}A^T) = (P^{-1} - AN^{-1}A^T). \quad (4.16)$$

Furthermore, we define the (singular) cofactor matrix from (4.13) as

$$Q := (I + K)[(I_n - AN^{-1}A^TP)P^{-1} \otimes (I_n - AN^{-1}A^TP)P^{-1}], \quad (4.17)$$

where the term $(I + K)$ is essentially a factor of 2 in Q . Let the g-inverse of matrix Q be called W , so that $QWQ = Q$. The following matrix satisfies this equation:

$$W := \frac{1}{4}(I + K)[P(I_n - AN^{-1}A^TP) \otimes P(I_n - AN^{-1}A^TP)]. \quad (4.18)$$

Note that multiplication of matrix W on both the right and left by Q yields

$$QWQ = \frac{1}{4}(I + K)^3[(I_n - AN^{-1}A^TP)P^{-1} \otimes (I_n - AN^{-1}A^TP)P^{-1}], \quad (4.19a)$$

but

$$\frac{1}{4}(I + K)^3 = \frac{1}{4}(I + 3K + 3K^2 + K^3) = \frac{1}{4}(I + 3K + 3I + IK) = I + K. \quad (4.19b)$$

So, indeed, W is a g-inverse of Q .

In the GMM we reach LESS by minimization of the target function $\Phi = (\mathbf{y} - A\xi)^T P(\mathbf{y} - A\xi)$. Now we are able to write an analogous LESS target function for the estimated variance component σ_0^2 using the g-inverse W derived above.

LESS target function:

$$\begin{aligned} & [(\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}) - E\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\}]^T W [(\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}) - E\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\}] = \\ & = \{(\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}) - \text{vec}[(I_n - AN^{-1}A^TP)P^{-1}]\sigma_0^2\}^T \cdot W \cdot \{(\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}) - \text{vec}[(I_n - AN^{-1}A^TP)P^{-1}]\sigma_0^2\} = \min_{\sigma_0^2}. \end{aligned} \quad (4.20)$$

Following the LESS approach, we write a system of normal equations directly, based on the target function (4.20).

LESS normal equations:

$$\begin{aligned} & \{\text{vec}[(I_n - AN^{-1}A^TP)P^{-1}]\}^T W \text{vec}[(I_n - AN^{-1}A^TP)P^{-1}] \cdot \hat{\sigma}_0^2 = \\ & = \{\text{vec}[(I_n - AN^{-1}A^TP)P^{-1}]\}^T W (\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}). \end{aligned} \quad (4.21)$$

To derive a solution for the estimated variance component σ_0^2 , we first simplify

$$\{\text{vec}[(I_n - AN^{-1}A^T P)P^{-1}]\}^T W,$$

since it appears in both sides of (4.21). Here we use (A.1) and the fact that $(I + K)$ is equivalent to a factor of 2 when multiplied in W .

$$\begin{aligned} & \{\text{vec}[(I_n - AN^{-1}A^T P)P^{-1}]\}^T W = \\ &= \frac{1}{2} \{\text{vec}[(I_n - PAN^{-1}A^T)P(I_n - AN^{-1}A^T P)P^{-1}P(I_n - AN^{-1}A^T P)]\}^T = \\ &= \frac{1}{2} [\text{vec}(P - PAN^{-1}A^T P)]^T = \{\text{vec}[(I_n - AN^{-1}A^T P)P^{-1}]\}^T W \quad (4.22a) \end{aligned}$$

Now we substitute (4.22a) into the right side of (4.21) to arrive at

$$\begin{aligned} & \{\text{vec}[(I_n - AN^{-1}A^T P)P^{-1}]\}^T W(\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}) = \frac{1}{2} [\text{vec}(P - PAN^{-1}A^T P)]^T (\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}) = \\ &= \frac{1}{2} \tilde{\mathbf{e}}^T (P - PAN^{-1}A^T P) \tilde{\mathbf{e}} = \frac{1}{2} \tilde{\mathbf{e}}^T P \tilde{\mathbf{e}}, \quad (4.22b) \end{aligned}$$

since $A^T P \tilde{\mathbf{e}} = \mathbf{0}$ according to (1.9b). Next we substitute (4.22a) into the left side of (4.21).

$$\begin{aligned} & \{\text{vec}[(I_n - AN^{-1}A^T P)P^{-1}]\}^T W \text{vec}[(I_n - AN^{-1}A^T P)P^{-1}] \hat{\sigma}_0^2 = \\ &= \frac{1}{2} [\text{vec}(P - PAN^{-1}A^T P)]^T \text{vec}[(I_n - AN^{-1}A^T P)P^{-1}] \hat{\sigma}_0^2 = \end{aligned}$$

(Continuing with help of (A.2), where matrices A and C are identity in that equation)

$$\begin{aligned} &= \frac{1}{2} \text{tr}[(I_n - AN^{-1}A^T P)P^{-1}(P - PAN^{-1}A^T P)] \hat{\sigma}_0^2 = \\ &= \frac{1}{2} \text{tr}[(I_n - AN^{-1}A^T P)(I_n - AN^{-1}A^T P)] \hat{\sigma}_0^2 = \\ &= \frac{1}{2} \text{tr}[(I_n - AN^{-1}A^T P)] \hat{\sigma}_0^2 = \frac{1}{2} \text{tr}\{(I_n - N^{-1}A^T P A)\} \hat{\sigma}_0^2 = \\ &= \frac{1}{2} (n - m) \hat{\sigma}_0^2. \quad (4.22c) \end{aligned}$$

Finally, we equate the left side (4.22b) and right side (4.22c) to obtain

$$\frac{1}{2} \tilde{\mathbf{e}}^T P \tilde{\mathbf{e}} = \frac{1}{2} (n - m) \hat{\sigma}_0^2, \quad (4.22d)$$

resulting in the following estimate for the variance component σ_0^2 :

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\mathbf{e}}^T P \tilde{\mathbf{e}}}{(n - m)}}. \quad (4.23)$$

We conclude that LESS for the model (4.13) is equivalent to BIQUUE in the GMM (3.1), which is evident from the respective formulas for the estimated variance component (4.23) and (3.18).

We can also use E-D correspondence to derive the *Best Linear Estimate*, BLE, of σ_0^2 . This is done by expressing the estimate $\hat{\sigma}_0^2$ as a linear function of $\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}$ and minimizing its MSE. The solution is equivalent to the BIQE mentioned in Chapter 3. The problem is setup below.

- Linear requirement:

$$\hat{\sigma}_0^2 = L^T(\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}). \quad (4.24)$$

- Best requirement:

$$\begin{aligned} \text{MSE}\{\hat{\sigma}_0^2\} &= D\{\hat{\sigma}_0^2\} + (E\{\hat{\sigma}_0^2\} - \sigma_0^2)^2 = \\ &= L^T D\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\} + L^T E\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\} E\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\}^T L - 2\sigma_0^2 L^T E\{\tilde{\mathbf{e}} \otimes \tilde{\mathbf{e}}\} + (\sigma_0^2)^2 = \min_L. \end{aligned} \quad (4.25)$$

- Solution:

$$\hat{\sigma}_0^2 = \frac{\tilde{\mathbf{e}}^T P \tilde{\mathbf{e}}}{n - m + 2} = \text{BLE}\{\sigma_0^2\} = \text{BIQE}\{\sigma_0^2\}. \quad (4.26)$$

Chapter 5

The Rank-Deficient Gauss-Markov Model

The rank-deficient Gauss-Markov Model (GMM) describes the case where the coefficient matrix A (also called design matrix or information matrix) does not have full column rank. As usual we speak of n observations and m parameters so that the (linearized) observation vector \mathbf{y} is of size $n \times 1$, while matrix A is of size $n \times m$. The model is stated as follows:

$$\mathbf{y} = A\boldsymbol{\xi} + \mathbf{e}, \quad \mathbf{e} \sim (\mathbf{0}, \Sigma = \sigma_0^2 P^{-1}), \quad \text{rk } A =: q < \min\{m, n\}. \quad (5.1)$$

The familiar least-squares normal equations are written as

$$N\hat{\boldsymbol{\xi}} = \mathbf{c}, \quad (5.2a)$$

where

$$[N, \mathbf{c}] := A^T P [A, \mathbf{y}]. \quad (5.2b)$$

The ranks of the $m \times m$ normal-equations matrix N and the $n \times m$ coefficient matrix A are related by

$$\text{rk } N = \dim \mathcal{R}(N) \leq \dim \mathcal{R}(A^T) = \text{rk } A^T = \text{rk } A = q < m, \quad (5.3)$$

where the symbol \mathcal{R} stands for range space (also called column space or kernel).

Here we have assumed that $m \leq n$, meaning that it is not necessarily a lack of observations that gives rise to the rank deficiency but that the system of observation equations does not carry enough information about the parameters to estimate all of them. In terms of the columns of matrix A , it can be said that only q of them are linearly independent and that each of the remaining $m - q$ of them can be expressed as a linear combination of the q independent ones.

The less-than-or-equals sign in (5.3) denotes a more general relationship than what is needed here. We may change it to the equality sign since the weight matrix P is positive definite, which means $\dim \mathcal{R}(N) = \dim \mathcal{R}(A^T)$. Therefore,

$$\text{rk } N = \text{rk } A = q < m. \quad (5.4)$$

Likewise, we can make a statement about the range spaces of matrices N and A^T as follows:

$$\mathcal{R}(N) \subset \mathcal{R}(A^T) \text{ and } \dim \mathcal{R}(N) = \text{rk } N = \text{rk } A^T = \dim \mathcal{R}(A^T) \Leftrightarrow \quad (5.5a)$$

$$\mathcal{R}(N) = \mathcal{R}(A^T). \quad (5.5b)$$

Question: Do solutions for $\hat{\xi}$ always exist? Yes, because

$$\mathbf{c} := A^T P \mathbf{y} \in \mathcal{R}(A^T) = \mathcal{R}(N). \quad (5.6)$$

In other words, the vector \mathbf{c} is in the range (column) space of N , which guarantees that we have solutions for $N\hat{\xi} = \mathbf{c}$.

Question: How many solutions for $\hat{\xi}$ are there and how do we represent them? The general solution $\hat{\xi}$ belongs to a solution hyperspace that is shifted out of the origin by a particular solution $\hat{\xi}_{\text{part}}$, where $\hat{\xi}_{\text{part}}$ is a solution to an inhomogeneous system of equations. Running parallel to the set of all particular solutions, and through the origin of the solution hyperspace, is the nullspace of N , which is comprised of all the solutions to the homogeneous system $N\hat{\xi} = \mathbf{0}$. Therefore we can write the general solution, as the sum of the particular solutions and the nullspace of N , denoted $\mathcal{N}(N)$, as in

$$\hat{\xi} = \hat{\xi}_{\text{part}} + \mathcal{N}(N). \quad (5.7)$$

Figure 5.1 shows a graphical representation of equation (5.7). Obviously there are infinite choices for the particular solution and thus infinitely many solutions for the unknown parameters within the rank deficient GMM.

Symbolically, we characterize the nullspaces of N and A by

$$\mathcal{N}(N) := \{\boldsymbol{\alpha} \mid N\boldsymbol{\alpha} = \mathbf{0}\} \text{ and } \mathcal{N}(A) := \{\boldsymbol{\alpha} \mid A\boldsymbol{\alpha} = \mathbf{0}\}, \quad (5.8)$$

respectively. When $A\boldsymbol{\alpha} = \mathbf{0}$ so does $N\boldsymbol{\alpha}$; therefore

$$\mathcal{N}(N) \subset \mathcal{N}(A). \quad (5.9a)$$

Also

$$\dim \mathcal{N}(N) = \dim \mathcal{N}(A) = m - q, \quad (5.9b)$$

since

$$\mathcal{N}(A) \overset{\perp}{\oplus} \mathcal{R}(A^T) = \mathbb{R}^m, \quad (5.9c)$$

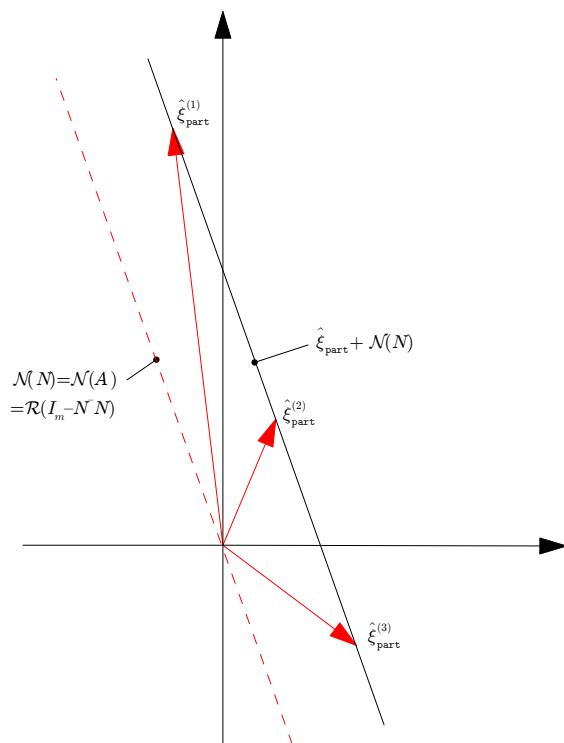
which says that the nullspace of matrix A and the range space of A^T are both complimentary and orthogonal subspaces of one another. Because of (5.9a) and (5.9b), we can state that

$$\mathcal{N}(N) = \mathcal{N}(A), \quad (5.9d)$$

allowing us to extend (5.7) to

$$\hat{\xi} = \hat{\xi}_{\text{part}} + \mathcal{N}(N) = \hat{\xi}_{\text{part}} + \mathcal{N}(A). \quad (5.10)$$

Thus we can generate all solutions $\hat{\xi}$ if we know how to find a particular solution $\hat{\xi}_{\text{part}}$ and if we know how to generate the nullspace of matrix A (or the nullspace of N). To find the nullspace, we must turn to the topic of generalized inverses.

Figure 5.1: A schematic representation of the solution space for $\hat{\xi}$

5.1 Generalized Inverses

Generalized inverses (g-inverses) are important for solving systems of equations that have singular coefficient matrices. Let G be the g-inverse of matrix N (with both G and N of size $m \times m$), then

$$NGN = N, \quad (5.11a)$$

implying that

$$N(I_m - GN) = 0, \quad (5.11b)$$

which further implies that

$$\mathcal{R}(I_m - GN) \subset \mathcal{N}(N). \quad (5.11c)$$

Question: are the two spaces shown in (5.11c) equivalent? The answer is yes, as shown below.

The matrix $(I_m - GN)$ is idempotent since

$$(I_m - GN)(I_m - GN) = (I_m - GN) - GN(I_m - GN) = (I_m - GN). \quad (5.12a)$$

Therefore $\text{rk}(I_m - GN) = \text{tr}(I_m - GN)$, and the dimension of the range space is

$$\dim \mathcal{R}(I_m - GN) = \text{rk}(I_m - GN) = \text{tr}(I_m - GN) = m - \text{tr}(GN). \quad (5.12b)$$

But, GN itself is also idempotent; therefore:

$$\begin{aligned} \dim \mathcal{R}(I_m - GN) &= m - \text{tr}(GN) = m - \dim \mathcal{R}(GN) = m - q = \\ &= m - \dim \mathcal{R}(N) = \dim \mathcal{N}(N). \end{aligned} \quad (5.12c)$$

Because of (5.11c) and (5.12c), we can indeed say that the spaces in (5.11c) are equivalent, i.e.,

$$\mathcal{R}(I_m - GN) = \mathcal{N}(N). \quad (5.12d)$$

Let the g-inverse matrix G be represented by the symbol N^- , then

$$\mathcal{R}(I_m - N^-N) = \mathcal{N}(N) \text{ for any } N^- \text{ of } N. \quad (5.13)$$

Using (5.13) together with (5.10), we are now ready to write the complete solution space of $\hat{\xi}$ as

$$\hat{\xi} = \hat{\xi}_{\text{part}} + (I_m - N^-N)\mathbf{z} \text{ for any } \mathbf{z} \in \mathbb{R}^m \text{ and any chosen g-inverse } N^-. \quad (5.14)$$

As an aside, we show the dimension of the range space of the idempotent matrix GN used in (5.12c). The rank of a product of matrices must be less than or equal to the rank of any factor. Therefore:

$$\text{rk}(NGN) \leq \text{rk}(GN) \leq \text{rk } N = \text{rk}(NGN) \quad (5.15a)$$

implying that

$$\text{rk}(GN) = \text{rk } N = q. \quad (5.15b)$$

5.2 Finding a Generalized Inverse

Note the following properties associated with the normal-equations matrix N and its g-inverse G :

1. GN is idempotent (and as such, is a projection matrix), and so is NG . That is, $GN \cdot GN = GN$, and $NG \cdot NG = NG$.
2. $I_m - GN$ is idempotent.
3. $m \geq \text{rk } G \geq \text{rk } N = q$.

Item 3 states that the g-inverse G of N will always have equal or greater rank than that of N itself.

An important subclass of g-inverses is the *reflexive g-inverse*. If G is a g-inverse of N , and if N is also a g-inverse of G , then we say that G is a reflexive g-inverse of N . Considering item 3 above, if $GNG = G$, then $\text{rk } N = q \geq \text{rk } G \Rightarrow \text{rk } G = q$. So, if we are given a g-inverse of rank q , it must be reflexive.

Another important g-inverse subclass is the Moore-Penrose inverse¹, which is also called the *pseudoinverse*. If the following four conditions are met, then the g-inverse G is the pseudoinverse of N denoted as N^+ .

$$\left. \begin{array}{l} NGN = N \\ GNG = G \\ NG = (NG)^T \\ GN = (GN)^T \end{array} \right\} \Leftrightarrow G = N^+ \quad (5.16)$$

The pseudoinverse is unique, and if N has full rank ($\text{rk } N = m$), the pseudoinverse is the same as the regular matrix inverse N^{-1} .

Note that g-inverses of N do not need to be symmetric. However, if G is a g-inverse of N , then G^T is as well. Proof:

$$(NGN)^T = N^T = N = N^T G^T N^T = NG^T N. \quad (5.17)$$

We also note that the pseudoinverse of a symmetric matrix is itself symmetric, and that $N^+ = N^+ N (N^+)^T$ is positive semidefinite (assuming N is singular). However, as already stated, an arbitrary g-inverse G might not be symmetric and also might not be positive semidefinite. However a reflexive symmetric (and therefore positive semidefinite) g-inverse defined as

$$N_{rs}^- = GNG^T \quad (5.18a)$$

is characterized by

$$\text{rk } N_{rs}^- = \text{rk } N = q \quad (5.18b)$$

and

$$N_{rs}^- = N_{rs}^- N (N_{rs}^-)^T = N_{rs}^- N N_{rs}^-. \quad (5.18c)$$

The g-inverse N_{rs}^- is in the class of reflexive symmetric g-inverses, which is a very important class for the work that follows. We note that a reflexive symmetric g-inverse can always be found from a given arbitrary g-inverse N^- by

$$N_{rs}^- = N^- N (N^-)^T. \quad (5.19)$$

There are many ways to construct a g-inverse of N . We show several examples below. In some of the examples we use a more generic symbol A in order to stress that the matrix does not have to be symmetric. For the discussion that follows, it is helpful to partition N so that the upper-left $q \times q$ block matrix N_{11} has rank q as follows:

$$N_{m \times m} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}, \quad \dim N_{11} = q \times q, \quad \text{rk } N_{11} = q = \text{rk } N. \quad (5.20)$$

¹According to Cross (1985), this g-inverse was first discovered by Moore in 1920 and then independently by Penrose in 1955.

The equations in (5.20) imply that the second column of the partitioned matrix is a linear combination of the first column. Therefore, for some $q \times (m - q)$ matrix L , we have

$$\begin{bmatrix} N_{12} \\ N_{22} \end{bmatrix} = \begin{bmatrix} N_{11} \\ N_{21} \end{bmatrix} \cdot L, \text{ or } N_{12} = N_{11}L \text{ and } N_{22} = N_{21}L. \quad (5.21)$$

In practice, the rows and columns of N might have to be reordered to ensure that N_{11} is full rank as shown in (5.20), but that is usually easy to do. Also note that since N is positive semidefinite, it can be decomposed as follows:

$$N = \begin{bmatrix} U^T D U & U^T D H \\ H^T D U & H^T D H \end{bmatrix} = \begin{bmatrix} U^T \\ H^T \end{bmatrix} D \begin{bmatrix} U & H \end{bmatrix} = \begin{bmatrix} (D^{1/2} U)^T \\ (D^{1/2} H)^T \end{bmatrix} \begin{bmatrix} D^{1/2} U & D^{1/2} H \end{bmatrix}. \quad (5.22)$$

Here, U is an upper triangular matrix of size $q \times q$; H is size $q \times (m - q)$, and D is a $q \times q$ diagonal matrix. Also note that $(D^{1/2} U)^T$ is the Cholesky factor of N_{11} .

g-inverse example 1: (with N defined as in (5.22))

$$N^- = \begin{bmatrix} U^T \\ H^T \end{bmatrix} (U U^T + H H^T)^{-1} D^{-1} (U U^T + H H^T)^{-1} \begin{bmatrix} U & H \end{bmatrix} \quad (5.23)$$

Check:

$$\begin{aligned} N N^- &= \begin{bmatrix} U^T \\ H^T \end{bmatrix} D \begin{bmatrix} U & H \end{bmatrix} \begin{bmatrix} U^T \\ H^T \end{bmatrix} (U U^T + H H^T)^{-1} D^{-1} (U U^T + H H^T)^{-1} \begin{bmatrix} U & H \end{bmatrix} = \\ &= \begin{bmatrix} U^T \\ H^T \end{bmatrix} (U U^T + H H^T)^{-1} \begin{bmatrix} U & H \end{bmatrix} \Rightarrow \\ N N^- N &= \begin{bmatrix} U^T \\ H^T \end{bmatrix} (U U^T + H H^T)^{-1} \begin{bmatrix} U & H \end{bmatrix} \begin{bmatrix} U^T \\ H^T \end{bmatrix} D \begin{bmatrix} U & H \end{bmatrix} = \\ &= \begin{bmatrix} U^T \\ H^T \end{bmatrix} D \begin{bmatrix} U & H \end{bmatrix} = N \end{aligned}$$

Since this g-inverse N^- has rank q , it is reflexive. Obviously it is also symmetric. Therefore, it could also be labeled N_{rs}^- . In this case it also satisfies all the properties of a pseudoinverse.

g-inverse example 2: (see Lütkepohl (1996), section 9.12.3, item (3))

$$N^- = \begin{bmatrix} (D^{1/2} U)^T \\ (D^{1/2} H)^T \end{bmatrix} \left(\begin{bmatrix} D^{1/2} U & D^{1/2} H \end{bmatrix} \begin{bmatrix} (D^{1/2} U)^T \\ (D^{1/2} H)^T \end{bmatrix} \right)^{-2} \begin{bmatrix} D^{1/2} U & D^{1/2} H \end{bmatrix} = N^+ \quad (5.24)$$

The properties (5.16) can be verified for (5.24) after some tedious matrix multiplication.

g-inverse example 3:

$$N^- = \begin{bmatrix} N_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} = N_{rs}^{-1} \quad (5.25)$$

Check:

$$\begin{aligned} NN^-N &= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} N_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ N_{21}N_{11}^{-1} & 0 \end{bmatrix} \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} = \\ &= \begin{bmatrix} N_{11} & N_{12} \\ N_{21}N_{11}^{-1}N_{11} & N_{21}N_{11}^{-1}N_{12} \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{21}N_{11}^{-1}N_{12} \end{bmatrix} \end{aligned}$$

From (5.21) we have $N_{12} = N_{11}L$ and $N_{22} = N_{21}L$ so that $N_{21}N_{11}^{-1}N_{12} = N_{21}L = N_{22}$, which completes the check. Reflexivity is easy to check also.

g-inverse example 4: By rank factorization, the $n \times m$ matrix A may be factored into the product of an $n \times q$ matrix F and a $q \times m$ matrix H , where $\text{rk } A = \text{rk } F = \text{rk } H = q$ and $A = FH$. Then a reflexive g-inverse of A may be obtained by

$$A_r^- = H^T(HH^T)^{-1}(F^T F)^{-1}F^T. \quad (5.26)$$

5.3 The Singular Value Decomposition

Given a matrix A of size $n \times m$ and $\text{rk } A = q$, the singular values of A are the positive square roots of the positive eigenvalues of $A^T A$ or AA^T , which are square, symmetric matrices with real eigenvalues. (Note that only the positive eigenvalues of the matrix products $A^T A$ and AA^T are the same.) Let the diagonal $n \times m$ matrix Λ contain q non-zero elements, being the singular values λ_j of A where $j = 1, \dots, q$. Let U be the orthogonal $n \times n$ matrix whose columns are the eigenvectors of AA^T , and let V be the orthogonal $m \times m$ matrix whose columns are the eigenvectors of $A^T A$. Then

$$A = U\Lambda V^T \quad (5.27)$$

is the *Singular Value Decomposition* (SVD) of matrix A . Note that if A is symmetric, $U = V$. The g-inverse examples 5 through 8 below are all based on the SVD.

g-inverse example 5: Define a $q \times q$ diagonal matrix as

$$\Delta^{-1} = \text{diag}(\lambda_j), \quad j = 1, \dots, q. \quad (5.28)$$

$$\{A^-\} = \left\{ V \begin{bmatrix} \Delta^{-1} & K \\ L & M \end{bmatrix} U^T \mid K, L, M \text{ arbitrary with suitable size} \right\} \quad (5.29)$$

The rank of the block matrix can vary between q and m depending on the choices for K , L , and M .

g-inverse example 6:

$$\{A_r^-\} = \left\{ V \begin{bmatrix} \Delta^{-1} & K \\ L & L\Delta K \end{bmatrix} U^T \mid K, L \text{ arbitrary with suitable size} \right\} \quad (5.30)$$

g-inverse example 7:

$$\{A_{rs}^-\} = \left\{ V \begin{bmatrix} \Delta^{-1} & L^T \\ L & L\Delta L^T \end{bmatrix} U^T \mid L \text{ arbitrary with suitable size} \right\} \quad (5.31)$$

g-inverse example 8:

$$A^+ = V \begin{bmatrix} \Delta^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T \quad (5.32)$$

g-inverse example 9: Zlobec's formula for the pseudoinverse is

$$N^+ = N(NNN)^-N, \quad (5.33)$$

where the g-inverse can be any g-inverse of N^3 . The invariance of N^+ with respect to the choice of the g-inverse in Zlobec's formula is due to the g-inverse's placement between the two occurrences of matrix N . Again we note that the pseudoinverse is unique, but there are a variety of ways to generate it.

