Least-Squares Variance Component Estimation

Theory and GPS Applications

Least-Squares Variance Component Estimation

Theory and GPS Applications

AliReza Amiri-Simkooei

Publications on Geodesy 64

NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission

Delft, March 2007

Least-squares variance component estimation: theory and GPS applications
AliReza Amiri-Simkooei
Publications on Geodesy 64
ISBN: 978 90 6132 301 3
ISSN 0165 1706

Published by: NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission, Delft, The Netherlands
Printed by: Optima Grafische Communicatie, Optima Graphic Communication, Rotterdam, The Netherlands
Cover illustration: Distribution of a covariance element

NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission
P.O. Box 5058, 2600 GB Delft, The Netherlands
T: +31 (0)15 278 28 19
F: +31 (0)15 278 17 75
E: info@ncg.knaw.nl
W: www.ncg.knaw.nl

The NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission is part of the Royal Netherlands Academy of Arts and Sciences (KNAW).

In memory of my father

Abstract

Data processing in geodetic applications often relies on the least-squares method, for which one needs a proper stochastic model of the observables. Such a realistic covariance matrix allows one first to obtain the best (minimum variance) linear unbiased estimator of the unknown parameters; second, to determine a realistic precision description of the unknowns; and, third, along with the distribution of the data, to correctly perform hypothesis testing and assess quality control measures such as reliability. In many practical applications the covariance matrix is only partly known. The covariance matrix is then usually written as an unknown linear combination of known cofactor matrices. The estimation of the unknown (co)variance components is generally referred to as *variance component estimation* (VCE).

In this thesis we study the method of least-squares variance component estimation (LS-VCE) and elaborate on theoretical and practical aspects of the method. We show that LS-VCE is a simple, flexible, and attractive VCE-method. The LS-VCE method is simple because it is based on the well-known principle of least-squares. With this method the estimation of the (co)variance components is based on a linear model of observation equations. The method is flexible since it works with a user-defined weight matrix. Different weight matrix classes can be defined which all automatically lead to unbiased estimators of (co)variance components. LS-VCE is attractive since it allows one to apply the existing body of knowledge of least-squares theory to the problem of (co)variance component estimation. With this method, one can 1) obtain measures of discrepancies in the stochastic model, 2) determine the covariance matrix of the (co)variance components, 3) obtain the minimum variance estimator of (co)variance components by choosing the weight matrix as the inverse of the covariance matrix, 4) take the a-priori information on the (co)variance component into account, 5) solve for a nonlinear (co)variance component model, 6) apply the idea of robust estimation to (co)variance components, 7) evaluate the estimability of the (co)variance components, and 8) avoid the problem of obtaining negative variance components.

LS-VCE is capable of unifying many of the existing VCE-methods such as MINQUE, BIQUE, and REML, which can be recovered by making appropriate choices for the weight matrix. An important feature of the LS-VCE method is the capability of applying hypothesis testing to the stochastic model, for which we rely on the w-test, v-test, and overall model test. We aim to find an appropriate structure for the stochastic model which includes the relevant noise components into the covariance matrix. The w-test statistic is introduced to see whether or not a certain noise component is likely to be present in the observations, which consequently can be included in the stochastic model. Based on the normal distribution of the original observables we determine the mean and the variance of the w-test statistic, which are zero and one, respectively. The distribution is a linear combination of mutually independent central chi-square distributions each with one degree

of freedom. This distribution can be approximated by the standard normal distribution for some special cases. An equivalent expression for the w-test is given by introducing the v-test statistic. The goal is to decrease the number of (co)variance components of the stochastic model by testing the significance of the components. The overall model test is introduced to generally test the appropriateness of a proposed stochastic model.

We also apply LS-VCE to real data of two GPS applications. LS-VCE is applied to the GPS geometry-free model. We present the functional and stochastic model of the GPS observables. The variance components of different observation types, satellite elevation dependence of GPS observables' precision, and correlation between different observation types are estimated by LS-VCE. We show that the precision of the GPS observables clearly depends on the elevation angle of satellites. Also, significant correlation between observation types is found. For the second application we assess the noise characteristics of time series of daily coordinates for permanent GPS stations. We apply LS-VCE to estimate white noise and power-law noise (flicker noise and random walk noise) amplitudes in these time series. The results confirm that the time series are highly time correlated. We also use the w-test statistic to find an appropriate stochastic model of GPS time series. A combination of white noise, autoregressive noise, and flicker noise in general best characterizes the noise in all three position components. Unmodelled periodic effects in the data are then captured by a set of harmonic functions, for which we rely on least-squares harmonic estimation (LS-HE) developed in the same framework as LS-VCE. The results confirm the presence of annual and semiannual signals, as well as other significant periodic patterns in the series. To avoid the biased estimation of the variance components, such sinusoidal signals should be included in the functional part of the model before applying LS-VCE.

Contents

1 Introduction

1.1 Background

Data processing in geodetic applications usually relies on the least-squares method, or equivalently, when the inverse of the covariance matrix of observables is taken as the weight matrix, the best linear unbiased estimation (BLUE). To that end we deal with two models, namely, 'functional model' and 'stochastic model'. The former is usually either well-known or subject to extensive research, and the latter, containing the second-order central moments of the observables receives far less attention. Statistical models in use for instance in the fields of Global Navigation Satellite Systems (GNSS) positioning applications are usually simple and rudimentary. For many applications, it is of importance to have information available on the covariance matrix of an observable random vector. Such information allows one to study the different contributing factors of the errors in observations, to properly describe the precision of functions of interest by means of application of the covariance propagation law, and to obtain minimum variance estimators of the parameters in a linear model. This will also allow one to correctly perform hypothesis testing and to assess other quality control measures such as reliability.

An adequate statistical model is thus needed to arrive at a proper description of the estimators' quality. Incomplete knowledge of the covariance matrix of the observables occurs in many of the GNSS applications. Often, however, the covariance matrix of the observables is only partly known, as a consequence of which the unknown part needs to be estimated from the redundant observables. The estimation of these unknown components of a covariance matrix is generally referred to as variance component estimation (VCE). Various VCE-studies have been conducted to improve our knowledge of the GNSS stochastic model. Variance component estimation is also an important issue in other geodetic fields of application, in particular in applications where heterogeneous data needs to be combined. An example where the heterogeneous data should be combined is the combination of InSAR and leveling data. Another example is the combination of classical geodesy networks and GPS networks.

Methods for estimating (co)variance components have been intensively investigated in the *statistical* and *geodetic* literature. There exist many different methods for variance component estimation. The methods differ in the estimation principle employed, as well as in the distributional assumptions that need to be made. Most methods have been devised for the linear model, for which one assumes that the covariance matrix of the observables can be written as an unknown linear combination of known cofactor matrices. The coefficients of this linear combination are then the unknown (co)variance components that need to be estimated. Of the leading variance component estimators, we mention the minimum norm quadratic unbiased estimator (MINQUE), the best invariant quadratic

unbiased estimator (BIQUE), the least-squares variance component estimator (LS-VCE), the restricted maximum likelihood estimator (REML), and the Bayesian method to VCE.

The MINQUE method is one of the more commonly used methods for estimation of variance components. Apart from the first and second order moments of the observables, this method does not require any distributional assumptions. The BIQUE, however, does require knowledge of some of the higher order moments. This minimum variance quadratic estimator has been derived and studied, under the assumption of normally distributed observables. The LS-VCE method is based on the least-squares principle and works with a user-defined weight matrix. As such the method only requires information on the first and second order moments. The REML method and the Bayesian method, both require in contrast to the other methods, complete information on the probability density function of the observables. Using the normal distribution, maximum likelihood estimators and Bayesian estimators have been derived and studied by different authors.

1.2 Objectives of thesis

In this thesis we further study and use the least-squares variance component estimation (LS-VCE) which was originally developed by Teunissen (1988). For a review see Teunissen and Amiri-Simkooei (2006). Since the method can be applied to many modern geodetic applications, this thesis elaborates this theory in much detail. Although the method is probably one of the lesser known VCE-methods, we will show that it is a simple, flexible and attractive method for the estimation of unknown variance and covariance components.

- The method is simple, since it is based on the principle of least-squares. This will lead us to use one unified estimation principle, namely the well-known and well-understood least-squares principle, for both the functional and stochastic model.

- The method is flexible, since it works with a user-defined weight matrix. The weighted LS-VCE does not need any distributional assumption for the observables. The weighted LS-VCE is formulated in a linear model and thus leads to unbiased (co)variance component estimators. In case of elliptical distributions which include for instance the normal distribution, the method can automatically produce minimum variance estimators.

- The method of LS-VCE is attractive, since it allows one to directly apply the existing body of knowledge of least-squares theory. In this thesis we present the LS-VCE method for different scenarios and explore its various properties. All other methods of VCE, for instance, concern only the estimation of (co)variance components. But, LS-VCE allows one also to apply hypothesis testing to the stochastic model which is considered to be a distinguished feature of this method.

Being a least-squares estimator, the LS-VCE automatically inherits all the well-known properties of a least-squares estimator. We show how the existing body of knowledge of least-squares theory can be used to one's advantage for studying various aspects of VCE. For example, since the method is based on the least-squares principle, the precision of (co)variance estimators can directly be obtained.

We include various examples to illustrate this theory at work and address implementation aspects. Application of LS-VCE to real GPS data will be considered as well. We will use

LS-VCE to study the stochastics of GPS code and carrier phase data and also of GPS coordinate time series.

1.3 Outline of thesis

This thesis is organized as follows:

Chapter 2 explains the least-squares estimation and validation in a general linear model of observation equations. Three estimation principles, which lead to the weighted least-squares estimation, the best linear unbiased estimation (BLUE), and the maximum likelihood estimation, will be discussed. Equivalent expressions for estimators are determined using the model of condition equations afterward. The last part of this chapter deals with hypotheses testing to find misspecifications (with respect to data) in a linear model. This includes two types of equivalent tests: the observation test and the parameter significance test. For this purpose the overall model test, the w-test statistic, and the v-test statistic will be addressed. Detection of observation outliers is a prerequisite for obtaining unbiased (co)variance estimators.

Chapter 3 briefly reviews various (co)variance component estimation principles. We start from elementary error sources and construct a variance component model. We then apply different principles like unbiasedness, minimum norm, minimum variance, and maximum likelihood to this model to obtain various estimators. This includes minimum norm quadratic unbiased estimators (MINQUE), best invariance quadratic unbiased estimators (BIQUE), the Helmert method to VCE, maximum likelihood estimators (MLE), the Bayesian method to VCE, and least-squares estimators. These methods differ in the estimation principles as well as in the distributional assumptions. We will present the underlying assumptions of each method. We then discuss simple and approximate VCE methods which need less computational load when compared to the rigorous methods.

Chapter 4 introduces the principle of weighted least-squares for the estimation of unknown (co)variance components. We formulate a linear (co)variance component model, define the least-squares variance component estimator and determine its covariance matrix. We consider the weighted LS-VCE method for a special class of weight matrices. Based on this weight matrix class we then show how the LS-VCE can be turned into a minimum variance VCE. We also show how the existing body of knowledge of least-squares theory can be used to one's advantage for studying and solving various aspects of the variance component estimation problem. Topics that are addressed are: measures of inconsistency, the use of a-priori variance component information, nonlinear variance component estimation, and robust and non-negative variance component estimation. Later, in this chapter we make some comments, supported by a few examples, on the estimability of (co)variance components.

In chapter 5 we apply hypothesis testing to the stochastic model. The w-test, the v-test, and the overall model test will be generalized for the stochastic model. It is aimed to find misspecifications in the stochastic model, to improve an existing stochastic model, and to judge whether or not (or which) additional (co)variance components are likely to be included in (or excluded from) the stochastic model. We will derive the distribution of the w-test and the v-test statistics under the normality assumption of the original observables. For the overall model test statistic, the distribution is complicated. We thus only obtain the first and the second order moments, instead of its complete distribution. However, using

an approximation we show how the overall model test of the stochastic model is converted to the overall model test of the functional model.

Chapter 6 deals with multivariate parameter and variance-covariance estimation and validation. We aim to apply the theory on least-squares estimation of (co)variance components introduced in chapter 4, and to apply the idea of detection and validation introduced in chapter 5, to a multivariate linear model of observation equations. We show that the estimators in the multivariate model can be considered as a generalization of the estimators in the univariate model. This holds in fact for the w-test and the v-test statistics as well as for their related distributions. We also show, when the redundancy of the model is large enough, that the distribution of the test statistics can be approximated by a normal distribution.

Chapters 7 and 8 present numerical results of application of LS-VCE to real GPS data. Chapter 7 deals with the GPS geometry-free observation model for which LS-VCE is used to assess the stochastics of GPS pseudorange and carrier-phase data. The purpose of this chapter is to come up with a realistic and adequate covariance matrix of GPS observables. Topics that will be addressed are: the precision of different code and phase observations, satellite elevation dependence of the observable's precision, and the correlation between different code and phase observations. Chapter 8 concerns coordinate time series analysis of permanent GPS stations. We discuss both the functional model and the stochastic model in detail. We will then introduce the least-squares harmonic estimation (LS-HE) to come up with an appropriate functional model by introducing a set of harmonic functions to compensate for unmodelled effects in the data series. We then apply the LS-VCE to estimate different noise components (white noise, flicker noise, and random walk noise) of the series. We also employ the w-test, in conjunction with LS-VCE, to come up with an appropriate stochastic model for GPS time series. Other topics like misspecifications in both the functional and the stochastic model, and duality between these two models are addressed as well.

Finally, chapter 9 reviews the conclusions of this work and presents recommendations for further research.

2 Least-Squares Estimation and Validation

2.1 Parameter estimation in linear models

2.1.1 Optimal properties of estimators

This chapter reviews in general the theory of least-squares estimation and validation in an inconsistent linear model where the inconsistency is caused by errors in the data. From experience we know that various uncertain phenomena can be modeled as a random variable (or a random vector), namely $\underline{y}$. An example is the uncertainty in instrument readings due to measurement errors. The randomness of $\underline{y}$ is expressed by its probability density function (PDF). In practice, our knowledge of the PDF is incomplete. The PDF can usually be indexed with one or more unknown parameters. The PDF of a random m-vector $\underline{y}$ is denoted as $f_y(y|x)$, in which x is an n-vector of unknown parameters to be estimated. The approach is to take an observation of the m-vector $\underline{y}$ and to use this observation vector to estimate the unknown n-vector x. The observation y as a realization of $\underline{y}$ with PDF $f_y(y|x)$ contains information about x which can be used to estimate its value.

We thus require to determine $\hat{x}$ from an observation vector y. The essential estimation problem is therefore to find a function $G : \mathbb{R}^m \mapsto \mathbb{R}^n$, such that $\hat{x} = G(y)$ can be taken as our estimate of x. In fact, if we apply G to $\underline{y}$, we obtain the random vector $\underline{\hat{x}} = G(\underline{y})$. The random vector $\underline{\hat{x}}$ is called the *estimator* of x and $\hat{x}$ is called the *estimate* of x. The estimator is a random vector which has its own probability density function (distribution), while, the estimate is a realized value of the estimator and thus a deterministic vector. The difference $\underline{\hat{\epsilon}} = \underline{\hat{x}} - x$ is called the *estimation error*. Since $\underline{\hat{x}}$ depends on the chosen function G, the estimation error depends on G as well. We list three desirable properties for $\underline{\hat{\epsilon}}$ which all express in some sense the closeness of $\underline{\hat{x}}$ to x. Such properties can be used as criteria for finding an 'optimal' function G.

Unbiasedness The estimator $\underline{\hat{x}}$ is said to be an *unbiased* estimator of x if and only if the mathematical expectation of the estimation error is zero. An estimator is therefore unbiased if the mean of its distribution equals x

$$\mathrm{E}\{\underline{\hat{x}}\} = x \quad \text{for all } x\,, \tag{2.1}$$

where $\mathrm{E}\{.\}$ denotes the expectation operator. This implies that the average of repeated realizations of $\underline{\hat{\epsilon}}$ will tend to zero on the long run. An estimator which is not unbiased is said to be *biased* and the difference $\mathrm{E}\{\underline{\hat{\epsilon}}\} = \mathrm{E}\{\underline{\hat{x}}\} - x$ is called the *bias* of the estimator. The size of the bias is therefore a measure of closeness of $\underline{\hat{x}}$ to x. The mean error $\mathrm{E}\{\underline{\hat{\epsilon}}\}$ is a measure of closeness that makes use of the *first* moment of the distribution of $\underline{\hat{x}}$.

Minimum variance (best) A second measure of closeness of the estimator to x is the *mean squared error* (MSE), which is defined as

$$\mathsf{MSE} = \mathsf{E}\left\{\|\underline{\hat{x}} - x\|^2\right\} \rightarrow \min, \tag{2.2}$$

where $\|.\|$ is a vector norm. If we were to compare different estimators by looking at their respective MSEs, we would prefer one with small or the smallest MSE. This is a measure of closeness that makes use also of the *second* moment of the distribution of $\underline{\hat{x}}$. The 'best' estimator, in the absence of biases, therefore is of minimum variance.

Maximum likelihood Rather than relying on the first two moments of a distribution, one can also define what closeness means in terms of the distribution itself. As a third measure of closeness we therefore consider the probability that the estimator $\underline{\hat{x}}$ resides in a small region centered at x. If we take this region to be a hypersphere with a given radius r, the measure is given as

$$\mathsf{P}\left(\|\underline{\hat{x}} - x\|^2 \leq r^2\right) \rightarrow \max . \tag{2.3}$$

If we were to compare different estimators by looking at their respective values for probability (2.3), we would prefer one with large or the largest such probability. Unfortunately it is rarely possible to derive an estimator which has the largest such probability for all x.

2.1.2 Model of observation equations

There are different estimation methods that we discuss in this chapter. The simplest method that one can apply needs information about the first moment of the distribution. Since the PDF depends on the unknown parameter x, the mean will generally depend on x as well. We will however assume that we know how the mean depends on x. The relation is through a mapping $A : \mathbb{R}^n \mapsto \mathbb{R}^m$. In the linear(ized) case A is an $m \times n$ matrix.

Redundancy Redundant measurements are often taken to increase the accuracy of the obtained results and to check for the presence of blunders (i.e. $m > n$). Due to intrinsic uncertainty in observations, redundant measurements generally lead to an inconsistent system of equations. For $m > n$ such an inconsistent linear system of equations has the form $y \approx Ax$. This linear system of equations for which $m > n = \mathrm{rank}(A)$ is also referred to as an *overdetermined* system. The number $b = m - \mathrm{rank}(A) = m - n$ is referred to as the *redundancy* of the system (or redundancy of the functional model).

Least-squares principle It is well known that an inconsistent system has no solution x that can reproduce y. As a first step one can make the system consistent by adding a measurement error vector e: $y = Ax + e$. But now we have $m+n$ unknowns in m equations, and therefore an infinite number of possible solutions for x and e (underdetermined system). It seems reasonable to select from this infinite set of possible solutions, the solution which in some sense gives the smallest value for e. This implies choosing the solution for x such that Ax is, in some sense, as close as possible to the measurement vector y. The (weighted) *least-squares* (LS) principle states to minimize the (weighted) norm of the residual vector e, namely $\|e\|_W^2 = e^T W e = (y - Ax)^T W (y - Ax)$, where $\|.\|$ denotes the norm of a vector and W is the weight matrix. Any symmetric and positive-definite matrix is considered to be in the class of admissible weight matrix W (see Teunissen et al., 2005).

Linear model From now on we will refer to the linear system of equations $y = Ax + e$ as the linear *model of observation equations* which is denoted as

$$\mathsf{E}\{\underline{y}\} = Ax\,, \quad W, \text{ or } \mathsf{D}\{\underline{y}\} = Q_y\,, \tag{2.4}$$

where $\underline{y}$ is the m-vector of (stochastic) observables, A is the $m \times n$ design matrix, x is the n-vector of unknown parameters, and W and Q_y are the $m \times m$ weight matrix and covariance matrix of the observables, respectively. The design matrix A is assumed to be of full column rank, i.e., $\mathrm{rank}(A) = n$, provided that $m \geq n$, and W and Q_y are symmetric and positive-definite. Again $\mathsf{E}\{.\}$ denotes the expectation operator, and $\mathsf{D}\{.\}$ the dispersion operator. The above parametric form of the functional model is referred to as a Gauss-Markov model when $\underline{y}$ is normally distributed, i.e. $\underline{y} \sim \mathsf{N}(Ax, Q_y)$.

Estimation methods Three different estimation methods will be treated in this chapter. They are: weighted least-squares estimation (WLSE), best linear unbiased estimation (BLUE) and maximum likelihood estimation (MLE). The methods differ not only in the estimation principles involved, but also in the information that is required about the PDF $f_y(y|x)$. WLSE is applied when we only have information about the first moment of the distribution. BLUE is a method which can be applied when we have information about the first two moments of the distribution. MLE is used if we know the complete structure of the PDF $f_y(y|x)$. An important example for which the complete structure of the PDF is known is the multivariate normal distribution, i.e. as $\underline{y} \sim \mathsf{N}_m(Ax, Q_y)$.

2.1.3 Weighted least-squares estimation

Definition 2.1 (Weighted least-squares) *Let $\mathsf{E}\{\underline{y}\} = Ax$, with A an $m \times n$ matrix of $\mathrm{rank}(A) = n$, be a possibly inconsistent linear model of observation equations and let W be a symmetric and positive-definite $m \times m$ weight matrix $(W = W^T > 0)$. Then the weighted least-squares solution of the system is defined as*

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} (y - Ax)^T W (y - Ax)\,. \tag{2.5}$$

The difference $\hat{e} = y - A\hat{x}$ is called the (weighted) least-squares residual vector. Its squared (weighted) norm $\|\hat{e}\|^2_W = \hat{e}^T W \hat{e}$ is a scalar measure for the inconsistency of the linear system. □

Estimator Since the mean of $\underline{y}$ depends on the unknown x, also the PDF of $\underline{y}$ depends on the unknown x. The problem of determining a value for x can thus now be seen as an estimation problem, i.e. as the problem of finding a function G such that $\hat{x} = G(y)$ can act as the estimate of x and $\underline{\hat{x}} = G(\underline{y})$ as the estimator of x. The weighted least-squares estimator (WLSE) is given as (Teunissen et al., 2005)

$$\boxed{\underline{\hat{x}} = (A^T W A)^{-1} A^T W \underline{y}\,,} \tag{2.6}$$

which is a linear estimator, since all the entries of $\underline{\hat{x}}$ are linear combinations of the entries of $\underline{y}$. The least-squares estimator $\underline{\hat{y}} = A\underline{\hat{x}}$ of observables and $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$ of residuals follow from equation $\underline{y} = Ax + \underline{e}$ as

$$\begin{cases} \underline{\hat{y}} &= P_A \underline{y}\,; \\ \underline{\hat{e}} &= P_A^{\perp} \underline{y}\,, \end{cases} \tag{2.7}$$

with $P_A = A(A^TWA)^{-1}A^TW$ and $P_A^\perp = I_m - P_A$ two (orthogonal) projectors. The projector P_A projects onto the range space of A (i.e. $\mathcal{R}(A)$), and along its orthogonal complement (i.e. $\mathcal{R}(A)^\perp$), while $P_A^\perp$ projects onto $\mathcal{R}(A)^\perp$ and along $\mathcal{R}(A)$. $\mathcal{R}(.)$ denotes the range space of a matrix. For some useful properties of these two projectors refer to Teunissen (2000a).

Unbiasedness To get some insight in the performance of an estimator, we need to know how the estimator relates to its target value. Based on the assumption $\mathsf{E}\{\underline{e}\} = 0$, the expectations of $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ follow as

$$\mathsf{E}\{\underline{\hat{x}}\} = x\,, \quad \mathsf{E}\{\underline{\hat{y}}\} = \mathsf{E}\{\underline{y}\} = Ax\,, \quad \mathsf{E}\{\underline{\hat{e}}\} = \mathsf{E}\{\underline{e}\} = 0\,. \tag{2.8}$$

This shows that the WLSE is an *linear unbiased estimator* (LUE). Unbiasedness is clearly a desirable property. It implies that on the average the outcomes of the estimator will be on target. Also $\underline{\hat{y}}$ and $\underline{\hat{e}}$ are on target on the average. Note that the unbiasedness of the WLSE is independent of the choice made for the weight matrix W.

Covariance matrix In order to obtain the covariance matrices of $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$, we need the covariance matrix of $\underline{e}$ or observables $\underline{y}$, namely Q_y. The covariance matrices of $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ will be denoted respectively as $Q_{\hat{x}}$, $Q_{\hat{y}}$ and $Q_{\hat{e}}$. Application of the *error propagation law* to equations (2.6) and (2.7) yields

$$\begin{cases} Q_{\hat{x}} = (A^TWA)^{-1}A^TWQ_yWA(A^TWA)^{-1}\,; \\ Q_{\hat{y}} = P_AQ_yP_A^T\,; \\ Q_{\hat{e}} = P_A^\perp Q_yP_A^{\perp T}. \end{cases} \tag{2.9}$$

These covariance matrices depend on the choice made for the weight matrix W.

Mean squared error The mean and the covariance matrix of an estimator come together in the mean squared error of the estimator. As before, let $\underline{\hat{\epsilon}} = \underline{\hat{x}} - x$ be the estimation error. Assume that we 'measure' the size of the estimation error by the expectation of the sum of squares of its entries, $\mathsf{E}\{\underline{\hat{\epsilon}}^T\underline{\hat{\epsilon}}\} = \mathsf{E}\{\|\underline{\hat{x}} - x\|^2\}$, which is called the *mean squared error* (MSE) of the estimator. It can easily be shown that the MSE is decomposed as $\mathsf{E}\{\|\underline{\hat{x}} - x\|^2\} = \mathsf{E}\{\|\underline{\hat{x}} - \mathsf{E}\{\underline{\hat{x}}\}\|^2\} + \mathsf{E}\{\|x - \mathsf{E}\{\underline{\hat{x}}\}\|^2\}$. The first term on the right-hand side is the trace of the covariance matrix of the estimator and the second term is the squared norm of the bias of the estimator. But since the WLSE is unbiased, the second term vanishes, as a result of which the MSE of the WLSE reads

$$\mathsf{E}\{\|\underline{\hat{x}} - x\|^2\} = \mathrm{tr}(Q_{\hat{x}})\,. \tag{2.10}$$

Quadratic form of residuals In weighted least-squares, one important criterion which shows the inconsistency of the linear model of observation equations is the quadratic form (squared norm) of the residuals which is given as

$$\|\underline{\hat{e}}\|_W^2 = \underline{\hat{e}}^TW\underline{\hat{e}} = \underline{y}^TW\underline{y} - \underline{y}^TWA(A^TWA)^{-1}A^TW\underline{y}\,. \tag{2.11}$$

NOTE 2.1 The results in equations (2.8), (2.9), (2.10) and (2.11) are all independent of the unspecified distribution of $\underline{y}$. The unbiasedness property (2.8) is even independent of the choice made for the weight matrix W, while the covariance matrices (2.9), the mean squared error (2.10), and the quadratic form (2.11) depend on W. One may therefore think of the 'best' weight matrix that minimizes the MSE (see next section). ⊡

Normality The random vectors $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ are all linear functions of $\underline{y}$. This implies that they have a Gaussian (normal) distribution whenever $\underline{y}$ has a normal distribution. Therefore, if $\underline{y}$ has a normal distribution, i.e. $\underline{y} \sim \mathsf{N}_m(Ax, Q_y)$ then

$$\underline{\hat{x}} \sim \mathsf{N}_n(x, Q_{\hat{x}}); \quad \underline{\hat{y}} \sim \mathsf{N}_m(Ax, Q_{\hat{y}}); \quad \underline{\hat{e}} \sim \mathsf{N}_m(0, Q_{\hat{e}})\,. \tag{2.12}$$

Note, since the covariance matrix $Q_{\hat{e}}$ is singular, that the least-squares residual vector $\underline{\hat{e}}$ has a singular normal distribution. Also note that the first two distributions differ from the third in the sense that their means are unknown. Therefore, if Q_y is given, only the shape of their PDFs is known, but not the location. The PDF of $\hat{e}$, however, is completely known once Q_y is given. This property will turn out to play an important role in section 2.2.

2.1.4 Best linear unbiased estimation

Minimum MSE The weighted least-squares approach was introduced as an appealing technique for solving an inconsistent system of equations. The method itself is a deterministic principle, since no concepts from probability theory are used in formulating the least-squares minimization problem. In order to select an optimal estimator from the class of linear unbiased estimators (LUE), we need to define the optimality criterion. As optimality criterion we choose the minimization of the mean squared error (MSE). The estimator which has the smallest mean squared error of all LUEs is called the best linear unbiased estimator (BLUE). Such a minimization problem results in the smallest possible variance for estimators, i.e. $\mathsf{E}\{\|\underline{\hat{x}} - x\|^2\} = \mathrm{tr}(Q_{\hat{x}}) \equiv \min$.

Estimator If the covariance matrix Q_y of the observables is known, one could use the best linear unbiased estimation (BLUE) by taking the weight matrix to be the inverse of the covariance matrix, namely taking $W = Q_y^{-1}$ in equations (2.6) and (2.7). With this the BLUE estimators of x, y, and e in equation $\underline{y} = Ax + \underline{e}$ read

$$\left\{ \begin{array}{rcl} \underline{\hat{x}} & = & (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y}\,; \\ \underline{\hat{y}} & = & P_A \underline{y}\,; \\ \underline{\hat{e}} & = & P_A^{\perp} \underline{y}\,. \end{array} \right. \tag{2.13}$$

where both $P_A = A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$ and $P_A^{\perp} = I_m - P_A$ are orthogonal projectors. Substitution of $W = Q_y^{-1}$ into equation (2.9) yields the covariance matrix of the BLUE estimators as

$$\left\{ \begin{array}{l} Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1}\,; \\ Q_{\hat{y}} = P_A Q_y\,; \\ Q_{\hat{e}} = P_A^{\perp} Q_y\,. \end{array} \right. \tag{2.14}$$

It can be shown that of all linear unbiased estimators, the BLUE-estimator has minimum variance. It is therefore a minimum variance linear unbiased estimator. The BLUE is also sometimes called the *probabilistic* least-squares estimator. The property of minimum variance is also independent of the distribution of $\underline{y}$ (like the unbiasedness property). In the literature, the choice of $W = Q_y^{-1}$, leading to the BLUE, is often made by default. In the thesis, in general, we will treat the weighted least-squares estimators and the BLUE to be different.

Quadratic form of residuals From the BLUE estimators, the inconsistency criterion of the linear model of observation equations, expressed by the quadratic form of the residuals, is given as

$$\|\underline{\hat{e}}\|^2_{Q_y^{-1}} = \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} = \underline{y}^T Q_y^{-1} \underline{y} - \underline{y}^T Q_y^{-1} A (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \,. \tag{2.15}$$

The preceding squared norm of the residuals will play an important role in section of detection and validation (section 2.2).

NOTE 2.2 In the weighted least-squares the weight matrix W plays the role of a *metric tensor* in a *vector space*. The BLUE estimators take the weight matrix as the inverse of the covariance matrix. Therefore, the covariance matrix of the observables is closely related to the metric tensor. We have thus some probabilistic interpretations in our vector space. For example if the covariances between observables are zero, this means that the standard basis vectors of the vector space are *orthogonal*; uncorrelated observables mean, for basis vectors, having no projection on each other. If in addition the variances are equal, this means that the basis vectors are *orthonormal*. Therefore, if we take the weight matrix as the inverse of covariance matrix, the definition of the minimum distance (minimum norm) in the vector space obtained from weighted least-squares will coincide with the definition of minimum variance in the stochastic model (space) obtained from BLUE. ⊡

2.1.5 Maximum likelihood estimation

So far we have seen two different estimation methods at work: WLSE and BLUE. These two methods are not only based on different principles, but they also differ in the type of information that is required of the PDF of $\underline{y}$. For WLSE we only need information about the first moment of the PDF, the mean of $\underline{y}$. For BLUE we need additional information. Apart from the first moment, we also need the second (central) moment of the PDF, the covariance matrix of $\underline{y}$. For the linear model, the two principles give identical results when the weight matrix is taken equal to the inverse of the covariance matrix. In this section we introduce the method of maximum likelihood estimation (MLE) which requires knowledge of the complete PDF.

The principle The maximum likelihood (ML) method is conceptually one of the simplest methods of estimation. It is only applicable however when the general structure of the PDF is known. Assume therefore that the PDF of $\underline{y} \in \mathbb{R}^m$, i.e. $f_y(y|x)$, is known apart from some n unknown parameters. Since the PDF will change when x changes, we in fact have a whole family of PDFs in which each member of the family is determined by the value taken by x. Since x is unknown, it is not known to which PDF an observed value of $\underline{y}$, i.e. y_0, belongs. The idea now is to select from the family of PDFs, the PDF which gives the best support of the observed data. For this purpose one considers $f_y(y_0|x)$ as function of x. This function is referred to as the *likelihood function* of y_0 which produces, as x varies, the probability densities of all the PDFs for the same sample value y_0. Would x be the correct value, then the probability of $\underline{y}$ being an element of an infinitesimal region centered at y_0 is given as $f_y(y_0|x)dy$. A reasonable choice for x, given the observed value y_0, is therefore the value which corresponds with the largest probability, $\max_x f_y(y_0|x)dy$, and thus with the largest value of the likelihood function. The *maximum likelihood estimator* (MLE) of x is therefore defined as follows:

Definition 2.2 (Maximum likelihood) *Let the PDF of the vector of observables $\underline{y} \in \mathbb{R}^m$ be parameterized as $f_y(y|x)$, with $x \in \mathbb{R}^n$ unknown. Then the MLE of x is given as*

$$\underline{\hat{x}} = \arg\max_{x \in \mathbb{R}^n} f_y(\underline{y}|x). \; \square \tag{2.16}$$

The computation of the maximum likelihood solution may not always be an easy task. If the likelihood function is sufficiently smooth, the two necessary and sufficient conditions for $\hat{x}$ to be a (local or global) maximizer are

$$\partial_x f_y(y|\hat{x}) = 0\,; \quad \text{and} \quad \partial^2_{xx^T} f_y(y|\hat{x}) < 0\,, \tag{2.17}$$

with ∂_x and $\partial^2_{xx^T}$ the first and the second order partial derivatives with respect to x, respectively. Therefore, the gradient has to be zero and the Hessian matrix (the symmetric matrix of second-order partial derivatives) has to be negative definite. For more information see Teunissen et al. (2005).

NOTE 2.3 (MLE AND BLUE) In case of normally distributed data (Gauss-Markov model), the MLE estimators are identical to the BLUE ones. Let $\underline{y} \sim \mathsf{N}_m(Ax, Q_y)$, with x the n-vector of unknown parameters. Then $f_y(y|x) = (\det(2\pi Q_y))^{-1/2} \exp\{-\frac{1}{2}\|y - Ax\|^2_{Q_y^{-1}}\}$, from which it follows that $\arg\max_{x\in\mathbb{R}^n} f_y(y|x) = \arg\min_{x\in\mathbb{R}^n} \|y - Ax\|^2_{Q_y^{-1}}$. Therefore, $\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y}$. The estimators $\underline{\hat{y}}$ and $\underline{\hat{e}}$ as well as their covariance matrices are also the same as those given for BLUE estimators (see previous section).

2.1.6 Model of condition equations

Relations between variables can be expressed in various *equivalent* forms. Two prime examples are the so-called *parametric form* and the *implicit form*. So far we have expressed the linear model in the form $\mathsf{E}\{\underline{y}\} = Ax$. This is the so-called parametric form of the linear model (model of observation equations). The mean of $\underline{y}$ is explicitly parameterized in terms of the unknown parameter vector x. Although this is the most common form of the linear model, other equivalent formulations exist. For example, one can also include some hard constraints into this model (see appendix D.1). We can enumerate at least other forms like the *conditioned* model and *combined* (mixed) model. The implicit form of the linear model is called the conditioned linear model. In this chapter we restrict ourselves only to the model of condition equations. This model describes the conditions which are satisfied by the entries of the mean vector $\mathsf{E}\{\underline{y}\}$. For this formulation we will give the corresponding expressions for the BLUE estimators.