Now that we have seen how to generate a g-inverse, the next question regarding our general solution (5.14) is "how do we represent the particular solution $\hat{\xi}_{\text{part}}$?" We claim that $\hat{\xi}_{\text{part}}$ is represented by N^-c (or equivalently $N^-A^T P y$) since $NN^-A^T P y = c$. To validate this claim, we must show that

$$NN^-A^T = A^T, \quad (5.34a)$$

which is done in the following: Because N begins with A^T , we can write

$$\mathcal{R}(NN^-A^T) \subset \mathcal{R}(A^T). \quad (5.34b)$$

Furthermore,

$$\dim \mathcal{R}(NN^-A^T) = \text{rk}(NN^-A^T) \geq \text{rk}([NN^-A^T]PA) = \quad (5.34c)$$

$$= \text{rk} NN^-N = \text{rk} N = q \geq \text{rk}(NN^-A^T), v \quad (5.34d)$$

or

$$\text{rk}(NN^-A^T) \geq q \geq \text{rk}(NN^-A^T), v \quad (5.34e)$$

implying that

$$\text{rk}(NN^-A^T) = q, \text{ but also } \dim \mathcal{R}(A^T) = q, \quad (5.34f)$$

which further implies

$$\mathcal{R}(NN^-A^T) = \mathcal{R}(A^T). \quad (5.34g)$$

Thus, we conclude that

$$NN^{-1}A^T = A^T, \quad (5.35a)$$

and, after transposing,

$$AN^{-1}N = A. \quad (5.35b)$$

Therefore it follows that

$$(NN^{-1}A^T)P\mathbf{y} = A^T P\mathbf{y} = \mathbf{c}. \quad (5.35c)$$

We can now write our general solution (5.14) in terms of $N^{-1}\mathbf{c}$ as follows:

$$\hat{\boldsymbol{\xi}} = \{N^{-1}\mathbf{c} + (I_m - N^{-1}N)\boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathbb{R}^m\} \text{ for any chosen g-inverse } N^{-1}, \quad (5.36)$$

where $\boldsymbol{\alpha}$ is an arbitrary, but non-random, $m \times 1$ vector.

From the law of error propagation, we find the dispersion of the general solution to be

$$D\{\hat{\boldsymbol{\xi}}\} = D\{N^{-1}\mathbf{c} + (I_m - N^{-1}N)\boldsymbol{\alpha}\} = N^{-1}D\{\mathbf{c}\}(N^{-1})^T = \sigma_0^2 N^{-1}N(N^{-1})^T = \sigma_0^2 N_{rs}^{-1}. \quad (5.37)$$

We now verify that the dispersion matrix in (5.37) is indeed a reflexive symmetric g-inverse.

1. Obviously the dispersion matrix $\sigma_0^2 N^{-1}N(N^{-1})^T$ is symmetric.
2. The matrix $N^{-1}N(N^{-1})^T$ is a g-inverse of N because:

$$N[N^{-1}N(N^{-1})^T]N = (NN^{-1}N)(N^{-1})^T N = N(N^{-1})^T N = N,$$

recalling that if N^{-1} is a g-inverse of N , so is $(N^{-1})^T$.

3. The matrix N is a g-inverse of $N^{-1}N(N^{-1})^T$ because:

$$\begin{aligned} [N^{-1}N(N^{-1})^T]N[N^{-1}N(N^{-1})^T] &= N^{-1}(N(N^{-1})^T N)[N^{-1}N(N^{-1})^T] = \\ &= N^{-1}N[N^{-1}N(N^{-1})^T] = N^{-1}(NN^{-1}N)(N^{-1})^T = N^{-1}N(N^{-1})^T. \end{aligned}$$

Because the dispersion in (5.37) is represented by a reflexive symmetric g-inverse of N , we may, without loss of generality, restrict ourselves to reflexive symmetric g-inverses in our search for a general solution $\hat{\boldsymbol{\xi}}$.

We have infinite choices for our particular solution $\hat{\boldsymbol{\xi}}_{\text{part}}$, but one of particular interest is that which is shortest in magnitude (i.e., smallest L_2 -norm). This particular solution can be derived by imposing a minimum norm condition on the parameter vector in the least-squares target function, and it is thus called MINimum Norm LEast-Squares Solution (MINOLESS).

5.4 Minimum Norm Least-Squares Solution

MINOLESS is an acronym for Minimum Norm LEast-Squares Solution. We know that, within the rank deficient GMM, $N\hat{\boldsymbol{\xi}} = \mathbf{c}$ has many solutions; we seek the shortest (minimum norm) of these. The idea is to minimize the norm (inner product), of $\hat{\boldsymbol{\xi}}$, according to

$$\boldsymbol{\xi}^T \boldsymbol{\xi} = \min_{\boldsymbol{\xi}} \text{ such that } N\boldsymbol{\xi} = \mathbf{c}. \quad (5.38)$$

Thus, if $\boldsymbol{\xi}$ is an “incremental” parameter vector, as it is under linearization, minimum norm means minimum change from the initial vector $\boldsymbol{\xi}^0$, e.g., the initial Taylor series expansion point.

The Lagrange target function to minimize is written as

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}) := \boldsymbol{\xi}^T \boldsymbol{\xi} + 2\boldsymbol{\lambda}^T (N\boldsymbol{\xi} - \mathbf{c}), \quad (5.39)$$

which must be made stationary for $\boldsymbol{\xi}$ and $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is an $m \times 1$ vector of Lagrange multipliers. Accordingly, the Euler-Lagrange (first-order) necessary conditions are then written as

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}^T + \hat{\boldsymbol{\lambda}}^T N \doteq \mathbf{0}, \quad (5.40a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \hat{\boldsymbol{\xi}}^T N - \mathbf{c}^T \doteq \mathbf{0}. \quad (5.40b)$$

The sufficient condition (i.e., that second partial derivative must be positive) is satisfied since $(1/2)\partial^2\Phi/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T = I_n$, which is positive definite.

Equations (5.40a) and (5.40b) lead to the solution $\hat{\boldsymbol{\xi}}$ as follows:

$$\hat{\boldsymbol{\xi}} = -N\hat{\boldsymbol{\lambda}} \text{ and } N\hat{\boldsymbol{\xi}} = \mathbf{c},$$

implying that

$$\mathbf{c} = -NN\hat{\boldsymbol{\lambda}} \Rightarrow \hat{\boldsymbol{\lambda}} = -(NN)^-\mathbf{c}.$$

Finally, we arrive at the solution for the unknown parameters $\boldsymbol{\xi}$ as

$$\boxed{\hat{\boldsymbol{\xi}} = N(NN)^-\mathbf{c}}. \quad (5.41)$$

Note that $N(NN)^-$ is a particular g-inverse of N , but also note that $(NN)^- \neq N^-N^-$. Equation (5.41) is one expression of MINOLESS. There are others, as will be shown later. Using variance propagation, the dispersion of $\hat{\boldsymbol{\xi}}$ is given by

$$D\{\hat{\boldsymbol{\xi}}\} = N(NN)^- \cdot D\{\mathbf{c}\} \cdot (NN)^-N = \sigma_0^2 N(NN)^-N(NN)^-N, \quad (5.42)$$

which implies that

$$N(NN)^-N(NN)^-N = N_{rs}^-. \quad (5.43)$$

Here, we have used the fact that any power of the symmetric matrix N is also symmetric, and the g-inverse of a symmetric matrix is also symmetric. The matrix

scaled by σ_0^2 in (5.42) is called *cofactor matrix*. We can always express a LESS as a product of such a cofactor matrix and the normal equation vector \mathbf{c} . That is,

$$\sigma_0^2 \cdot \hat{\boldsymbol{\xi}} = D\{\hat{\boldsymbol{\xi}}\} \cdot \mathbf{c} \text{ for any } \hat{\boldsymbol{\xi}}. \quad (5.44)$$

Therefore, we can also express MINOLESS in terms of the matrix N_{rs}^- in (5.43) by writing

$$\hat{\boldsymbol{\xi}} = [N(NN)^-N(NN)^-N]\mathbf{c}. \quad (5.45)$$

Here, we note that the product AN^-A^T is invariant with respect to the chosen g-inverse N^- . Also, not only is $N(NN)^-$ a g-inverse of N , according to (5.17) its transpose $(NN)^-N$ is also a g-inverse of N . Based on these relations, and by expressing (5.45) alternatively as

$$\hat{\boldsymbol{\xi}} = [A^T P A (NN)^- N (NN)^- N] A^T P \mathbf{y}, \quad (5.46)$$

it is seen that (5.45) is unique regardless of the choice of the g-inverse.

Typically, in geodetic science applications the estimated parameter vector $\hat{\boldsymbol{\xi}}$ is a vector of corrections to initial approximations (non-linear case). As noted above, using MINOLESS in this case guarantees that changes from the initial approximations are a minimum, in terms of the L_2 -norm. This minimum-norm solution is shown schematically in Figure 5.2.

5.5 Partial Minimum Norm Least-Squares Solution

In some cases we may only want a certain subset of the initial parameter vector to change in a minimum-norm sense. For example, we may know the locations of some geodetic network points to a high level of accuracy, while locations of the remaining network points may not be known as well or may even be known only approximately. In this case, we may wish to employ partial-MINOLESS, which is based on using a selection matrix to choose a subset of the parameters for norm minimization.

The minimization problem is then stated as

$$\hat{\boldsymbol{\xi}}^T S \hat{\boldsymbol{\xi}} = \min_{\hat{\boldsymbol{\xi}}} \{N \hat{\boldsymbol{\xi}} = \mathbf{c}\}, \quad S := \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix}. \quad (5.47)$$

The size of the identity matrix I_s corresponds to the number of selected parameters. Note that we can always construct S with I_s in the upper-left block, as shown in (5.47), by reordering the parameter vector if necessary.

The Lagrange target function to minimize in this case is given by

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}) := \boldsymbol{\xi}^T S \boldsymbol{\xi} - 2\boldsymbol{\lambda}^T (N\boldsymbol{\xi} - \mathbf{c}), \quad (5.48)$$

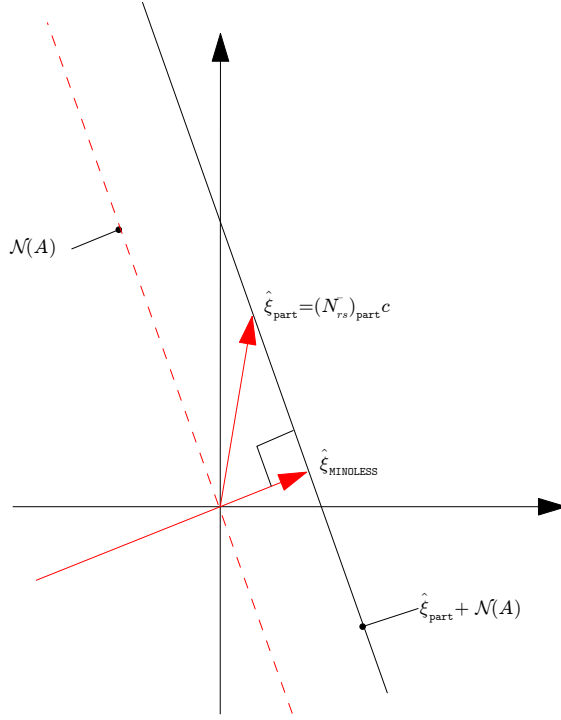


Figure 5.2: Schematic representation of the solution space for MINOLESS

which must be made stationary with respect to ξ and λ . Accordingly, The Euler-Lagrange necessary conditions are written as

$$\frac{1}{2} \frac{\partial \Phi}{\partial \xi} = \hat{\xi}^T S - \hat{\lambda}^T N \doteq \mathbf{0} \Rightarrow S \hat{\xi} - N \hat{\lambda} \doteq \mathbf{0}, \quad (5.49a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \lambda} = \hat{\xi}^T N - \mathbf{c}^T \doteq \mathbf{0} \Rightarrow N \hat{\xi} - \mathbf{c} \doteq \mathbf{0}. \quad (5.49b)$$

The sufficient condition for minimization is satisfied since $(1/2) \partial^2 \Phi / \partial \xi \partial \xi^T = S$, which is positive (semi) definite.

Obviously matrix S is singular, but we choose S so that $(S + N)$ is invertible, requiring that S selects at least $m - \text{rk } A$ parameters, or equivalently requiring $\text{rk } S \geq m - q$. We solve the above system of normal equations as follows:

By adding equations (5.49a) and (5.49b) we obtain

$$(S + N) \hat{\xi} = N \hat{\lambda} + \mathbf{c}, \quad (5.50a)$$

leading to

$$\hat{\xi} = (S + N)^{-1} (N \hat{\lambda} + \mathbf{c}). \quad (5.50b)$$

Substituting (5.50b) into (5.49b) yields

$$N(S + N)^{-1}(N\hat{\lambda} + \mathbf{c}) - \mathbf{c} = \mathbf{0},$$

or

$$\begin{aligned} N(S + N)^{-1}N\hat{\lambda} &= \mathbf{c} - N(S + N)^{-1}\mathbf{c} = \\ &= [(S + N) - N](S + N)^{-1}\mathbf{c} = S(S + N)^{-1}\mathbf{c}, \end{aligned}$$

leading to

$$\hat{\lambda} = [N(S + N)^{-1}N]^{-1}S(S + N)^{-1}\mathbf{c} \quad (5.50c)$$

as an estimate for the vector of Lagrange multipliers. Then, substituting (5.50c) into (5.50a) yields

$$(S + N)\hat{\xi} = \mathbf{c} + N[N(S + N)^{-1}N]^{-1}S(S + N)^{-1}\mathbf{c}. \quad (5.50d)$$

We use the identity $NN^{-}\mathbf{c} = \mathbf{c}$ to write an equivalent equation

$$\begin{aligned} (S + N)\hat{\xi} &= N[N(S + N)^{-1}N]^{-1}N(S + N)^{-1}\mathbf{c} + \\ &\quad + N[N(S + N)^{-1}N]^{-1}S(S + N)^{-1}\mathbf{c} = \\ &= N[N(S + N)^{-1}N]^{-1}[N(S + N)^{-1} + S(S + N)^{-1}]\mathbf{c} = \\ &= N[N(S + N)^{-1}N]^{-}\mathbf{c}, \end{aligned} \quad (5.50e)$$

finally leading to the partial-MINOLESS

$$\hat{\xi} = \hat{\xi}_{\text{P-MINOLESS}} = (S + N)^{-1}N[N(S + N)^{-1}N]^{-}\mathbf{c}. \quad (5.50f)$$

Using the law of covariance propagation, we write the partial-MINOLESS dispersion matrix as

$$D\{\hat{\xi}\} = \sigma_0^2(S + N)^{-1}N[N(S + N)^{-1}N]^{-}N[N(S + N)^{-1}N]^{-}N(S + N)^{-1}. \quad (5.51)$$

We may rewrite the partial-MINOLESS solution, replacing the matrix multiplying \mathbf{c} in (5.50f) with the cofactor matrix appearing in the dispersion (5.51), resulting in

$$\hat{\xi}_{\text{P-MINOLESS}} = (S + N)^{-1}N[N(S + N)^{-1}N]^{-}N[N(S + N)^{-1}N]^{-}N(S + N)^{-1}\mathbf{c}. \quad (5.52)$$

Now, what happens if we replace S by I_m , i.e., all parameters are selected for norm minimization? Obviously partial-MINOLESS becomes MINOLESS itself as shown below.

$$\begin{aligned} \hat{\xi}_{I_m\text{-MINOLESS}} &= (I_m + N)^{-1}N[N(I_m + N)^{-1}N]^{-}N \cdot \\ &\quad \cdot [N(I_m + N)^{-1}N]^{-}N(I_m + N)^{-1}\mathbf{c} = \\ &= (I_m + N)^{-1}N[N(I_m + N)^{-1}N]^{-}\mathbf{c} \Rightarrow \\ &\quad \hat{\xi}_{\text{MINOLESS}} = N^+\mathbf{c} \end{aligned} \quad (5.53)$$

The dispersion for MINOLESS is computed by

$$D\{\hat{\boldsymbol{\xi}}_{\text{MINOLESS}}\} = D\{N^+ \mathbf{c}\} = N^+ D\{\mathbf{c}\} N^+ = \sigma_0^2 N^+. \quad (5.54)$$

It is interesting to compare (5.53) with the form of MINOLESS found earlier in (5.41). Once again we note that regardless of the form of MINOLESS (or similarly the form of N^+), the MINOLESS is unique. However, there is no connection between MINOLESS and BLUUE, as there is no unbiased estimate for this LESS. That is, the rank deficient GMM has no unbiased solution for the unknown parameters. This fact is easily demonstrated by attempting to derive a LUUE (Linear Uniformly Unbiased Estimate) as follows.

The Linear Uniformly Unbiased Estimate (LUUE) requires that

$$\hat{\boldsymbol{\xi}} = L\mathbf{y},$$

with L being an $m \times n$ matrix. Then

$$\boldsymbol{\xi} = E\{\hat{\boldsymbol{\xi}}\} = LE\{\mathbf{y}\} = LA\boldsymbol{\xi},$$

with the size of LA being $m \times m$ and $\text{rk}(LA) \leq \text{rk} A = q < m$. But, LA is singular, and therefore $LA \neq I_m$. Thus, LUUE = \emptyset ; i.e., there is no unbiased solution for $\boldsymbol{\xi}$.

5.6 Best Least-Squares Solution

We expect the least-squares solution to be *best* in a certain class. By best we mean that the trace of its dispersion matrix is minimum. We already found that the dispersion is based on a reflexive symmetric g-inverse, i.e.,

$$D\{\hat{\boldsymbol{\xi}}_{\text{LESS}}\} = \sigma_0^2 N_{rs}^- \text{ for all } N_{rs}^-. \quad (5.55)$$

Our task now is to compare the trace of the dispersion matrix from MINOLESS to that of a general LESS, recalling that the best LESS must satisfy the normal equations $N\boldsymbol{\xi} = \mathbf{c}$. We start by expressing the estimate as a linear combination of the observations as follows:

$$\hat{\boldsymbol{\xi}} = L\mathbf{y}, \text{ with } NL = A^T P, \quad (5.56a)$$

which implies that

$$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 LP^{-1}L^T, \quad (5.56b)$$

permitting us to write

$$\sigma_0^{-2} \text{tr} D\{\hat{\boldsymbol{\xi}}\} = \text{tr}(LP^{-1}L^T) = \min_{L^T} \{L^T N = PA\}. \quad (5.56c)$$

So we see that minimizing the trace of the dispersion matrix is tantamount to minimizing the $m \times n$ matrix L (under the specified conditions) since the weight matrix P is fixed. Analogous to (2.9), we make use of the vec operator and the Kronecker-Zehfuss product to form the following Lagrange target function, where $\mathbf{l} := \text{vec} L^T$:

$$\Phi(\mathbf{l}, \boldsymbol{\lambda}) := \mathbf{l}^T (I_m \otimes P^{-1}) \mathbf{l} + 2\boldsymbol{\lambda}^T [N \otimes I_n \mathbf{l} - \text{vec}(PA)], \quad (5.57)$$

which must be made stationary with respect to \mathbf{l} and $\boldsymbol{\lambda}$. To clarify the form of the target function, we note that the first product comes from applying (A.2) to $\text{tr}(LP^{-1}L^T) = \text{tr}(P^{-1}L^T I_n L)$. The term containing the Lagrange multiplier $\boldsymbol{\lambda}$ comes from the constraint $L^T N - PA = 0$ with application of the vec operator such that $\text{vec}(I_n L^T N) - \text{vec}(PA) = \mathbf{0}$, followed by the application of (A.1).

The Euler-Lagrange necessary conditions are

$$\frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{l}} = (I_m \otimes P^{-1}) \hat{\mathbf{l}} + (N \otimes I_n) \hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \quad (5.58a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = (N \otimes I_n) \hat{\mathbf{l}} - \text{vec}(PA) \doteq \mathbf{0}. \quad (5.58b)$$

The sufficient condition for minimization is satisfied since $(1/2) \cdot \partial^2 \Phi / (\partial \mathbf{l} \partial \mathbf{l}^T) = (I_n \otimes P^{-1})$ is positive definite. The system of normal equations is solved as follows:

$$\hat{\mathbf{l}} = -(I_m \otimes P^{-1})^{-1} (N \otimes I_n) \hat{\boldsymbol{\lambda}} = -(N \otimes P) \hat{\boldsymbol{\lambda}}, \text{ from (5.58a)}. \quad (5.59a)$$

And, by substituting the preceding equation into (5.58b),

$$\text{vec}(PA) = -(NN \otimes P) \hat{\boldsymbol{\lambda}}. \quad (5.59b)$$

We need $(N \otimes P) \hat{\boldsymbol{\lambda}}$; so we exploit the fact that $N(N^2)^{-1} N^2 = N$ by multiplying both sides of (5.59b) by $(N(NN)^{-1} \otimes I_n)$, which gives

$$\begin{aligned} (N(NN)^{-1} \otimes I_n) \text{vec}(PA) &= -(N(NN)^{-1} \otimes I_n) (NN \otimes P) \hat{\boldsymbol{\lambda}} = \\ &= (N(NN)^{-1} \otimes I_n) \text{vec}(PA) = -(N \otimes P) \hat{\boldsymbol{\lambda}} = \hat{\mathbf{l}}, \end{aligned}$$

leading to

$$\hat{\mathbf{l}} = \text{vec}[I_n PA(N(NN)^{-1})^T] = \text{vec}(PA(NN)^{-1} N) = \text{vec} L^T, \text{ (using (A.1))}, \quad (5.59c)$$

implying that

$$L^T = PA(NN)^{-1} N \Rightarrow L = N(NN)^{-1} A^T P. \quad (5.59d)$$

Finally, we substitute (5.59d) into (5.56a) to get (compare to (5.41))

$$\hat{\boldsymbol{\xi}} = \mathbf{L} \mathbf{y} = N(NN)^{-1} A^T P \mathbf{y} = N(NN)^{-1} \mathbf{c} = \hat{\boldsymbol{\xi}}_{\text{MINOLESS}}. \quad (5.59e)$$

Thus we find that MINOLESS is best among all LESS with minimum trace. How can we prove this directly? We start by showing that $(N^+ N) N_{rs}^- (NN^+) = N^+$ and then exploit this relationship in the proof that follows.

$$(N^+ N) N_{rs}^- (NN^+) = N^+ (N N_{rs}^- N) N^+ = N^+ N N^+ = N^+$$

implying that

$$\text{tr} N^+ = \text{tr} [(N^+ N) N_{rs}^- (NN^+)].$$

We continue by exploiting symmetry and applying a cyclic transformation.

$$\begin{aligned}
\text{tr } N^+ &= \text{tr}[(NN^+)^T(N^+N)N_{rs}^-] = \\
&= \text{tr}[N^+NN_{rs}^-] = \quad (\text{because } N^+N \text{ is idempotent}) \\
&= \text{tr}[N_{rs}^- - N_{rs}^- + N^+NN_{rs}^-] = \\
&= \text{tr}[N_{rs}^- - (I_m - N^+N)N_{rs}^-] = \\
&= \text{tr } N_{rs}^- - \text{tr}[(I_m - N^+N)(I_m - N^+N)N_{rs}^-] = \\
&\quad (\text{because } (I_m - N^+N) \text{ is idempotent}) \\
&= \text{tr } N_{rs}^- - \text{tr}[(I_m - N^+N)^T(I_m - N^+N)N_{rs}^-] = \\
&\quad (\text{transpose due to symmetry}) \\
&= \text{tr } N_{rs}^- - \text{tr}[(I_m - N^+N)N_{rs}^-(I_m - N^+N)^T] = \quad (\text{cyclic transformation})
\end{aligned}$$

But, the triple product is positive semi-definite; therefore we can state that

$$\text{tr } N^+ = \text{tr } N_{rs}^- - \text{tr}[(I_m - N^+N)N_{rs}^-(I_m - N^+N)^T] \leq \text{tr } N_{rs}^-. \quad (5.60)$$

Thus we have proved directly that the pseudoinverse N^+ provides a minimum trace in the class of cofactor matrices

$$\text{tr } N^+ \leq \text{tr } N_{rs}^- \text{ for all } N_{rs}^-. \quad (5.61)$$

Can we make a similar characterization of partial-MINOLESS? Is it *partially best* in terms of having the smallest (partial) trace of the cofactor matrix? In other words, is the sum of the cofactor diagonal elements corresponding to the selected points smallest? The answer is yes, as we show below.