Linear model Each parametric linear model $\mathsf{E}\{\underline{y}\} = Ax$ has its equivalent conditioned linear model. The model of parametric equations can be rewritten in terms of the model of *condition equations* as

$$B^T \underline{y} = \underline{t}\,, \quad \mathsf{E}\{\underline{t}\} = 0\,, \quad Q_t = \mathsf{D}\{\underline{t}\} = B^T Q_y B\,, \tag{2.18}$$

with B a given $m \times b$ matrix, and Q_t the $b \times b$ covariance matrix of misclosure b-vector $\underline{t}$. The vector of misclosures provides a direct measure for inconsistency. The matrix B has full column rank, i.e., $\operatorname{rank}(B) = b$ provided that $b \leq m$, which is always true since b, the redundancy of the functional model, is given as $b = m - n$. The matrix Q_t is assumed to be symmetric and positive-definite.

Relation with A Each additional observation on top of the n-number of observations which are needed for the unique determination of the unknown parameters x in the linear model $\mathsf{E}\{\underline{y}\} = Ax$ leads to an extra condition equation. For example, the height differences of a closed loop in leveling should sum up to zero. Or three interior angles of a triangle should always add up to π. The matrix B is constructed on the basis of these independent conditions which the redundant observations have to fulfill. It should be pointed out that obtaining the condition equations and the B matrix is sometimes difficult. The following relation is always true between two matrices A and B

$$B^T A = 0\,, \tag{2.19}$$

which means that the columns of the matrix B are complementary and orthogonal to the columns of the matrix A. That is, $\mathcal{R}(B) = \mathcal{N}(A^T)$ which means that $\mathcal{R}(A) \oplus \mathcal{R}(B) = \mathbb{R}^m$ and $\mathcal{R}(A) \perp \mathcal{R}(B)$, where $\mathcal{R}(.)$ denotes the range space of a matrix, $\mathcal{N}(.)$ the null space and $\oplus$ denotes the direct sum of the two subspaces.

Estimator In order to obtain the BLUE estimator, we will use the following identity:

$$A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} + Q_y B (B^T Q_y B)^{-1} B^T = I_m\,. \tag{2.20}$$

For a proof see Teunissen et al. (2005). According to this matrix identity, we have for the orthogonal projector $P_A = A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$ and $P_A^\perp = I - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$ equivalent expressions in terms of the model of condition equations (B-model)

$$\begin{array}{rcl} P_A & = & P_{Q_y B}^\perp \;=\; I - Q_y B (B^T Q_y B)^{-1} B^T\,; \\ P_A^\perp & = & P_{Q_y B} \;=\; Q_y B (B^T Q_y B)^{-1} B^T\,, \end{array} \tag{2.21}$$

where $P_{Q_y B}^\perp$ is an orthogonal projector which projects onto $\mathcal{R}(Q_y B)^\perp$ and along $\mathcal{R}(B)$ and $P_{Q_y B}$ projects onto $\mathcal{R}(B)$ and along $\mathcal{R}(Q_y B)^\perp$. From equation (2.18), the least-squares estimator (BLUE) of the observables and residuals is obtained as

$$\underline{\hat{y}} = P_{Q_y B}^\perp \underline{y}\,; \quad \underline{\hat{e}} = P_{Q_y B} \underline{y}\,, \tag{2.22}$$

with the covariance matrices of the form

$$Q_{\hat{y}} = P_{Q_y B}^\perp Q_y, \quad Q_{\hat{e}} = P_{Q_y B} Q_y. \tag{2.23}$$

Quadratic form of misclosures The vector of misclosures has a direct link with the BLUE's residual vector. The BLUE's residual vector $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$ and its squared norm $\|\underline{\hat{e}}\|^2_{Q_y^{-1}}$ can be expressed in the vector of misclosures as

$$\underline{\hat{e}} = Q_y B Q_t^{-1} \underline{t}\,; \quad \|\underline{\hat{e}}\|^2_{Q_y^{-1}} = \|\underline{t}\|^2_{Q_t^{-1}} = \underline{t}^T Q_t^{-1} \underline{t}\,. \tag{2.24}$$

NOTE 2.4 The formulas of the weighted least-squares estimators of section 2.1.3 can also be obtained for the model of condition equations. For this purpose we need to define the projectors P_A and $P_A^\perp$ as $P_{W^{-1}B}^\perp$ and $P_{W^{-1}B}$, respectively, where Q_y is substituted by W^{-1} in projectors (2.21) .

2.2 Detection and validation in linear models

2.2.1 Testing simple and composite hypotheses

Most powerful test The simple likelihood ratio (SLR) test is derived based on the Neyman-Pearson testing principle. This principle states to choose, among all tests or critical regions possessing the same size type I error, α, the one for which the size of the type II error, β, is as small as possible. Such a test with the smallest possible type II error is called the *most powerful test*.

Simple hypotheses We consider two simple hypotheses. Using the Neyman-Pearson principle, we will test the null hypothesis against the alternative one. When the m-vector $\underline{y}$ has the probability density function (PDF) $f_y(y|x)$, we may define two simple hypotheses as $\mathrm{H}_o:\ x = x_o$ versus $\mathrm{H}_a:\ x = x_a$. Both hypotheses each pertain to a single distinct point in the parameter space. The objective is to decide, based on observations y of observables $\underline{y}$, from which of the two distributions the observations originated, either from $f_y(y|x_o)$ or from $f_y(y|x_a)$. The simple likelihood ratio (SLR) test as a decision rule reads (Teunissen et al., 2005)

$$\text{reject } \mathrm{H}_o \text{ if } \ \frac{f_y(y|x_o)}{f_y(y|x_a)} < a\,, \tag{2.25}$$

and accept otherwise, with a a *positive* constant (threshold). It can be proved that the simple likelihood ratio test is a most powerful test.

In practice we usually deal with composite hypotheses. We deal with the general problem of testing a composite hypothesis against another composite hypothesis. The *generalized likelihood ratio* (GLR) test is therefore defined. The fact that a composite hypothesis represents more than just a single distinct point in the parameter space complicates the notion of the power of a test. We can therefore address the *uniformly most powerful* (UMP) property of a test. In case of testing a simple hypothesis against a composite hypothesis, it is possible to derive indeed an UMP test, but most tests in practice (e.g. composite H_o versus composite H_a) are unfortunately not uniformly most powerful.

Composite hypotheses The probability density function (PDF) of observable vector $\underline{y}$ is $f_y(y|x)$ with $x \in \Phi$, where parameters in x may pertain to the location and shape of the PDF (think for instance of mean and variance). The parameter space Φ which contains all possible values for x is divided into two parts. The hypotheses then read

$$\mathrm{H}_o:\ x \in \Phi_o\,; \ \text{ and } \ \mathrm{H}_a:\ x \in \Phi \backslash \Phi_o\,. \tag{2.26}$$

The set $\Phi\backslash\Phi_o$ is the subset of Φ that is complementary to Φ_o. Therefore, $\Phi\backslash\Phi_o = \{x \in \Phi \mid x \notin \Phi_o\}$. The null and alternative hypotheses together cover the whole parameter space.

GLR test The PDF of the vector of observables $\underline{y}$ specified by $f_y(y|x)$ is a function of x. Therefore, the specification implies a whole family of PDFs. For the (given) observed y one considers $f_y(y|x)$ as a function of x. This is referred to as the *likelihood function* of y; see section 2.1.5. When x varies, the likelihood function produces the probability densities of all possible PDFs for the observed sample vector y. We maximize the likelihood function through the method of maximum likelihood estimation (section 2.1.5). This holds for x

restricted to the set Φ_o under the null hypothesis as well as for the unrestricted case $x \in \Phi$. The generalized likelihood ratio test is defined as

$$\text{reject } \mathrm{H}_o \text{ if } \frac{\max_{x\in\Phi_o} f_y(y|x)}{\max_{x\in\Phi} f_y(y|x)} < a\,, \tag{2.27}$$

and accept otherwise, with $a \in (0, 1)$. The GLR-test (2.27) yields a binary decision. The numerator implies maximization of the likelihood *within* the subset $\Phi_o \subset \Phi$ put forward by the null hypothesis H_o. The denominator amounts to a maximization over the *whole* parameter space Φ. When the numerator is, to a certain extent (specified by a), smaller than the denominator, the null hypothesis H_o is likely not true and therefore is rejected.

2.2.2 Hypothesis testing in linear models

In this section we address an important practical application of the generalized likelihood ratio (GLR) test, namely hypotheses testing in linear models. In many applications, observed data are treated with a linear model; validation of data and model together. The goal is to make a correct decision to be able to eventually compute estimates for the unknown parameters of interest. This also provides us with the criteria such as reliability to control the quality of our final estimators.

The observables are assumed to have a *normal* distribution. In addition, different hypotheses differ only in the specification of the *functional* model. Misspecifications in the functional model have to be handled prior to variance component estimation. In this chapter, the stochastic model of the observables is not subject to discussion or decision (see next chapters instead). When testing hypotheses on misspecifications in the functional model, we consider two types of *equivalent* tests: the *observation* test and the *parameter significance* test. These types of hypothesis testing using the GLR are dealt with when the covariance matrix Q_y of the observables is completely known. This is called 'σ known'. When the covariance matrix is known up to the variance of unit weight, i.e. $Q_y = \sigma^2 Q$, we will give some comments. This is referred to as 'σ unknown'.

Observation testing The general approach to handling observed data is that a nominal or default model is usually available. One wants to verify whether the observed data 'obey' the basis model also this time. There could have been disturbances or anomalies that invalidate the nominal model. Testing for gross errors and anomalies in the observations is referred to as *observation testing*. In this section the m-vector of observables $\underline{y}$ is assumed to be normally distributed

$$\underline{y} \sim \mathsf{N}_m(Ax, Q_y)\,, \tag{2.28}$$

with (for the moment) known covariance matrix Q_y. The following two composite hypotheses on the expectation of $\underline{y}$ are put forward:

$$\mathrm{H}_o : \mathsf{E}\{\underline{y}\} = Ax\,; \quad \text{and} \quad \mathrm{H}_a : \mathsf{E}\{\underline{y}\} = Ax + C_y\nabla,\ \nabla \neq 0\,, \tag{2.29}$$

with the $m \times n$ design matrix A and the n-vector x of unknown parameters. In the alternative hypothesis q additional unknown parameters in vector ∇ are related to the expectation of $\underline{y}$ by $m \times q$ matrix C_y. Matrix C_y is assumed to be of full column rank, i.e. $\mathrm{rank}(C_y) = q$. Columns of A and C_y are also assumed to be linearly independent, i.e.

$\mathrm{rank}(A\ C_y) = n + q$. The matrix C_y prescribes how unmodeled effects translate into the individual observations, i.e. into all elements of vector y.

The number of parameters which describes the expectation of the observables is extended in the alternative hypothesis H_a. The goal is to accommodate extraordinary effects in the observed process, e.g. disturbances, systematic effects, and gross errors of the observables. The generalized likelihood ratio (GLR) test will provide us with a decision whether or not the additional explanatory variables in ∇ should be taken into account.

Test statistic The probability density functions read, under H_o: $f_y(y|x) = \mathsf{N}(Ax, Q_y)$ and under H_a: $f_y(y|x) = \mathsf{N}(Ax + C_y\nabla, Q_y)$. The two parts of the GLR are $\max_x \mathsf{N}(Ax, Q_y)$ with $x \in \mathbb{R}^n$, and $\max_{x,\nabla} \mathsf{N}(Ax + C_y\nabla, Q_y)$ with $x \in \mathbb{R}^n$ and $\nabla \in \mathbb{R}^q$. Maximizing these likelihoods leads to the maximum likelihood estimators (MLE); or here BLUE as it is identical to MLE for a normal distribution. One can show that the test statistic related to this testing problem is given as (Teunissen et al., 2005)

$$\underline{T}_q = \underline{\hat{e}}_o^T Q_y^{-1} C_y (C_y^T Q_y^{-1} Q_{\hat{e}_o} Q_y^{-1} C_y)^{-1} C_y^T Q_y^{-1} \underline{\hat{e}}_o \,, \tag{2.30}$$

where $\underline{\hat{e}}_o$ is the least-squares residuals under the null hypothesis H_o. Therefore, practically one does not need to compute quantities under the alternative hypothesis. The index q refers to the additional degrees of freedom by vector ∇ in the alternative hypothesis. The preceding test statistic is distributed as

$$\mathrm{H}_o: \ \underline{T}_q \sim \chi^2(q, 0)\,; \quad \text{and} \quad \mathrm{H}_a: \ \underline{T}_q \sim \chi^2(q, \lambda)\,, \tag{2.31}$$

where the noncentrality parameter λ is given as

$$\lambda = \nabla^T C_y^T P_A^{\perp\,T} Q_y^{-1} P_A^{\perp} C_y \nabla = \|P_A^{\perp} C_y \nabla\|^2_{Q_y^{-1}} \,, \tag{2.32}$$

which equals the squared norm of the vector $P_A^{\perp} C_y \nabla$. The GLR-test is therefore used to decide between the extended model under the alternative hypothesis and the default model under the null hypothesis H_o.

Special cases The number of parameters in vector ∇ in alternative hypothesis (2.29) can range from 1 to $m - n$, i.e. $1 \leq q \leq m - n$. With m observations, the maximum number of estimable unknown parameters is m as well. If we have already n parameters in vector x, there are $q = m - n$ parameters left. This is considered as the upper bound for vector ∇ in the alternative hypothesis H_a. In this chapter we consider two special cases of the GLR-test in linear models, namely the case $q = m - n$ that yields the overall model test, and the case $q = 1$ that leads to the so-called w-test. For the parameter significance test we will only consider a special case that is equivalent to the w-test. This is referred to as the v-test.

2.2.3 Overall model test

Test statistic In the limiting case of $q = m - n$, there is no redundancy in the alternative hypothesis H_a in equation (2.29). This implies immediately that $\hat{e}_a = 0$, where $\hat{e}_a$ is the least-squares (BLUE) residual vector under the alternative hypothesis. One can show that the test statistic (2.30) yields

$$\underline{T}_{q=m-n} = \|\underline{\hat{e}}_o\|^2_{Q_y^{-1}} = \underline{\hat{e}}_o^T Q_y^{-1} \underline{\hat{e}}_o \,, \tag{2.33}$$

which is expressed as the quadratic form (squared norm) of the residuals (cf. equation (2.15)). The test of H_a with $q = m - n$ versus H_o by means of the above test statistic, is referred to as the *overall model* test.

Note that both A and C_y should be of full column rank and therefore matrix $(A\ C_y)$ square and invertible. According to the alternative hypothesis H_a, the vector of observations y is allowed to lie anywhere in the observation space $\mathbb{R}^m$. The vector of observations will *always* satisfy the alternative hypothesis, i.e. $y \in \mathbb{R}^m$. This vector is therefore checked for the validity all in one go *without* having yet a specific error signature in mind.

Distribution The test statistic value of (2.33) equals the squared norm of vector $\hat{e}_o$ and provides an overall inconsistency measure. The test statistic $\underline{T}_{q=m-n}$ is distributed as

$$\mathrm{H}_o:\ \underline{T}_q \sim \chi^2(m-n,0)\,;\ \text{ and }\ \mathrm{H}_a:\ \underline{T}_q \sim \chi^2(m-n,\lambda)\,, \tag{2.34}$$

where the noncentrality parameter λ in the alternative hypothesis is stated in equation (2.32).

Variance of unit weight When the covariance matrix of the observables is decomposed as $Q_y = \sigma^2 Q$ with Q the $m \times m$ cofactor matrix, then

$$\underline{\hat{\sigma}}^2 = \frac{\underline{\hat{e}}_o^T Q^{-1} \underline{\hat{e}}_o}{m-n} \tag{2.35}$$

is an *unbiased* estimator for the (un)known variance of unit weight since $\mathrm{E}\{\underline{\hat{e}}_o^T Q^{-1} \underline{\hat{e}}_o\} = \sigma^2(m-n)$; see example 4.7. The relation with the above overall model test statistic becomes clear from

$$\frac{\underline{\hat{\sigma}}^2}{\sigma^2} = \frac{\underline{\hat{e}}_o^T Q^{-1} \underline{\hat{e}}_o}{\sigma^2(m-n)}\,, \tag{2.36}$$

which equals $\underline{T}_{q=m-n}/(m-n)$, and is the estimator for the variance *factor* of unit weight. It indicates how we have to scale up or down the a-priori taken σ^2 to achieve the value for the overall model test statistic $\underline{T}_{q=m-n}$ being equal to its expectation value $m-n$. Next to gross errors and anomalies an incorrect stochastic model through covariance matrix Q_y may also cause rejection of the null hypothesis H_o in the overall model test. When for instance the elements of matrix Q_y are taken too small, the value for the overall model test will become too large. Therefore, it is (more) likely that the null hypothesis H_o will be rejected. In this case the estimate obtained from equation (2.36) measures how to scale up the a-priori variance σ^2.

NOTE 2.5 With just the overall model test through test statistic (2.33) we cannot allocate a rejection of the null hypothesis. Rejection can be caused by large (gross) errors in the observed data, an inappropriate (default) functional model for the data at hand, and/or by a poor specification of the observables' noise characteristics in the stochastic model (through matrix Q_y). Just the overall model test by itself cannot provide the answer. ⊡

2.2.4 The w-test statistic

In the data snooping method which was originally proposed by Baarda (1968), each individual observation is screened for the presence of an outlier. The number of parameters in

vector ∇ in alternative hypothesis (2.29) can range from 1 to $m-n$, i.e. $1 \leq q \leq m-n$. Let us now consider the lower bound of $q = 1$. We will test for specific single observational errors (single in the sense of one dimensional, $q = 1$). One popular application of this special test is to check every observation for a large error (outlier). This will be done for all m observations, implying that we will perform an m number of tests.

Test statistic When $q = 1$, the matrix C_y actually reduces to a vector. We will denote this m-vector by c_y. Introducing this vector in equation (2.30) leads to

$$\underline{T}_{q=1} = \frac{(c_y^T Q_y^{-1} \underline{\hat{e}}_o)^2}{c_y^T Q_y^{-1} Q_{\hat{e}_o} Q_y^{-1} c_y}\,, \tag{2.37}$$

where the numerator and denominator are both scalar quantities. The test statistic $\underline{T}_{q=1}$ is distributed as

$$\mathrm{H}_o:\ \underline{T}_q \sim \chi^2(1,0)\,;\quad \text{and}\quad \mathrm{H}_a:\ \underline{T}_q \sim \chi^2(1,\lambda)\,, \tag{2.38}$$

with the noncentrality parameter λ expressed in equation (2.32). In practice, it is more convenient to use the square root of the test statistic (2.37), namely

$$\underline{w} = \frac{c_y^T Q_y^{-1} \underline{\hat{e}}_o}{\left[c_y^T Q_y^{-1} Q_{\hat{e}_o} Q_y^{-1} c_y\right]^{1/2}}\,, \tag{2.39}$$

which is *normally distributed*, and even has standard normal distribution under H_o

$$\mathrm{H}_o:\ \underline{w} \sim \mathsf{N}(0,1)\,;\quad \text{and}\quad \mathrm{H}_a:\ \underline{w} \sim \mathsf{N}(\nabla w,1)\,, \tag{2.40}$$

with $\nabla w = [c_y^T Q_y^{-1} Q_{\hat{e}_o} Q_y^{-1} c_y]^{1/2} \nabla$. The noncentrality parameter in equation (2.32) is related to ∇w as $\lambda = \nabla w^2$. Random variable $\underline{w}$ is referred to as the *w-test* statistic.

Data snooping An important application of the w-test is *blunder* detection. A blunder, or outlier affects just a single observation. To screen the observations, in order to identify those that are grossly falsified by outliers, we formulate m alternative hypotheses. And they are all tested against the default or nominal model, represented by H_o. The vector c_y is taken a canonical unit vector, i.e. a vector with all zeros and a one at the ith position $c_{y_i} = [0,\ldots,1,\ldots,0]^T$, and i ranges from 1 to m. This screening of the observations with equation (2.39) is also referred to as *data snooping*. When the test for observation i is rejected, it is concluded that observation i is affected by some extraordinary large errors.

Normalized residuals When dealing with data snooping, if the covariance matrix Q_y of observables is diagonal, the expression for the w-test statistic reduces to a very simple form. The simple expression for the w-test statistic then reads

$$\underline{w}_i = \frac{\underline{\hat{e}}_i}{\sigma_{\hat{e}_i}}\,, \tag{2.41}$$

with $\sigma_{\hat{e}_i} = (Q_{\hat{e}_o})_{ii}^{1/2}$ the standard deviation of the least-squares residual i, for $i = 1,\ldots,m$. This quantity is also referred to as the *normalized* residual.

Geometric interpretation To conclude this section, we consider the geometric interpretation of the w-test statistic in general. With $\hat{\underline{e}}_o = P_A^\perp \underline{y}$ and $Q_y^{-1} P_A^\perp = P_A^{\perp T} Q_y^{-1} P_A^\perp$, the numerator of equation (2.39) equals the the inner product of the projected vector of observations $\hat{e}_o = P_A^\perp y$ with vector $P_A^\perp c_y$ in the observation space $\mathbb{R}^m$. The inner product equals $\langle P_A^\perp c_y,\, P_A^\perp \underline{y} \rangle = \|P_A^\perp c_y\| . \|P_A^\perp y\| \cos\varphi$ with φ the angle between the two vectors. Also with $Q_y^{-1} Q_{\hat{e}_o} Q_y^{-1} = P_A^{\perp T} Q_y^{-1} P_A^\perp$, the denominator is the length (norm) of m-vector $P_A^\perp c_y$ which equals $\|P_A^\perp c_y\|_{Q_y^{-1}}$. Assembling all parts leads to the conclusion that the test statistic (2.39) reads

$$\underline{w} = \frac{\langle P_A^\perp c_y,\, P_A^\perp \underline{y} \rangle_{Q_y^{-1}}}{\|P_A^\perp c_y\|_{Q_y^{-1}}} = \|\hat{\underline{e}}_o\|_{Q_y^{-1}} \cos\varphi\,, \tag{2.42}$$

which is equal to the *length* of the projection of vector $\hat{e}_o$ onto vector $P_A^\perp c_y$. This expression shows that the value of the w-test statistic is undefined, when vector c_y lies in the range space of matrix A, $c_y \in \mathcal{R}(A)$ or $\varphi = \pi/2$, and hence $P_A^\perp c_y = 0$. The occurrence of an error with a signature represented by this vector c_y can never be found or detected with statistical testing in this set up.

2.2.5 The v-test statistic

In the previous sections the default or nominal model H_o was tested against alternatives which compared to H_o were extended with one or more parameters in ∇. It is because we want to account for unmodeled effects as disturbances and gross errors in the observations. In this section we address the question whether the full nominal model is needed to appropriately describe the observed data. That is, we will test the model against a more restrictive one (*fewer* degrees of freedom). We consider a testing problem that, though mathematically equivalent to the above given testing problem, occurs when one wants to test the significance of parameters. The resulting test is thus referred to as a *parameter significance test*. The idea is to test whether or not it is possible to reduce the number of unknowns, e.g. by introducing hard constraints on the parameters. One goal for instance is to leave out parameters that are not significant (e.g. when they are statistically indistinguishable from zero).

General form We can derive the appropriate generalized likelihood ratio test of size α. Based on this GLR test we can obtain the corresponding test statistic as well as its distribution. The m-vector of observables $\underline{y}$ is assumed to be normally distributed. The following two composite hypotheses on the expectation of $\underline{y}$ are put forward:

$$\mathrm{H}_o:\ \mathsf{E}\{\underline{y}\} = Ax \quad \text{with} \quad C_x^T x = c_o\,; \quad \text{versus} \quad \mathrm{H}_a:\ \mathsf{E}\{\underline{y}\} = Ax, \tag{2.43}$$

with $n \times d$ constraint matrix C_x, and d-vector c_o with known values. The constrained linear model is introduced in appendix D.1. The $m \times n$ design matrix A is again assumed to be of full rank. The constraint matrix C_x is also assumed to be of full column rank, i.e. $\mathrm{rank}(C_x) = d$ where $d \leq n$. A row of matrix C_x^T generally implies a linear combination of the parameters in vector x.

The alternative hypothesis H_a is actually the default model, without constraints on the parameters. In the null hypothesis H_o the vector x of unknown parameters is subject to d (hard) constraints. In practice, generally by default, the computations are carried out

using the *alternative* hypothesis, i.e. the model without constraints in equation (2.43). The resulting quantities can be indexed here with $\cdot_a$. But just for convenience we will drop the index a.

v-test One may wonder whether all parameters currently included in vector x in the functional model are really necessary to describe the expectation of the observables. As a special case we consider the case with a single constraint ($d = 1$). The constraint in the null hypothesis becomes $c_x^T x = c_o$ with n-vector c_x and scalar c_o. The test statistic obtained from the generalized likelihood ratio test is given as (see appendix D.2)

$$\underline{v} = \frac{c_x^T \underline{\hat{x}} - c_o}{\sqrt{c_x^T Q_{\hat{x}} c_x}}\,, \tag{2.44}$$

which is referred to as the *v-test* statistic and is distributed as

$$\mathrm{H}_o:\ \underline{v} \sim \mathsf{N}(0,1)\,;\quad \text{and}\quad \mathrm{H}_a:\ \underline{v} \sim \mathsf{N}(\nabla v, 1)\,, \tag{2.45}$$

with the noncentrality parameter $\nabla v = (c_a - c_o)/\sqrt{c_x^T Q_{\hat{x}} c_x}$ where $c_a \neq c_o$ under the alternative hypothesis. Again, note that both $\hat{x}$ and $Q_{\hat{x}}$ are given under the unrestrictive *alternative* hypothesis H_a. One can give a similar geometric interpretation of the v-test statistic to that previously given for the w-test statistic. One can also obtain the w-test statistic from the v-test statistic and vice versa (see appendix D.3).

Example 2.1 In case one is interested in a single unknown parameter, the vector $c_x = c_i$ can be taken to be the corresponding canonical unit vector. The test statistic (2.44) then simplifies to

$$\underline{v} = \frac{c_i^T \underline{\hat{x}} - c_o}{\sqrt{c_i^T Q_{\hat{x}} c_i}} = \frac{\underline{\hat{x}}_i - c_o}{\sigma_{\hat{x}_i}}\,. \tag{2.46}$$

One can in addition test whether or not two unknown parameters x_j and x_i are equal. In this case, vector $c_x = c_j - c_i$ and scalar $c_o = 0$ which lead to the v-test statistic as

$$\underline{v} = \frac{\underline{\hat{x}}_j - \underline{\hat{x}}_i}{\sqrt{\sigma_{\hat{x}_i}^2 + \sigma_{\hat{x}_j}^2 - 2\sigma_{\hat{x}_i\hat{x}_j}}}\,, \tag{2.47}$$

with $\sigma_{\hat{x}_i}^2$ and $\sigma_{\hat{x}_i\hat{x}_j}$, respectively, the variance and covariance elements of $\hat{x}$ obtained from $Q_{\hat{x}}$ under alternative hypothesis H_a. ■

2.2.6 Unknown variance of unit weight

So far the parameter estimation and testing of statistical hypotheses in linear models concerned the case that the covariance matrix Q_y of the observables is completely known. This sometimes turns out not to be the case. We can address the case that the covariance matrix is decomposed into a known cofactor matrix Q and an unknown variance of unit weight, i.e. $Q_y = \sigma^2 Q$. This is referred to as σ unknown. This is the simplest form of an unknown covariance matrix, which can occur in many practical applications.

One can simply show that the BLUE estimators are not affected by the unknown variance factor of unit weight. This is however not the case for their covariance matrices as they are directly dependent on σ^2. For hypotheses testing, the difference here with the case 'σ known' is that the overall model test does not exist when σ is unknown. We can just

estimate the unknown variance factor of unit weight and testing is not possible. But one can derive the corresponding w-test and v-test statistics. In this case both test statistics have a Student t distribution; see Teunissen et al. (2005).

In this thesis we will assume a more complicated form of the covariance matrix. In the next chapters we consider the problem of estimating the stochastic model. The covariance matrix can in general be written as a linear combination of some known cofactor matrices. This linear combination is, however, not known and some variance and covariance components are to be estimated. We are going to estimate the unknown (co)variance components of the stochastic model based on the least-squares principle that was introduced in this chapter.

Variance Component Estimation: A Review

3

3.1 Introduction

So far we assumed that the covariance matrix Q_y is known. Therefore, in the estimation process one is usually concerned about the optimal estimation of the unknown parameters in the functional model. In this chapter we consider the problem of estimating the stochastic model. The importance of specifying a correct stochastic model is evidenced by formulas (2.13). What we really need in these equations is a realistic covariance matrix of observables. A realistic description of the measurement noise characteristics through the observation covariance matrix is required to yield minimum variance (best) estimators (BLUE). Often in geodetic literature, the covariance matrix of the observables $\underline{y}$ consists of one simple component, namely

$$\mathrm{E}\{\underline{y}\} = Ax; \quad Q_y = \sigma^2 Q\,, \tag{3.1}$$

where σ^2 is the *unknown* variance of unit weight and Q is the *known* $m \times m$ symmetric and positive-definite cofactor matrix. This is in fact the simplest form of an unknown covariance matrix in a linear model. The BLUE expressions $\hat{x}$, $\hat{y}$, and $\hat{e}$ are invariant against a change in the variance σ^2, but in general not against other changes like different weighting and covariances. More importantly, the covariance matrices $Q_{\hat{x}}$, $Q_{\hat{y}}$, and $Q_{\hat{y}}$ describing the precision of the least-squares estimators, directly depend on Q_y. In addition, test statistics highly depend on the a-priori assumption for the stochastic model.

This chapter reviews the principles and the resulted formulas for the (co)variance component estimation problem. We start from the same point of departure, but each time with a different line of thought. We will show how these formulas give identical results under certain conditions. There are different estimation methods based on optimality criteria as unbiasedness, best (minimum variance), minimum norm, and maximum likelihood (see previous chapter). Over the past years several variance component estimation (VCE) techniques have been developed. A complete review of all methods would be beyond the scope of this thesis. A relatively comprehensive review of several rigorous and simplified methods for VCE is given by Grafarend (1985); Crocetto et al. (2000); Fotopoulos (2003). In this chapter we review the methods that were previously developed for estimating the (co)variance components. We will outline the underlying assumptions of each method. The focus is placed on identifying and discussing the key developments related to geodetic research and applications.

An effective way for characterizing different VCE procedures is to list them according to certain distinguishable features. Most approaches for VCE within a least-squares estimation framework can be categorized after Crocetto et al. (2000)

- functional model (Gauss-Markov, condition model, Gauss-Helmert, etc.);

- stochastic model (block diagonal, block-structured, etc.);
- estimation principles (MINQUE, BIQUE, MLE, Bayesian method, etc.);
- rigorous or simplified methods (as almost unbiased estimation (AUE), etc.).

Using such a categorical scheme, a timeline outlining the key developments in VCE theory applicable to geodetic problems is tabulated by Fotopoulos (2003).

Functional model The relation between variables can be expressed in different equivalent forms. In chapter 2 we gave two important formulations of the functional model, namely the model of observation equations (Gauss-Markov model, see equation (2.4)) and the model of condition equations (condition model, see equation (2.18)). Another important formulation is called mixed model which involves both the observables and the unknown parameters in an implicit form

$$B^T \mathsf{E}\{\underline{y}\} = Ax; \quad \mathsf{D}\{\underline{y}\} = Q_y \,. \tag{3.2}$$

In the geodetic literature, the preceding model is also known as Gauss-Helmert model. A desirable property of an estimator should be its independence from the type of the formulation of the functional model. For example, the least-squares residuals of $\underline{y}$ are invariant to the sort of the formulation of the functional model. Given the (co)variance estimators in one of the functional models, it is not difficult to obtain estimators for other formulations (see e.g. section 4.7). In this chapter we will only consider the model of observation equations.

The stochastic model consists of variance and covariance components. In the next section we will consider a general form of the stochastic model. Our emphasis in a later section is on the major existing estimation principles.

3.2 Variance component model

The concept of a (co)variance component model can now be introduced as an extension of the linear model of observation (or condition) equations where the covariance matrix is written as a linear combination of p components. The division of the covariance matrix in components should always be the result of a detailed analysis of the stochastic model. Preferably each component should have a natural explanation in the measurement process or in the physical limitations of the measurements. In many cases the basis of a variance component model can be found in an analysis of the $m \times 1$ vector of stochastic errors. The resulting observational errors can often be considered as a function of several elementary error sources. A relevant decomposition of the stochastic model can then be derived from an analysis of the influence of these elementary errors on the observations. Consider the observational errors $\underline{e}$ as a linear function of $\bar{p}$ groups of errors $\underline{\epsilon}_k$, $k = 1, ..., \bar{p}$

$$\underline{e} = U_1\underline{\epsilon}_1 + ... + U_{\bar{p}}\underline{\epsilon}_{\bar{p}} = \sum_{k=1}^{\bar{p}} U_k\underline{\epsilon}_k = U\underline{\epsilon}\,, \tag{3.3}$$

with the $m \times m_k$ transformation matrices U_k describing the influence of the k^{th} error group on the observables, and the $m_k \times 1$ vectors $\underline{\epsilon}_k$ of error sources for different groups. Note also that the $m \times \sum_{k=1}^{\bar{p}} m_k$ matrix U and the $\sum_{k=1}^{\bar{p}} m_k \times 1$ vector $\underline{\epsilon}$ are respectively given

as: $U = [U_1, ..., U_{\bar{p}}]$ and $\underline{\epsilon} = [\underline{\epsilon}_1^T, ..., \underline{\epsilon}_{\bar{p}}^T]^T$. We assume that the error sources are zero on average, they have different variance components, and in general they are mutually correlated. In other words

$$\mathsf{E}\{\underline{\epsilon}_k\} = 0; \quad \mathsf{D}\{\underline{\epsilon}_k\} = \sigma_{kk} Q_{\epsilon_k}; \quad \mathsf{C}\{\underline{\epsilon}_k, \underline{\epsilon}_l\} = \sigma_{kl} Q_{\epsilon_k \epsilon_l}; \quad k, l = 1, ..., \bar{p}\,. \tag{3.4}$$

Application of the error propagation law to equation (3.3) leads to the (co)variance component model with $p = \frac{\bar{p}(\bar{p}+1)}{2}$ variance and covariance components

$$Q_y = \sum_{k=1}^{\bar{p}} \sigma_{kk} U_k Q_{\epsilon_k} U_k^T + \sum_{k=1}^{\bar{p}-1} \sum_{l=k+1}^{\bar{p}} \sigma_{kl} (U_k Q_{\epsilon_k \epsilon_l} U_l^T + U_l Q_{\epsilon_l \epsilon_k} U_k^T)\,. \tag{3.5}$$

Just for the sake of notation convenience the above stochastic model can be reformulated as $Q_y = \sum_{k=1}^{p} \sigma_k Q_k$ with p-number of unknown variance and/or covariance components σ_k and their known cofactor matrices Q_k of the form either $U_k Q_{\epsilon_k} U_k^T$ or $U_k Q_{\epsilon_k \epsilon_l} U_l^T + U_l Q_{\epsilon_l \epsilon_k} U_k^T$. The Gauss-Markov model with p (co)variance components is then given as

$$\mathsf{E}\{\underline{y}\} = Ax; \quad Q_y = \sum_{k=1}^{p} \sigma_k Q_k\,, \tag{3.6}$$

where $\sigma_1, ..., \sigma_p$ are *unknown* (co)variance components, and $Q_1, ..., Q_p$ are *known* symmetric and positive definite $m \times m$ cofactor matrices. The estimation of the (co)variance components shall therefore be considered as a generalization of the estimation of the variance of unit weight. Variance components are estimated, for example, if sets of different observations like those of the carrier-phase and pseudorange GPS observables are given. If in addition different observation types are statistically dependent, then covariance components are estimated to express the degree of dependence of the observables. The above unknown covariance matrix is called the *(co)variance component model* or simply the *stochastic model*. This by default also includes $\mathsf{E}\{\underline{y} - Ax\} = \mathsf{E}\{\underline{e}\} = 0$.

As a *necessary* condition, the cofactor matrices $Q_1, ..., Q_p$ should be linearly independent (see NOTE 4.2). One can show that the system of equations in the stochastic model becomes singular if at least one of the cofactor matrices is written as a linear combination of the others, i.e. if $Q_l = \sum_{k=1, k \neq l}^{p} \alpha_k Q_k$. The (co)variance component model, therefore, needs careful consideration before we make any inference on the unknown (co)variance components (see section 4.8).