Analogous to (5.56c) we minimize a subset of the trace of the parameter dispersion matrix as

$$\sigma_0^{-2} \text{tr}(SD\{\hat{\boldsymbol{\xi}}\}) = \min_{L^T} \text{ such that } L^T N = PA. \quad (5.62)$$

From (5.50f) we already know that matrix L should satisfy $L := (S+N)^{-1}N[N(S+N)^{-1}N]^{-1}A^T P$. Analogous to (5.57), we write the following Lagrange target function:

$$\Phi(\mathbf{l}, \boldsymbol{\lambda}) := \mathbf{l}^T(S \otimes P^{-1})\mathbf{l} + 2\boldsymbol{\lambda}^T[(N \otimes I_n)\mathbf{l} - \text{vec}(PA)], \quad (5.63)$$

which must be made stationary with respect to \mathbf{l} and $\boldsymbol{\lambda}$. Again we have defined $\mathbf{l} := \text{vec } L^T$. Accordingly, the Euler-Lagrange necessary conditions are written as (compare to (5.58a) and (5.58b))

$$\frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{l}} = (S \otimes P^{-1})\hat{\mathbf{l}} + (N \otimes I_n)\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \quad (5.64)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = (N \otimes I_n)\hat{\mathbf{l}} - \text{vec}(PA) \doteq \mathbf{0}. \quad (5.65)$$

Note that $(S \otimes P^{-1})$ is not invertible due to the singularity of S , in general. We solve the system of equations (5.64) and (5.65) as follows: Multiplying (5.64) by $(I_n \otimes P)$ gives

$$(S \otimes I_n)\hat{\mathbf{l}} + (N \otimes P)\hat{\boldsymbol{\lambda}} = \mathbf{0}. \quad (5.66a)$$

Adding this to (5.65) yields

$$[(N + S) \otimes I_n]\hat{\mathbf{l}} = \text{vec}(PA) - (N \otimes P)\hat{\boldsymbol{\lambda}}, \quad (5.66b)$$

leading to

$$\begin{aligned} \hat{\mathbf{l}} &= [(N + S) \otimes I_n]^{-1} \text{vec}(PA) - [(N + S) \otimes I_n]^{-1} (N \otimes P)\hat{\boldsymbol{\lambda}} = \\ &= [(N + S)^{-1} \otimes I_n] \text{vec}(PA) - [(N + S)^{-1} \otimes I_n] (N \otimes P)\hat{\boldsymbol{\lambda}} = \\ &= \text{vec}[PA(N + S)^{-1}] - [(N + S)^{-1}N \otimes P]\hat{\boldsymbol{\lambda}}, \quad \text{using (A.2)}. \end{aligned} \quad (5.66c)$$

Now substitute $\hat{\mathbf{l}}$ from (5.66c) into (5.65) to obtain

$$\begin{aligned} \text{vec}(PA) &= (N \otimes I_n) \{ \text{vec}[PA(N + S)^{-1}] - [(N + S)^{-1}N \otimes P]\hat{\boldsymbol{\lambda}} \} = \\ &= \text{vec}[PA(N + S)^{-1}N] - [N(N + S)^{-1}N \otimes P]\hat{\boldsymbol{\lambda}}, \quad \text{using (A.1)}. \end{aligned} \quad (5.66d)$$

Now, the product that includes $\hat{\boldsymbol{\lambda}}$ in (5.66a) can be expressed as follows:

$$(N \otimes P)\hat{\boldsymbol{\lambda}} = \{N[N(N + S)^{-1}N]^{-} \otimes I_n\} [N(N + S)^{-1}N \otimes P]\hat{\boldsymbol{\lambda}}. \quad (5.66e)$$

Combining (5.66d) and (5.66e) gives

$$(N \otimes P)\hat{\boldsymbol{\lambda}} = -\{N[N(N + S)^{-1}N]^{-} \otimes I_n\} [\text{vec}(PA) - \text{vec}[PA(N + S)^{-1}N]]. \quad (5.66f)$$

Multiplying the right side through and using (A.1) yields

$$\begin{aligned} (N \otimes P)\hat{\boldsymbol{\lambda}} &= -\text{vec}\{PA[N(N + S)^{-1}N]^{-}N\} + \\ &+ \text{vec}\{PA(N + S)^{-1}N[N(N + S)^{-1}N]^{-}N\}. \end{aligned} \quad (5.66g)$$

Recalling that $N = A^T PA$, the matrix A in the last term of the preceding line can be replaced by $A = AN^{-}N$, see (5.35a), which permits reduction of said term to $\text{vec}(PA)$.

$$(N \otimes P)\hat{\boldsymbol{\lambda}} = -\text{vec}\{PA[N(N + S)^{-1}N]^{-}N\} + \text{vec}(PA) \quad (5.66h)$$

Now we can substitute the preceding line into (5.66b) in order to solve for $\hat{\mathbf{l}}$.

$$\text{vec}(PA) - [(N + S) \otimes I_n]\hat{\mathbf{l}} = -\text{vec}\{PA[N(N + S)^{-1}N]^{-}N\} + \text{vec}(PA), \quad (5.66i)$$

implying that

$$\begin{aligned} \hat{\mathbf{l}} &= [(N + S)^{-1} \otimes I_n] \text{vec}\{PA[N(N + S)^{-1}N]^{-}N\} = \\ &= \text{vec}\{PA[N(N + S)^{-1}N]^{-}N(N + S)^{-1}\} = \text{vec}(\hat{L}^T), \end{aligned} \quad (5.66j)$$

which finally leads to an expression for the $m \times m$ matrix L as

$$\hat{L} = (N + S)^{-1}N[N(N + S)^{-1}N]^{-1}A^T P, \text{ with } \hat{\xi}_{\text{P-MINOLESS}} = \hat{L}\mathbf{y}. \quad (5.66\text{k})$$

This agrees with our solution in (5.50f) and shows that partial-MINOLESS indeed yields the minimum partial trace of the dispersion matrix amongst all other estimators; thus it is *partially best*.

5.7 Best Linear Uniformly Minimum Bias Estimate

Analogous to BLUUE in the GMM of full rank, we take a statistical approach here to derive an estimator in the rank-deficient GMM. We already stated that all estimates $\hat{\xi}$ in the rank-deficient GMM are biased by our treatment of the rank deficiency. We wish to minimize this bias by finding the Best Linear Uniformly Minimum Bias Estimate (*BLUMBE*). The attributes of BLUMBE are described below.

- (i) Linear: The estimate is required to be linear in the observation vector \mathbf{y} .

$$\hat{\xi} = L\mathbf{y}, \text{ where the } m \times n \text{ matrix } L \text{ is to be determined.} \quad (5.67)$$

- (ii) Minimum bias:

$$E\{\hat{\xi}\} = (LA)\xi, \text{ with } \text{rk}(LA) \leq \text{rk } A = q \Rightarrow LA \neq I_m. \quad (5.68)$$

v

We see that the matrix product LA cannot be the identity matrix I_m because it has rank $q < m$. But the product LA would need to be equal to I_m in order for the estimate to be uniformly unbiased. We call the difference $LA - I_m$ the *bias matrix*, and we wish to make it as small as possible by minimizing its L_2 -norm, or rather by minimizing the square of the norm as follows (see definition of the Euclidean norm of a matrix in the appendix):

$$\Phi(L) = \|LA - I_m\|_2^2 = \text{tr}[(LA - I_m)(LA - I_m)^T] = \min_L. \quad (5.69)$$

The (first-order) necessary conditions lead to

$$(AA^T)L^T - A \doteq 0 \text{ or } (LA - I_m)A^T = 0. \quad (5.70)$$

See the appendix for derivatives of the trace. Of course the sufficiency conditions are satisfied for minimization since AA^T is positive semi-definite. From (5.70) we have the geometric interpretation that $\mathcal{R}(A^T L^T - I_m) \in N(A^T)$.

(iii) Best: The trace of dispersion matrix must be minimum.

We wish to minimize the dispersion matrix $D\{\hat{\xi}\} = \sigma_0^2 LP^{-1}L^T$. Dropping the constant σ_0^2 and considering (ii) leads to the following target function to minimize:

$$\Phi(L^T, \Lambda) = \text{tr}(LP^{-1}L^T) + 2 \text{tr}(LA - I_m)A^T\Lambda, \quad (5.71)$$

which must be made stationary with respect to L^T and Λ . Accordingly, the Euler-Lagrange necessary conditions are written as

$$\frac{1}{2} \frac{\partial \Phi}{\partial L^T} = P^{-1}\hat{L}^T + AA^T\hat{\Lambda} \doteq 0, \quad (5.72a)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \Lambda} = AA^T\hat{L}^T - A \doteq 0. \quad (5.72b)$$

Note that we could also check for the sufficient condition; however, this requires the vec operator and Kronecker products to do so. For the sake of brevity, we simply state that the necessary conditions do indeed lead to a minimization of (5.71). The above system (5.72a) and (5.72b) is solved as follows: From (5.72a) we can write

$$\hat{L}^T = -PAA^T\hat{\Lambda}. \quad (5.73a)$$

Then, by substituting (5.73a) into (5.72b), we obtain

$$A = -AA^T PAA^T \hat{\Lambda}.$$

Multiplying by $A^T P$ from the left results in

$$N = -NNA^T \hat{\Lambda}.$$

Then multiplying by $N(NN)^-$ from the left gives

$$N(NN)^-N = -N(NN)^-NNA^T \hat{\Lambda}.$$

Noting that $N(NN)^-NN = N$ means

$$N(NN)^-N = -NA^T \hat{\Lambda}. \quad (5.73b)$$

We seek an expression for $PAA^T \hat{\Lambda}$ in terms of known quantities to substitute into (5.73a). So we multiply both sides of (5.73b) by AN^- to get

$$AN^-N(NN)^-N = -AN^-NA^T \hat{\Lambda},$$

which, together with using (5.35b), implies

$$A(NN)^-N = -AA^T \hat{\Lambda}.$$

Then, multiplying on the left by P results in

$$PA(NN)^-N = -PAA^T \hat{\Lambda},$$

and, by substitution into (5.73a), we get

$$\hat{L}^T = PA(NN)^-N,$$

leading to

$$\hat{L} = N(NN)^-A^T P. \quad (5.73c)$$

Finally, upon substituting (5.73c) for \hat{L} into (5.67) and comparing to (5.59e) we get

$$\boxed{\hat{\boldsymbol{\xi}}_{\text{BLUMBE}} = \hat{L}\mathbf{y} = N(NN)^-\mathbf{c} = \hat{\boldsymbol{\xi}}_{\text{MINOLESS}}.} \quad (5.73d)$$

We have just shown that the BLUMBE and the MINOLESS are equivalent. This equivalency makes these solutions very appealing for the rank deficient GMM, as together they fulfill the following properties:

- Minimum norm of parameter vector.
- Smallest trace of dispersion matrix.
- Smallest norm of bias matrix.

A relevant question at this point is “what is the bias associated with BLUMBE?” The BLUMBE bias vector $\boldsymbol{\beta}$ is derived as follows:

$$\boldsymbol{\beta} := E\{\hat{\boldsymbol{\xi}}\} - \boldsymbol{\xi} = \quad (5.74a)$$

$$= \hat{L}E\{\mathbf{y}\} - \boldsymbol{\xi} \doteq LA\boldsymbol{\xi} - \boldsymbol{\xi} \Rightarrow \quad (5.74b)$$

$$\boldsymbol{\beta} = (\hat{L}A - I_m)\boldsymbol{\xi} = [N(NN)^-N - I_m]\boldsymbol{\xi}. \quad (5.74c)$$

Equation (5.74c) in the above derivation reveals the bias matrix $\hat{L}A - I_m$ that is minimized by BLUMBE.

We now make a few additional comments about the least-squares solution within the rank-deficient GMM. In addition to the vector of estimated parameters $\hat{\boldsymbol{\xi}}$, we can also generate a predicted residual vector $\tilde{\mathbf{e}}$, a vector of adjusted observations $\hat{\boldsymbol{\mu}}_y$, and an estimated variance component $\hat{\sigma}_0^2$. Each of their formulas are summarized below.

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= \{N_{rs}^-\mathbf{c} + (I_m - N_{rs}^-N)\boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in \mathbb{R}^m\} = \\ &= \{N_{rs}^-\mathbf{c} \mid N_{rs}^- \text{ is a reflexive symmetric g-inverse of } N\} \end{aligned} \quad (5.75a)$$

$$\hat{\boldsymbol{\mu}}_y = A\hat{\boldsymbol{\xi}} \quad (5.75b)$$

$$\tilde{\mathbf{e}} = [I_n - (AN_{rs}^-A^T)P]\mathbf{y}, \quad (5.75c)$$

where $AN_{rs}^-A^T$ is invariant with respect to the chosen g-inverse

$$\hat{\sigma}_0^2 = \tilde{\mathbf{e}}^T P \tilde{\mathbf{e}} / (n - \text{rk } A). \quad (5.75d)$$

The denominator in (5.75d) is the redundancy of the model. It is equal to the number of observations minus the number of estimable parameters. It is only equivalent to the number of observations minus the number of parameters, or unknowns, if the design matrix A has full column rank, in which case the redundancy is $n - m$.

An important point to make is that $\tilde{\mathbf{e}}$, $\hat{\boldsymbol{\mu}}_y$, and $\hat{\sigma}_0^2$ are all unique. That is, they do not depend on the chosen g-inverse, N_{rs}^- , for the solution (5.75a). However, $\hat{\boldsymbol{\xi}}$ itself is not unique; that is, it does depend directly on the chosen g-inverse.

Suppose we have two different elements of the solution space, namely $\hat{\boldsymbol{\xi}}^{(1)}$ and $\hat{\boldsymbol{\xi}}^{(2)}$, that take the forms

$$\hat{\boldsymbol{\xi}}^{(1)} = G_1 \mathbf{c} \quad \text{and} \quad \hat{\boldsymbol{\xi}}^{(2)} = G_2 \mathbf{c} \quad \text{where} \quad G_1, G_2 \in \{N_{rs}^-\}. \quad (5.76)$$

Now suppose we would like to transform from one solution to the other. The transformations are written as

$$\hat{\boldsymbol{\xi}}^{(1)} = (G_1 N) \hat{\boldsymbol{\xi}}^{(2)} \quad \text{and} \quad \hat{\boldsymbol{\xi}}^{(2)} = (G_2 N) \hat{\boldsymbol{\xi}}^{(1)}, \quad (5.77)$$

with respective dispersion matrices

$$D\{\hat{\boldsymbol{\xi}}^{(1)}\} = (G_1 N) \cdot D\{\hat{\boldsymbol{\xi}}^{(2)}\} \cdot (G_1 N)^T \quad \text{and} \quad D\{\hat{\boldsymbol{\xi}}^{(2)}\} = (G_2 N) \cdot D\{\hat{\boldsymbol{\xi}}^{(1)}\} \cdot (G_2 N)^T. \quad (5.78)$$

The relations in (5.77) hold because the normal equations $N \hat{\boldsymbol{\xi}}^{(i)} = \mathbf{c}$ are fulfilled for all $\hat{\boldsymbol{\xi}}^{(i)}$. These transformations are called “ S -transformations.” They have practical use in datum transformation problems, and they are discussed further in section 5.9.

It is often costly to compute N_{rs}^- . How then can we represent the product $G_i N$? The only difference between various $G_i N$ is in the dimension of their nullspaces. Thus, in the following we look at different bases for the nullspace of A (or, equivalently, the nullspace of N) to solve our rank deficient problem.

5.8 Minimum and Inner Constraints

In the context of minimum constraints, the term “minimum” is used to mean the minimum number of constraints required to overcome the rank deficiency of the system. The constraints are given in the form of linear equations in the unknown parameters.

5.8.1 Restricted LEast-Squares Solution (RLESS)

The minimal constraint equation is written as

$$K \boldsymbol{\xi} = \boldsymbol{\kappa}_0. \quad (5.79)$$

where, K is an $l \times m$ matrix with

$$\text{rk } K = l = m - q \quad \text{and} \quad \mathcal{R}(K^T) \cup \mathcal{R}(A^T) = \mathbb{R}^m, \quad (5.80a)$$

implying that

$$\mathcal{R}(K^T) \oplus \mathcal{R}(A^T) = \mathbb{R}^m, \quad (5.80b)$$

which further implies that

$$\text{rk}[A^T, K^T] = \text{rk } A + \text{rk } K = m. \quad (5.80c)$$

Both the *constraint matrix* K and the right-side vector $\boldsymbol{\kappa}_0$ are known, constant quantities. In practice, $\boldsymbol{\kappa}_0$ is often a vector of zeros, especially in the case of linearized observation equations.

The above equations tell us that the row space of matrix A combined with the row space of K (i.e., their union) span all of \mathbb{R}^m . Even more, the union forms a basis for \mathbb{R}^m . And since the union forms a basis, the matrix K provides only the minimum number of constraints needed. Combining equation (5.79) with the observation equations (1.3), allows us to write the following system of normal equations:

$$\begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \boldsymbol{\kappa}_0 \end{bmatrix}. \quad (5.81)$$

The normal-equation matrix on the left side is indeed regular (non-singular) due to the rank relations of (5.80c). The normal equations can be solved as follows: Adding $K^T \times$ row 2 to row 1 results in

$$(N + K^T K) \hat{\boldsymbol{\xi}} = \mathbf{c} + K^T (\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}), \quad (5.82a)$$

leading to

$$\hat{\boldsymbol{\xi}} = (N + K^T K)^{-1} \mathbf{c} + (N + K^T K)^{-1} K^T (\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}). \quad (5.82b)$$

Now we combine the preceding line with row 2 to obtain

$$\boldsymbol{\kappa}_0 = K \hat{\boldsymbol{\xi}} = K(N + K^T K)^{-1} \mathbf{c} + K(N + K^T K)^{-1} K^T (\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}), \quad (5.82c)$$

which leads to

$$\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}} = [K(N + K^T K)^{-1} K^T]^{-1} [\boldsymbol{\kappa}_0 - K(N + K^T K)^{-1} \mathbf{c}]. \quad (5.82d)$$

Finally, upon substituting (5.82d) into (5.82b), we can write the Restricted Least-Squares Solution (RLESS) as

$$\hat{\boldsymbol{\xi}}_{\text{RLESS}} = (N + K^T K)^{-1} \mathbf{c} + (N + K^T K)^{-1} K^T \cdot [K(N + K^T K)^{-1} K^T]^{-1} [\boldsymbol{\kappa}_0 - K(N + K^T K)^{-1} \mathbf{c}]. \quad (5.83)$$

If $\boldsymbol{\kappa}_0$ turns out to be zero, and if we factor out the vector \mathbf{c} , the solution (5.83) reduces to

$$\hat{\boldsymbol{\xi}}_{\text{RLESS}} = [(N + K^T K)^{-1} - (N + K^T K)^{-1} K^T [K(N + K^T K)^{-1} K^T]^{-1} \cdot K(N + K^T K)^{-1}] \mathbf{c}, \text{ if } \boldsymbol{\kappa}_0 = \mathbf{0}. \quad (5.84)$$

Now, for convenience in further analysis, denote the matrix on the right side of 5.84 as G , viz.

$$G := [(N + K^T K)^{-1} - (N + K^T K)^{-1} K^T [K(N + K^T K)^{-1} K^T]^{-1} K(N + K^T K)^{-1}]. \quad (5.85)$$

5.8.2 Reflexive Symmetric G-Inverse

Question: is the matrix G in (5.85) a reflexive symmetric g-inverse of N ? We claim that it is. The proof that follows is rather lengthy, but out of it comes a representation of matrix K that leads to the so called inner-constraint solution. The symmetry of G is obvious from inspection. The reflexivity can be confirmed by checking the rank. Because the rank of G will not change when premultiplied by the full-rank matrix $N + K^T K$, and because this multiplication results in a idempotent matrix $(N + K^T K)G$, we make use of this multiplication as follows:

$$\begin{aligned} \text{rk } G &= \text{rk}[(N + K^T K)G] = \\ &= \text{tr}[(N + K^T K)G] = \quad (\text{due to its idempotent property}) \\ &= \text{tr } I_m - \text{tr}\{[K(N + K^T K)^{-1}K^T]^{-1}K(N + K^T K)^{-1}K^T\} = \\ &= m - (m - q) = q = \text{rk } N. \end{aligned}$$

Since $\text{rk } G = \text{rk } N$, if the symmetric matrix G is a g-inverse of N , it is also a reflexive symmetric g-inverse. What is left is to show that G is indeed a g-inverse of N . We start by forming the product NGN .

$$\begin{aligned} NGN &= N(N + K^T K)^{-1}N - N(N + K^T K)^{-1}K^T \\ &\quad \cdot [K(N + K^T K)^{-1}K^T]^{-1}K(N + K^T K)^{-1}N \end{aligned} \quad (5.86)$$

Our aim is to show that $K(N + K^T K)^{-1}N = 0$, which would cancel what follows the minus sign on the right side of (5.86). Then we must show that the remaining term $N(N + K^T K)^{-1}N$ equals N , which implies that $(N + K^T K)^{-1}$ is a g-inverse of N and thereby proves that G is as well. This is done in the following section.

5.8.3 (Partial) Minimum Norm Least-Squares Solution (MINOLESS)

Recalling that m is the number of unknown parameters and l is the number of constraints, we introduce an $l \times m$ matrix E such that

$$AE^T = 0, \quad (5.87a)$$

and

$$\text{rk } E = l = m - q, \quad (5.87b)$$

implying that

$$\mathcal{R}(A^T) \perp \mathcal{R}(E^T) = \mathbb{R}^m. \quad (5.87c)$$

The above relations mean that the columns of E^T (or rows of E) form a basis for the nullspace of A , and thus also for the nullspace of N . Every row of E is perpendicular to every row of A , and though the rows of E do not have to be perpendicular to one another, they are linearly independent of each other. So, we could construct matrix E with eigenvectors corresponding to the zero eigenvalues of N . But this is only one choice for constructing E ; the matrix E is not unique.

Because of (5.87a) we have the relation

$$(N + K^T K)E^T = K^T (KE^T). \quad (5.88)$$

Now we assert that the $l \times l$ matrix KE^T is invertible.

Proof: Suppose KE^T is not invertible. This implies that there exists a linear combination of the rows of K that is orthogonal to a column of E^T ; or in mathematical terms $\mathcal{R}(K^T) \subset \mathcal{R}(E^T)^\perp$. This would mean that a vector in $\mathcal{R}(K^T)$ is contained in $\mathcal{R}(A^T)$ since $\mathcal{R}(A^T) = \mathcal{R}(E^T)^\perp$. But this contradicts the direct sum in (5.80b). In other words, there exists no linear combination of the rows of K that is perpendicular to a column of E^T , and therefore, KE^T is invertible. We continue by pre- and post-multiplying (5.88) by appropriate inverses as follows:

$$E^T (KE^T)^{-1} = (N + K^T K)^{-1} K^T, \quad (5.89a)$$

implying that

$$N(N + K^T K)^{-1} K^T = NE^T (KE^T)^{-1}. \quad (5.89b)$$

But, $NE^T = 0$ due to (5.87a), therefore

$$N(N + K^T K)^{-1} K^T = 0. \quad (5.89c)$$

Thus (5.86) does reduce to $NGN = N(N + K^T K)^{-1}N$. Now, using two successive applications of the rule for the inverse of a sum (see equation (A.15) in the appendix) we can check to see if this product further reduces to N .

$$N(N + K^T K)^{-1}N = N[N^- - N^- K^T (I_l + KN^- K^T)^{-1} KN^-]N = \quad (5.90a)$$

$$= NN^- N - NN^- K^T [I_l - K(N + K^T K)^{-1} K^T] KN^- N = \quad (5.90b)$$

$$= N - NN^- K^T (I_l - I_l) KN^- N = N \quad (5.90c)$$

Here we have used the relationship $K(N + K^T K)^{-1} K^T = I_l$, which is obvious from (5.89a).

Thus we have shown that the matrix G of (5.85) is indeed a reflexive symmetric g-inverse for N , given any arbitrary matrix K satisfying the conditions (5.80a)–(5.80c). We summarize by listing three important relations between the normal-equations matrix N and the minimal-constraint matrix K .

$$N(N + K^T K)^{-1}N = N \quad (5.91a)$$

$$N(N + K^T K)^{-1} K^T = 0 \quad (5.91b)$$

$$K(N + K^T K)^{-1} K^T = I_l \quad (5.91c)$$

As we have said already, the minimum-constraint matrix K must satisfy conditions (5.80a)–(5.80c); the matrix K is otherwise arbitrary. The matrix E introduced above satisfies these conditions and may be used in place of K . In this case we get MINOLESS. Rewriting (5.84), which used $\kappa_0 = \mathbf{0}$, with E instead of K gives

$$\begin{aligned} \hat{\xi}_{\text{MINOLESS}} &= \{ (N + E^T E)^{-1} - (N + E^T E)^{-1} E^T \\ &\cdot [E(N + E^T E)^{-1} E^T]^{-1} E(N + E^T E)^{-1} \} \mathbf{c}. \end{aligned} \quad (5.92)$$

Using relations (5.89a) and (5.91c), with K replaced by E , allows us to rewrite (5.92) as

$$\hat{\xi}_{\text{MINOLESS}} = [(N + E^T E)^{-1} - E^T (E E^T)^{-1} (E E^T)^{-1} E] c. \quad (5.93)$$

The diagram in Figure 5.3 shows the geometric relationships between the range spaces of A^T , E^T , and K^T , together with the MINOLESS.

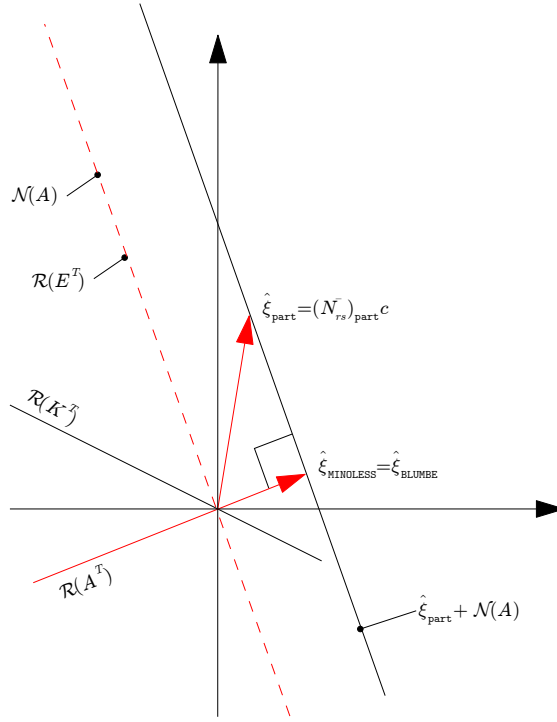


Figure 5.3: Schematic representation of solution space with RLESS and MINOLESS

Now we prove that (5.93), or equivalently (5.92), is in fact MINOLESS. To do so we must show that the matrix on the right side of (5.93), which we define here as G , is the pseudoinverse N^+ of N .

$$G := [(N + E^T E)^{-1} - E^T (E E^T)^{-1} (E E^T)^{-1} E] \stackrel{?}{=} N^+. \quad (5.94)$$

We already know that $G \in N_{rs}^-$ from the above derivation of RLESS. We only have to show the two remaining properties of the pseudoinverse; see (5.16).