3.3 Rigorous methods

3.3.1 MINQUE method

Rao (1971a, 1973) derived the minimum norm quadratic unbiased estimators (MINQUE) for which he avoids any distributional assumption. The MINQUE method is one of the commonly used methods for the estimation of (co)variance components. This approach was extended to the condition model and to the general Gauss-Helmert model by Sjöberg (1983). In the following we will give the underlying assumptions and the final results of the MINQUE method. Rao (1971a) derived the MINQUE estimators first for variance components and then extended it for covariance components. Here we consider only

the problem of variance component estimation. Let an estimator of a linear function $f^T\sigma$ of variance components be defined as a quadratic form of the observables $\underline{y}$, namely $f^T\hat{\underline{\sigma}} = \underline{y}^T M \underline{y}$. This estimator is now subject to the following three conditions:

- The estimator should be unbiased (i.e. $A^T MA = 0, \;\; \mathrm{tr}(MQ_k) = f_k, \; k = 1, ..., p$),
- The estimator is invariant under translations of the parameters x (i.e. $MA = 0$),
- The difference between estimator and a natural estimator in sense of Euclidean norm should be minimum.

The estimator is unbiased if $\mathsf{E}\{f^T\hat{\underline{\sigma}}\} = \mathsf{E}\{\underline{y}^T M \underline{y}\} = \mathrm{tr}(MQ_y) + x^T A^T MAx = f^T\sigma$. The unbiasedness property can therefore be achieved if the second term is zero, i.e. $A^T MA = 0$ and the first term equals $f^T\sigma$, i.e. $\mathrm{tr}(MQ_k) = f_k$. The estimators should also be invariant under translation of the parameters x (i.e. transformation $\underline{y} \rightarrow \underline{y} + Ax$ for arbitrary x). One can simply show that this can be achieved if $MA = 0$. To understand the last criterion let us refer to equation (3.5) when the covariance components are absent. If the stochastic model $\underline{\epsilon}$ were known, a *natural* estimator for the linear function $f^T\sigma$ is given as

$$f^T\hat{\underline{\sigma}} = \sum_{k=1}^{p} f_k \hat{\underline{\sigma}}_k = \sum_{k=1}^{p} \frac{f_k}{m_k} \underline{\epsilon}_k^T Q_{\epsilon_k}^{-1} \underline{\epsilon}_k = \underline{\epsilon}^T D \underline{\epsilon}, \text{ with } D \text{ known}, \tag{3.7}$$

where $\underline{\epsilon}_k$ is a hypothetical random vector with mean zero and dispersion matrix $\sigma_k Q_{\epsilon_k}$. The preceding equation is written as a quadratic form of $\underline{\epsilon} = [\underline{\epsilon}_1^T, ..., \underline{\epsilon}_p^T]^T$. Another estimator is $f^T\hat{\underline{\sigma}} = \underline{y}^T M \underline{y}$ which can also be written in a quadratic form of $\underline{\epsilon}$ as $\underline{\epsilon}^T U^T MU \underline{\epsilon}$ (note that $\underline{y} = A\underline{x} + U\underline{\epsilon}$). To ensure the estimator to be optimal, the difference between these two estimators should be minimum

$$\min_M \underline{\epsilon}^T (U^T MU - D) \underline{\epsilon}\,. \tag{3.8}$$

Rao proposed to minimize the Euclidean norm (sum of squares of all elements of a matrix) of the matrix $U^T MU - D$. This leads to the following minimization problem:

$$\min_M \|U^T MU - D\|^2 = \min_M \mathrm{tr}(MQ_y MQ_y)\,, \tag{3.9}$$

subject to the invariance and unbiasedness conditions

$$MA = 0, \;\; \mathrm{tr}(MQ_k) = f_k, \; k = 1, ..., p\,. \tag{3.10}$$

The preceding minimization is solved for M using a *minimum trace problem*. This leads to the following normal equations for estimation of the (co)variance components

$$\hat{\underline{\sigma}} = N^{-1}\underline{l}\,, \tag{3.11}$$

with the $p \times p$ matrix N and the p-vector $\underline{l}$ as

$$n_{kl} = \mathrm{tr}(Q_k Q_y^{-1} P_A^{\perp} Q_l Q_y^{-1} P_A^{\perp}); \;\; \text{and} \;\; \underline{l}_k = \underline{y}^T P_A^{\perp T} Q_y^{-1} Q_k Q_y^{-1} P_A^{\perp} \underline{y}, \tag{3.12}$$

respectively, and $k, l = 1, 2, ..., p$. In a later work, Rao (1971b) derived the minimum variance quadratic unbiased estimators of variance components that under the normality condition coincides with the MINQUE estimators. This can also be considered as an alternative derivation for BIQUE estimators (see next subsection).

3.3.2 BIQUE method

The BIQUE method as a quadratic-based estimation in the stochastic model is the counterpart to the BLUE method as a linear-based estimation in the functional model. They both give minimum variance estimators. Koch (1978, 1999) uses Lagrange multipliers to solve the best invariant quadratic unbiased estimators (BIQUE) when observables are normally distributed. The term *best* indicates that the estimator has minimum variance compared to all other quadratic unbiased estimators. Caspary (1987) assumes the normal distribution and considers the covariance matrix as a sum of some variance components. He defines the BIQUE and derives an iterative procedure for these estimates. Sjöberg (1984) assumes the normal distribution and a stochastic model consisting of variances. He has suggested an iterative procedure for the BIQUE estimates and shows the coincidence of these estimates with the MINQUE. With regard to the more general functional model (Gauss-Helmert with constraints on the parameters), Yu (1992) used a block-structured stochastic model. He derives the BIQUE theory of (co)variance component estimation, showing that the MINQUE of Rao (1971a) and BIQUE of Sjöberg (1984) are special cases of his estimate. In addition, he shows that, in the case of a block-structured covariance matrix, the Helmert estimators of (co)variance components are identical with BIQUE.

In the following we will give the underlying assumptions and the final results of the BIQUE method. We again consider an estimator of a linear function $f^T\sigma$ of (co)variance components as a quadratic form of the observables $\underline{y}$, namely $f^T\hat{\underline{\sigma}} = \underline{y}^T M \underline{y}$. Strictly speaking the first two conditions are the same as Rao's. The third condition is rewritten as

- The estimator should be of minimum variance ('best' property).

Note that the unbiasedness and invariance conditions are independent of the distribution of the observations. The best property is achieved when we minimize the variance of a quadratic form, which can only be specified if up to and including the fourth central moment of the distribution of the observables is known. Based on the normal distribution this leads to the following minimization problem:

$$\min_M \mathsf{D}\{\underline{y}^T M \underline{y}\} = \min_M \mathrm{tr}(MQ_yMQ_y)\,, \tag{3.13}$$

subject to the invariance and unbiasedness conditions

$$MA = 0, \quad \mathrm{tr}(MQ_k) = f_k,\ k = 1, ..., p\,. \tag{3.14}$$

Koch (1978, 1999) minimizes the variance of the estimator by minimizing a Lagrange function consisting of the unbiasedness and translational invariance constraints

$$g(M) = 2\mathrm{tr}(MQ_yMQ_y) - 4\mathrm{tr}(MA\Lambda^T) - 4\sum_{k=1}^{p}\lambda_k(\mathrm{tr}(MQ_k) - f_k)\,, \tag{3.15}$$

where -4Λ denotes the $m \times n$ matrix of Lagrange multipliers for the constraints $MA = 0$ and $-4\lambda_k$ are the p Lagrange multipliers for the constraints $\mathrm{tr}(MQ_k) = f_k$. Solving for M from the above minimization problem leads again to equation (3.11) with equation (3.12). Schaffrin (1983) obtained the BIQUE by the E-D-correspondence method.

3.3.3 Helmert method

In geodetic applications, variance component estimation originates from Helmert (1907) who used the least-squares residuals to estimate heterogeneous variance components. Using the Gauss-Markov model (model of observation equations), Helmert (1924) proposed a stepwise method for unbiased variance estimates. Grafarend (1984) extended Helmert's method to the models of condition equations and Gauss-Helmert (mixed model). As an intuitive approach the method is closely related to a derivation from the expectation of the shifting variate proposed by Förstner (1979). The derivation of the (co)variance components is in fact only based on the unbiasedness and invariance properties.

The point of departure here is the least-squares residuals $\underline{\hat{e}} = P_A^{\perp}\underline{y} = P_A^{\perp}\underline{e}$ which are invariant with respect to the transformation $\underline{y} \rightarrow \underline{y} + Ax$. The shifting variate is the squared norm of the least-squares residuals. Its expectation is given as

$$\mathsf{E}\{\underline{\hat{e}}^T Q_y^{-1}\underline{\hat{e}}\} = \mathsf{E}\{\underline{e}^T P_A^{\perp T} Q_y^{-1} P_A^{\perp}\underline{e}\} = \mathrm{tr}(P_A^{\perp T} Q_y^{-1} P_A^{\perp} Q_y)\,. \tag{3.16}$$

Grafarend (1985) used a block-structured covariance matrix: $Q_y = \sum_{l=1}^{p} \sigma_l Q_l$ with the multinomial inverse of the form $Q_y^{-1} = \sum_{k=1}^{p} E_k$. He gave a simple example how to obtain E_k's. When the covariance matrix Q_y has a block-diagonal structure one can also simply obtain E_k's. Substituting these two terms in the preceding equation yields

$$\sum_{k=1}^{p} \mathsf{E}\{\underline{\hat{e}}^T E_k \underline{\hat{e}}\} = \sum_{k=1}^{p}\sum_{l=1}^{p} \mathrm{tr}(P_A^{\perp T} E_k P_A^{\perp}\sigma_l Q_l)\,. \tag{3.17}$$

From the expectation of the k^{th} term one obtains

$$\mathsf{E}\{\underline{\hat{e}}^T E_k \underline{\hat{e}}\} = \sum_{l=1}^{p} \mathrm{tr}(P_A^{\perp T} E_k P_A^{\perp} Q_l)\sigma_l, \quad k = 1, ..., p\,. \tag{3.18}$$

The preceding equation can be written in a compact form as $\mathsf{E}\{\underline{q}\} = H\sigma$ with the $p \times p$ matrix H and p-vector $\underline{q}$ as

$$h_{kl} = \mathrm{tr}(P_A^{\perp T} E_k P_A^{\perp} Q_l); \quad \underline{q}_k = \underline{\hat{e}}^T E_k \underline{\hat{e}} = \underline{y}^T P_A^{\perp T} E_k P_A^{\perp}\underline{y}; \quad k, l = 1, ..., p\,. \tag{3.19}$$

respectively. If H is regular, an unbiased estimator of (co)variance components reads

$$\underline{\hat{\sigma}} = H^{-1}\underline{q}\,. \tag{3.20}$$

One can also show that the above estimators are also translational invariant. Grafarend (1984, 1985) did not specify a general way of obtaining the multinomial submatrices E_k. From $Q_y^{-1} = Q_y^{-1} Q_y Q_y^{-1} = \sum_{k=1}^{p} E_k$ with $Q_y = \sum_{k=1}^{p} \sigma_k Q_k$ it follows that we may choose $E_k = \sigma_k Q_y^{-1} Q_k Q_y^{-1}$. If we now substitute E_k in equation (3.19) we obtain $h_{kl} = \sigma_k n_{kl}$ and $\underline{q}_k = \sigma_k \underline{l}_k$, where n_{kl} and $\underline{l}_k$ are given in equation (3.12). This confirms the equality of the estimators with the MINQUE and BIQUE estimators.

3.3.4 MLE method

Maximum likelihood estimation (MLE) as one of the conceptually simplest methods is applicable only when the general structure of the probability density function is known.

Most of the efforts in the field of (co)variance component estimation is restricted to the normal distribution. There are two methods of maximum likelihood, namely (unrestricted) ML and restricted ML. The maximum likelihood estimator of variance components in a linear model can be biased downwards because it does not account for the degrees of freedom lost by estimating the unknown parameters x in the linear model. Restricted maximum likelihood (REML) corrects this problem by maximizing the likelihood of a set of residual contrasts and is generally considered superior.

Using the Gauss-Helmert model, Kubik (1970) considers a simplified stochastic model. He used the maximum likelihood method for estimating weight ratios in a hybrid distance-direction network. Koch (1986), assuming the Gauss-Markov model and that the observables are normally distributed, derives an iterative procedure for the maximum-likelihood estimates of the (co)variance components using the orthogonal complement likelihood function. This approach is equivalent to the restricted maximum likelihood estimation (REML). He also shows that these estimators are identical to the best invariant quadratic unbiased estimators (BIQUE) of (co)variance components and also to the MINQUE estimators. Ou (1989), with the same functional model and with normally distributed data, assumes a block-structured covariance matrix. He shows that the iterative maximum-likelihood estimates are identical to the Helmert (1924) and Koch (1986) estimates. Making the same assumptions about the functional and stochastic models, Yu (1996) derives the maximum likelihood estimates of the (co)variance components. He expands the formulas given by Kubik (1970) and Koch (1986). In addition, he proves the equivalence of the maximum-likelihood estimator to that of the Helmert type and to the BIQUE.

Assume that the observables $\underline{y}$ have a multivariate normal distribution, namely $\underline{y} \sim \mathsf{N}(Ax, \sum_{k=1}^{p} \sigma_k Q_k)$. Then the log-likelihood function associated with $\underline{y}$ is

$$\ln L(y; x, \sigma) = -\frac{m}{2} \ln 2\pi - \frac{1}{2} \ln \det(Q_y) - \frac{1}{2}(y - Ax)^T Q_y^{-1}(y - Ax)\,. \tag{3.21}$$

The unknown parameters of the log-likelihood function (i.e. x and σ) are then solved by setting the partial derivatives of the unknown parameters equal to zero. Maximization of $\ln L$ with respect to x yields the well-known normal equations of the maximum likelihood estimators which gives identical results with the method of least-squares or best linear unbiased estimation (BLUE); see equation (2.13). One should however note that the maximum likelihood estimator of σ is biased downwards. To handle the problem the restricted maximum likelihood is employed (Patterson and Thompson, 1971; Harville, 1977). The method takes care of the "loss in redundancy" due to estimating the unknown parameters, in which we maximize the likelihood of a set of $m - n$ linearly independent residual contrasts. In fact the transfered observables $\underline{t} = B^T \underline{y}$ are the residual contrasts. In geodetic literature this is equivalent to reformulating the model of observation equations in terms of the model of condition equations. With doing so, for normally distributed data, we will get the same results as the MINQUE and BIQUE methods (see section 4.6). Such estimators are therefore unbiased, of minimum variance, and restrictedly of maximum likelihood.

3.3.5 Bayesian method

Another approach sometimes used for estimating the unknown (co)variance components in linear models is the Bayesian method (Koch, 1987). Both the maximum likelihood and the Bayesian methods belong to the family of distribution-based methods as we need

to assign a distribution function to the data. The key difference is that the Bayesian method requires some prior information (e.g. provided by equipment manufacturers) about the vector of (co)variance components in the form of a prior probability density function (PDF). Application of the Bayes theorem results in the a posterior PDF which is the product of the likelihood function of the data and the priori PDF of the (co)variance components. To find the Bayes estimate, classical techniques can be applied to the posterior PDF. A common technique is the maximum likelihood which was introduced in the previous section. The principle of maximum a posteriori (MAP), (see e.g. Koch, 1987; Ou and Koch, 1994) or also called generalized maximum likelihood (GML) (see e.g. Grodecki, 1999, 2001) relies on maximizing the a posterior PDF.

Using the normal distribution, Bayes estimators and confidence intervals for the (co)variance components are obtained by Koch (1987). Ou (1991) proposes an approximate and a strict Bayes estimation method and the Bayes confidence limits for variance components in the Gauss-Markov model for a block-diagonal covariance matrix (disjunctive uncorrelated groups). A few years later, Ou and Koch (1994) gave an analytical expression for Bayes estimates of variance components. Based on informative and non-informative priors, they derive posterior density functions of variance components using series expansion of the likelihood function. Grodecki (1999) derived the generalized maximum-likelihood (GML) estimator using the inverted gamma distribution as the prior information. In a later work, Grodecki (2001) applied the principle of generalized maximum-likelihood (GML) estimation to the a posterior probability density function with no prior information for the estimation of (co)variance components. He also showed that his estimator agrees numerically with the MAP estimator of Ou and Koch (1994).

Since a full explanation of this method is beyond the scope of this work, we only give an example of variance estimators derived by the Bayesian method. For non-informative priors, the Bayes and the MAP estimators of variance components of a disjunctive uncorrelated group model are respectively of the form (Koch, 1990; Ou and Koch, 1994)

$$\underline{\hat{\sigma}}_k^B = \frac{b_k}{b_k - 2}\underline{\hat{\sigma}}_k; \quad \underline{\hat{\sigma}}_k^M = \frac{b_k}{b_k + 2}\underline{\hat{\sigma}}_k; \quad k = 1, ..., p\,, \tag{3.22}$$

where $\underline{\hat{\sigma}}_k$ is the BIQUE estimator of σ_k and b_k is the so-called redundancy number for the i^{th} set of observations. Using this method, one can also obtain a confidence interval for variance components (see Ou, 1991).

3.3.6 Non-negative methods

A drawback of the VCE techniques is that the estimate of variance components may turn out to be negative, which is obviously not acceptable. Sjöberg (1984), Caspary (1987) and Kubik (1970) affirm that a negative BIQUE estimate may be due to either an insufficient number of observations (low redundancy in functional model) or an improper stochastic model. As a quadratic-based estimation scheme, there have been suggested two non-negative methods, namely

- best quadratic minimum bias non-negative estimation (BQMBNE)
- best quadratic unbiased non-negative estimation (BQUNE)

Sjöberg (1995) demonstrates that the two conditions of *non-negativeness* and *unbiasedness* are incompatible in an additive two-variance component stochastic model. In an

earlier work, Sjöberg (1984) derived explicit formulas for the BQMBNE estimators in a (co)variance model consisting of uncorrelated disjunctive groups (each group consists of two variance factors). He also gives the BQUNE estimators of individual variance components. The BQUNE exists in the case of the disjunctive uncorrelated groups with only one variance factor for each group. In other words, for such a model the variance components are always positive and therefore they are automatically unbiased. The BQMBNE coincides with the BQUNE if it exists. Moreover, Sjöberg notes that the computational effort for the BQMBNE is much larger than that for the BIQUE.

There are some methods that intrinsically give non-negative variance components. One of such methods is given in section 3.4. In section 4.8.1 we will suggest a method which enforces the variance components to be non-negative.

3.3.7 LS method

Teunissen (1988) developed a *least-squares* formula for estimation of (co)variance components. Since the approach is based on the least-squares principle, the estimators are unbiased. If we take the weight matrix as the inverse of the covariance matrix of the observables we can in addition obtain the minimum variance estimators. The derivation given of the least-squares VCE formula is based on the model of condition equations rather than the model of observation equations. We can however present the LS-VCE (co)variance estimators for both the model of observation equations and the model of condition equations. We can derive the least-squares (co)variance component estimation formula by rewriting the (co)variance component model into a linear model of observation equations. Motivated by the fact that the method of least squares is one of the leading principles in parameter estimation, in this thesis, we introduce and develop the method of least-squares variance component estimation (LS-VCE) and apply the theory to a few GPS applications.

It can be shown that five different methods of VCE, namely MINQUE, BIQUE, Helmert, REML and LS, give identical estimators if the normal distribution is assumed. Least-squares variance component estimation has however many attractive features that will be elaborated upon in this thesis. It provides a unified least-squares framework for estimating the unknown parameters of both the functional and stochastic models. Since the method is based on the least-squares principle, our existing body of knowledge on least-squares theory is directly applicable to LS-VCE. We can at least mention the following features:

- we can find a general class of unbiased estimators which are independent of the distribution of the data,
- we obtain minimum variance estimators for a class of elliptically contoured distributions. For special examples we obtain the minimum variance estimators which are independent of the distribution,
- as with the functional model, the covariance matrix of (co)variance estimators can directly be obtained,
- LS-VCE has a similar insightful geometric interpretation as standard least squares,
- properties of the normal matrix and the orthogonal projectors can easily be established,

- measures of inconsistency such as the quadratic form of residuals and the w-test statistic can directly be given.

The theory of LS-VCE will be treated in chapters 4, 5, and 6.

3.4 Simplified methods

The use of variance component estimation formulas is often time consuming and results in a vast amount of computational work. In general, an iterative procedure, starting from a set of initial values for (co)variance components, is needed (see next chapter, figures 4.1 and 4.2). In geodetic literature to reduce the computational load of (co)variance component estimation the following ways are recommended:

- Approximate variance component estimation formulas
  Because of the considerable computational effort required for even moderately sized models, approximate methods have been developed. Almost unbiased estimation (AUE) is intended to reduce the large computational load required in VCE (Horn et al., 1975; Förstner, 1979). To derive the almost unbiased estimation formula of variance components, let us rewrite equation (3.16) as

  $$\mathsf{E}\{\hat{\underline{e}}^T Q_y^{-1} Q_y Q_y^{-1} \hat{\underline{e}}\} = \mathrm{tr}(Q_y^{-1} P_A^{\perp} Q_y)\,. \tag{3.23}$$

  If we substitute $Q_y = \sum_{k=1}^{p} \sigma_k Q_k$, it follows that

  $$\sum_{k=1}^{p} \mathsf{E}\{\hat{\underline{e}}^T Q_y^{-1} \sigma_k Q_k Q_y^{-1} \hat{\underline{e}}\} = \sum_{k=1}^{p} \mathrm{tr}(Q_y^{-1} P_A^{\perp} \sigma_k Q_k)\,. \tag{3.24}$$

  By equating the corresponding terms in the preceding equation, the variance components are estimated using the following iterative formula:

  $$\hat{\underline{\sigma}}_k = \frac{\hat{\underline{e}}^T Q_y^{-1} \sigma_k Q_k Q_y^{-1} \hat{\underline{e}}}{\mathrm{tr}(Q_y^{-1} P_A^{\perp} Q_k)},\ k = 1, ..., p, \tag{3.25}$$

  which is not in general an unbiased estimator of variance components but almost. If the ratio of the a priori values for the variance factors is correct (when they are known up to a factor α) this formula gives *unbiased* estimators. An alternative formula for almost unbiased estimation of variance factors is given by Hsu (1999). One can show that the above estimator leads, at the point of convergence, to the results of the iterated estimates of the local MINQUE, BIQUE or REML estimates (see Koch, 1986, 1987; Ou, 1989; Koch, 1999). That is, the iterated almost unbiased estimators (IAUE) are also *unbiased* at the point of convergence. This can also be verified by Hsu (2001) who showed that the Helmert method is identical with the IAUE method. It can easily be shown that a variance component obtained from the IAUE is always positive provided that the resulting covariance matrix is positive definite.

- Simple formulation of functional and stochastic models
  Different authors have proposed approximate methods with low computational cost (Kusche, 2003a; Satirapod et al., 2002; Lucas and Dillinger, 1998; Sjöberg, 1984;

Caspary, 1987; Ou, 1989, 1991; Barbarella and Pieri, 1983). Most of these methods assume more simplified stochastic models which are frequently used in geodetic applications. They assume for instance that the covariance matrix is block-diagonal

$$Q_y = \mathrm{blkdiag}(\sigma_k Q_k),\ k = 1, ..., p\,, \tag{3.26}$$

where $\mathrm{blkdiag}$ is a block-diagonal operator and σ_k are some variance components. This stochastic model corresponds to the observable vectors that are referred to as *disjunctive* group models. Such observations are uncorrelated between groups but can be correlated within the same group. The limitation of this simplified stochastic model is that only one unknown variance component is associated with each group of observations but has the advantage of having less computational load.

- Monte-Carlo technique
  Kusche (2003a) employed the Monte-Carlo technique to estimate variance components in satellite geodesy without the need for repeated inversion of the matrices involved.

3.5 Applications to geodetic data

Examples of recent geodetic applications in which different variance-component estimation methods or noise assessment techniques were used include

- assessment of classical triangulation and trilateration network for monitoring tectonic activity observed with different electromagnetic distance measuring instruments and theodolites, and estimating error components and weighting of GPS observations (Chen et al., 1990)
- assessment of noise characteristics, namely, white noise, random walk noise and flicker noise, in daily coordinate time series of permanent GPS stations (see e.g. Zhang et al., 1997; Mao et al., 1999; Williams et al., 2004; Amiri-Simkooei et al., 2006)
- assessment of stochastic model for satellite laser ranging and very long baseline interferometry data (see Sahin et al., 1992; Lucas and Dillinger, 1998, respectively)
- studying the elevation dependence of the GPS observables precision (see e.g. Euler and Goad, 1991; Gerdan, 1995; Gianniou, 1996; Jin and de Jong, 1996)
- estimation of covariance matrix for precise GPS observables to improve precision, reliability, efficiency of positioning results, and/or increasing success rate for GPS ambiguity resolution (see e.g. Wang et al., 1998; Jonkman, 1998; Teunissen, 1998)
- determination of the GPS stochastic model (e.g. stochastic characteristics of multipath error and receiver noise) using for instance the GPS signal to noise ratio (SNR)(see e.g. Barnes et al., 1998; Barnes and Cross, 1998; Barnes, 2002)
- estimation of a stochastic model for processing of GPS code and phase data that incorporates time correlation and/or cross correlation of GPS observables (Satirapod et al., 2002; Tiberius and Kenselaar, 2000a,b, 2003; Tiberius et al., 1999; Bona, 2000; Amiri-Simkooei and Tiberius, 2007)

- determination of the stochastic model for long-baseline kinematic GPS positioning derived from observation time series (Kim and Langley, 2001a,b)
- employing a statistical test procedure based on uncorrelated least-squares residuals to show that the assumption of constant variances for GPS observables is not appropriate, and giving a procedure to estimate an individual variance function for a pair of satellites and to check the appropriateness of the estimated variances (Bischoff et al., 2005, 2006)
- development of a stochastic model for GPS phase observations from a single receiver under the assumption of dominance of noise caused by variations in atmospheric delays and estimation of its time-dependent scale factor by VCE as a measure for atmospheric turbulence (strength)(Kleijer et al., 2004)
- estimation of variance components for weighting data of different types in gravity field models (Koch and Kusche, 2002; Kusche, 2003b)
- assessment of the stochastic model of CHAMP data using the energy balance in an iterative VCE method (Kusche and van Loon, 2005; van Loon and Kusche, 2005)
- application of bias-corrected methods for estimating variance components in linear ill-posed models and the effect of regularization on VCE (Xu et al., 2006)
- estimation of covariance matrix for the adjustment of combined height data types, i.e. ellipsoidal, orthometric and geoid heights (Fotopoulos, 2003, 2005).

Least-Squares Variance Component Estimation 4

4.1 Introduction and background

This thesis formulates a unified framework for both the estimation and validation problem of the stochastic model. In this chapter we concentrate on the problem of estimating components of the stochastic model. The method which was originally proposed by Teunissen (1988) is based on the least-squares principle. The idea is to investigate whether it is possible to use the method of least-squares estimation, introduced in chapter 2, also for the problem of variance component estimation. This turns out to be the case. As a consequence, we will have the possibility of applying one estimation principle, namely our well-known and well understood method of least-squares, to both the problem of estimating the functional model and the stochastic model. In this chapter the probability density function (PDF) of a random m-vector $\underline{y}$ is written as $f_y(y|x,\sigma)$, in which x is an n-vector of unknown parameters in the functional model, and σ is a p-vector of unknown (co)variance components in the stochastic model. In this case the first two moments of the observables $\underline{y}$, namely E$\{\underline{y}\}$ and D$\{\underline{y}\}$, are partially unknown. We will therefore employ the least-squares principle to estimate both x and σ.

We derive the deterministic (weighted) least-squares (co)variance component estimation (LS-VCE) formula for which an *arbitrary* symmetric and positive-definite *weight* matrix can be used. We will give the formulation both in terms of the model of condition equations and the model of observation equations. The LS-VCE method gives *unbiased* estimators. One advantage of the method over other methods of (co)variance component estimation is that the (weighted) least-squares estimators are *independent* from the distribution of observables, namely from $f_y(y|x,\sigma)$.

Based on the normal distribution of original observations, we derive the covariance matrix of the observables in the stochastic model. We can therefore obtain the minimum variance estimators of previous chapter by taking the weight matrix as the inverse of the covariance matrix. This corresponds to the BLUE estimator of unknown parameter x in the functional model (see chapter 2). These 'probabilistic' least-squares estimators are both *unbiased* and of *minimum variance*. The property of minimum variance is however restricted to normally distributed data. As a generalization, Teunissen and Amiri-Simkooei (2006) derived such estimators for a class of elliptical distributions.

NOTE 4.1 In the sequel, we will sometimes use the term 'functionally known quantity' in a few examples. We mean that the functional model is of the form $E\{\underline{y}\} = \mu_y$ (known) where $A = \{\}$ (empty). This leads to the model of condition equations as $B^T(\underline{y} - \mu_y) = \underline{t},\ B = I,\ b = m$ and the orthogonal projector as $P_A^\perp = I$. A zero-mean stationary noise process is considered as an example for which $\mu_y = 0$. The goal is to arrive at a clear and simple expression in our formulation in examples. ⊡

The remainder of this section deals with some basic mathematical concepts that we will frequently use in the rest of the chapter. Before going further into the details, let us consider the following five definitions:

Definition 4.1 (Kronecker product) *Consider a matrix $S = (s_{ij})$ of order $u \times v$ and a matrix $T = (t_{ij})$ of order $q \times r$. The Kronecker product of the two matrices, denoted by $S \otimes T$ is defined as the partitioned matrix*

$$S \otimes T = \begin{bmatrix} s_{11}T & s_{12}T & \cdots & s_{1v}T \\ s_{21}T & s_{22}T & \cdots & s_{2v}T \\ \vdots & \vdots & \ddots & \vdots \\ s_{u1}T & s_{u2}T & \cdots & s_{uv}T \end{bmatrix} . \tag{4.1}$$

The matrix $S \otimes T$ is a matrix of order $uq \times vr$. It has uv number of blocks; the ij^{th} block is the matrix $s_{ij}T$ of the order $q \times r$. □

Definition 4.2 (vec-operator) *Let $S = [s_1\ s_2 \dots s_v]$ be an arbitrary matrix of size $u \times v$, with s_i its i^{th} column (vector) of size u. Then the* vec *operator on this matrix is defined as*

$$\text{vec}(S) = \text{vec}([s_1\ s_2 \dots s_v]) = \left[s_1^T\ s_2^T \dots s_v^T\right]^T . \tag{4.2}$$

Therefore, the vec *operator creates a column vector of size $uv \times 1$ from the matrix S by stacking the column vectors of S below one another.* □

The transpose of S, namely S^T, contains the same uv elements as S, but in a different pattern. Therefore, there exists a unique $uv \times uv$ permutation matrix which transforms $\text{vec}(S)$ into $\text{vec}(S^T)$. This is called the *commutation matrix* and is denoted by K_{uv}.

Definition 4.3 (Commutation matrix) *The commutation matrix K_{uv} is the $uv \times uv$ matrix with the property that*

$$K_{uv}\text{vec}(S) = \text{vec}(S^T) , \tag{4.3}$$

for every $u \times v$ matrix S. When $u = v$, we use the notation K instead of K_{uv}. In this case, the commutation matrix has the following form:

$$K = \sum_{i=1}^{u}\sum_{k=1}^{u} c_i c_k^T \otimes c_k c_i^T , \tag{4.4}$$

where $c_i = (0 \cdots 0\ 1\ 0 \cdots 0)^T$ is the canonical unit vector which contains zeros except a one *at i^{th} position.* □

Definition 4.4 (vh-operator) *Let $S = (s_{ij})$ be an arbitrary square matrix of size u. The* vh *operator of S is obtained in a similar way as $\text{vec}(S)$ is defined, but it starts each column at its diagonal element. That is, $\text{vh}(S)$ contains the $\frac{1}{2}u(u+1)$ elements $s_{ij},\ i \geq j$, only the elements on and below the diagonal of S.* □

If S is symmetric, then $\text{vh}(S)$ contains precisely the $\frac{1}{2}u(u+1)$ distinct elements of S and the elements of $\text{vec}(S)$ are those of $\text{vh}(S)$ with the off-diagonal elements occurring twice. Therefore, there exists a unique $u^2 \times \frac{1}{2}u(u+1)$ matrix which transforms, for symmetric S, $\text{vh}(S)$ into $\text{vec}(S)$. This transformation matrix is called the *duplication matrix*.

Definition 4.5 (Duplication matrix) *Let S be a symmetric matrix of size u. The duplication matrix D is the $u^2 \times \frac{1}{2}u(u+1)$ matrix with the property that*

$$D\mathrm{vh}(S) = \mathrm{vec}(S). \tag{4.5}$$

The duplication matrix D is of full column rank, namely $\mathrm{rank}(D) = \frac{1}{2}u(u+1)$. □

The reader is referred to appendix A, for properties of the vec and vh operators, the Kronecker product, and the commutation and duplication matrices. For a complete reference on the properties and the theorems among the vec and vh operators, the Kronecker product, and the commutation and duplication matrices see Magnus (1988). The most useful properties of the Kronecker product have also been collected by Langville and Stewart (2004). The preceding operators and transformation matrices will be frequently used in the sequel.

4.2 Weighted least-squares estimators

Consider the following linear model of observation equations (Gauss-Markov model):

$$\mathsf{E}\{\underline{y}\} = Ax\,; \quad \mathsf{D}\{\underline{y}\} = Q_y = \mathsf{E}\{(\underline{y} - Ax)(\underline{y} - Ax)^T\} = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k\,, \tag{4.6}$$

where, again, A is assumed to have full rank, Q_0 is the known part of the covariance matrix, and $Q_k,\ k = 1, \ldots, p$ are the cofactor matrices such that the sum $Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$ is non-negative definite. The matrices $Q_k,\ k = 1, \ldots, p$ should be linearly *independent* and *symmetric*. Symmetry of the cofactor matrices is on account of the symmetry of the covariance matrix of the observables. Linear dependency of the cofactor matrices will however cause the normal matrix of the stochastic model to be singular. In this case, at least one of the cofactor matrices can be written as a linear combination of the others, i.e., $Q_l = \sum_{k=1, k\neq l}^{p} \alpha_k Q_k$. Note that the linear independency of cofactor matrices is only a *necessary* condition and *not* sufficient (see Note 4.2).