Note that $NG = N(N + E^T E)^{-1}$, because NE^T contains the product AE^T , which is zero by (5.87a). If G is in fact equal to N^+ , then NG must satisfy $NG =$

$(NG)^T$ according to (5.16).

$$\begin{aligned} NG &= N(N + E^T E)^{-1} = \\ &= (N + E^T E - E^T E)(N + E^T E)^{-1} = \\ &= (N + E^T E)(N + E^T E)^{-1} - E^T E(N + E^T E)^{-1} \end{aligned}$$

And now using the transpose of (5.89a) with K replaced by E leads to

$$NG = I_m - E^T (EE^T)^{-1} E. \quad (5.95)$$

The matrix in (5.95) is obviously symmetric so that $NG = (NG)^T$. Also, since G and N are both symmetric, $NG = (NG)^T = G^T N^T = GN$ so that all conditions for the pseudoinverse have been satisfied, and thus it is proved that (5.92) is indeed MINOLESS. Note also that due to the orthogonality relation (5.87a), we can write

$$\hat{\boldsymbol{\xi}}_{\text{MINOLESS}} = (N + E^T E)^{-1} \mathbf{c} = N^+ \mathbf{c}. \quad (5.96)$$

Note, however, that $(N + E^T E)^{-1} \neq N^+$. The solution for $\boldsymbol{\xi}$ based on matrix E is a particular type of minimum-constraint solution, which has been called the *inner-constraint* solution. Note that the constraint equation (5.79) has, in essence, been replaced by $E\boldsymbol{\xi} = 0$ and that MINOLESS can actually be obtained by the following extended normal equations system, analogously to (5.81):

$$\begin{bmatrix} N & E^T \\ E & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix}. \quad (5.97)$$

One form of the dispersion matrix for MINOLESS was already shown in (5.54). Applying covariance propagation to (5.96) leads to the equivalent formula

$$D\{\hat{\boldsymbol{\xi}}_{\text{MINOLESS}}\} = \sigma_0^2 (N + E^T E)^{-1} N (N + E^T E)^{-1} = \sigma_0^2 N^+. \quad (5.98)$$

Also, analogous to (5.74c), we write the bias vector for the inner constraint solution (5.96) as

$$\boldsymbol{\beta} = [(N + E^T E)^{-1} N - I_m] \boldsymbol{\xi}. \quad (5.99)$$

By introduction of the selection matrix S into the extended normal equations (5.97), one may also derive partial MINOLESS and its dispersion matrix as

$$\hat{\boldsymbol{\xi}}_{\text{P-MINOLESS}} = (N + SE^T ES)^{-1} \mathbf{c}, \quad (5.100a)$$

$$D\{\hat{\boldsymbol{\xi}}_{\text{P-MINOLESS}}\} = \sigma_0^2 (N + SE^T ES)^{-1} N (N + SE^T ES)^{-1}. \quad (5.100b)$$

We end this section by noting that as a consequence of equations (5.91a) through (5.91c), and because the leading N in (5.91b) can be replaced by A , the formula (5.83) for RLESS can be rewritten in the following simplified form:

$$\hat{\boldsymbol{\xi}}_{\text{RLESS}} = (N + K^T K)^{-1} (\mathbf{c} + K^T \boldsymbol{\kappa}_0). \quad (5.101)$$

Applying covariance propagation to equation (5.101) yields an alternate form for the RLESS dispersion as

$$D\{\hat{\boldsymbol{\xi}}_{\text{RLESS}}\} = \sigma_0^2(N + K^T K)^{-1}N(N + K^T K)^{-1}. \quad (5.102)$$

5.8.4 Summary Formulas for Minimally Constrained LESS

Regarding partial MINOLESS, if the selection matrix S is the identity matrix, all parameters are selected, and partial MINOLESS becomes MINOLESS. On the other hand, if S selects only the minimum number of parameters necessary to overcome the datum deficiency, then partial MINOLESS is equivalent to RLESS (if $\boldsymbol{\kappa}_0 = \mathbf{0}$). The following table list commonly used formulas for the three minimally constrained solutions RLESS, MINOLESS, and partial MINOLESS.

Table 5.1: Summary of formulas for minimally constrained least-squares estimators

Type	Estimator	Dispersion matrix
RLESS	$\hat{\boldsymbol{\xi}} = (N + K^T K)^{-1}(\mathbf{c} + K^T \boldsymbol{\kappa}_0)$	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2(N + K^T K)^{-1}N(N + K^T K)^{-1}$
MINOLESS	$\hat{\boldsymbol{\xi}} = (N + E^T E)^{-1}\mathbf{c} = N^+\mathbf{c}$	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2(N + E^T E)^{-1}N(N + E^T E)^{-1}$
Partial MINOLESS	$\hat{\boldsymbol{\xi}} = (N + SE^T ES)^{-1}\mathbf{c}$	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2(N + SE^T ES)^{-1}N(N + SE^T ES)^{-1}$

5.9 S-Transformations

In equation (5.77) we introduced the so called S -transformation. We now express the S -transformation in terms of the minimum-constraint matrix K from (5.79) and the inner-constraint matrix E from (5.87a). From (5.84) we have the following reflexive symmetric g -inverse for the (singular) normal-equations matrix N .

$$\begin{aligned} N_{rs}^- &= \{(N + K^T K)^{-1} - (N + K^T K)^{-1}K^T [K(N + K^T K)^{-1}K^T]^{-1} \\ &\quad \cdot K(N + K^T K)^{-1}\} = \\ &= (N + K^T K)^{-1} - (N + K^T K)^{-1}K^T K(N + K^T K)^{-1} = \quad (\text{using (5.91c)}) \\ &= (N + K^T K)^{-1}[(N + K^T K) - K^T K](N + K^T K)^{-1} \Rightarrow \\ &\quad \boxed{N_{rs}^- = (N + K^T K)^{-1}N(N + K^T K)^{-1}} \quad (5.103) \end{aligned}$$

Now according to (5.77) we must multiply N_{rs}^- on the right by N to form an S -transformation.

$$\begin{aligned} N_{rs}^- N &= (N + K^T K)^{-1} N (N + K^T K)^{-1} N = (N + K^T K)^{-1} N = \quad (\text{using (5.91a)}) \\ &= (N + K^T K)^{-1} (N + K^T K - K^T K) = I_m - (N + K^T K)^{-1} K^T K = \\ &= I_m - E^T (K E^T)^{-1} K = N_{rs}^- N \quad (\text{using (5.89a)}) \end{aligned} \quad (5.104)$$

Thus, given any RLESS solution $\hat{\xi}^{(2)}$ we can compute a different RLESS solution $\hat{\xi}^{(1)}$ that is based on its associated constraint matrix K using (5.104) as follows:

$$\boxed{\hat{\xi}^{(1)} = N_{rs}^- N \hat{\xi}^{(2)} = [I_m - E^T (K E^T)^{-1} K] \hat{\xi}^{(2)}}. \quad (5.105)$$

Note that the matrix to invert in (5.105) might be very small compared to the dimension of N .

5.9.1 Example S -Transformation

Here, an example is presented where the preservation of sparsity in the extended normal equation matrix is the motivation to use an S -transformation. Consider the case of a 3D network adjustment comprised of GPS vectors as the only type of observation. Such observations provide scale and orientation information about the network, but they provide no information on its origin. Thus, the network adjustment problem has a datum deficiency of three, which also means the normal equation matrix $N = A^T P A$ has a rank deficiency of three. Now, further suppose that MINOLESS is the type of adjustment that must be computed. Then, the $3 \times m$ matrix E would be defined by

$$E := [I_3, \dots, I_3], \quad (5.106a)$$

and thus the $m \times m$ product $E^T E$ results in

$$E^T E = \begin{bmatrix} I_3 & \dots & I_3 \\ \vdots & \ddots & \vdots \\ I_3 & \dots & I_3 \end{bmatrix}. \quad (5.106b)$$

Obviously, adding $E^T E$ to N in this case may greatly reduce the sparsity of N , especially if N had most of its nonzero elements near its diagonal. On the other hand, if partial MINOLESS is computed with the $m \times m$ selection matrix S defined by

$$S := \text{diag}([I_3 \ 0_3 \ \dots \ 0_3]), \quad (5.106c)$$

then the sparsity of $N + S E^T E S$ would be the same as that of N itself. Now, let $\hat{\xi}^{(2)}$ be the partial MINOLESS computed using S as just defined. The solution can

be transformed to MINOLESS by substituting E for K , where $K := SE$ in this case, in (5.105), resulting in

$$\hat{\xi}^{(1)} = [I_m - E^T(EE^T)^{-1}E]\hat{\xi}^{(2)}. \quad (5.106d)$$

Then, we must only invert the 3×3 matrix EE^T to convert partial MINOLESS $\hat{\xi}^{(2)}$ to MINOLESS $\hat{\xi}^{(1)}$. This is important, because large systems of equations can be solved more efficiently when the coefficient matrix is sparse.

Obviously, the corresponding dispersion matrix could be computed by

$$D\{\hat{\xi}^{(1)}\} = [I_m - E^T(EE^T)^{-1}E] \cdot D\{\hat{\xi}^{(2)}\} \cdot [I_m - E^T(EE^T)^{-1}E]. \quad (5.106e)$$

The respective residual vectors corresponding to $\hat{\xi}^{(1)}$ and $\hat{\xi}^{(2)}$ are identical, a property of minimally constrained solutions reiterated in the next section.

5.10 Concluding Remarks about the Restricted Least-Squares Solution

In addition to (5.14) and (5.36), we have an alternative way to represent the solution space for the rank deficient GMM via RLESS.

$$\hat{\xi} = \{\hat{\xi} \mid \hat{\xi} = \hat{\xi}_{\text{RLESS}} \text{ subject to minimum constraints } K\xi = \kappa_0\} \quad (5.107)$$

For convenience we have set $\kappa_0 := K\xi = \mathbf{0}$ in some of the derivations above. The zero-vector always applies to the case where we linearize, as $\hat{\xi}$ becomes a vector of estimated corrections to the initial parameter values. For purely linear observation equations (5.1), we may have a non-zero vector κ_0 , in which case it must be included in the solution formula as in (5.91b).

If we base the reflexive symmetric matrix N_{rs}^- for RLESS on the singular value decomposition of N , as in (5.31), we only need to replace the arbitrary matrix L in that formula with the constraint matrix K in order to reach a minimum-constraint solution satisfying $K\xi = \kappa_0$.

Finally, we reiterate that no matter what minimum-constraint conditions we impose, the residual vector will be the same. This is how we determine if two adjustment models are the same; they should produce the same residual vector. This is true because in the equation for the predicted random errors (residuals)

$$\tilde{e} = (I_n - AN_{rs}^-A^T)P\mathbf{y} \quad (5.108)$$

the term $AN_{rs}^-A^T$ is invariant with respect to the choice of N_{rs}^- . Applying covariance propagation to (5.108) leads to the dispersion matrix

$$D\{\tilde{e}\} = \sigma_0^2(P^{-1} - AN_{rs}^-A^T). \quad (5.109)$$

Chapter 6

The Variance Component Model

The variance component model (VCM) is used for the case where more than one variance component must be estimated. The functional part of the model (6.1) looks like that of the Gauss-Markov model (GMM), except that the dispersion matrix Σ of the random error vector \mathbf{e} (or equivalently the observation vector \mathbf{y}) is now expressed as a linear combination of (known) cofactor matrices Q_i , each multiplied by a unique, unknown variance component σ_i^2 . In the following, we restrict the index on the variance components to $i = 2$. It is certainly possible to add additional variance components to the model, but such an increase could become computationally intensive and require a rather large redundancy in the system of equations. Note also that the model (6.1) shows a full-rank design matrix A .

$$\mathbf{y} = \underset{n \times m}{A} \boldsymbol{\xi} + \mathbf{e}, \quad \text{rk } A = m < n, \quad \mathbf{e} \sim (\mathbf{0}, \Sigma = \sigma_1^2 Q_1 + \sigma_2^2 Q_2) \quad (6.1)$$

The Best Linear Uniformly Unbiased Estimate (BLUUE) of the unknown parameters $\boldsymbol{\xi}$ within model (6.1) is given by

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{\text{BLUUE}} &= (A^T \Sigma^{-1} A)^{-1} A^T \Sigma^{-1} \mathbf{y} = \\ &= [A^T (\sigma_1^2 Q_1 + \sigma_2^2 Q_2)^{-1} A]^{-1} A^T (\sigma_1^2 Q_1 + \sigma_2^2 Q_2)^{-1} \mathbf{y}. \end{aligned} \quad (6.2)$$

We see from (6.2) that the parameter estimates $\hat{\boldsymbol{\xi}}_{\text{BLUUE}}$ depend upon the unknown variance components σ_i^2 , and thus we cannot actually compute BLUUE for this model. So we are left with the option of replacing the unknown variance components with their estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$; then we have the functional dependency $\hat{\boldsymbol{\xi}} := \hat{\boldsymbol{\xi}}(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$.

At this point, one may naturally ask whether such an estimator retains the *linear* and *unbiased* properties of BLUUE. Obviously, $\hat{\boldsymbol{\xi}}(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$ is non-linear in \mathbf{y} , since the unknown variance components $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are functions of \mathbf{y} and they also multiply \mathbf{y} in the normal equations. Therefore, the estimator is not BLUUE.

However, under certain assumptions, it can be shown that $\hat{\boldsymbol{\xi}}$ is unbiased. In order to show this we make use of E-D correspondence (see chapter 4). Let us now proceed with the derivation of $\hat{\boldsymbol{\xi}}(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$.

Our approach will be to develop a model linear in the unknown variance components that has the same structure as the Gauss Markov Model (GMM). We will then derive estimators for the unknown parameters of the model by application of a least-squares solution (LESS), as was done within the GMM.

We begin by introducing approximations (or initial values) to Σ , σ_1^2 , σ_2^2 as Σ_0 , $\sigma_{1,0}^2$, $\sigma_{2,0}^2$, respectively, which are defined as follows:

$$\Sigma_0 := \sigma_{1,0}^2 Q_1 + \sigma_{2,0}^2 Q_2, \quad (6.3a)$$

which implies

$$\text{vec } \Sigma_0 = \begin{bmatrix} \text{vec } Q_1 & \text{vec } Q_2 \end{bmatrix} \begin{bmatrix} \sigma_{1,0}^2 \\ \sigma_{2,0}^2 \end{bmatrix} = V \boldsymbol{\vartheta}_0. \quad (6.3b)$$

The variables in (6.3) are defined as follows:

$\text{vec } \Sigma_0$ is an $n^2 \times 1$ vector.

V is defined as $V := [\text{vec } Q_1, \text{vec } Q_2]$ and is size $n^2 \times c$, where c is the number of unknown variance components.

$\boldsymbol{\vartheta}_0$ is defined as $\boldsymbol{\vartheta}_0 := [\sigma_{1,0}^2, \sigma_{2,0}^2]^T$, which is size $c \times 1$.

The transformation of (6.3a) into (6.3b) by use of the vec operator is key to arriving at a model that is linear in the unknown variance components. Note that in the following, we restrict the number of variance components to two ($c := 2$) for the sake of brevity.

Using the above approximations leads to the following vectors of estimated parameters and predicted random errors (residuals):

$$\hat{\boldsymbol{\xi}}_0 = \hat{\boldsymbol{\xi}}(\sigma_{1,0}^2, \sigma_{2,0}^2) = (A^T \Sigma_0^{-1} A)^{-1} A^T \Sigma_0^{-1} \mathbf{y}, \quad (6.4a)$$

$$\tilde{\mathbf{e}}_0 = \mathbf{y} - A \hat{\boldsymbol{\xi}}_0 = [I_n - A(A^T \Sigma_0^{-1} A)^{-1} A^T \Sigma_0^{-1}] \mathbf{y} = [I_n - S_0] \mathbf{e}. \quad (6.4b)$$

Here, we define the similarity-transformation matrix as $S_0 := A(A^T \Sigma_0^{-1} A)^{-1} A^T \Sigma_0^{-1}$ and note that A is in the nullspace of $[I_n - S_0]$, which is apparent from

$$[I_n - S_0] A = [I_n - A(A^T \Sigma_0^{-1} A)^{-1} A^T \Sigma_0^{-1}] A = 0. \quad (6.5a)$$

Thus, we can replace \mathbf{y} with the true random error vector \mathbf{e} to arrive at

$$[I_n - S_0] \mathbf{y} = [I_n - S_0] (A \boldsymbol{\xi} + \mathbf{e}) = [I_n - S_0] \mathbf{e}, \quad (6.5b)$$

which is theoretically accurate even though \mathbf{e} is unknown. From here we can write a new model based on the Kronecker product $\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0$.

$$E\{\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0\} = [(I_n - S_0) \otimes (I_n - S_0)] \cdot E\{\mathbf{e} \otimes \mathbf{e}\} \quad (6.6)$$

Again, the motivation for the new model is to eventually arrive at a model that is linear in our unknown variance components $\sigma_{1,0}^2$ and $\sigma_{2,0}^2$. Using (2.7), we have $\mathbf{e} \otimes \mathbf{e} = \text{vec}(\mathbf{e}\mathbf{e}^T)$, which implies that

$$E\{\mathbf{e} \otimes \mathbf{e}\} = E\{\text{vec}(\mathbf{e}\mathbf{e}^T)\} = \text{vec} E\{\mathbf{e}\mathbf{e}^T\} = \text{vec} D\{\mathbf{e}\} = \text{vec} \Sigma. \quad (6.7)$$

This means that (6.6) can be rewritten as

$$E\{\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0\} = [(I_n - S_0) \otimes (I_n - S_0)]V\boldsymbol{\vartheta}, \quad (6.8)$$

where $\boldsymbol{\vartheta}$ is analogous to $\boldsymbol{\vartheta}_0$ but is based on the true (unknown) variance components.

Equation (6.8) is now a linear form in $\boldsymbol{\vartheta} = [\sigma_1^2, \sigma_2^2]^T$, which is precisely the quantity that we want to estimate. Keep in mind that the initial approximation $\boldsymbol{\vartheta}_0$ is used in the computation of $\tilde{\mathbf{e}}_0$. This is because $\tilde{\mathbf{e}}_0$ is defined through S_0 , and S_0 depends on Σ_0 , which depends on the approximation $\boldsymbol{\vartheta}_0$. Numerically, this means that we must iterate the solution of $\boldsymbol{\vartheta}_0$ until it converges to $\hat{\boldsymbol{\vartheta}}$. This is the so-called *reproducing property*.

Now we show the dispersion of the Kronecker product $\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0$, which, under the assumption of quasi-normality, reads

$$D\{\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0\} = (I_{n^2} + K)[(I_n - S_0)\Sigma(I_n - S_0)^T \otimes (I_n - S_0)\Sigma(I_n - S_0)^T]. \quad (6.9a)$$

Here, K is a commutation matrix. Equation (6.9a) is more complicated than (4.17) within the GMM, since it is based on both the true matrix Σ and the approximate matrix S_0 . However, by substituting the approximation Σ_0 , exploiting the symmetry of $\Sigma_0(I_n - S_0)^T$ and the idempotent property of $(I_n - S_0)$, we can write an approximate dispersion matrix as

$$D_0\{\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0\} = (I_{n^2} + K)[(I_n - S_0)\Sigma_0 \otimes (I_n - S_0)\Sigma_0]. \quad (6.9b)$$

Combining (6.4b) and (6.8), and including (6.9b), the analogy of the model (6.6) to the GMM (i.e., $E\{\mathbf{y}\} = A\boldsymbol{\xi}$, $D\{\mathbf{y}\} = \Sigma = \sigma_0^2 Q$) is shown in the following schematic:

<p>New model—linear in $\boldsymbol{\vartheta} = [\sigma_1^2, \sigma_2^2]^T$</p> <p>Expectation:</p> $E\{\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0\} = E\{\underbrace{(I_n - S_0)\mathbf{y} \otimes (I_n - S_0)\mathbf{y}}_{\text{Analogous to } \mathbf{y} \text{ in the GMM}}\} =$ $= \underbrace{[(I_n - S_0) \otimes (I_n - S_0)]V}_{\text{Analogous to } A \text{ in GMM}} \underbrace{\boldsymbol{\vartheta}}_{\text{Analogous to } \boldsymbol{\xi} \text{ in the GMM}}$	(6.10a)
<p>Dispersion:</p> $D_0\{\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0\} = \underbrace{(I_{n^2} + K)[(I_n - S_0)\Sigma_0 \otimes (I_n - S_0)\Sigma_0]}_{\text{Analogous to } \Sigma = \sigma_0^2 Q \text{ in the GMM}}$	(6.10b)

Based on the analogy to the GMM, we require a weight matrix G_0 (analogous to P in the GMM) to compute the weighted LEast-Squares Solution (LESS) of the

variance component vector $\boldsymbol{\vartheta}$. In full analogy to (4.18), we define an ‘‘approximate weight matrix’’ as follows:

$$G_0 = (I_{n^2} + K)[\Sigma_0^{-1}(I_n - S_0) \otimes \Sigma_0^{-1}(I_n - S_0)]. \quad (6.11)$$

Note that up to a factor of 1/4, G_0 turns out to be a g-inverse of the approximate dispersion matrix (6.9b) (cf. (4.17) and (4.18)). Once again, we note that K is a commutation matrix. We can now write the normal equations for the weighted LESS. First we form the right-side vector, analogously to $A^T P \mathbf{y}$ in the GMM.

Right-side:

$$\begin{aligned} V^T [(I_n - S_0)^T \otimes (I_n - S_0)^T] (I_{n^2} + K) [\Sigma_0^{-1}(I_n - S_0) \otimes \Sigma_0^{-1}(I_n - S_0)] \cdot \\ \cdot [(I_n - S_0) \mathbf{y} \otimes (I_n - S_0) \mathbf{y}] = \end{aligned} \quad (6.12a)$$

$$\begin{aligned} V^T (I_{n^2} + K) [(I_n - S_0)^T \otimes (I_n - S_0)^T] [\Sigma_0^{-1}(I_n - S_0) \otimes \Sigma_0^{-1}(I_n - S_0)] \cdot \\ \cdot [(I_n - S_0) \mathbf{y} \otimes (I_n - S_0) \mathbf{y}] = \end{aligned} \quad (6.12b)$$

$$\begin{aligned} = V^T (I_{n^2} + K) [(I_n - S_0)^T \Sigma_0^{-1}(I_n - S_0) \otimes (I_n - S_0)^T \Sigma_0^{-1}(I_n - S_0)] \cdot \\ \cdot [(I_n - S_0) \mathbf{y} \otimes (I_n - S_0) \mathbf{y}] = \end{aligned} \quad (6.12c)$$

$$\begin{aligned} = V^T (I_{n^2} + K) [(I_n - S_0)^T \Sigma_0^{-1}(I_n - S_0)^2 \mathbf{y} \otimes (I_n - S_0)^T \Sigma_0^{-1}(I_n - S_0)^2 \mathbf{y}] = \end{aligned} \quad (6.12d)$$

$$= V^T (I_{n^2} + K) [(I_n - S_0)^T \Sigma_0^{-1}(I_n - S_0) \mathbf{y} \otimes (I_n - S_0)^T \Sigma_0^{-1}(I_n - S_0) \mathbf{y}]. \quad (6.12e)$$

We used (A.10) in going from (6.12a) to (6.12b), and the idempotency of $(I_n - S_0)$ was exploited from (6.12d) to (6.12e). Note that $(I_n - S_0)^T \Sigma_0^{-1}$ is symmetric, so by using its transpose and considering that $I_n - S_0$ is idempotent, we can further reduce the right side to

$$\begin{aligned} V^T (I_{n^2} + K) [\Sigma_0^{-1}(I_n - S_0) \mathbf{y} \otimes \Sigma_0^{-1}(I_n - S_0) \mathbf{y}] = V^T (I_{n^2} + K) [\Sigma_0^{-1} \tilde{\mathbf{e}}_0 \otimes \Sigma_0^{-1} \tilde{\mathbf{e}}_0] = \\ = 2V^T (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) (\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0). \end{aligned} \quad (6.12f)$$

In the second line we used the fact that K is a vec permutation matrix, so that with symmetric Q_i ($i = 2$ in this case) and use of (A.11), we can rewrite $V^T (I_{n^2} + K)$ as

$$V^T (I_{n^2} + K) = \begin{bmatrix} (\text{vec } Q_1)^T \\ (\text{vec } Q_2)^T \end{bmatrix} (I_{n^2} + K) = 2V^T. \quad (6.12g)$$

Note that the factor of 2 is independent of the number of variance components. Now we can successively apply (A.1) in its transposed form to the last line of (6.12f), resulting in

$$\begin{aligned} 2 \begin{bmatrix} [\text{vec}(\Sigma_0^{-1} Q_1 \Sigma_0^{-1})]^T \\ [\text{vec}(\Sigma_0^{-1} Q_2 \Sigma_0^{-1})]^T \end{bmatrix} (\tilde{\mathbf{e}}_0 \otimes \tilde{\mathbf{e}}_0) = 2 \begin{bmatrix} [\text{vec}(\tilde{\mathbf{e}}_0^T \Sigma_0^{-1} Q_1 \Sigma_0^{-1} \tilde{\mathbf{e}}_0)]^T \\ [\text{vec}(\tilde{\mathbf{e}}_0^T \Sigma_0^{-1} Q_2 \Sigma_0^{-1} \tilde{\mathbf{e}}_0)]^T \end{bmatrix} = \\ = 2 \begin{bmatrix} \tilde{\mathbf{e}}_0^T \Sigma_0^{-1} Q_1 \Sigma_0^{-1} \tilde{\mathbf{e}}_0 \\ \tilde{\mathbf{e}}_0^T \Sigma_0^{-1} Q_2 \Sigma_0^{-1} \tilde{\mathbf{e}}_0 \end{bmatrix}. \end{aligned} \quad (6.12h)$$

Finally, by use of (6.4b) and introduction of the singular matrix $W_0 := \Sigma_0^{-1}(I_n - S_0)$, we may write

$$2 \begin{bmatrix} \tilde{\mathbf{e}}_0^T \Sigma_0^{-1} Q_1 \Sigma_0^{-1} \tilde{\mathbf{e}}_0 \\ \tilde{\mathbf{e}}_0^T \Sigma_0^{-1} Q_2 \Sigma_0^{-1} \tilde{\mathbf{e}}_0 \end{bmatrix} = 2 \begin{bmatrix} \mathbf{y}^T (I_n - S_0) \Sigma_0^{-1} Q_1 \Sigma_0^{-1} (I_n - S_0) \mathbf{y} \\ \mathbf{y}^T (I_n - S_0) \Sigma_0^{-1} Q_2 \Sigma_0^{-1} (I_n - S_0) \mathbf{y} \end{bmatrix} = 2 \begin{bmatrix} \mathbf{y}^T W_0 Q_1 W_0 \mathbf{y} \\ \mathbf{y}^T W_0 Q_2 W_0 \mathbf{y} \end{bmatrix}. \quad (6.12i)$$

as an expression of the right side of normal equations.

Now we work out the left side of the normal equations, analogous to $(A^T P A) \hat{\boldsymbol{\xi}}$ in the GMM. Much of this work has already been done since the left side begins with the same terms as the right side (analogous to $A^T P$ in the GMM); these steps will not be repeated.

Left side:

$$V^T [(I_n - S_0)^T \otimes (I_n - S_0)^T] (I_n^2 + K) [\Sigma_0^{-1} (I_n - S_0) \otimes \Sigma_0^{-1} (I_n - S_0)] \cdot [(I_n - S_0) \otimes (I_n - S_0)] V \hat{\boldsymbol{\vartheta}} = \quad (6.13a)$$

$$= 2V^T [\Sigma_0^{-1} (I_n - S_0) \otimes \Sigma_0^{-1} (I_n - S_0)] V \hat{\boldsymbol{\vartheta}} = \quad (6.13b)$$

(see (6.12) for more details)

$$= 2 \left[\frac{\text{tr}[\Sigma_0^{-1} (I_n - S_0) Q_1 \Sigma_0^{-1} (I_n - S_0) Q_1]}{\text{tr}[\Sigma_0^{-1} (I_n - S_0) Q_2 \Sigma_0^{-1} (I_n - S_0) Q_1]} \mid \frac{\text{tr}[\Sigma_0^{-1} (I_n - S_0) Q_1 \Sigma_0^{-1} (I_n - S_0) Q_2]}{\text{tr}[\Sigma_0^{-1} (I_n - S_0) Q_2 \Sigma_0^{-1} (I_n - S_0) Q_2]} \right] \begin{bmatrix} \hat{\sigma}_{1,0}^2 \\ \hat{\sigma}_{2,0}^2 \end{bmatrix} = \quad (6.13c)$$

$$= 2 \left[\frac{\text{tr}[W_0 Q_1 W_0 Q_1]}{\text{tr}[W_0 Q_2 W_0 Q_1]} \mid \frac{\text{tr}[W_0 Q_1 W_0 Q_2]}{\text{tr}[W_0 Q_2 W_0 Q_2]} \right] \begin{bmatrix} \hat{\sigma}_{1,0}^2 \\ \hat{\sigma}_{2,0}^2 \end{bmatrix}. \quad (6.13d)$$

Before combining the left (6.12i) and right (6.13d) sides into one system of equations we introduce subscripts to express the dependence of the $(j+1)$ th solution on the j th solution, and we drop the leading factor of 2 from both sides. Then the system of normal equations for the $(j+1)$ th solution is given by

$$\left[\frac{\text{tr}[W_{0,j} Q_1 W_{0,j} Q_1]}{\text{tr}[W_{0,j} Q_2 W_{0,j} Q_1]} \mid \frac{\text{tr}[W_{0,j} Q_1 W_{0,j} Q_2]}{\text{tr}[W_{0,j} Q_2 W_{0,j} Q_2]} \right] \begin{bmatrix} \hat{\sigma}_{1,0}^2 \\ \hat{\sigma}_{2,0}^2 \end{bmatrix}^{(j+1)} = \begin{bmatrix} \mathbf{y}^T W_{0,j} Q_1 W_{0,j} \mathbf{y} \\ \mathbf{y}^T W_{0,j} Q_2 W_{0,j} \mathbf{y} \end{bmatrix}. \quad (6.14)$$

The solution for (6.14) is iterated until, for some prescribed level of precision δ , we arrive at

$$\|\hat{\boldsymbol{\vartheta}}_{j+1} - \hat{\boldsymbol{\vartheta}}_j\|^2 < \delta^2. \quad (6.15)$$

Thus, the solution $\hat{\boldsymbol{\vartheta}}$ is called the reproducing Best Invariant Quadratic Uniformly Unbiased Estimate (reproBIQUUE) of $\boldsymbol{\vartheta}$.

There is always a solution to the system of equations (6.14) since they represent normal equations. However, the solution may not be unique, and often it is not. As stated previously, we consider the system to be non-linear in $\boldsymbol{\vartheta}$. In summary, we write the normal equations for the weighted LESS of $\boldsymbol{\vartheta}$ in its non-linear form, i.e., without the zero-subscripts denoting approximation and without iteration subscripts.

$$\left[\begin{array}{c|c} \text{tr}(\hat{W}Q_1\hat{W}Q_1) & \text{tr}(\hat{W}Q_1\hat{W}Q_2) \\ \hline \text{tr}(\hat{W}Q_2\hat{W}Q_1) & \text{tr}(\hat{W}Q_2\hat{W}Q_2) \end{array} \right] \begin{bmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} \mathbf{y}^T \hat{W}Q_1 \hat{W} \mathbf{y} \\ \mathbf{y}^T \hat{W}Q_2 \hat{W} \mathbf{y} \end{bmatrix} \quad (6.16a)$$

$$\hat{W} := \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} A (A^T \hat{\Sigma}^{-1} A)^{-1} A^T \hat{\Sigma}^{-1}, \quad \text{vec } \hat{\Sigma} = V \hat{\boldsymbol{\vartheta}}. \quad (6.16b)$$

Question: Should we solve the problem by aiming for a local BIQUUE at every iteration step? This is an open question. Dr. Schaffrin does not believe it is the best way, but it is the way it is often done in practice. The best algorithm may not produce a local minimum at each iteration, but we are not interested in these local minimums. Our objective is to convergence to a minimum.

The solutions may or may not depend on the initial approximations. Usually we know which solution to choose if we do find multiple solutions. The larger problem is that the system is “blind” to the non-negativity requirement of the estimates (i.e. the variance components must be positive). In practice, the cofactor matrices, Q_i , are usually revised if the solution yields negative variance component estimates. This is because the negative values are likely an indicator that there is something wrong with the model, i.e., the model is not consistent with the the observations. And we would not change the observations. However, we may be inclined to disregard a few observations (if we deem them to be outliers). Another approach would be to introduce an additional variance component to estimate.

Another question that one might ask is why the variance component estimates sometimes turn out negative. It is easy to see this in our case of two variance components. The normal matrix in (6.16a) is positive in each block. Therefore, the off-diagonal elements of its inverse are negative (think of the familiar formula for the inverse of a 2×2 matrix). So depending on the relative magnitudes of Q_1 and Q_2 , we may or may not end up with positive estimates.

What about the precision of our estimates? The estimated dispersion matrix is simply the inverse of the matrix on the left side of 6.16a, i.e.

$$\hat{D}\{\hat{\boldsymbol{\vartheta}}\} = \left[\begin{array}{c|c} \text{tr}(\hat{W}Q_1\hat{W}Q_1) & \text{tr}(\hat{W}Q_1\hat{W}Q_2) \\ \hline \text{tr}(\hat{W}Q_2\hat{W}Q_1) & \text{tr}(\hat{W}Q_2\hat{W}Q_2) \end{array} \right]^{-1}. \quad (6.17)$$

This is already the estimated dispersion of $\hat{\boldsymbol{\vartheta}}$ due to the use of the estimated matrix \hat{W} . It is hard to express the true dispersion $D\{\hat{\boldsymbol{\vartheta}}\}$ because of the iteration

process. Let us check this formula for the case of a single variance component, which we simply label as $\hat{\sigma}_0^2$.

Start with $\text{tr}(\hat{W}Q\hat{W}Q)$, where

$$\hat{W} = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}A(A^T\hat{\Sigma}^{-1}A)^{-1}A^T\hat{\Sigma}^{-1} = (\hat{\sigma}_0^2)^{-1}(P - PAN^{-1}A^TP). \quad (6.18a)$$

Multiplication by the cofactor matrix Q from the right yields

$$\hat{W}Q = (\hat{\sigma}_0^2)^{-1}(I_n - PAN^{-1}A^T), \quad (6.18b)$$

which implies that

$$\begin{aligned} \text{tr}(\hat{W}Q\hat{W}Q) &= (\hat{\sigma}_0^2)^{-2} \text{tr}[(I_n - PAN^{-1}A^T)^2] = \\ &= (\hat{\sigma}_0^2)^{-2} \text{tr}(I_n - PAN^{-1}A^T) = (\hat{\sigma}_0^2)^{-2}(n - m), \end{aligned} \quad (6.18c)$$

from which it follows

$$2[\text{tr}(\hat{W}Q\hat{W}Q)]^{-1} = 2(\hat{\sigma}_0^2)^2/(n - m). \quad (6.18d)$$

Note that (6.18d) is identical to (3.23), showing the consistency of univariate and multivariate variance component estimators. In general, we should replace m with $\text{rk } A$ to account for possible rank deficiency in matrix A . Compare this result to the solution found in (3.23).

In general, variance component estimation requires a relatively large redundancy in the model. For comparison, when estimating the $m \times 1$ vector of unknowns $\boldsymbol{\xi}$ in the GMM we might like to have a redundancy of about m . However, for variance component estimation we probably would like to have roughly the square of m . It may even require a redundancy of over 100 to estimate as few as five variance components.

Now we return to our earlier question regarding the unbiasedness of the parameter estimates. Specifically, is $\hat{\boldsymbol{\xi}}$ still unbiased when we replace the “true” variance components σ_1^2 and σ_2^2 with their reproBIQUUE estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$?

Formally we can equate

$$\begin{bmatrix} \mathbf{y}^T \hat{W}Q_1 \hat{W} \mathbf{y} \\ \mathbf{y}^T \hat{W}Q_2 \hat{W} \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{e}^T \hat{W}Q_1 \hat{W} \mathbf{e} \\ \mathbf{e}^T \hat{W}Q_2 \hat{W} \mathbf{e} \end{bmatrix}, \quad (6.19)$$

though we cannot actually compute the right side because of the unknown random error vector \mathbf{e} . However, let us assume that \mathbf{e} is symmetrically distributed with $E\{\mathbf{e}\} = \mathbf{0}$. This assumption means that we have an equal chance of any element of \mathbf{e} being positive or negative. So $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ (or more generally $\hat{\boldsymbol{\theta}}$) do not change when $+\mathbf{e}$ is replaced by $-\mathbf{e}$, because we base our estimation on a quadratic form in \mathbf{e} . Formally we can write a difference between the estimate and the true parameter vector as follows:

$$\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} = (A^T\hat{\Sigma}^{-1}A)^{-1}A^T\hat{\Sigma}^{-1}(\mathbf{y} - A\boldsymbol{\xi}) = [(A^T\hat{\Sigma}^{-1}A)^{-1}A^T\hat{\Sigma}^{-1}]\mathbf{e}. \quad (6.20)$$

We see that the difference $\hat{\xi} - \xi$ is linear in e and note that the term multiplying e will not change in sign when e does. Due to our symmetric distribution assumption, there is an equal chance of e being positive or negative; therefore there is also an equal chance of $\hat{\xi} - \xi$ being positive or negative. As a formality, we also assume that $E\{\hat{\xi}\}$ exists. And since $\hat{\xi} - \xi$ changes sign whenever e does, this implies that

$$E\{\hat{\xi} - \xi\} = \mathbf{0} \Rightarrow E\{\hat{\xi}\} = \xi, \quad (6.21)$$

which means that $\hat{\xi}$ is uniformly unbiased under reproBIQUUE.

Chapter 7

Prior Information

In this chapter we investigate the topic of prior information on the unknown parameters. More specifically, we decompose the parameter vector $\boldsymbol{\xi}$ into two parts, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, where we assume that prior information, in the form of *pseudo-observations* \mathbf{b}_0 , is available only for $\boldsymbol{\xi}_1$. Furthermore, we associate a random error vector \mathbf{e}_0 with the prior information and assume that it is uncorrelated with the random error vector \mathbf{e} associated with the observations \mathbf{y} , i.e., we assume that $C\{\mathbf{e}, \mathbf{e}_0\} = \mathbf{0}$. In practice, the prior information may come from a previous adjustment with its dispersion matrix provided by the covariance matrix of the estimated parameters from that adjustment. The data model with prior-information can be written as an extended Gauss-Markov model (GMM) as follows:

$$\begin{aligned} \mathbf{y} &= A_1 \boldsymbol{\xi}_1 + A_2 \boldsymbol{\xi}_2 + \mathbf{e}, & (7.1a) \\ \xi_1 \in \mathbb{R}^{r \times 1}, \quad \xi_2 \in \mathbb{R}^{(m-r) \times 1}, \quad \text{rk } A_1 &\leq r, \quad \text{rk } A_2 = m - r & (7.1b) \\ \mathbf{b}_0 &= \boldsymbol{\xi}_1 + \mathbf{e}_0, & (7.1c) \\ \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_0 \end{bmatrix} &\sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & Q_0 \end{bmatrix} \right) & (7.1d) \end{aligned}$$

Here the full design matrix and parameter vector are denoted by

$$A_{n \times m} := \left[A_1 \mid A_2 \right] \quad \text{and} \quad \boldsymbol{\xi}_{m \times 1} := \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix}. \quad (7.2)$$

The variables in the model are described by

\mathbf{y} is a given $n \times 1$ vector of observations.

$\boldsymbol{\xi}_1$ is an $r \times 1$ vector of unknown parameters.

$\boldsymbol{\xi}_2$ is an $(m - r) \times 1$ vector of unknown parameters.

A_1 is a given $n \times r$ coefficient (design) matrix.

A_2 is a given $n \times (m - r)$ coefficient (design) matrix.

\mathbf{e} is an $n \times 1$ vector of unknown random errors associated with the observations \mathbf{y} .

\mathbf{b}_0 is a given $r \times 1$ vector of (random) pseudo-observations called prior information.

\mathbf{e}_0 is an $r \times 1$ vector of unknown random errors associated with the pseudo-observations \mathbf{b}_0 .

P is a given $n \times n$ positive-definite weight matrix for the observations \mathbf{y} .

Q_0 is a given $r \times r$ positive-definite cofactor matrix for the pseudo-observations \mathbf{b}_0 .

σ_0^2 is an unknown variance component.

Note that matrix A_2 is assumed to have full column rank, i.e., $\text{rk } A_2 = m - r$, while A_1 does not necessarily have full column rank. Typically, \mathbf{b}_0 is a vector of zeros due to linearization (though still a random vector). If \mathbf{b}_0 is not zero, then it contains the bias of the prior information with respect to the initial approximations for the parameters $\boldsymbol{\xi}$ (assuming linearization). Finally, we note that the model uses a single variance component σ_0^2 , multiplying both cofactor matrices P^{-1} and $Q_0 =: P_0^{-1}$, where P and P_0 are called weight matrices.

7.1 Pseudo-observations

The extended GMM includes pseudo-observations, which are considered to be direct observations of the unknown parameters $\boldsymbol{\xi}_1$. Since the model uses only a single variance component, it indeed belongs to the class of Gauss-Markov models. The following expression summarizes the model in a more compact form than does (7.1):

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{b}_0 \end{bmatrix} \sim \left(\begin{bmatrix} A_1 & A_2 \\ I_r & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & P_0^{-1} \end{bmatrix} \right). \quad (7.3)$$

Because the model is a type of GMM, we can immediately write the LEast-Squares Solution (LESS) for the unknown parameters $\boldsymbol{\xi}$ and the associated disper-

sion matrix, in accordance with the “addition theory of normal equations.”

$$\begin{aligned} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} &= \left(\begin{bmatrix} A_1^T & I_r \\ A_2^T & 0 \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & P_0 \end{bmatrix} \begin{bmatrix} A_1 & A_2 \\ I_r & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} A_1^T P & P_0 \\ A_2^T P & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{b}_0 \end{bmatrix} = \\ &= \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_1 + P_0 \mathbf{b}_0 \\ \mathbf{c}_2 \end{bmatrix}, \end{aligned} \quad (7.4a)$$

$$D\left\{ \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} \right\} = \sigma_0^2 \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1}, \quad (7.4b)$$

$$\text{with } N_{ij} := A_i^T P A_j, \quad i, j \in \{1, 2\}. \quad (7.4c)$$

It is evident from the upper-left block of the dispersion matrix in (7.4b) that the magnitude of the variances of $\hat{\boldsymbol{\xi}}_1$ have been reduced due to the prior information on $\boldsymbol{\xi}_1$.

Now we want to find an equivalent estimator and dispersion matrix in terms of previous estimates made within a model without prior information. A solution of this form is more revealing of what is gained by adding the prior information to the model. For simplicity, we assume that the complete design matrix A has full column rank, though it does not have to in general. We start with the cofactor matrix $Q_{\hat{\boldsymbol{\xi}}}$ (inverted matrix on right side of (7.4b)), and rewrite it as follows:

$$Q_{\hat{\boldsymbol{\xi}}} := \sigma_0^{-2} D\left\{ \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} \right\} = \left(\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} + \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} = \quad (7.5a)$$

$$= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \left(\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} + \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \quad (7.5b)$$

$$= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} - \\ - \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1}. \quad (7.5c)$$

Now, introducing the first Schur complement of the partitioned matrix of N as $S_1 := N_{11} - N_{12} N_{22}^{-1} N_{21}$, and then using the rules for inverting a partitioned matrix,

we may write

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \left[\begin{array}{c|c} S_1^{-1} & -S_1^{-1}N_{12}N_{22}^{-1} \\ \hline -N_{22}^{-1}N_{21}S_1^{-1} & N_{22}^{-1} + N_{22}^{-1}N_{21}S_1^{-1}N_{12}N_{22}^{-1} \end{array} \right]. \quad (7.5d)$$

Now we continue manipulating the cofactor matrix from (7.5c) to result in

$$Q_{\hat{\xi}} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} - \left[\begin{array}{c|c} S_1^{-1}P_0 & 0 \\ \hline -N_{22}^{-1}N_{21}S_1^{-1}P_0 & 0 \end{array} \right]. \quad (7.5e)$$

$$\cdot \left[\begin{array}{c|c} (S_1 + P_0)^{-1} & -(S_1 + P_0)^{-1}N_{12}N_{22}^{-1} \\ \hline -N_{22}^{-1}N_{21}(S_1 + P_0)^{-1} & N_{22}^{-1} + N_{22}^{-1}N_{21}(S_1 + P_0)^{-1}N_{12}N_{22}^{-1} \end{array} \right].$$

This result implies that the dispersion matrix of $\hat{\xi}$ can be written as

$$D\left\{ \begin{array}{c} \hat{\xi}_1 \\ \hat{\xi}_2 \end{array} \right\} = \sigma_0^2 \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} -$$

$$- \sigma_0^2 \left[\begin{array}{c|c} S_1^{-1}P_0(S_1 + P_0)^{-1} & -S_1^{-1}P_0(S_1 + P_0)^{-1}N_{12}N_{22}^{-1} \\ \hline -N_{22}^{-1}N_{21}S_1^{-1}P_0(S_1 + P_0)^{-1} & N_{22}^{-1}N_{21}S_1^{-1}P_0(S_1 + P_0)^{-1}N_{12}N_{22}^{-1} \end{array} \right]. \quad (7.6)$$

Note that (7.6) is still symmetric since $S_1^{-1}P_0(S_1 + P_0)^{-1}$ is symmetric as shown in the following:

$$\begin{aligned} & S_1^{-1}P_0(S_1 + P_0)^{-1} = \\ & = S_1^{-1}P_0[S_1(I_r + S_1^{-1}P_0)]^{-1} = S_1^{-1}P_0(I_r + S_1^{-1}P_0)^{-1}S_1^{-1} = \end{aligned}$$

(now applying (A.21a))

$$\begin{aligned} & = S_1^{-1}(I_r + P_0S_1^{-1})^{-1}P_0S_1^{-1} = \\ & = [(I_r + P_0S_1^{-1})S_1]^{-1}P_0S_1^{-1} = \\ & = (S_1 + P_0)^{-1}P_0S_1^{-1}. \end{aligned}$$

An interesting observation from the dispersion $D\{\hat{\xi}\}$ given in (7.6) is that though prior information is only provided for ξ_1 , we also gain an improvement in the dispersion of $\hat{\xi}_2$.

We are now ready to express the estimator for the parameter vector in terms of the estimator within the model that does not include prior information. For

convenience, we write the latter using cup symbols. We then make use of the cofactor matrix appearing in (7.6).

$$\begin{aligned} \begin{bmatrix} \check{\xi}_1 \\ \check{\xi}_2 \end{bmatrix} &:= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} = \\ &= \left[\begin{array}{c|c} S_1^{-1} & -S_1^{-1}N_{12}N_{22}^{-1} \\ \hline -N_{22}^{-1}N_{21}S_1^{-1} & N_{22}^{-1} + N_{22}^{-1}N_{21}S_1^{-1}N_{12}N_{22}^{-1} \end{array} \right] \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} \Rightarrow \end{aligned} \quad (7.7a)$$

$$\begin{aligned} \begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix} &= \begin{bmatrix} \check{\xi}_1 \\ \check{\xi}_2 \end{bmatrix} - \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}P_0(S_1 + P_0)^{-1}[-I_r, N_{12}N_{22}^{-1}] \cdot \\ &\quad \cdot \begin{bmatrix} \mathbf{c}_1 + P_0\mathbf{b}_0 \\ \mathbf{c}_2 \end{bmatrix} + \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} P_0\mathbf{b}_0 \\ \mathbf{0} \end{bmatrix} \end{aligned} \quad (7.7b)$$

Note that it is important not to ignore the prior information vector \mathbf{b}_0 even if it is numerically zero. This is because it is a random variable and thus its impact will not be zero in the dispersion matrix $D\{\hat{\xi}\}$. Making use of (7.5d), and performing certain algebraic manipulations, we can further modify (7.7b) with the objective of

reaching a vector of “parameter improvements.”

$$\begin{aligned}
\begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix} - \begin{bmatrix} \check{\xi}_1 \\ \check{\xi}_2 \end{bmatrix} &= \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}P_0(I_r + S_1^{-1}P_0)^{-1}S_1^{-1}[\mathbf{c}_1 + P_0\mathbf{b}_0 - N_{12}N_{22}^{-1}\mathbf{c}_2] + \\
&+ \left(\begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}[-I_r, N_{12}N_{22}^{-1}] + \begin{bmatrix} 0 & 0 \\ 0 & N_{22}^{-1} \end{bmatrix} \right) \begin{bmatrix} P_0\mathbf{b}_0 \\ \mathbf{0} \end{bmatrix} = \\
&= \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}(I_r + P_0S_1^{-1})^{-1}[-P_0S_1^{-1}(N_{12}N_{22}^{-1}\mathbf{c}_2 - \mathbf{c}_1) + P_0S_1^{-1}P_0\mathbf{b}_0] + \\
&\quad + \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1} \left(\begin{bmatrix} -I_r, N_{12}N_{22}^{-1} \end{bmatrix} \begin{bmatrix} P_0\mathbf{b}_0 \\ \mathbf{0} \end{bmatrix} \right) = \\
&= \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}(I_r + P_0S_1^{-1})^{-1}[-P_0S_1^{-1}(N_{12}N_{22}^{-1}\mathbf{c}_2 - \mathbf{c}_1) + \\
&\quad + P_0S_1^{-1}P_0\mathbf{b}_0 - (I_r + P_0S_1^{-1})P_0\mathbf{b}_0] = \\
&= \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}(I_r + P_0S_1^{-1})^{-1}[P_0S_1^{-1}(\mathbf{c}_1 - N_{12}N_{22}^{-1}\mathbf{c}_2) - P_0\mathbf{b}_0] = \\
&= \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} S_1^{-1}(I_r + P_0S_1^{-1})^{-1}P_0(\check{\xi}_1 - \mathbf{b}_0)
\end{aligned}$$

In summary, we can express the vector of parameter improvements as

$$\begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix} - \begin{bmatrix} \check{\xi}_1 \\ \check{\xi}_2 \end{bmatrix} = \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} (I_r + S_1^{-1}P_0)^{-1}S_1^{-1}P_0(\check{\xi}_1 - \mathbf{b}_0) \Rightarrow \quad (7.8a)$$

$$\begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix} - \begin{bmatrix} \check{\xi}_1 \\ \check{\xi}_2 \end{bmatrix} = \begin{bmatrix} -I_r \\ N_{22}^{-1}N_{21} \end{bmatrix} (I_r + P_0^{-1}S_1)^{-1}(\check{\xi}_1 - \mathbf{b}_0). \quad (7.8b)$$

Equation (7.8a) may be used if P_0 is not invertible, and equation (7.8b) may be used if P_0 is invertible. The vector $\check{\xi}_1 - \mathbf{b}_0$ is the discrepancy vector between the prior information and what would have been estimated using the new data set without the prior information. Since the matrix $(I_r + P_0^{-1}S_1)$ has positive eigenvalues, multiplication by its inverse reduces the discrepancy vector.

If we had introduced a second variance component σ_1^2 associated with the new data set, this would only have had a second-order effect on the estimates and would have required the ratio σ_0^2/σ_1^2 in front of the prior information weight matrix P_0 .

How does the prior information change the predicted random error (residual) vector \tilde{e} ? We want to express the change as an update to the residual vector \tilde{e} and also as a function of $\tilde{\xi}$, which would be predicted, respectively, estimated within a GMM without prior information.

$$\begin{aligned} \tilde{e} &= \mathbf{y} - A_1 \hat{\xi}_1 - A_2 \hat{\xi}_2 = \\ &= [\mathbf{y} - A_1 \tilde{\xi}_1 - A_2 \tilde{\xi}_2] - A_1 (\hat{\xi}_1 - \tilde{\xi}_1) - A_2 (\hat{\xi}_2 - \tilde{\xi}_2) = \\ &= \tilde{e} - A_1 (\hat{\xi}_1 - \tilde{\xi}_1) - A_2 (\hat{\xi}_2 - \tilde{\xi}_2) = \\ &= \tilde{e} + (A_1 - A_2 N_{22}^{-1} N_{21}) (I_r + P_0^{-1} S_1)^{-1} (\tilde{\xi}_1 - \mathbf{b}_0) \end{aligned} \quad (7.9)$$

We note that the product $(I_r + P_0^{-1} S_1)^{-1} (\tilde{\xi}_1 - \mathbf{b}_0)$ appears frequently in the above equations and so in practice it may be worth computing it once at the outset and then saving it for subsequent use.