Stochastic model In the preceding model, we have two sets of unknowns: the parameter vector x and the (co)variance components (or even sometimes correlation coefficients) $\sigma_k,\ k = 1, 2, ..., p$. The idea of the least-squares approach to (co)variance component estimation is now to interpret the matrix equation in the second part of equation (4.6), which represents the covariance matrix of $\underline{y}$, as a set of m^2-number of observation equations. Therefore, just like we interpret the functional model $\mathsf{E}\{\underline{y}\} = Ax$ as a set of m-number of observation equations with the observation vector $\underline{y}$, we are going to interpret the stochastic model $\mathsf{E}\{(\underline{y} - Ax)(\underline{y} - Ax)^T\} = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$ as a set of m^2-number of observation equations with the observation matrix $(\underline{y} - Ax)(\underline{y} - Ax)^T$. There is however one complication: The matrix $(\underline{y} - Ax)(\underline{y} - Ax)^T$ is not completely *observable* since the vector x is unknown apriori. This problem can be circumvented by taking the term Axx^TA^T to the right-hand side. Therefore, the second part of equation (4.6) can also be rewritten as

$$\mathsf{E}\{\underline{y}\,\underline{y}^T - Q_0\} = Axx^TA^T + \sum_{k=1}^{p} \sigma_k Q_k\,, \tag{4.7}$$

which looks like the model of observation equations, namely $\mathsf{E}\{\underline{y}\} = Ax$; on the left hand side, the observation matrix $\underline{y}\,\underline{y}^T - Q_0$ and on the right hand side, the unknown parameters x and σ. Let us, for the moment, assume that the covariance matrix Q_y is known and therefore remains on the right-hand side. Assume that the observable vector $\underline{y}$ is transferred into a new observable vector $\underline{y}'$ using the following $m \times m$ regular and invertible transformation matrix T:

$$\underline{y}' = T\underline{y} = \begin{bmatrix} (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \\ B^T \end{bmatrix} \underline{y}\,, \tag{4.8}$$

where the $m \times (m-n)$ matrix B satisfies $B^T A = 0$, with $\mathrm{rank}(B) = m - n = b$ the redundancy of the functional model (see section 2.1.6). Since T is square and invertible, reparameterization of the observables $\underline{y}$ does not lose or add any information. Therefore, an equivalent expression for equation (4.7) is given as

$$\mathsf{E}\{T\underline{y}\,\underline{y}^T T^T\} = TAxx^T A^T T^T + TQ_y T^T\,, \tag{4.9}$$

If we now substitute T from equation (4.8) into the preceding equation, it will follow that

$$\mathsf{E}\left\{\begin{bmatrix} \underline{\hat{x}}\,\underline{\hat{x}}^T & \underline{\hat{x}}\,\underline{t}^T \\ \underline{t}\,\underline{\hat{x}}^T & \underline{t}\,\underline{t}^T \end{bmatrix}\right\} = \begin{bmatrix} x\,x^T & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} (A^T Q_y^{-1} A)^{-1} & 0 \\ 0 & B^T Q_y B \end{bmatrix}, \tag{4.10}$$

with $\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y}$ and $\underline{t} = B^T \underline{y}$ the misclosure vector. The above equation consists of three separate parts, namely

$$\mathsf{E}\{\underline{\hat{x}}\,\underline{\hat{x}}^T\} = x\,x^T + (A^T Q_y^{-1} A)^{-1}; \quad \mathsf{E}\{\underline{\hat{x}}\,\underline{t}^T\} = 0; \quad \mathsf{E}\{\underline{t}\,\underline{t}^T\} = B^T Q_y B\,. \tag{4.11}$$

The first part is trivial as we have the same number of observables as unknowns. Therefore, the observables $\underline{\hat{x}}\,\underline{\hat{x}}^T$ are just enough to estimate the unknowns $x\,x^T$ (we recognize $\underline{\hat{x}}$ as the BLUE of x). In other words, there is no more redundancy left to estimate any (co)variance components. The second part $\mathsf{E}\{\underline{\hat{x}}\,\underline{t}^T\} = 0$ has no interaction with (co)variance components and shows that $\underline{\hat{x}}$ and $\underline{t}$ are uncorrelated. They are independent if $\underline{y}$ is normally distributed. We may therefore focus on the third part for the problem of variance component estimation. Substituting $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$, the third part reads $\mathsf{E}\{\underline{t}\,\underline{t}^T\} - B^T Q_0 B = \sum_{k=1}^{p} \sigma_k B^T Q_k B$. The same transformation can also be applied to the first part of equation (4.6), namely $\mathsf{E}\{T\underline{y}\} = TAx$. Therefore, equation (4.6) can be reformulated in terms of the model of condition equations as

$$\mathsf{E}\{\underline{t}\} = 0; \quad \mathsf{E}\{\underline{t}\,\underline{t}^T\} - B^T Q_0 B = \sum_{k=1}^{p} \sigma_k B^T Q_k B\,. \tag{4.12}$$

Note that the first part of equation (4.12), i.e., the functional part, consists of all redundant observations as there exists no unknown in this model. The adjustment of this part is trivial because $\hat{\underline{t}} = 0$. We may therefore concentrate on the second part, i.e., the stochastic model. Note also that the condition $\mathsf{E}\{\underline{t}\} = 0$, which implies that there is no misspecification in the functional model, has been used in the second part by default because $Q_t = \mathsf{E}\{\underline{t}\,\underline{t}^T\} - \mathsf{E}\{\underline{t}\}\mathsf{E}\{\underline{t}\}^T$. The matrix equation in the second part of equation (4.12) can now be recast into a set of b^2-number of observation equations by stacking the b-number of $b \times 1$ column vectors of $\mathsf{E}\{\underline{t}\,\underline{t}^T\}$ into a $b^2 \times 1$ observation vector.

Since the matrix of observables $\underline{t}\,\underline{t}^T$ is symmetric, its upper triangular elements do not provide new information. There are only $\frac{b(b+1)}{2}$ distinct (functionally independent) elements. We can therefore apply the vh operator to the second part of equation (4.12). This results in the following equation (note that both the vh and the E operators are linear):

$$\mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)\} = \sum_{k=1}^{p} \sigma_k \mathrm{vh}(B^T Q_k B)\,. \tag{4.13}$$

Equation (4.13) can be rewritten, in a matrix form, as the *linear model of observation equations* for (co)variance component model:

$$\mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)\} = A_{\mathrm{vh}}\sigma, \quad W_{\mathrm{vh}} \quad \text{or} \quad Q_{\mathrm{vh}}\,, \tag{4.14}$$

where A_{vh} is a $\frac{b(b+1)}{2} \times p$ (design) matrix and has the following form:

$$\boxed{A_{\mathrm{vh}} = \left[\mathrm{vh}(B^T Q_1 B)\ \mathrm{vh}(B^T Q_2 B)\ \cdots\ \mathrm{vh}(B^T Q_p B)\right],} \tag{4.15}$$

and σ is a p-vector as $\sigma = [\sigma_1\ \sigma_2\ \cdots\ \sigma_p]^T$. The $\frac{b(b+1)}{2} \times \frac{b(b+1)}{2}$ matrix Q_{vh} is the *covariance* matrix of the *observables* $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ and the $\frac{b(b+1)}{2} \times \frac{b(b+1)}{2}$ matrix W_{vh} is accordingly the *weight* matrix. This is therefore a standard form of the linear model of observation equations with a $\frac{b(b+1)}{2}$-vector of observables, a $\frac{b(b+1)}{2} \times p$ design matrix and a p-vector of unknown (co)variance components.

Redundancy In an analogous way to the functional model in which one deals with redundancy $b = m-n$, one can define the redundancy or here the *degrees of freedom* df in the stochastic model. From equation (4.15) it follows that:

$$df = \frac{b(b+1)}{2} - p\,, \tag{4.16}$$

when the design matrix A_{vh} of the stochastic model is assumed to be of full rank, and with p, as before, being the number of unknowns in stochastic model, namely the number of (co)variance components. One point which is obvious from this formula is that the relation between functional and stochastic redundancies is not a linear but a second order function. When for instance b changes from 100 to 101, the redundancy of the stochastic model changes from $5050 - p$ to $5151 - p$. That is, the redundancy of the stochastic model can in one way or another be deceptive. The worst case occurs when $p = \frac{b(b+1)}{2}$ which gives $df = 0$, meaning that all elements of the covariance matrix Q_t are assumed to be unknown (see example 4.1). The best situation, however, occurs when $p = 0$, $df = \frac{b(b+1)}{2}$, meaning that all elements of the covariance matrix or the stochastic model are known. In the special case that we deal with only one variance component, i.e. $Q_y = \sigma^2 Q$, the redundancy is given as $df = \frac{(b-1)(b+2)}{2}$, which is zero when $b = 1$. This implies that the least-squares estimation of the variance of unit-weight in case of a redundancy of one in the functional model is reduced to the simple unique estimation.

NOTE 4.2 As previously mentioned in chapter 3, the cofactor matrices $Q_k,\ k = 1, ..., p$ should be linearly independent. This is indeed the necessary condition and not sufficient. With our formulation in equation (4.15) it becomes clear that the variance component model has a unique solution *if and only if* (*necessary and sufficient condition*) the matrices

$B^TQ_kB,\ k=1,...,p$ are linearly independent. From the definition of the redundancy df of the stochastic model it goes without saying that the maximum number of estimable (co)variance components in the stochastic model is $\frac{b(b+1)}{2}$ which leads to $df=0$ (unique solution). That is, in general an $\frac{m(m+1)}{2}-\frac{b(b+1)}{2}$ number of distinct elements of Q_y remains inestimable as $m \geq b$. Xu et al. (2007) also prove that all variance and covariances are not estimable. They propose a method to show that at most only a $\frac{b(b+1)}{2}$ number of independent parameters are estimable. $\boxdot$

(Co)variance estimators (1) Having established these results, we can now apply the estimation methods of chapter 2. That is, if the weight matrix $W_{\rm vh}$ is known (or assumed), we can obtain the weighted (deterministic) least-squares estimators of the (co)variance components. The *weighted least-squares estimators* of the (co)variance components then read (cf. equation (2.6))

$$\hat{\underline{\sigma}} = (A_{\rm vh}^T W_{\rm vh} A_{\rm vh})^{-1} A_{\rm vh}^T W_{\rm vh} {\rm vh}(\underline{t}\,\underline{t}^T - B^TQ_0B)\,. \tag{4.17}$$

If we denote $N = A_{\rm vh}^T W_{\rm vh} A_{\rm vh}$ and $\underline{l} = A_{\rm vh}^T W_{\rm vh} {\rm vh}(\underline{t}\,\underline{t}^T - B^TQ_0B)$ it follows that $\hat{\underline{\sigma}} = N^{-1}\underline{l}$. This equation can be written out in full as

$$\boxed{\begin{bmatrix} \hat{\underline{\sigma}}_1 \\ \hat{\underline{\sigma}}_2 \\ \vdots \\ \hat{\underline{\sigma}}_p \end{bmatrix} = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1p} \\ n_{12} & n_{22} & \cdots & n_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ n_{1p} & n_{2p} & \cdots & n_{pp} \end{bmatrix}^{-1} \begin{bmatrix} \underline{l}_1 \\ \underline{l}_2 \\ \vdots \\ \underline{l}_p \end{bmatrix}}\,, \tag{4.18}$$

where

$$n_{kl} = {\rm vh}(B^TQ_kB)^T W_{\rm vh} {\rm vh}(B^TQ_lB)\,, \tag{4.19}$$

and

$$\underline{l}_k = {\rm vh}(B^TQ_kB)^T W_{\rm vh} {\rm vh}(\underline{t}\,\underline{t}^T - B^TQ_0B)\,, \tag{4.20}$$

with $k,l = 1,2,\cdots,p$. Any symmetric and positive-definite matrix $W_{\rm vh}$ can play the role of the weight matrix. An important feature of the preceding estimators is the *unbiasedness* property. One can obtain the BLUE estimators (or in fact BIQUE estimators as they results in a quadratic form of the original observables) by taking the weight matrix as the inverse of $Q_{\rm vh}$. Unfortunately, in practice it is not always possible to obtain $Q_{\rm vh}$ as it depends on the *fourth* central moments of the observables (see equation (4.53)) which may not be obtainable all the times, but only for special cases. In the following, we will however show two examples as limiting cases that no supposition on the distribution is required.

Example 4.1 Consider the model of condition equations $\mathsf{E}\{B^T\underline{y}\} = \mathsf{E}\{\underline{t}\} = 0$ with a completely unknown covariance matrix Q_t (all elements are unknown). It is possible to obtain the solution of these unknown (co)variance components. For this purpose we can write the unknown matrix Q_t as

$$Q_t = \sum_{k=1}^{\frac{b(b+1)}{2}} \sigma_k C_k, \quad \text{with} \quad C_k = \begin{cases} c_i c_j^T, & \text{if} \quad \sigma_k = \sigma_{ii},\ i=j\,, \\ \\ c_i c_j^T + c_j c_i^T, & \text{if} \quad \sigma_k = \sigma_{ij},\ i \neq j\,, \end{cases} \tag{4.21}$$

where σ_k is respectively $\sigma_{11}, \sigma_{12}, ..., \sigma_{1b}, \sigma_{22}, \sigma_{23}, ..., \sigma_{bb}$ and $c_i = (0 \cdots 1 \cdots 0)^T$ is the canonical unit vector of size b containing zeros except a *one* at i^{th} position. With the preceding formulation, the matrices $B^T Q_k B$ in equation (4.15) are of the form $C_k = c_i c_i^T$ or $C_k = c_i c_j^T + c_j c_i^T$. Applying the vh operator to these matrices gives canonical unit vectors in $\mathbb{R}^{\frac{b(b+1)}{2}}$. That is, A_{vh} in equation (4.14) is given as $A_{\text{vh}} = I_{\frac{b(b+1)}{2}}$. This means that the (co)variance components σ can be uniquely estimated as

$$\underline{\hat{\sigma}} = \text{vh}(\underline{\hat{Q}}_t) = A_{\text{vh}}^{-1}\text{vh}(\underline{t}\,\underline{t}^T) = \text{vh}(\underline{t}\,\underline{t}^T)\,, \tag{4.22}$$

which is independent of the weight matrix W_{vh} or covariance matrix Q_{vh}, and therefore of distribution of $\underline{t}$. This expression can be rewritten as $\hat{\underline{Q}}_t = \underline{t}\,\underline{t}^T$. One can also obtain the covariance matrix of these estimators (see example 4.4). ■

Example 4.2 If we measure a zero-mean quantity (see NOTE 4.1), then $\mathsf{E}\{B^T\underline{y}\} = \mathsf{E}\{\underline{y}\} = \mathsf{E}\{\underline{t}\} = 0,\ b = m$. Let the observables be independent and identically distributed (in general unknown distribution). In this case $Q_t = \sigma^2 I$ with σ^2 an unknown variance component to be estimated. It is not difficult to show that only the following central moments (up to degree four) are non-zero

$$\mathsf{E}\{\underline{t}_i^2\} = \mu_2; \quad \mathsf{E}\{\underline{t}_i^4\} = \mu_4; \quad \mathsf{E}\{\underline{t}_i^2\underline{t}_j^2\} = \begin{cases} \mu_4 & \text{if} \quad i = j\,, \\ \mu_2^2 & \text{if} \quad i \neq j\,. \end{cases} \tag{4.23}$$

With these non-zero elements it becomes clear that Q_{vh} is diagonal with the entries

$$\text{diag}(Q_{\text{vh}}) = (\mu_4 - \mu_2^2,\ \overbrace{\mu_2^2,\ ...,\ \mu_2^2}^{m-1 \text{ times}},\ \mu_4 - \mu_2^2,\ \overbrace{\mu_2^2,\ ...,\ \mu_2^2}^{m-2 \text{ times}},\ ...,\ ...,\ \mu_4 - \mu_2^2). \tag{4.24}$$

After a few simple mathematical operations one obtains the inverse of the matrix Q_{vh} as

$$Q_{\text{vh}}^{-1} = \frac{1}{\mu_4 - \mu_2^2} D^T(I \otimes I)D + \frac{\mu_4 - 3\mu_2^2}{\mu_2^2(\mu_4 - \mu_2^2)}(D^T D - I_{\frac{m(m+1)}{2}}). \tag{4.25}$$

In equation (4.17) we have $A_{\text{vh}} = \text{vh}(I)$. If observables are normally distributed, then $\mu_4 = 3\sigma^4$ and therefore the second term of Q_{vh}^{-1} vanishes. But even though the distribution is not specified, the second term when pre-multiplied with $\text{vh}(I)^T$ becomes zero. Therefore, one obtains

$$N = n_{11} = \frac{m}{\mu_4 - \mu_2^2}; \quad \underline{l} = \underline{l}_1 = \frac{\underline{t}^T\underline{t}}{\mu_4 - \mu_2^2}\,, \tag{4.26}$$

which gives the variance estimator $\underline{\hat{\sigma}}^2$ independent of μ_4 as

$$\underline{\hat{\sigma}}^2 = \frac{\underline{t}^T\underline{t}}{m}. \tag{4.27}$$

This estimator is therefore independent of the unspecified distribution of $\underline{t}$. Such an estimator is not only *unbiased* but also of *minimum variance*. Note however that the variance of this estimator is given as $N^{-1} = n_{11}^{-1}$ which is obtainable when we specify the distribution. ■

Weight matrix As mentioned, any symmetric and positive-definite matrix $W_{\rm vh}$ can play the role of weight matrix in equation (4.17). From a numerical point of view such arbitrary weight matrices may not be advisable as they are as large as by the second power of the sample size. Now we will restrict ourselves to those weight matrices which computationally are more efficient. To find different classes of admissible weight matrices $W_{\rm vh}$ we first come up with different classes of admissible weight matrices for $W_{\rm vec}$. Pre-multiplying equation (4.14) with the duplication matrix D yields that

$$\mathsf{E}\{\mathrm{vec}(\underline{t}\,\underline{t}^T - B^T Q_0 B)\} = D A_{\rm vh}\sigma,\ W_{\rm vec}\,. \tag{4.28}$$

We now recall the property of *invariance* of the estimators under a regular transformation matrix T. Let us assume that the original observation vector $\underline{t}$ is transformed into $\underline{v}$ using the non-singular $b \times b$ matrix M, i.e. $\underline{v} = M\,\underline{t}$ or $\underline{t} = M^{-1}\underline{v}$. In this case, one obtains

$$\mathrm{vec}(\underline{t}\,\underline{t}^T) = \mathrm{vec}(M^{-1}\underline{v}\,\underline{v}^T M^{-T})\,. \tag{4.29}$$

Using equation (A.13), in appendix A, the preceding equation reads

$$\mathrm{vec}(\underline{t}\,\underline{t}^T) = (M^{-1}\otimes M^{-1})\mathrm{vec}(\underline{v}\,\underline{v}^T) = T\mathrm{vec}(\underline{v}\,\underline{v}^T)\,. \tag{4.30}$$

To make our estimators independent of the type of reparameterization of observables, we can use the identity $W^t_{\rm vec} = T^{-T} W^v_{\rm vec} T^{-1}$, which yields

$$W^t_{\rm vec} = (M^T\otimes M^T) W^v_{\rm vec}(M\otimes M)\,. \tag{4.31}$$

From the definition of vec and vh operators, it follows that for any symmetric matrix one gets $\mathrm{vec} = DD^+\mathrm{vec}$ where D^+ is the pseudo inverse (Moore-Penrose inverse) of the matrix D and is given as $D^+ = (D^TD)^{-1}D^T$ (see appendix A). Since the matrix DD^+ has rank $\frac{b(b+1)}{2}$, this implies that we will only deal with $\frac{b(b+1)}{2}$ number of distinct (functionally independent) elements. Therefore, in order to specify a class of weight matrices for the vh operator, we can argue that both the vec and vh operators should give identical estimators. Comparing the least-squares solution of equation (4.14) with that of equation (4.28) simply with equation (4.31) follows that

$$W_{\rm vh} = D^T W_{\rm vec} D = D^T(M^T\otimes M^T) W^v_{\rm vec}(M\otimes M)D\,. \tag{4.32}$$

In order for the weight matrix $W_{\rm vh}$ to be simple, we may take simple forms for $W^v_{\rm vec}$ in the preceding equation. Here we consider two choices as follows:

- The identity matrix $W^v_{\rm vec} = I_{b^2} = I\otimes I$ is positive-definite and therefore an element of this class. It follows then

  $$W_{\rm vh} = D^T(M^T\otimes M^T)I(M\otimes M)D = D^T(M^TM\otimes M^TM)D\,. \tag{4.33}$$

  Since M^TM is a positive-definite symmetric matrix, a class of admissible weight matrices for $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ is given as

  $$W_{\rm vh} = D^T(W_t\otimes W_t)D\,, \tag{4.34}$$

  where W_t is an arbitrary positive-definite symmetric matrix. Using the properties of the Kronecker product one can show that $W_{\rm vh}$ is also positive-definite and therefore can play the role of the weight matrix.