7.2 Alternative Normal Equations

In this section we introduce an alternative system of normal equation to accommodate prior information. The resulting solution is identical to that presented in the preceding section, however this alternative form allows for a singular cofactor matrix Q_0 .

The normal equations are written as

$$(N_{11} + P_0) \hat{\xi}_1 + N_{12} \hat{\xi}_2 = \mathbf{c}_1 + P_0 \mathbf{b}_0, \quad (7.10a)$$

$$N_{21} \hat{\xi}_1 + N_{22} \hat{\xi}_2 = \mathbf{c}_2, \quad (7.10b)$$

$$\hat{\lambda} = P_0 (\hat{\xi}_1 - \mathbf{b}_0). \quad (7.10c)$$

The preceding three equations can be combined in matrix form as follows:

$$\begin{bmatrix} N_{11} & N_{12} & I_r \\ N_{21} & N_{22} & 0 \\ I_r & 0 & -P_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{b}_0 \end{bmatrix}. \quad (7.11)$$

Here, $\hat{\lambda}$ is an $r \times 1$ vector of estimated Lagrange multipliers. The *normal-equations matrix* on the left side of (7.11) is of size $(m+r) \times (m+r)$. We could use the Cholesky algorithm to reduce the upper 2×2 sub-matrix block and then proceed with Gaussian elimination.

The inverse of the normal-equations matrix yields the cofactor matrix of the estimates. However, we only need to concern ourselves with the upper 2×2 sub-matrix block of the inverse in order to find the dispersion of the parameter estimates $\hat{\xi}_1$ and $\hat{\xi}_2$. In the equation that follows, the other terms of no special interest have been replaced with the symbol X .

$$\begin{aligned} \begin{bmatrix} N_{11} & N_{12} & I_r \\ N_{21} & N_{22} & 0 \\ I_r & 0 & -P_0^{-1} \end{bmatrix}^{-1} &= \left[\begin{array}{c|c} \left(\begin{bmatrix} N_{11} & N_{11} \\ N_{11} & N_{11} \end{bmatrix} + \begin{bmatrix} I_r \\ 0 \end{bmatrix} P_0 \begin{bmatrix} I_r & 0 \end{bmatrix} \right)^{-1} & X \\ \hline X & X \end{array} \right] = \\ &= \left[\begin{array}{c|c} \frac{\sigma_0^{-2} D\{\hat{\xi}\}}{X} & X \\ \hline X & X \end{array} \right] = \left[\begin{array}{c|c} \begin{bmatrix} N_{11} + P_0 & N_{11} \\ N_{11} & N_{11} \end{bmatrix}^{-1} & X \\ \hline X & X \end{array} \right] \end{aligned} \quad (7.12)$$

It is interesting to investigate the consequences of diminishing the weight of the prior information. Suppose the prior information weight matrix is defined as $P_0 := \varepsilon P_0$ and we have the situation where $\varepsilon \rightarrow 0$. This means that the prior information loses its influence, resulting in $\hat{\xi}_1 \rightarrow \check{\xi}_1$ with degrees of freedom (redundancy) $n + r - m$. However, if $\varepsilon = 0$ then we have $\hat{\xi}_1 = \check{\xi}_1$ with degree of freedom $n - m$. In other words, as ε approaches zero, the resulting estimate numerically approaches what would be obtained if prior information were not included in the model. However, the degrees of freedom of the model with prior information is larger than that of the model without prior information by a constant r , which is the number of parameters that we supposedly have prior information for. This has an unsatisfactory result on our estimated variance component $\hat{\sigma}_0^2$; it makes it look better than what it is. We might rather specify redundancy as a function of ε , but exactly how best to do that is still an open question.

Suppose we are given values for $\check{\xi}_1$ and $\check{\xi}_2$, together with the prior information \mathbf{b}_0 and associated weights P_0 , and suppose we want to find the solution for $\hat{\xi}_1$ and $\hat{\xi}_2$. From row 1 of (7.8b) we can solve

$$(P_0 + S_1)(\hat{\xi}_1 - \check{\xi}_1) = P_0(\mathbf{b}_0 - \check{\xi}_1). \quad (7.13)$$

Then, by substitution of the first row of (7.13) into the second row, we can write

$$\hat{\xi}_2 - \check{\xi}_2 = -N_{22}^{-1} N_{21}(\hat{\xi}_1 - \check{\xi}_1). \quad (7.14)$$

The update for the dispersion is then given by

$$D\left\{\begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix}\right\} - D\left\{\begin{bmatrix} \check{\xi}_1 \\ \check{\xi}_2 \end{bmatrix}\right\} = -\sigma_0^2 \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1}, \quad (7.15)$$

which already was evident from (7.5c).

Suppose we are given only the prior information \mathbf{b}_0 and we would like to find the solutions for $\hat{\xi}_1$ and $\hat{\xi}_2$. The solution is developed by starting with (7.4a) and using the relationship $Q_0 = P_0^{-1}$ as follows:

$$\begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix} = \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{c}_1 - N_{11}\mathbf{b}_0 \\ \mathbf{c}_2 - N_{21}\mathbf{b}_0 \end{bmatrix} + \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{0} \end{bmatrix} \right),$$

which implies that

$$\begin{aligned} \begin{bmatrix} \hat{\xi}_1 - \mathbf{b}_0 \\ \hat{\xi}_2 \end{bmatrix} &= \begin{bmatrix} N_{11} + P_0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_1^T P(\mathbf{y} - A_1 \mathbf{b}_0) \\ A_2^T P(\mathbf{y} - A_1 \mathbf{b}_0) \end{bmatrix} = \\ &= \left(\begin{bmatrix} I_r + N_{11}Q_0 & N_{12} \\ N_{21}Q_0 & N_{22} \end{bmatrix} \begin{bmatrix} P_0 & 0 \\ 0 & I_{m-r} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_1^T P(\mathbf{y} - A_1 \mathbf{b}_0) \\ A_2^T P(\mathbf{y} - A_1 \mathbf{b}_0) \end{bmatrix} = \\ \begin{bmatrix} \hat{\xi}_1 - \mathbf{b}_0 \\ \hat{\xi}_2 \end{bmatrix} &= \begin{bmatrix} Q_0 & 0 \\ 0 & I_{m-r} \end{bmatrix} \begin{bmatrix} I_r + N_{11}Q_0 & N_{12} \\ N_{21}Q_0 & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_1^T P(\mathbf{y} - A_1 \mathbf{b}_0) \\ A_2^T P(\mathbf{y} - A_1 \mathbf{b}_0) \end{bmatrix}. \quad (7.16) \end{aligned}$$

The first matrix on the right side of (7.16) is singular if Q_0 is singular, but this is of no consequence since we do not need to invert it. The second matrix on the right side is regular (non-singular) even if Q_0 is singular. The dispersion is given by

$$D\left\{\begin{bmatrix} \hat{\xi}_1 \\ \hat{\xi}_2 \end{bmatrix}\right\} = \sigma_0^2 \begin{bmatrix} Q_0 & 0 \\ 0 & I_{m-r} \end{bmatrix} \begin{bmatrix} I_r + N_{11}Q_0 & N_{12} \\ N_{21}Q_0 & N_{22} \end{bmatrix}^{-1}. \quad (7.17)$$

In order to confirm the consistency between the current data and the prior information, we can test the validity of the null hypothesis

$$H_0 : E\{\hat{\xi}_1 - \check{\xi}_1\} = \mathbf{0}. \quad (7.18)$$

The test statistic T is defined as

$$T := \frac{(\hat{\xi}_1 - \check{\xi}_1)^T \cdot D\{\hat{\xi}_1 - \check{\xi}_1\}^{-1} \cdot (\hat{\xi}_1 - \check{\xi}_1)}{r(\hat{\sigma}_0^2/\sigma_0^2)} \sim F(r, n - m). \quad (7.19)$$

7.3 Mixed Linear Model (Helmert's Knack)

The idea underlying the mixed linear model is that some of the parameters are random, while others are fixed. This is different than all models presented up to this point, where we have consistently defined the unknown parameters to be fixed (non-random). Here we introduce a non-random analogue to the vector of prior information \mathbf{b}_0 , denoted by $\boldsymbol{\beta}_0$. Numerically, \mathbf{b}_0 and $\boldsymbol{\beta}_0$ are equivalent, but stochastically their equivalence is obtained only by addition of a *random zero-vector*, denoted by $\mathbf{0}$, as follows:

$$\mathbf{b}_0 = \boldsymbol{\beta}_0 + \mathbf{0} = \boldsymbol{\xi}_1 + \mathbf{e}_0, \quad \mathbf{e}_0 \sim (\mathbf{0}, \sigma_0^2 P_0^{-1} = \sigma_0^2 Q_0), \quad (7.20a)$$

$$\boldsymbol{\beta}_0 = (\boldsymbol{\xi}_1 - \mathbf{0}) + \mathbf{e}_0 = \mathbf{x}_1 + \mathbf{e}_0, \quad \text{where } \mathbf{x}_1 := \boldsymbol{\xi}_1 - \mathbf{0}. \quad (7.20b)$$

Equation (7.20b) is known as Helmert's knack. It is used to transform the non-random parameter vector $\boldsymbol{\xi}_1$ to a random parameter vector \mathbf{x}_1 . Some explanation about the notation might be helpful. As usual, we use Greek letters for non-random variables and Latin letters for random variables. In this case we have also placed a tilde beneath the zero to denote a random vector of zeros associated with the unknown parameters. The expectation and dispersion of the unknown, random parameters \mathbf{x}_1 are

$$E\{\mathbf{x}_1\} = E\{\boldsymbol{\beta}_0 - \mathbf{e}_0\} = \boldsymbol{\beta}_0 - E\{\mathbf{e}_0\} = \boldsymbol{\beta}_0, \quad (7.21a)$$

$$D\{\mathbf{x}_1\} = D\{\boldsymbol{\beta}_0 - \mathbf{e}_0\} = D\{\mathbf{e}_0\} = \sigma_0^2 Q_0. \quad (7.21b)$$

Since we have used the random vector $\mathbf{0}$ in the pseudo-observations, we need to modify the original observation equations given in (7.1a) by subtracting $A_1 \mathbf{0}$ from both sides of the equation. This does not change the numerical values on the left side, but it does make it a different vector in terms of its stochastic properties. We denote the revised left-side vector as $\bar{\mathbf{y}}$.

$$\begin{aligned} \bar{\mathbf{y}} &= \mathbf{y} - A_1 \mathbf{0} = \\ &= A_1 \boldsymbol{\xi}_1 - A_1 \mathbf{0} + A_2 \boldsymbol{\xi}_2 + \mathbf{e} = \\ &= A_1 (\boldsymbol{\xi}_1 - \mathbf{0}) + A_2 \boldsymbol{\xi}_2 + \mathbf{e} \Rightarrow \\ \bar{\mathbf{y}} &= A_1 \mathbf{x}_1 + A_2 \boldsymbol{\xi}_2 + \mathbf{e} \end{aligned} \quad (7.22)$$

Again we note that $\bar{\mathbf{y}}$ contains the same numerical values as \mathbf{y} , but now with dispersion matrix

$$D\{\bar{\mathbf{y}}\} = \sigma_0^2 (A_1 Q_0 A_1^T + P^{-1}). \quad (7.23)$$

On the right side of (7.22), we have a random parameter-vector \mathbf{x}_1 and a non-random parameter-vector $\boldsymbol{\xi}_2$; the equation is linear in these unknowns. This is why we call the model a *mixed* linear model (MLM); it has a mix of fixed and random unknown parameters. We summarize the MLM in the box below.

$$\bar{\mathbf{y}} := (\mathbf{y} - A_1 \mathbf{0}) = A_1 \mathbf{x}_1 + A_2 \boldsymbol{\xi}_2 + \mathbf{e}, \quad (7.24a)$$

$$\mathbf{x}_1 = \boldsymbol{\beta}_0 - \mathbf{e}_0, \quad \mathbf{x}_1 \sim (\boldsymbol{\beta}_0, \sigma_0^2 Q_0), \quad \text{rk } A_2 = m - r, \quad (7.24b)$$

$$\begin{bmatrix} \mathbf{e} \\ \mathbf{e}_0 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & Q_0 \end{bmatrix} \right). \quad (7.24c)$$

In going from the extended GMM (7.1) to the MLM (7.24), we have changed from a model that has no a-priori information about the *non-random* parameters $\boldsymbol{\xi}_1$ to a model that has a-priori information about the *random* parameters \mathbf{x}_1 . In either case, we know nothing a priori about the parameters in $\boldsymbol{\xi}_2$. We claim that the MLM is more flexible, in general, than the extended GMM. The following discussion supports this claim.

In the extended GMM, the class of linear estimators is represented by

$$[\hat{\boldsymbol{\xi}}_1^T, \hat{\boldsymbol{\xi}}_2^T]^T = L_1 \mathbf{y} + L_2 \mathbf{b}_0 + \boldsymbol{\gamma}_0, \quad (7.25)$$

where L_1 and L_2 are unknown matrices. In contrast, in the MLM the class of linear predictors/estimators is represented by

$$[\tilde{\mathbf{x}}_1^T, \hat{\boldsymbol{\xi}}_2^T]^T = L \bar{\mathbf{y}} + \boldsymbol{\gamma}, \quad (7.26)$$

where L is unknown and the vector $\boldsymbol{\beta}_0$, that \mathbf{x}_1 depends on, could be non-linear. So we see that the linear class is larger for the MLM than for the extended GMM, which makes the MLM more flexible. However, it might be that the optimal estimate found in the MLM could also be found in the extended GMM; it depends on the linearity of $\boldsymbol{\beta}_0$.

H. Moritz used the MLM in the 1970's to introduce least-squares collocation. The collocation solution was linear for both $\bar{\mathbf{y}}$ and $\boldsymbol{\beta}_0$; so it could be described by the extended GMM. Schaffrin prefers the MLM to the extended GMM because it permits non-linear forms of $\boldsymbol{\beta}_0$.

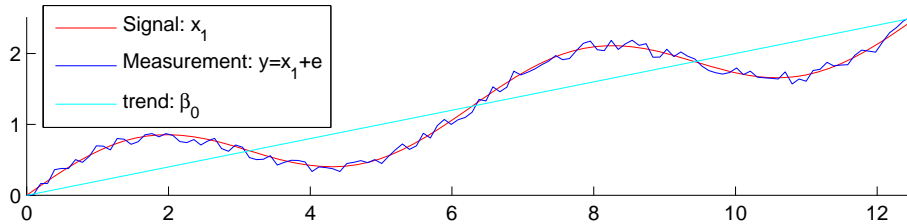


Figure 7.1: Sinusoidal signal containing random noise and linear trend

We now list some practical examples for the use of the MLM.

Example 1:

A typical application comes from signal theory. Here we are interested in a signal \mathbf{x}_1 , which may include a linear or non-linear component $\boldsymbol{\beta}_0$. Assuming no fixed

parameters $\hat{\boldsymbol{\xi}}_2$ and $A_1 = I$, the observation equations become $\mathbf{y} = \mathbf{x}_1 + \mathbf{e}$. Figure 7.1 illustrates this example.

Example 2:

The MLM can be applied to deformation analysis, for example the monitoring of bridges or dams. In this case, we have a-priori information about how we believe the structure should deform under load, but we are most interested in the actual deformation. We observe the signal plus noise; we must remove the noise (i.e., extract the signal from the noise).

Example 3:

Moritz applied the MLM to the gravity field problem. The normal gravity field is the prior information, which is the (known) expectation of a physical phenomenon that has random behavior. This physical phenomenon is called the disturbing gravity field.

7.4 Solutions for the Mixed Linear Model

To obtain solutions for the unknown parameters of the MLM, we start by deriving the BLUE for the non-random parameters $\boldsymbol{\xi}_2$. Substituting the equation for \mathbf{x}_1 from (7.24b) into the observation equation of (7.24) allows us to write the Mixed Linear Model (MLM) in an alternative form as

$$\bar{\mathbf{y}} - A_1\boldsymbol{\beta}_0 = A_2\boldsymbol{\xi}_2 + (\mathbf{e} - A_1\mathbf{e}_0), \quad (7.27a)$$

$$(\mathbf{e} - A_1\mathbf{e}_0) \sim (\mathbf{0}, \sigma_0^2 [P^{-1} + A_1Q_0A_1^T]). \quad (7.27b)$$

The MLM in (7.27) appears in the form of a GMM. The left side of (7.27a) is known and so are the characteristics of the combined error vector $\mathbf{e} - A_1\mathbf{e}_0$ on the right side. So, we can estimate $\boldsymbol{\xi}_2$ using least-squares principles via the following formula:

$$\hat{\boldsymbol{\xi}}_2 = [A_2^T(P^{-1} + A_1Q_0A_1^T)^{-1}A_2]^{-1}A_2^T(P^{-1} + A_1Q_0A_1^T)^{-1}(\bar{\mathbf{y}} - A_1\boldsymbol{\beta}_0). \quad (7.28)$$

The first inverted matrix in (7.28) is the cofactor matrix for $\hat{\boldsymbol{\xi}}_2$ so that the dispersion matrix of $\hat{\boldsymbol{\xi}}_2$ is given by

$$D\{\hat{\boldsymbol{\xi}}_2\} = \sigma_0^2 [A_2^T(P^{-1} + A_1Q_0A_1^T)^{-1}A_2]^{-1}. \quad (7.29)$$

An alternative form of the dispersion matrix is obtained as follows: By use of (A.15) we obtain

$$(P^{-1} + A_1Q_0A_1^T)^{-1} = P - PA_1(Q_0^{-1} + A_1^TPA_1)^{-1}A_1^TP, \quad (7.30a)$$

with

$$(Q_0^{-1} + A_1^TPA_1)^{-1} = (I_r + Q_0A_1^TPA_1)^{-1}Q_0. \quad (7.30b)$$

Upon substitution of (7.30a), together with (7.30b), into the dispersion formula (7.29) we get

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}_2\} &= \sigma_0^2 \{A_2^T [P - PA_1(I_r + Q_0A_1^T PA_1)^{-1}Q_0A_1^T P]A_2\}^{-1} = \\ &= \sigma_0^2 [N_{22} - N_{21}(I_r + Q_0N_{11})^{-1}Q_0N_{12}]^{-1} = \sigma_0^2 [N_{22} - N_{21}Q_0^{-1} + N_{11}]^{-1}N_{12}]^{-1} = \\ &= \sigma_0^2 N_{22}^{-1} + \sigma_0^2 N_{22}^{-1}N_{21}[(Q_0^{-1} + N_{11})N_{12}N_{22}^{-1}N_{21}]^{-1}N_{12}N_{22}^{-1}, \end{aligned}$$

or

$$\boxed{D\{\hat{\boldsymbol{\xi}}_2\} = \sigma_0^2 N_{22}^{-1} + \sigma_0^2 N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}Q_0N_{12}N_{22}^{-1}}, \quad (7.31)}$$

where

$$S_1 := N_{11} - N_{12}N_{22}^{-1}N_{21}. \quad (7.32)$$

Also, we have used the familiar relations

$$N_{ij} := A_i^T PA_j \quad \text{and} \quad \bar{\mathbf{c}}_i := A_i^T P\bar{\mathbf{y}}, \quad (7.33)$$

where the symbol $\bar{\mathbf{c}}_i$ is used below. To reach an alternative expression for $\boldsymbol{\xi}_2$, we use (7.30a) through (7.31) to modify (7.28) as follows:

$$\begin{aligned} \hat{\boldsymbol{\xi}}_2 &= [N_{22}^{-1} + N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}Q_0N_{12}N_{22}^{-1}] \cdot \\ &\cdot [A_2^T P - N_{21}(I_r + Q_0N_{11})^{-1}Q_0A_1^T P](\bar{\mathbf{y}} - A_1\boldsymbol{\beta}_0). \end{aligned} \quad (7.34)$$

For convenience, and for future reference, we also write

$$\hat{\boldsymbol{\xi}}_2 = G_2(\bar{\mathbf{y}} - A_1\boldsymbol{\beta}_0), \quad (7.35a)$$

with

$$\begin{aligned} G_2 &:= [N_{22}^{-1} + N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}Q_0N_{12}N_{22}^{-1}] \cdot \\ &\cdot [A_2^T P - N_{21}(I_r + Q_0N_{11})^{-1}Q_0A_1^T P]. \end{aligned} \quad (7.35b)$$

Expanding (7.35a) leads to

$$\begin{aligned} \hat{\boldsymbol{\xi}}_2 &= N_{22}^{-1}(\bar{\mathbf{c}}_2 - N_{21}\boldsymbol{\beta}_0) - \underline{N_{22}^{-1}N_{21}(I_r + Q_0N_{11})^{-1}Q_0(\bar{\mathbf{c}}_1 - N_{11}\boldsymbol{\beta}_0)} + \\ &\quad + N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}Q_0N_{12}N_{22}^{-1}(\bar{\mathbf{c}}_2 - N_{21}\boldsymbol{\beta}_0) - \\ &\quad - \underline{N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}Q_0N_{12}N_{22}^{-1}N_{21}(I_r + Q_0N_{11})^{-1}Q_0(\bar{\mathbf{c}}_1 - N_{11}\boldsymbol{\beta}_0)}. \end{aligned} \quad (7.36)$$

The single and double underlines in the second and fourth lines of the above equation are used to highlight similar terms. We may insert the identity matrix

$$(I_r + Q_0S_1)^{-1}(I_r + Q_0N_{11} - Q_0N_{12}N_{22}^{-1}N_{21}) = I_r \quad (7.37)$$

between the underlined terms in the second line, which, after some algebraic manipulation, leads to

$$\hat{\xi}_2 = N_{22}^{-1}(\bar{c}_2 - N_{21}\beta_0) - N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}Q_0(\bar{c}_1 - N_{12}N_{22}^{-1}\bar{c}_2 - S_1\beta_0). \quad (7.38)$$

After further algebraic manipulation, we can also write

$$\hat{\xi}_2 = N_{22}^{-1}\bar{c}_2 - N_{22}^{-1}N_{21}(I_r + Q_0S_1)^{-1}[Q_0(\bar{c}_1 - N_{12}N_{22}^{-1}\bar{c}_2) + \beta_0]. \quad (7.39)$$

In summary, we began with equation (7.27), which has the form of the GMM, and we applied least-squares criteria to reach a solution for ξ_2 . We know that LESS within the (full-rank) GMM is equivalent to BLUE. So, we claim that the various expressions of $\hat{\xi}_2$ above, beginning with (7.28), give the BLUE within the mixed linear model for the non-random (fixed) parameter vector ξ_2 .

From (7.27) we see that our solution will only lead to a prediction for $e - A_1e_0$. But what we need a prediction for e_0 so that we can predict x_1 . We can arrive at LESS for \tilde{e}_0 based on the following *Model of Condition Equations with Parameters* (see first set of Adjustment Notes):

$$\bar{y} - A_1\beta_0 = A_2\xi_2 + [I_n, -A_1] \begin{bmatrix} e \\ e_0 \end{bmatrix}, \quad (7.40a)$$

$$\begin{bmatrix} e \\ e_0 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & Q_0 \end{bmatrix} \right). \quad (7.40b)$$

This model leads to the following solution for the predicted random errors:

$$\begin{aligned} \begin{bmatrix} \tilde{e} \\ \tilde{e}_0 \end{bmatrix} &= \begin{bmatrix} P^{-1} & 0 \\ 0 & Q_0 \end{bmatrix} \begin{bmatrix} I_n \\ -A_1^T \end{bmatrix} \left([I_n, -A_1] \begin{bmatrix} P^{-1} & 0 \\ 0 & Q_0 \end{bmatrix} \begin{bmatrix} I_n \\ -A_1^T \end{bmatrix} \right)^{-1} \\ &\quad \cdot (\bar{y} - A_1\beta_0 - A_2\hat{\xi}_2) = \\ &= \begin{bmatrix} P^{-1} \\ -Q_0A_1^T \end{bmatrix} (P^{-1} + A_1Q_0A_1^T)^{-1} (\bar{y} - A_1\beta_0 - A_2\hat{\xi}_2) \Rightarrow \\ \begin{bmatrix} \tilde{e} \\ \tilde{e}_0 \end{bmatrix} &= \begin{bmatrix} P^{-1} \\ -Q_0A_1^T \end{bmatrix} [P - PA_1(I_r + Q_0N_{11})^{-1}Q_0A_1^T P] (\bar{y} - A_1\beta_0 - A_2\hat{\xi}_2). \quad (7.41) \end{aligned}$$

The second row of (7.41) provides the following formula for \tilde{e}_0 :

$$\begin{aligned}
\tilde{e}_0 &= -Q_0(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2) + (Q_0N_{11} + I_r - I_r)(I_r + Q_0N_{11})^{-1}Q_0A_1^T P \cdot \\
&\quad \cdot (\bar{y} - A_1\beta_0 - A_2\hat{\xi}_2) = \\
&= -Q_0(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2) + Q_0(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2) - \\
&\quad - (I_r + Q_0N_{11})^{-1}Q_0(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2) = \\
&= -(I_r + Q_0N_{11})^{-1}Q_0(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2) \Rightarrow \\
\tilde{e}_0 &= -Q_0(I_r + N_{11}Q_0)^{-1}(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2). \tag{7.42}
\end{aligned}$$

By comparing the first and second rows of (7.41), we immediately see \tilde{e}_0 as a function of \tilde{e} :

$$\tilde{e}_0 = -Q_0A_1^T P\tilde{e}. \tag{7.43}$$

Now, it is obvious from the MLM that we have $\tilde{x}_1 = \beta_0 - \tilde{e}_0$, which upon substitution of (7.42) yields

$$\boxed{\tilde{x}_1 = \beta_0 + Q_0(I_r + N_{11}Q_0)^{-1}(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2),} \tag{7.44}$$

or, alternatively,

$$\tilde{x}_1 = (I_r + Q_0N_{11})^{-1}Q_0A_1^T P(\bar{y}_1 - A_2\hat{\xi}_2) + [I_r - (I_r + Q_0N_{11})^{-1}Q_0N_{11}]\beta_0 = \tag{7.45a}$$

$$= (I_r + Q_0N_{11})^{-1}Q_0A_1^T P(\bar{y}_1 - A_2\hat{\xi}_2) + (I_r + Q_0N_{11})^{-1}\beta_0. \tag{7.45b}$$

Here we used the general relationship $(I + A)^{-1} = I - (I + A)^{-1}A$ in the last step to reach (7.45b). Note that we have arrived at the prediction \tilde{x}_1 strictly by least-squares principles. However, in this model we have the equivalence of LESS to the inhomBLIP (Best inhomogeneous Linear Predictor). The idea behind inhomBLIP is given in the following section.

7.5 Best Inhomogeneous Linear Predictor

The idea behind inhomBLIP is that in the class of linear predictors

$$\{L(\bar{y} - A_2\hat{\xi}_2) + \gamma \mid L \text{ is an } r \times n \text{ matrix, } \gamma \text{ is an } r \times 1 \text{ vector}\} \tag{7.46a}$$

the predictor

$$\begin{aligned}
\tilde{x}_1 &= (I_r + Q_0N_{11})^{-1}Q_0A_1^T P(\bar{y}_1 - A_2\hat{\xi}_2) + (I_r + Q_0N_{11})^{-1}\beta_0 = \\
&= L_1(\bar{y}_1 - A_2\hat{\xi}_2) + \gamma_1 \tag{7.46b}
\end{aligned}$$

has minimum mean square prediction error (MSPE). That is,

$$\begin{aligned} \text{tr MSPE}\{\tilde{\mathbf{x}}_1\} &= \text{tr } E\{(\tilde{\mathbf{x}}_1 - \mathbf{x}_1)(\tilde{\mathbf{x}}_1 - \mathbf{x}_1)^T\} = \\ &= \text{tr } E\{[L_1(\bar{\mathbf{y}}_1 - A_2\hat{\boldsymbol{\xi}}_2) + \boldsymbol{\gamma}_1 - \mathbf{x}_1][L_1(\bar{\mathbf{y}}_1 - A_2\hat{\boldsymbol{\xi}}_2) + \boldsymbol{\gamma}_1 - \mathbf{x}_1]^T\} = \min_{L_1, \boldsymbol{\gamma}_1} \end{aligned}$$

(7.46c)

The variables L_1 and $\boldsymbol{\gamma}_1$ are defined as follows:

$$L_1 := (I_r + Q_0 N_{11})^{-1} Q_0 A_1^T P = Q_0 A_1^T (P^{-1} + A_1 Q_0 A_1^T)^{-1}, \quad (7.47a)$$

$$\boldsymbol{\gamma}_1 := (I_r + Q_0 N_{11})^{-1} \boldsymbol{\beta}_0 = \boldsymbol{\beta}_0 - L_1 A_1 \boldsymbol{\beta}_0. \quad (7.47b)$$

The minimization of (7.46c) is not developed further here. However, we do note that the predictor $\tilde{\mathbf{x}}_1$ is automatically weakly unbiased in the sense that

$$E\{\tilde{\mathbf{x}}_1\} = (I_r + Q_0 N_{11})^{-1} [Q_0 A_1^T P \cdot E\{\bar{\mathbf{y}}_1 - A_2 \hat{\boldsymbol{\xi}}_2\} + E\{\boldsymbol{\beta}_0\}] = \quad (7.48a)$$

$$= (I_r + Q_0 N_{11})^{-1} [Q_0 A_1^T P A_1 \boldsymbol{\beta}_0 + \boldsymbol{\beta}_0] = \boldsymbol{\beta}_0 \quad (7.48b)$$

for the given vector $\boldsymbol{\beta}_0$. Note that (7.48b) does not necessarily hold for any arbitrary vector $\boldsymbol{\beta}_0$, but rather for the given $\boldsymbol{\beta}_0$, hence the term *weakly unbiased*.

Let us now consider in detail the mean-square prediction error MSPE of $\{\tilde{\mathbf{x}}_1\}$. Because of unbiasedness, we can write

$$\text{MSPE}\{\tilde{\mathbf{x}}_1\} = D\{\tilde{\mathbf{x}}_1 - \mathbf{x}_1\}. \quad (7.49a)$$

Also, because the vector differences $\tilde{\mathbf{x}}_1 - \mathbf{x}_1$ and $\mathbf{e}_0 - \tilde{\mathbf{e}}_0$ only differ by $\boldsymbol{\beta}_0$, we have

$$\text{MSPE}\{\tilde{\mathbf{x}}_1\} = D\{\mathbf{e}_0 - \tilde{\mathbf{e}}_0\} = D\{\mathbf{e}_0\} - C\{\mathbf{e}_0, \tilde{\mathbf{e}}_0\} - C\{\tilde{\mathbf{e}}_0, \mathbf{e}_0\} + D\{\tilde{\mathbf{e}}_0\}. \quad (7.49b)$$

Let us compute the last four terms of (7.49b) individually.

$$D\{\mathbf{e}_0\} = \sigma_0^2 Q_0 \quad (7.50a)$$

In computing $D\{\tilde{\mathbf{e}}_0\}$, we first write the dispersion for the term $\bar{\mathbf{y}} - \widetilde{A_1 \boldsymbol{\beta}_0} - A_2 \hat{\boldsymbol{\xi}}_2$. This term, as we have already seen, is equivalent to the prediction $\mathbf{e} - A_1 \mathbf{e}_0$. Also, equation (7.40b) implies no covariance between $\bar{\mathbf{y}} - A_1 \boldsymbol{\beta}_0$ and $A_2 \hat{\boldsymbol{\xi}}_2$, i.e.

$$D\{\bar{\mathbf{y}} - A_1 \boldsymbol{\beta}_0 - A_2 \hat{\boldsymbol{\xi}}_2\} = D\{\widetilde{\mathbf{e} - A_1 \mathbf{e}_0}\} = D\{\bar{\mathbf{y}} - A_1 \boldsymbol{\beta}_0\} - D\{A_2 \hat{\boldsymbol{\xi}}_2\}. \quad (7.50b)$$

Now making use of (7.42), we can write

$$D\{\tilde{\mathbf{e}}_0\} = Q_0 (I_r + N_{11} Q_0)^{-1} A_1^T P D\{\bar{\mathbf{y}} - A_1 \boldsymbol{\beta}_0 - A_2 \hat{\boldsymbol{\xi}}_2\} P A_1 Q_0 (I_r + N_{11} Q_0)^{-1}. \quad (7.50c)$$

7.6. ALTERNATIVE NORMAL EQUATIONS FOR THE MIXED LINEAR MODEL 83

For the covariance terms, we have $C\{\mathbf{e}_0, \tilde{\mathbf{e}}_0\} = C\{\tilde{\mathbf{e}}_0, \mathbf{e}_0\}^T$, and with the help of (7.35b), we write the covariance $C\{\mathbf{e}_0, \tilde{\mathbf{e}}_0\}$ as follows:

$$\begin{aligned} C\{\mathbf{e}_0, \tilde{\mathbf{e}}_0\} &= Q_0(I_r + N_{11}Q_0)^{-1}A_1P \cdot C\{\bar{\mathbf{y}} - A_1\boldsymbol{\beta}_0 - A_2\hat{\boldsymbol{\xi}}_2, \mathbf{e}_0\} = \\ &= Q_0(I_r + N_{11}Q_0)^{-1}A_1P \cdot C\{(I_n - A_2G_2)(\bar{\mathbf{y}} - A_1\boldsymbol{\beta}_0), \mathbf{e}_0\} = \\ &= Q_0(I_r + N_{11}Q_0)^{-1}A_1P(I_n - A_2G_2) \cdot C\{(A_2\boldsymbol{\xi}_2 + \mathbf{e} - A_1\mathbf{e}_0), \mathbf{e}_0\} = \\ &= -[Q_0(I_r + N_{11}Q_0)^{-1}A_1P](I_n - A_2G_2)A_1(\sigma_0^2Q_0) = C\{\tilde{\mathbf{e}}_0, \mathbf{e}_0\}^T \end{aligned} \quad (7.50d)$$

To recap, equation (7.49b) is comprised of equations (7.50a) through (7.50d). The way we would actually form the dispersion matrix is as follows:

$$\begin{aligned} D\left\{\begin{bmatrix} \tilde{\mathbf{x}}_1 - \mathbf{x}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix}\right\} &= \begin{bmatrix} I_r + Q_0N_{11} & Q_0N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} Q_0 & 0 \\ 0 & I_{m-r} \end{bmatrix} = \\ &= \begin{bmatrix} \text{MSPE}\{\tilde{\mathbf{x}}_1\} & C\{\tilde{\mathbf{x}}_1 - \mathbf{x}_1, \hat{\boldsymbol{\xi}}_2\} \\ C\{\hat{\boldsymbol{\xi}}_2, \tilde{\mathbf{x}}_1 - \mathbf{x}_1\} & D\{\hat{\boldsymbol{\xi}}_2\} \end{bmatrix}. \end{aligned} \quad (7.51)$$

Here we stress that we are not interested in the dispersion $D\{\tilde{\mathbf{x}}_1\}$, since this is an indicator of variation between $\tilde{\mathbf{x}}_1$ and $E\{\tilde{\mathbf{x}}_1\}$. Rather we are interested in the variation between $\tilde{\mathbf{x}}_1$ and the true variable \mathbf{x}_1 , a concept that the following formula makes clear:

$$\text{MSPE}\{\tilde{\mathbf{x}}_1\} = E\{(\tilde{\mathbf{x}}_1 - \mathbf{x}_1)(\tilde{\mathbf{x}}_1 - \mathbf{x}_1)^T\} = D\{\tilde{\mathbf{x}}_1 - \mathbf{x}_1\}, \quad (7.52)$$

since $E\{\tilde{\mathbf{x}}_1 - \mathbf{x}_1\} = \mathbf{0}$.

7.6 Alternative Normal Equations for the Mixed Linear Model

In the previous section we showed different, but equivalent, expressions for the predicted parameter vector $\tilde{\mathbf{x}}_1$. All of these expressions depended on the estimate $\hat{\boldsymbol{\xi}}_2$ for the fixed parameters. Our goal in this section is to find a system of normal equations that will permit the random parameters $\tilde{\mathbf{x}}_1$ to be predicted without the need to compute the fixed parameters $\hat{\boldsymbol{\xi}}_2$. With reference to (7.28), we begin with the following orthogonality relations, which are analogous to $A^T P \tilde{\mathbf{e}} = \mathbf{0}$ in the GMM:

$$A_2^T(P^{-1} + A_1I_rQ_0A_1^T)^{-1}(\tilde{\mathbf{e}} - A_1\tilde{\mathbf{e}}_0) = \quad (7.53a)$$

$$= A_2^T[P - PA_1(I_r + Q_0A_1^TPA_1)^{-1}Q_0A_1^TP](\tilde{\mathbf{e}} - A_1\tilde{\mathbf{e}}_0) = \quad (7.53b)$$

$$\begin{aligned} &= A_2^TP\tilde{\mathbf{e}} - N_{21}\tilde{\mathbf{e}}_0 - N_{21}(I_r + Q_0N_{11})^{-1}Q_0A_1^TP\tilde{\mathbf{e}} + \\ &\quad + N_{21}(I_r + Q_0N_{11})^{-1}Q_0N_{11}\tilde{\mathbf{e}}_0 = \mathbf{0}. \end{aligned} \quad (7.53c)$$

Now we make use of the relations

$$\tilde{\mathbf{e}} = (\bar{\mathbf{y}} - A_1 \tilde{\mathbf{x}}_1 - A_2 \hat{\boldsymbol{\xi}}_2) \quad (7.54a)$$

and

$$\tilde{\mathbf{e}}_0 = \boldsymbol{\beta}_0 - \tilde{\mathbf{x}}_1 \quad (7.54b)$$

in order to write

$$\begin{aligned} \bar{\mathbf{c}}_2 &= N_{21} \tilde{\mathbf{x}}_1 + N_{22} \hat{\boldsymbol{\xi}}_2 + N_{21}(\boldsymbol{\beta}_0 - \tilde{\mathbf{x}}_1) + N_{21}(I_r + Q_0 N_{11})^{-1} Q_0 A_1^T P(\tilde{\mathbf{e}} - A_1 \tilde{\mathbf{e}}_0) = \\ &= N_{21} \boldsymbol{\beta}_0 + N_{22} \hat{\boldsymbol{\xi}}_2 + A_2^T P \tilde{\mathbf{e}} - N_{21} \tilde{\mathbf{e}}_0. \end{aligned} \quad (7.55)$$

Also, multiplying the residuals in (7.54) by $A_1^T P$ leads to

$$\bar{\mathbf{c}}_1 = N_{11} \boldsymbol{\beta}_0 + N_{12} \hat{\boldsymbol{\xi}}_2 + A_1^T P \tilde{\mathbf{e}} - N_{11} \tilde{\mathbf{e}}_0. \quad (7.56)$$

Now we introduce a new symbol

$$\hat{\boldsymbol{\nu}} := A_1^T P \tilde{\mathbf{e}} \quad (7.57)$$

and note that $A_2^T P \tilde{\mathbf{e}} = \mathbf{0}$. Combining equations (7.55) through (7.57) into a single systems of equations yields the normal equations

$$\begin{bmatrix} N_{11} & N_{12} & I_r \\ N_{21} & N_{22} & 0 \\ I_r & 0 & -Q_0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \hat{\boldsymbol{\xi}}_2 \\ \hat{\boldsymbol{\nu}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{c}}_1 \\ \bar{\mathbf{c}}_2 \\ \boldsymbol{\beta}_0 \end{bmatrix}. \quad (7.58)$$

The solution to (7.58) yields both $\tilde{\mathbf{x}}_1$ and $\hat{\boldsymbol{\xi}}_2$; it also allows us to invert the normal-equations matrix when Q_0 is singular. If Q_0 is *regular* (non-singular), we may reduce the size of the system as follows:

$$\begin{bmatrix} P_0 + N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{c}}_1 + P_0 \boldsymbol{\beta}_0 \\ \bar{\mathbf{c}}_2 \end{bmatrix}. \quad (7.59)$$

Consistent with previous claims, the solution to (7.59) yields inhomBLIP for $\tilde{\mathbf{x}}_1$ and BLUE for $\hat{\boldsymbol{\xi}}_2$. It also leads to $\hat{\boldsymbol{\nu}} = P_0(\tilde{\mathbf{x}}_1 - \boldsymbol{\beta}_0)$.

After inverting the matrix on the left side of (7.59) (see Appendix A for inversion formula), we can write the prediction for \mathbf{x}_1 as follows:

$$\begin{aligned} \tilde{\mathbf{x}}_1 &= (P_0 + N_{11})^{-1}(\bar{\mathbf{c}}_1 + P_0 \boldsymbol{\beta}_0) + (P_0 + N_{11})^{-1} \cdot \\ &\quad \cdot N_{12} [N_{22} - N_{21} (P_0 + N_{11})^{-1} N_{12}]^{-1} \cdot \\ &\quad \cdot [N_{21} (P_0 + N_{11})^{-1} (\bar{\mathbf{c}}_1 + P_0 \boldsymbol{\beta}_0) - \bar{\mathbf{c}}_2]. \end{aligned} \quad (7.60a)$$

7.6. ALTERNATIVE NORMAL EQUATIONS FOR THE MIXED LINEAR MODELS 85

Likewise, the estimation for ξ_2 is given by

$$\hat{\xi}_2 = -[N_{22} - N_{21}(P_0 + N_{11})^{-1}N_{12}]^{-1}[N_{21}(P_0 + N_{11})^{-1}(\bar{c}_1 + P_0\beta_0) - \bar{c}_2]. \quad (7.60b)$$

Note that (7.60b) is equivalent to (7.28), which can be seen by confirming the following two equivalences:

$$[A_2^T(P^{-1} + A_1Q_0A_1^T)^{-1}A_2]^{-1} = [N_{22} - N_{21}(P_0 + N_{11})^{-1}N_{12}]^{-1}, \quad (7.61a)$$

$$N_{21}(P_0 + N_{11})^{-1}(\bar{c}_1 + P_0\beta_0) = \bar{c}_2 + A_2^T(P^{-1} + A_1Q_0A_1^T)^{-1}(A_1\beta_0 - \bar{y}). \quad (7.61b)$$

Combining (7.60a) and (7.60b) yields the following expression for the predicted random effects vector \tilde{x}_1 as a function of the estimated fixed parameters $\hat{\xi}_2$:

$$\tilde{x}_1 = \beta_0 + (P_0 + N_{11})^{-1}(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2), \quad (7.62a)$$

which agrees with (7.44).

Recall that \bar{y} , \bar{c}_1 , and \bar{c}_2 are numerically equivalent to y , c_1 , and c_2 , respectively, but they have different stochastic properties due to the randomness of x_1 (see (7.33) and (7.22)). If we factor out the term $A_1^T P$ from second parenthetical expression in (7.62a), we get

$$\tilde{x}_1 = \beta_0 + (P_0 + N_{11})^{-1}A_1^T P(\bar{y} - A_1\beta_0 - A_2\hat{\xi}_2), \quad (7.62b)$$

which can be re-written as

$$\tilde{x}_1 = \beta_0 + C\{x_1, y\}[D\{y\}]^{-1}(\bar{c}_1 - N_{11}\beta_0 - N_{12}\hat{\xi}_2), \quad (7.63)$$

since

$$C\{x_1, y\} = \sigma_0^2 P_0^{-1} A_1^T \quad (7.64a)$$

and

$$D\{y\} = \sigma_0^2 (A_1 P_0^{-1} A_1^T + P^{-1}). \quad (7.64b)$$

We conclude this chapter by commenting that LESS from the extended GMM yields the same numerical results as LESS from the MLM, but the interpretation is completely different. In the mixed linear model, \tilde{x}_1 is predicted, while $\hat{\xi}_1$ is estimated within the extended GMM. Therefore, we are not interested in the dispersion of \tilde{x}_1 itself but rather its MSPE.

$$D\left\{\begin{bmatrix} \tilde{x}_1 - x_1 \\ \hat{\xi}_2 \end{bmatrix}\right\} = \sigma_0^2 \begin{bmatrix} P_0 + N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \neq D\left\{\begin{bmatrix} \tilde{x}_1 \\ \hat{\xi}_2 \end{bmatrix}\right\}, \quad (7.65a)$$

where

$$D\{\tilde{x}_1 - x_1\} = \text{MSPE}\{\tilde{x}_1\}. \quad (7.65b)$$

Finally, we state that the estimated variance component as shown below is the Best Invariant Quadratic Uniformly Unbiased Estimate within the MLM.

$$\hat{\sigma}_0^2 = \frac{(\tilde{\mathbf{e}} - A_1 \tilde{\mathbf{e}}_0)^T (P^{-1} + A_1 Q_0 A_1^T) (\tilde{\mathbf{e}} - A_1 \tilde{\mathbf{e}}_0)}{(n - m + r)} \quad (7.66a)$$

Or, alternatively, using (7.40b), we can write

$$\hat{\sigma}_0^2 = \frac{\tilde{\mathbf{e}}^T (I_n + Q_0 N_{11})^T (P^{-1} + A_1 Q_0 A_1^T) (I_n + Q_0 N_{11}) \tilde{\mathbf{e}}}{(n - m + r)}. \quad (7.66b)$$

Here, we have assumed that $\text{rk } A = m$. Also, recall that for the MLM, r is the dimension of \mathbf{x}_1 .

Chapter 8

The Dynamic Linear Model

The Dynamic Linear Model (DLM) is a linearized model that consists of an *initial value problem (IVP)* and *observed variables*. It can be viewed as a model of *observation equations with differential constraints*. The constraints are not imposed on the parameters but rather on the parameter changes (i.e., changes that occur in time).

After linearizing and discretizing the original differential equations, we arrive at the following (differential) *state equation* at epoch 1 as a function of the *state variables* \mathbf{x}_0 at epoch 0:

$$\mathbf{x}_1 = \Phi_0 \mathbf{x}_0 + \mathbf{u}_1. \quad (8.1)$$

The symbols are defined as follows:

\mathbf{x}_i is an $m \times 1$ unknown *state vector* at epoch $i = 0, 1$.

Φ_0 is an $m \times m$ given *state transition matrix*.

\mathbf{u}_1 is an $m \times 1$ random noise vector.

From (8.1) we can write the following *stochastic constraints* for \mathbf{x}_0 and \mathbf{x}_1 :

$$[I_m, -\Phi_0] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_0 \end{bmatrix} = \mathbf{u}_1. \quad (8.2)$$

We mentioned already that the DLM is an initial value problem. The initial conditions are expressed by

$$\tilde{\mathbf{x}}_0 = \mathbf{x}_0 + \mathbf{e}_0^0. \quad (8.3)$$

Here the subscript 0 denotes epoch 0, while the superscript 0 denotes the initial condition. We note that the subscript for \mathbf{u}_1 is sometimes shown as 0 rather than 1 in the literature. This is merely a convention, as the variable \mathbf{u}_1 represents the

noise (random error) of the difference of the states \mathbf{x}_1 and $\Phi_0\mathbf{x}_0$, between epochs 1 and 0, respectively. Also, for any epoch i , the state transition matrix Φ_i is unique; that is, it changes from epoch to epoch. Our knowledge of the initial state vector $\tilde{\mathbf{x}}_0$ can be improved by using a *backward filter*, but its use is not possible in real-time applications.

Equations (8.2) and (8.3) constitute the IVP in discrete form. Their stochastic properties (expectation and dispersion) are written as

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{e}_0^0 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Theta_1 & 0 \\ 0 & \Sigma_0^0 \end{bmatrix} \right). \quad (8.4)$$

At this stage we have $2m$ unknowns and $2m$ equations (owing to the unknown $m \times 1$ vectors \mathbf{x} and \mathbf{u}). Since there is no redundancy in the model, we cannot determine the unknowns. The redundancy enters the model via the following observation equations:

$$\mathbf{y}_1 = A_1\mathbf{x}_1 + \mathbf{e}_1, \quad \mathbf{y}_1 \in \mathbb{R}^n, \quad A_1 \in \mathbb{R}^{n \times m}, \quad (8.5a)$$

with the stochastic model

$$\mathbf{e}_1 \sim (\mathbf{0}, \Sigma_1), \quad C\{\mathbf{e}_1, \mathbf{u}_1\} = \mathbf{0}, \quad C\{\mathbf{e}_1, \mathbf{e}_0^0\} = \mathbf{0}. \quad (8.5b)$$

Thus we see that the DLM is comprised of three components: observation equations (8.5a), state equations (8.1), and initial conditions (8.3). We may combine all three parts of the model into one succinct statement as follows:

$$\left. \begin{array}{l} \mathbf{y}_1 = A_1\mathbf{x}_1 + \mathbf{e}_1 \\ \mathbf{x}_1 = \Phi_0\mathbf{x}_0 + \mathbf{u}_1 \\ \tilde{\mathbf{x}}_0 = \mathbf{x}_0 + \mathbf{e}_0^0 \end{array} \right\} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{u}_1 \\ \mathbf{e}_0^0 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \Theta_1 & 0 \\ 0 & 0 & \Sigma_0^0 \end{bmatrix} \right). \quad (8.6)$$

Note that a (common) variance component, σ_0^2 , is embedded within the covariance matrices Σ_1 , Θ_1 , and Σ_0^0 .

Our goal is to predict the unknown state vector \mathbf{x}_1 and determine its mean squared error (MSE) matrix. The relations between the predicted variables (with tildes) and true variables (without tildes) are described by

$$\tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \mathbf{e}_1^0, \quad (8.7a)$$

and

$$D\{\mathbf{e}_1^0 = \tilde{\mathbf{x}}_1 - \mathbf{x}_1\} = \text{MSPE}\{\tilde{\mathbf{x}}_1\} = \Sigma_1^0. \quad (8.7b)$$

We may also wish to use “backward filtering” to compute the prediction $\tilde{\tilde{\mathbf{x}}}_0$ for the initial state vector \mathbf{x}_0 . However, as mentioned previously, this is not feasible, or even possible, in real-time problems.

We now introduce a new prediction variable $\tilde{\mathbf{x}}_1$ by combining the state equation and initial condition. This variable represents our prior knowledge about the state vector

$$\mathbf{x}_1 = \Phi_0 \mathbf{x}_0 + \mathbf{u}_1 = \Phi_0(\tilde{\mathbf{x}}_0 - \mathbf{e}_0^0) + \mathbf{u}_1, \quad (8.8a)$$

which leads to the predictor

$$\tilde{\mathbf{x}}_1 := \Phi_0 \tilde{\mathbf{x}}_0 = \mathbf{x}_1 - (\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0). \quad (8.8b)$$

We call the term in parenthesis in (8.8b) the *combined error*. Note that

$$E\{\tilde{\mathbf{x}}_1\} = E\{\mathbf{x}_1\}, \quad (8.9a)$$

since

$$E\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\} = \mathbf{0}. \quad (8.9b)$$

We note that the ‘‘prior information’’ in the DLM is more complicated than in the Mixed Linear Model (MLM). Here, we must determine the predicted state vector $\tilde{\mathbf{x}}_1$ (which is different than $\tilde{\mathbf{x}}_1$) based on the new observations. The variable $\tilde{\mathbf{x}}_1$ is the best prediction based on the state equation and the initial condition only. We essentially blend the prior knowledge $\tilde{\mathbf{x}}_1$ with the observations \mathbf{y}_1 . With this *fusion* of information we are able to determine the prediction $\tilde{\mathbf{x}}_1$. Note that all of the redundancy in the model comes from the observation equations. The initial value problem is just uniquely solvable.

This fusion process is called *Kalman filtering*. It can be done in real time, in which case the number of state parameters may be restricted by the speed of the computer processor. The key is to have good information about the state equation, not only the state transition matrix Φ_0 but also the associated covariance matrix, Θ_1 , of the state equation. The information contained in matrices Φ_0 and Θ_1 describes how we think the dynamic system behaves. Our knowledge of the system is introduced as a differential equation, which is *linearized* and *discretized* to form the state equation. This work must be done before the adjustment stage.

With the introduction of (8.8b), we may write an equivalent version of the DLM as follows:

$$\left. \begin{array}{l} \mathbf{y}_1 = A_1 \mathbf{x}_1 + \mathbf{e}_1 \\ \tilde{\mathbf{x}}_1 = \mathbf{x}_1 - (\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0) \end{array} \right\} \left[\begin{array}{c} \mathbf{e}_1 \\ -(\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0) \end{array} \right] \sim \left(\left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array} \right], \left[\begin{array}{cc} \Sigma_1 & 0 \\ 0 & \Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T \end{array} \right] \right) \quad (8.10)$$

The model (8.10) essentially combines observation equations (in \mathbf{y}_1) with pseudo-observation equations (in $\tilde{\mathbf{x}}_1$). But here we are dealing with *random effects*, so the DLM is not an extended GMM but rather is essentially an *extended random effects model* (REM).

The LEast-Squares Solution (LESS) within the DLM is equivalent to the inhomBLIP of \mathbf{x}_1 . Based on the model (8.10), we can write the least-squares normal equations directly as follows:

$$\begin{aligned} \left[\begin{array}{c|c} A_1^T & I_m \end{array} \right] \left[\begin{array}{c|c} \Sigma_1^{-1} & 0 \\ \hline 0 & (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1} \end{array} \right] \begin{bmatrix} A_1 \\ I_m \end{bmatrix} \tilde{\mathbf{x}}_1 = \\ = \left[\begin{array}{c|c} A_1^T \Sigma_1^{-1} & (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1} \end{array} \right] \begin{bmatrix} \mathbf{y}_1 \\ \tilde{\mathbf{x}}_1 \end{bmatrix}. \end{aligned} \quad (8.11)$$

Solving the normal equations results in

$$\tilde{\mathbf{x}}_1 = [A_1^T \Sigma_1^{-1} A_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1}]^{-1} [A_1^T \Sigma_1^{-1} \mathbf{y}_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1} \tilde{\mathbf{x}}_1]. \quad (8.12)$$

Then, the following steps lead to the isolation of $\tilde{\mathbf{x}}_1$:

$$\begin{aligned} \tilde{\mathbf{x}}_1 &= [A_1^T \Sigma_1^{-1} A_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1}]^{-1} \\ &\cdot [A_1^T \Sigma_1^{-1} \mathbf{y}_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1} \tilde{\mathbf{x}}_1 + A_1^T \Sigma_1^{-1} A_1 \tilde{\mathbf{x}}_1 - A_1^T \Sigma_1^{-1} A_1 \tilde{\mathbf{x}}_1] \Rightarrow \end{aligned} \quad (8.13a)$$

$$\tilde{\mathbf{x}}_1 = \tilde{\mathbf{x}}_1 + [A_1^T \Sigma_1^{-1} A_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1}]^{-1} A_1^T \Sigma_1^{-1} (\mathbf{y}_1 - A_1 \tilde{\mathbf{x}}_1), \quad (8.13b)$$

or

$$\tilde{\mathbf{x}}_1 = \tilde{\mathbf{x}}_1 + K_1 \mathbf{z}_1. \quad (8.13c)$$

Here, the $m \times n$ matrix

$$K_1 := [A_1^T \Sigma_1^{-1} A_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1}]^{-1} A_1^T \Sigma_1^{-1} \quad (8.13d)$$

is called *Kalman gain matrix*, and the $n \times 1$ vector

$$\mathbf{z}_1 := (\mathbf{y}_1 - A_1 \tilde{\mathbf{x}}_1) \quad (8.13e)$$

is called the *innovation*.

The form of the Kalman gain matrix in (8.13d) is useful for the case where the dimension, m , of the state vector is smaller than the number of observations n .

We may write alternative forms of the solution as follows:

$$\begin{aligned} \tilde{\mathbf{x}}_1 - \tilde{\mathbf{x}}_1 &= [I_m + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T \Sigma_1^{-1} A_1]^{-1} \\ &\cdot (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T \Sigma_1^{-1} (\mathbf{y}_1 - A_1 \tilde{\mathbf{x}}_1) = \end{aligned} \quad (8.14a)$$

$$= (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T [\Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} (\mathbf{y}_1 - A_1 \tilde{\mathbf{x}}_1). \quad (8.14b)$$

In (8.14a) we have used the the relations (A.22a) and (A.22c), and in (8.14b) we have used the relations (A.21a) and (A.21e). Both equations (8.14a) and (8.14b) are in the form of an *update*. However, equation (8.14a) requires the inversion of an $m \times m$ matrix, whereas equation (8.14b) requires the inversion of an $n \times n$ matrix. Oftentimes, in real-time applications, the number of observations n at a given epoch is small (perhaps only 1) compared to the number of state parameters m . In such a case, equation (8.14b) would be preferred over equation (8.14a).

We note that in the technical literature $\tilde{\mathbf{x}}_1$ is called the *filtered state*, while $\check{\mathbf{x}}_1$ is called the *predicted state*. However, in the statistical literature, $\tilde{\mathbf{x}}_1$ represents the *best prediction*. It is this best prediction $\tilde{\mathbf{x}}_1$ that we are interested in.

We summarize the various forms of the Kalman gain matrix appearing in the above formulas as follows:

$$K_1 = [A_1^T \Sigma_1^{-1} A_1 + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T)^{-1}]^{-1} A_1^T \Sigma_1^{-1} = \quad (8.15a)$$

$$= [I_m + (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T \Sigma_1^{-1} A_1]^{-1} (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T \Sigma_1^{-1} = \quad (8.15b)$$

$$= (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T [\Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1}. \quad (8.15c)$$

By combining the two equations in (8.10), we can alternatively express the innovation vector as

$$\mathbf{z}_1 := (\mathbf{y}_1 - A_1 \check{\mathbf{x}}_1) = \mathbf{e}_1 + A_1 (\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0) = \left[I_n \mid A_1 \mid -A_1 \Phi_0 \right] \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{u}_1 \\ \mathbf{e}_0^0 \end{bmatrix}. \quad (8.16)$$

The dispersion of the innovation vector is readily apparent from (8.16) as

$$\begin{aligned} D\{\mathbf{z}_1\} &= \left[I_n \mid A_1 \mid -A_1 \Phi_0 \right] \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \Theta_1 & 0 \\ 0 & 0 & \Sigma_0^0 \end{bmatrix} \begin{bmatrix} I_n \\ A_1^T \\ -\Phi_0^T A_1^T \end{bmatrix} = \\ &= \Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T. \end{aligned} \quad (8.17)$$

We may express the stochastic properties of the innovation \mathbf{z}_1 more concisely as

$$\mathbf{z}_1 \sim (\mathbf{0}, D\{\mathbf{z}_1\}) \text{ and } C\{\mathbf{z}_i, \mathbf{z}_j\} = \mathbf{0} \text{ for } i \neq j. \quad (8.18)$$

The statement of zero correlation means that the innovative sequence (from epoch to epoch) is uncorrelated. The expectation $E\{\mathbf{z}_i\} = \mathbf{0}$ should be tested for. If, through statistical testing, the expectation is found to be non-zero, this means that the state equations are inconsistent with the observation equations, and it means that the state equations might need to be modified.

The familiar *model of condition equations* (see first set of Adjustment Notes), along with the LESS for the residual vector is given by

$$\mathbf{w} := B\mathbf{e}, \quad (8.19)$$

and

$$\tilde{\mathbf{e}} = P^{-1}B^T(BP^{-1}B^T)^{-1}\mathbf{w}. \quad (8.20)$$

Comparing (8.16) with the first part of (8.20), we see that the innovation vector is in the form of the model of condition equations. Thus, in accordance with the LESS within that model, we can immediately write the solution to the vector of predicted errors as

$$\begin{aligned} \begin{bmatrix} \tilde{\mathbf{e}}_1 \\ \tilde{\mathbf{u}}_1 \\ \tilde{\mathbf{e}}_0^0 \end{bmatrix} &= \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \Theta_1 & 0 \\ 0 & 0 & \Sigma_0^0 \end{bmatrix} \begin{bmatrix} I_n \\ A_1^T \\ -\Phi_0^T A_1^T \end{bmatrix} \\ &\cdot \left(\begin{bmatrix} I_n & | & A_1 & | & -A_1 & | & \Phi_0 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \Theta_1 & 0 \\ 0 & 0 & \Sigma_0^0 \end{bmatrix} \begin{bmatrix} I_n \\ A_1^T \\ -\Phi_0^T A_1^T \end{bmatrix} \right)^{-1} \mathbf{z}_1, \end{aligned} \quad (8.21a)$$

or

$$\begin{bmatrix} \tilde{\mathbf{e}}_1 \\ \tilde{\mathbf{u}}_1 \\ \tilde{\mathbf{e}}_0^0 \end{bmatrix} = \begin{bmatrix} \Sigma_1 \\ \Theta_1 A_1^T \\ -\Sigma_0^0 \Phi_0^T A_1^T \end{bmatrix} [\Sigma_1 + A_1(\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} \mathbf{z}_1. \quad (8.21b)$$

If we substitute the predicted errors from (8.21b) into the second equation of (8.10), we arrive at

$$\tilde{\mathbf{x}}_1 = \tilde{\tilde{\mathbf{x}}}_1 + (\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0), \quad (8.22)$$

which leads to the same update formula found in (8.14b).

We mentioned earlier that backwards filtering can be used to obtain a better prediction of the initial state vector $\tilde{\mathbf{x}}_0$, though this is usually not feasible in real-time applications. Substituting the predicted random error vector $\tilde{\mathbf{e}}_0^0$ of (8.21b) into the third equation of (8.6), and making use of (8.13e) for \mathbf{z}_1 , allows us to write the backwards filter in the form of an update to $\tilde{\mathbf{x}}_0$ as follows:

$$\tilde{\tilde{\mathbf{x}}}_0 = \tilde{\mathbf{x}}_0 + \tilde{\mathbf{e}}_0^0 = \tilde{\mathbf{x}}_0 - \Sigma_0^0 \Phi_0^T A_1^T [\Sigma_1 + A_1(\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} (\mathbf{y}_1 - A_1 \Phi_0 \tilde{\mathbf{x}}_0). \quad (8.23)$$

In order to form the model for the next interval, we need the covariance matrix Σ_1^0 . This matrix is defined as

$$\Sigma_1^0 := \text{MSPE}\{\tilde{\mathbf{x}}_1\} = D\{\tilde{\mathbf{x}}_1 - \mathbf{x}_1\} = D\{(\tilde{\mathbf{x}}_1 - \tilde{\tilde{\mathbf{x}}}_1) - (\mathbf{x}_1 - \tilde{\tilde{\mathbf{x}}}_1)\}. \quad (8.24)$$

Referring to (8.8b), we may write a vector difference depending on unknown vectors \mathbf{x}_1 , \mathbf{u}_1 , and \mathbf{e}_0^0 as

$$\mathbf{x}_1 - \tilde{\mathbf{x}}_1 = \mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0, \quad (8.25a)$$

which implies the following vector difference based on corresponding predicted variables:

$$\tilde{\mathbf{x}}_1 - \tilde{\tilde{\mathbf{x}}}_1 = \tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0. \quad (8.25b)$$

So, with help of (8.25b), we may replace the differences in (8.24) with linear combinations of the residual vectors as in the following:

$$\begin{aligned} D\{(\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0) - (\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0)\} = \\ D\{\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0\} - C\{(\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0), (\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0)\} - \\ - C\{(\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0), (\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0)\} + D\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\}. \end{aligned} \quad (8.26)$$

We now determine each of the four terms on the right side of (8.26) before combining them into a single equation. Comparing (8.14b) and (8.22) we see that

$$\begin{aligned} D\{\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0\} = (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T [\Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} D\{\mathbf{z}_1\} \\ \cdot [\Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T), \end{aligned} \quad (8.27a)$$

which, upon substitution of (8.17), leads to

$$\begin{aligned} D\{\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0\} = (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T [\Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} \\ \cdot A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T). \end{aligned} \quad (8.27b)$$

From the given model (8.10) we can write

$$D\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\} = \Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T. \quad (8.27c)$$

Using (8.27c), we can rewrite (8.27b) as

$$\begin{aligned} D\{\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0\} = (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T [\Sigma_1 + A_1 (\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) A_1^T]^{-1} \\ \cdot A_1 D\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\} = K_1 A_1 D\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\}, \end{aligned} \quad (8.27d)$$

which leads to the following covariance terms:

$$C\{\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0, \mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\} = K_1 A_1 D\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0\} = \quad (8.27e)$$

(and, due to symmetry,)

$$= C\{\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0, \tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0\} = D\{\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0\}. \quad (8.27f)$$

Summing the individual components (8.27c) through (8.27f) yields

$$D\{(\tilde{\mathbf{u}}_1 - \Phi_0 \tilde{\mathbf{e}}_0^0) - (\mathbf{u}_1 - \Phi_0 \mathbf{e}_0^0)\} = (I_m - K_1 A_1)(\Theta_1 + \Phi_0 \Sigma_0^0 \Phi_0^T) =: \Sigma_1^0. \quad (8.28)$$

With the covariance matrix Σ_1^0 , we are ready to process the data at epoch 2, and we can continue in a like manner with any epochs that follow.

Notice that the variance component σ_0^2 has not been included in the dispersion formulas. This is because we try to avoid extra computations in real-time applications. However, we may wish to test our hypothesis that the innovation vector \mathbf{z}_1 is zero. To do so, we form the test statistic

$$\Omega := \mathbf{z}_1^T (D\{\mathbf{z}_1\})^{-1} \mathbf{z}_1, \quad (8.29)$$

which has redundancy n . Our hypothesis test (at each epoch) is

$$H_0 : E\{\mathbf{z}_1\} = \mathbf{0} \text{ versus } H_a : E\{\mathbf{z}_1\} \neq \mathbf{0}. \quad (8.30)$$

The distribution of the test statistic is

$$\Omega \sim \chi_n^2 \text{ under } H_0. \quad (8.31)$$

For some chosen level of significance α , we reject the null hypothesis H_0 if $\Omega > \chi_\alpha^2$. Note that we could also test the expectations of $\tilde{\mathbf{e}}_1$, $\tilde{\mathbf{u}}_1$, and $\tilde{\mathbf{e}}_0^0$ separately if the null hypothesis in (8.30) is rejected.

Appendix A

Useful Matrix Relations and Identities

Kronecker product The Kronecker-Zehfuss product of matrices is often simply called the Kronecker product. Its definition and several computational rules associated with it are given below.

Definition: let $G = [g_{ij}]$ be a $p \times q$ matrix and $H = [h_{ij}]$ be an $r \times s$ matrix, then

$$G \otimes H := [g_{ij} \cdot H] \quad (\text{A.0})$$

gives the Kronecker-Zehfuss product $G \otimes H$, which is of size $pr \times qs$.

Kronecker-Zehfuss computational rules:

$$(1) \quad \text{vec } ABC^T = (C \otimes A) \text{vec } B \quad (\text{A.1})$$

$$(2) \quad \text{tr } ABC^T D^T = \text{tr } D^T ABC^T = (\text{vec } D)^T (C \otimes A) \text{vec } B \quad (\text{A.2})$$

$$(3) \quad (G \otimes H)^T = G^T \otimes H^T \quad (\text{A.3})$$

$$(4) \quad (G \otimes H)^{-1} = G^{-1} \otimes H^{-1} \quad (\text{A.4})$$

$$(5) \quad \alpha(G \otimes H) = \alpha G \otimes H = G \otimes \alpha H \quad \text{for } \alpha \in \mathbb{R} \quad (\text{A.5})$$

$$(6) \quad (F + G) \otimes H = (F \otimes H) + (G \otimes H) \quad (\text{A.6})$$

$$(7) \quad G \otimes (H + J) = (G \otimes H) + (G \otimes J) \quad (\text{A.7})$$

$$(8) \quad (A \otimes B)(G \otimes H) = AG \otimes BH \quad (\text{A.8})$$

$$(9) \quad (H \otimes G) = K(G \otimes H)K \quad \text{for "commutation matrices" of suitable size} \quad (\text{A.9})$$

$$(10) \quad K^T \text{ is also a commutation matrix with } KK^T = I = K^T K \Rightarrow \\ K(H \otimes G) = (G \otimes H)K \quad (\text{A.10})$$

(Note that K is a generic symbol; the two K matrices could be different.)

specially: $K(H \otimes g) = g \otimes H$ for any vector g

- (11) $K \otimes K$ is also a commutation matrix $\Rightarrow \text{vec}(G^T) = \text{vec}(KK^T G^T) = (G \otimes I)K \text{vec} I = K(I \otimes G) \text{vec} I = K \text{vec} G$. Hence, K is called a “vec-permutation matrix.” (A.11)
- (12) Let λ_G and λ_H be vectors with the respective eigenvalues of the matrices G and H ; then the vector $x(\lambda_G \otimes \lambda_H)$ contains exactly the eigenvalues of the matrix $(G \otimes H)$. (A.12)
- (13) $\text{tr}(G \otimes H) = \text{tr} G \text{tr} H$ (A.13)
- (14) G and H positive (semi) definite $\Rightarrow G \otimes H$ positive (semi) definite (A.14)

The four fundamental matrix subspaces Let A be a matrix of size $n \times m$ with $\text{rk} A =: q$. The four fundamental matrix subspaces are

The *column space* of A (also range of A) is denoted by $\mathcal{R}(A)$.

The *nullspace* of A (also the kernel of A) is denoted by $\mathcal{N}(A)$.

The *row space* of A , which is $\mathcal{R}(A^T)$.

The *left nullspace* of A , which is $\mathcal{N}(A^T)$.

The subspaces are elements of larger spaces, the sizes of which are determined by the dimension of A .

$$\mathcal{N}(A) \subset \mathbb{R}^m, \quad \mathcal{R}(A^T) \subset \mathbb{R}^m$$

$$\mathcal{N}(A^T) \subset \mathbb{R}^n, \quad \mathcal{R}(A) \subset \mathbb{R}^n$$

The dimensions of the subspaces are a function of the rank of A , which we denote by q .

$$\dim \mathcal{R} = q$$

$$\dim \mathcal{N}(A) = m - q \text{ (also called the nullity of } A)$$

$$\dim \mathcal{R}(A^T) = q$$

$$\dim \mathcal{N}(A^T) = n - q$$

Sherman-Morrison-Woodbury-Schur formula

$$(T - UW^{-1}V)^{-1} = T^{-1} + T^{-1}U(W - VT^{-1}U)^{-1}VT^{-1} \quad (\text{A.15})$$

As a consequence, we also have:

$$(I \pm W^{-1}V)^{-1} = I \mp (W \pm V)^{-1}V, \quad (\text{A.16a})$$

$$(I \pm V)^{-1} = I \mp (I \pm V)^{-1}V, \quad (\text{A.16b})$$

$$(I \pm W^{-1})^{-1} = I \mp (W \pm I)^{-1}. \quad (\text{A.16c})$$

Inverse of the partitioned normal equation matrix Assume the matrix N is of full rank and is partitioned as

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}. \quad (\text{A.17})$$

The following row reductions lead to the inverse of N , expressed in terms of the partitioned blocks:

$$\begin{aligned} & \left[\begin{array}{cc|cc} N_{11} & N_{12} & I & 0 \\ N_{21} & N_{22} & 0 & I \end{array} \right] \rightarrow \left[\begin{array}{c|cc|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline N_{21} & N_{22} & 0 & I \end{array} \right] \rightarrow \\ & \rightarrow \left[\begin{array}{c|cc|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline 0 & N_{22} - N_{21}N_{11}^{-1}N_{12} & -N_{21}N_{11}^{-1} & I \end{array} \right] \rightarrow \\ & \rightarrow \left[\begin{array}{cc|c|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline 0 & I & -(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1} & (N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1} \end{array} \right] \rightarrow \\ & \rightarrow \left[\begin{array}{cc|c|c} I & 0 & \frac{N_{11}^{-1} + N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1}}{-(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1}} & \\ \hline 0 & I & \frac{-N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}}{(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}} & \end{array} \right] \end{aligned}$$

Finally we may write

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \left[\begin{array}{c|c} \frac{N_{11}^{-1} + N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1}}{-(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1}} & \\ \hline \frac{-N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}}{(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}} & \end{array} \right]. \quad (\text{A.18})$$

Note that other equivalent representations of this inverse exist.

Euclidean norm Also called l_2 -norm, Frobenius norm, Hilbert-Schmidt norm, or Schur Norm: (see LÜTKEPOHL, pg. 103).

$$\|A\| \equiv \sqrt{\text{tr}(A^T A)} \text{ for an } m \times n \text{ real matrix } A. \quad (\text{A.19})$$

Derivatives of the trace (for additional formulas see Lütkepohl, pp. 177-179)

$$X(m \times n), A(n \times m) : \frac{\partial \operatorname{tr}(AX)}{\partial X} = \frac{\partial \operatorname{tr}(XA)}{\partial X} = A^T \quad (\text{A.20a})$$

$$X(m \times n), A(m \times n) : \frac{\partial \operatorname{tr}(X^T A)}{\partial X} = \frac{\partial \operatorname{tr}(AX^T)}{\partial X} = A \quad (\text{A.20b})$$

$$X(m \times n) : \frac{\partial \operatorname{tr}(X^T X)}{\partial X} = \frac{\partial \operatorname{tr}(XX^T)}{\partial X} = 2X \quad (\text{A.20c})$$

$$X(m \times n), A(m \times m) : \frac{\partial \operatorname{tr}(X^T AX)}{\partial X} = (A + A^T)X \quad (\text{A.20d})$$

$$X(m \times n), A(m \times m) \text{ symmetric: } \frac{\partial \operatorname{tr}(X^T AX)}{\partial X} = 2AX \quad (\text{A.20e})$$

$$X(m \times n), A(n \times n) : \frac{\partial \operatorname{tr}(XAX^T)}{\partial X} = X(A + A^T) \quad (\text{A.20f})$$

$$X(m \times n), A(n \times n) \text{ symmetric: } \frac{\partial \operatorname{tr}(XAX^T)}{\partial X} = 2XA \quad (\text{A.20g})$$

$$X, A(m \times m) : \frac{\partial \operatorname{tr}(XAX)}{\partial X} = X^T A^T + A^T X^T \quad (\text{A.20h})$$

$$X(m \times n), A(p \times m) : \frac{\partial \operatorname{tr}(AXX^T A^T)}{\partial X} = 2A^T AX \quad (\text{A.20i})$$

Useful Matrix Equivalents (handout from Prof. Schaffrin, possibly originating from Urho A. Uotila)

$$DC(A + BDC)^{-1} = (D^{-1} + CA^{-1}B)^{-1}CA^{-1} = \quad (\text{A.21a})$$

$$= D(I + CA^{-1}BD)^{-1}CA^{-1} = \quad (\text{A.21b})$$

$$= DC(I + A^{-1}BDC)^{-1}A^{-1} = \quad (\text{A.21c})$$

$$= DCA^{-1}(I + BDCA^{-1})^{-1} = \quad (\text{A.21d})$$

$$= (I + DCA^{-1}B)^{-1}DCA^{-1} \quad (\text{A.21e})$$

We may expand the above UME's by setting, in turn, each matrix equal to the identity matrix, thus generating four new sets of identities, as follows:

Let $A = I$:

$$DC(I + BDC)^{-1} = (D^{-1} + CB)^{-1}C = \quad (\text{A.22a})$$

$$= D(I + CBD)^{-1}C = \quad (\text{A.22b})$$

$$= (I + DCB)^{-1}DC. \quad (\text{A.22c})$$

Let $B = I$:

$$DC(A + DC)^{-1} = (D^{-1} + CA^{-1})^{-1}CA^{-1} = \quad (\text{A.23a})$$

$$= D(I + CA^{-1}D)^{-1}CA^{-1} = \quad (\text{A.23b})$$

$$= DC(I + A^{-1}DC)^{-1}A^{-1} = \quad (\text{A.23c})$$

$$= DCA^{-1}(I + DCA^{-1})^{-1} = \quad (\text{A.23d})$$

$$= (I + DCA^{-1})^{-1}DCA^{-1}. \quad (\text{A.23e})$$

Let $C = I$:

$$D(A + BD)^{-1} = (D^{-1} + A^{-1}B)^{-1}A^{-1} = \quad (\text{A.24a})$$

$$= D(I + A^{-1}BD)^{-1}A^{-1} = \quad (\text{A.24b})$$

$$= DA^{-1}(I + BDA^{-1})^{-1} = \quad (\text{A.24c})$$

$$= (I + DA^{-1}B)^{-1}DA^{-1}. \quad (\text{A.24d})$$

Let $D = I$:

$$C(A + BC)^{-1} = (I + CA^{-1}B)^{-1}CA^{-1} = \quad (\text{A.25a})$$

$$= C(I + A^{-1}BC)^{-1}A^{-1} = \quad (\text{A.25b})$$

$$= CA^{-1}(I + BCA^{-1})^{-1}. \quad (\text{A.25c})$$

References: LÜTKEPOHL, H. (1996). Handbook of Matrices, John Wiley & Sons Ltd., West Sussex, England.

Bibliography

- Bjerhammar, A. (1973). *Theory of Errors and Generalized Matrix Inverses*. Elsevier Scientific Publishing Company, Amsterdam.
- Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer Texts in Statistics. Springer, New York, fourth edition.
- Cross, P. (1985). Numerical Methods in Network Design. In Grafarend, E. and Sansò, F., editors, *Optimization and Design of Geodetic Networks*, pages 132–168. Springer-Verlag, Berlin/Heidelberg.
- Grafarend, E. W. and Schaffrin, B. (1993). *Ausgleichsrechnung in linearen Modellen*. Bibliographisches Institut, Mannheim.
- Harville, D. A. (2000). *Matrix Algebra From a Statistician's Perspective*. Springer, corrected edition.
- Heumann, C., Nittner, T., Rao, C., Scheid, S., and Toutenburg, H. (2013). *Linear Models: Least Squares and Alternatives*. Springer Series in Statistics. Springer New York.
- Koch, K.-R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag, Berlin, second edition. Translated from the third German edition (1997).
- Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, Chichester.
- Mikhail, E. and Ackermann, F. (1982). *Observations and least squares*. University Press of America.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572.
- Pope, A. (1972). Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. In *Proc. of the 38th Annual Meeting of the ASPRS*, pages 449–477, Falls Church, VA. American Society for Photogrammetry & Remote Sensing.
- Snow, K. (2002). Applications of parameter estimation and hypothesis testing to GPS network adjustments. Master's thesis, The Ohio State University.

- Stigler, S. M. (1981). Gauss and the invention of least squares. *The Annals of Statistics*, 9(3):465–474.