- The second weight matrix includes, in addition to identity matrix, one more simple term as $W^v_{\mathrm{vec}} = I \otimes I + \alpha \mathrm{vec}(I)\mathrm{vec}(I)^T$. For some α, namely $\alpha > \frac{-1}{b}$, this matrix is positive-definite. When substituted in equation (4.33), the second term simplifies to (see equation (A.13)): $(M^T \otimes M^T)(\mathrm{vec}(I)\mathrm{vec}(I)^T)(M \otimes M) = \mathrm{vec}(W_t)\mathrm{vec}(W_t)^T$. This results in the following weight matrix:

$$W_{\mathrm{vh}} = D^T(W_t \otimes W_t + \alpha\, \mathrm{vec}(W_t)\mathrm{vec}(W_t)^T)D\,. \tag{4.35}$$

(Co)variance estimators (2) In this thesis we consider only the simplest weight matrix $W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D$ (first choice). But we will as well give a few comments on the second case. Equation (4.14) can therefore be rewritten as

$$\boxed{\mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)\} = A_{\mathrm{vh}}\sigma, \quad W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D,} \tag{4.36}$$

with W_t any positive-definite symmetric matrix and D the duplication matrix. Substitution of vh with $D^+\mathrm{vec}$ and $W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D$ in equation (4.19) follows

$$n_{kl} = \mathrm{vec}(B^T Q_k B)^T D^{+T} D^T (W_t \otimes W_t) D D^+ \mathrm{vec}(B^T Q_l B)\,. \tag{4.37}$$

Using identities (A.21) and (A.22) with (A.15) and (A.18), in appendix A, the terms $D^{+T}D^T$ and DD^+ in the preceding equation can be left out, i.e.

$$n_{kl} = \mathrm{vec}(B^T Q_k B)^T (W_t \otimes W_t)\mathrm{vec}(B^T Q_l B)\,. \tag{4.38}$$

This also confirms that the operators vh and vec, on symmetric matrices, give identical results. Using equation (A.14), the preceding equation reads

$$\boxed{n_{kl} = \mathrm{tr}(B^T Q_k B W_t B^T Q_l B W_t).} \tag{4.39}$$

In a similar manner, $\underline{l}_k$ of equation (4.20) simplifies

$$\underline{l}_k = \mathrm{tr}(B^T Q_k B W_t [\underline{t}\,\underline{t}^T - B^T Q_0 B] W_t) \tag{4.40}$$

or finally

$$\boxed{\underline{l}_k = \underline{t}^T W_t B^T Q_k B W_t \underline{t} - \mathrm{tr}(B^T Q_k B W_t B^T Q_0 B W_t).} \tag{4.41}$$

The weighted least-squares (co)variance component estimation was formulated by rewriting the (co)variance component model into a linear model of observation equations. Note that the above formulation to (co)variance component estimation is based on the deterministic least-squares technique for which an arbitrary weight matrix W_{vh} (e.g. in form of equation (4.34)) can be used. The (co)variance components are therefore estimated as a linear unbiased estimator (LUE) of the quadratic misclosures $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ which can then be rewritten as an invariant quadratic unbiased estimator (IQUE) of the misclosures $\underline{t}$.

The following example shows that the empirical autocovariance function is a weighted LS-VCE, if the weight matrix is chosen from the given weight matrix class in equation (4.34).

Example 4.3 (unbiased empirical autocorrelation function) Let y_i, $i = 1, \ldots$, be a zero-mean stationary random process, with unknown covariance function $\sigma_{ij} = \sigma_\tau$ ($\tau = |i-j|$). If m random variables are available, we have, with $y = (\underline{y}_1, \ldots, \underline{y}_m)^T$, that

$$\mathsf{E}\{\underline{y}\} = 0 \;\text{ and }\; \mathsf{D}\{\underline{y}\} = Q_y = \sum_{\tau=0}^{m-1} \sigma_\tau Q_\tau \,, \tag{4.42}$$

with the $m \times m$ cofactor matrices

$$Q_\tau = a_\tau \sum_{i=1}^{m-\tau} (c_i c_{i+\tau}^T + c_{i+\tau} c_i^T) \,, \tag{4.43}$$

for $\tau = 0, 1, \ldots, m-1$, with $a_0 = \frac{1}{2}$ and $a_\tau = 1$ for $\tau \neq 0$. Note that σ_0 is the variance of y_i and that σ_τ is the covariance between y_i and $y_{i+\tau}$.

We can now apply the above formulation to estimate $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_{m-1})^T$. In our case we have $\underline{t} = B^T \underline{y}$, with $B = I$. We will use $W_t = W_y = I$ as weight matrix. Substitution of $B = I$ and $W_t = W_y = I$ into equation (4.39), gives, with equation (4.43) (see below)

$$n_{\tau,\kappa} = \mathrm{tr}(Q_\tau Q_\kappa) = \begin{cases} 2a_\tau(m-\tau) & \text{if} \quad \tau = \kappa \,; \\ 0 & \text{if} \quad \tau \neq \kappa \,. \end{cases} \tag{4.44}$$

Therefore, the normal matrix is a diagonal matrix. In a similar manner, and with $Q_0 = 0$, we find for the entries of the right-hand side vector, $\underline{l}_\tau = 2a_\tau \sum_{i=1}^{m-\tau} \underline{y}_i \underline{y}_{i+\tau}$ (see below). The LS-VCE solution for the unknown (co)variance components follows therefore as

$$\underline{\hat{\sigma}}_\tau = \frac{\sum_{i=1}^{m-\tau} \underline{y}_i \underline{y}_{i+\tau}}{m-\tau} \;,\; \tau = 0, 1, \ldots, m-1. \tag{4.45}$$

This solution is known as the unbiased sample autocovariance function of a zero-mean stationary time-series y_i, see e.g. Priestley (1981). We thus have shown that the sample autocovariance function is a LS-VCE, if the weight matrix is chosen as the identity matrix. If one has a functionally known quantity with mean μ, one will use the residuals $\underline{\hat{e}}_i = \underline{y}_i - \mu$ in the preceding formula. One can also compute the correlation coefficients that together represent the empirical autocorrelation function (ACF)

$$\underline{\hat{\rho}}_\tau = \frac{\underline{\hat{\sigma}}_\tau}{\underline{\hat{\sigma}}_0}, \;\tau = 1, ..., m-1. \tag{4.46}$$

where $\underline{\hat{\sigma}}_0 = \underline{\hat{\sigma}}^2$ is the variance of the noise process, and $\underline{\hat{\sigma}}_\tau$ the covariances. ■

Proof. ($n_{\tau,\kappa}$ and $\underline{l}_\tau$) Substitution of $Q_\tau = a_\tau \sum_{i=1}^{m-\tau} (c_i c_{i+\tau}^T + c_{i+\tau} c_i^T)$ and $Q_\kappa = a_\kappa \sum_{j=1}^{m-\kappa} (c_j c_{j+\kappa}^T + c_{j+\kappa} c_j^T)$ into $n_{\tau,\kappa} = \mathrm{tr}(Q_\tau Q_\kappa)$ results in the following equation

$$n_{\tau\kappa} = 2a_\tau a_\kappa \sum_{i=1}^{m-\tau} \sum_{j=1}^{m-\kappa} \mathrm{tr}(c_i c_{i+\tau}^T c_j c_{j+\kappa}^T + c_i c_{i+\tau}^T c_{j+\kappa} c_j^T), \tag{4.47}$$

or

$$n_{\tau\kappa} = 2a_\tau a_\kappa \sum_{i=1}^{m-\tau} \sum_{j=1}^{m-\kappa} \delta_{i+\tau,j} \delta_{i,j+\kappa} + \delta_{i+\tau,j+\kappa} \delta_{i,j}, \tag{4.48}$$

where $\delta_{i,j}$ is the Kronecker delta. The first term of the above equation is identical to zero except for the case that $\tau = \kappa = 0$. Also, the second term is zero when $\tau \neq \kappa$. It then follows that $n_{\tau,\tau} = 2a_\tau(m-\tau)$ and $n_{\tau,\kappa} = 0$ if $\tau \neq \kappa$. For the right-hand side vector, substitution of $Q_\tau = a_\tau \sum_{i=1}^{m-\tau}(c_i c_{i+\tau}^T + c_{i+\tau}c_i^T)$ in $\underline{l}_\tau = \underline{y}^T Q_\tau \underline{y}$ gives

$$\underline{l}_\tau = a_\tau \sum_{i=0}^{m-\tau} \underline{y}^T c_i c_{i+\tau}^T \underline{y} + \underline{y}^T c_{i+\tau} c_i^T \underline{y}, \tag{4.49}$$

or, finally

$$\underline{l}_\tau = 2a_\tau \sum_{i=0}^{m-\tau} \underline{y}_i \underline{y}_{i+\tau}. \quad \blacktriangledown \tag{4.50}$$

Before we enumerate the attractive features of the least-squares approach to (co)variance component estimation, let us first derive the covariance matrix of the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$.

4.3 Covariance matrix of observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$

In order to evaluate the covariance matrix of (co)variance components, i.e. $Q_{\hat{\sigma}}$, we need to know the $\frac{b(b+1)}{2} \times \frac{b(b+1)}{2}$ covariance matrix of $\mathrm{vh}(\underline{t}\,\underline{t}^T)$, namely Q_{vh}. In addition, one can in particular take the weight matrix W_{vh} to be the inverse of Q_{vh} to obtain the BLUE estimators (or better say BIQUE estimators, as our 'observations' are in quadratic form of the original ones). Let us first derive the covariance matrix of $\mathrm{vec}(\underline{t}\,\underline{t}^T)$ which is based on the following theorem:

Theorem 4.1 (Central moments) *Let the stochastic vector $\underline{t}$ be normally distributed with mean zero and covariance matrix Q_t, i.e. $\underline{t} \sim \mathsf{N}(0, Q_t)$, then the odd-order central moments of $\underline{t}$ are zero, and the second and fourth central moments are of the form*

$$\begin{array}{lcl} \mathsf{E}\{\underline{t}_i\underline{t}_j\} & = & q_{ij}\,; \\ \mathsf{E}\{\underline{t}_i\underline{t}_j\underline{t}_k\underline{t}_l\} & = & q_{ij}q_{kl} + q_{ik}q_{jl} + q_{jk}q_{il}\,; \\ i,j,k,l & = & 1,2,\cdots,b\,, \end{array} \tag{4.51}$$

where q_{ij} represents Q_t in index notation, i.e., $q_{ij} = Q_{t_i,t_j}$. $\boxtimes$

Proof. For a proof of the preceding theorem we refer to appendix C. $\blacktriangledown$

NOTE 4.3 The multivariate normal distribution is classified as an elliptical distribution. Another example of elliptical distributions is the multivariate Student t distribution. For elliptically contoured distributed data one needs to define a *kurtosis parameter* κ in order to obtain the *moments* of different orders, see Berkane and Bentler (1986); Maruyama and Seo (2003); Teunissen and Amiri-Simkooei (2006). The kurtosis parameter for normally distributed data is zero. $\boxdot$

Covariance matrix Q_{vec} The elements of the covariance matrix Q_{vec} are by definition given as

$$Q_{\mathrm{vec}}^{ijkl} = \mathsf{E}\{(\underline{t}_i\underline{t}_j - \mathsf{E}\{\underline{t}_i\underline{t}_j\})(\underline{t}_k\underline{t}_l - \mathsf{E}\{\underline{t}_k\underline{t}_l\})\}, \quad i,j,k,l = 1,2,\cdots,b\,. \tag{4.52}$$

This shows the covariance between elements of observable vector $\mathrm{vec}(\underline{t}\,\underline{t}^T)$, i.e. between $\underline{t}_i\underline{t}_j$ and $\underline{t}_k\underline{t}_l$. If we factor the right-hand side we obtain

$$Q_{\mathrm{vec}}^{i\,j\,k\,l} = \mathsf{E}\{\underline{t}_i\underline{t}_j\underline{t}_k\underline{t}_l\} - \mathsf{E}\{\underline{t}_i\underline{t}_j\}\mathsf{E}\{\underline{t}_k\underline{t}_l\},\ i,j,k,l = 1,2,\cdots,b\,. \tag{4.53}$$

This result shows that we need the second and the fourth multivariate central moments of the random vector $\underline{t}$. Using Theorem 4.1, equation (4.53) as a fourth-order tensor, can be written as

$$Q_{\mathrm{vec}}^{ijkl} = \mathsf{C}\{\underline{t}_i\underline{t}_j, \underline{t}_k\underline{t}_l\} = q_{ik}q_{jl} + q_{jk}q_{il},\ i,j,k,l = 1,2,...,b\,. \tag{4.54}$$

One can rewrite the preceding equation in terms of a second-order tensor (a matrix). To follow consider the following theorem:

Theorem 4.2 *Let $Q_t = q_{ij}$ be a symmetric and positive-definite matrix of size b. Then $Q_{\mathrm{vec}}^{ijkl} = q_{ik}q_{jl} + q_{jk}q_{il},\ i,j,k,l = 1,2,...,b$, as a fourth-order tensor, is equivalent to the following second-order tensor (matrix):*

$$Q_{\mathrm{vec}} = 2\,DD^+(Q_t \otimes Q_t)D^{+T}D^T\,, \tag{4.55}$$

which represents the covariance matrix of the observable vector $\mathrm{vec}(\underline{t}\,\underline{t}^T)$, where the $b^2 \times \frac{b(b+1)}{2}$ matrix D is the duplication matrix. ☒

Proof. From $Q_{\mathrm{vec}}^{ijkl} = q_{ik}q_{jl} + q_{jk}q_{il},\ i,j,k,l = 1,2,...,b$ it follows that the $b^2 \times b^2$ covariance matrix Q_{vec}, as a fourth order tensor, is composed of b^2-number of $b \times b$ submatrices, i.e.

$$Q_{\mathrm{vec}} = \begin{pmatrix} Q^{1.1.} & Q^{1.2.} & \cdots & Q^{1.b.} \\ Q^{2.1.} & Q^{2.2.} & \cdots & Q^{2.b.} \\ \vdots & \vdots & Q^{i.k.} & \vdots \\ Q^{b.1.} & Q^{b.2.} & \cdots & Q^{b.b.} \end{pmatrix}, \tag{4.56}$$

where the $b \times b$ submatrix $Q^{i.k.}$ is of the form $Q^{i.k.} = c_i^T Q_t c_k Q_t + Q_t c_k c_i^T Q_t$, with c_i the canonical unit vector. Equation (4.56) can be rewritten as follows:

$$Q_{\mathrm{vec}} = \sum_{i=1}^{b}\sum_{k=1}^{b} c_i c_k^T \otimes Q^{i.k.}. \tag{4.57}$$

Substituting $Q^{i.k.} = c_i^T Q_t c_k Q_t + Q_t c_k c_i^T Q_t$ in the preceding equation yields (note that the term $c_i^T Q_t c_k = q_{ik}$ is a scalar)

$$Q_{\mathrm{vec}} = \sum_{i=1}^{b}\sum_{k=1}^{b} c_i^T Q_t c_k c_i c_k^T \otimes Q_t + c_i c_k^T \otimes Q_t c_k c_i^T Q_t. \tag{4.58}$$

Using the properties of the Kronecker product, after a few simple matrix operations, the preceding equation yields

$$Q_{\mathrm{vec}} = [Q_t \otimes Q_t + (I \otimes Q_t)K(I \otimes Q_t)],\ \text{ with }\ K = \sum_{i=1}^{b}\sum_{k=1}^{b} c_i c_k \otimes c_k c_i^T, \tag{4.59}$$

the commutation matrix (see definition 4.3). Using equation (A.24) and then equations (A.5) and (A.27), the preceding equation reads

$$Q_{\mathrm{vec}} = [I + K](Q_t \otimes Q_t) = 2DD^+(Q_t \otimes Q_t), \tag{4.60}$$

which with equations (A.22) and (A.18) completes the proof. ▼

Covariance matrix Q_{vh} One can show that the covariance matrix Q_{vec} suffers from a rank deficiency of $\frac{b(b-1)}{2}$. This is what we would expect. The $\frac{b(b-1)}{2}$ number of elements of matrix $\underline{t}\,\underline{t}^T$ located in the upper triangle do not introduce new information. That is why they are one-to-one correlated with those located in the lower triangle. This causes the matrix Q_{vec} to be rank deficient. Applying the error propagation law to $\mathrm{vh}(\underline{t}\,\underline{t}^T) = D^+\mathrm{vec}(\underline{t}\,\underline{t}^T)$ results in the covariance matrix of the observable vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$

$$\boxed{Q_{\mathrm{vh}} = 2\,D^+(Q_t \otimes Q_t)D^{+T},} \tag{4.61}$$

which is a symmetric and positive-definite matrix provided that Q_t is positive-definite; $\mathrm{rank}(Q_{\mathrm{vh}}) = \frac{b(b+1)}{2}$. Using the properties of the duplication matrix, the inverse of Q_{vh} is obtained as (see identity (A.23))

$$Q_{\mathrm{vh}}^{-1} = \frac{1}{2}D^T(Q_t^{-1} \otimes Q_t^{-1})D\,. \tag{4.62}$$

NOTE 4.4 For normally distributed data, Q_{vh}^{-1} is an element of the class of admissible weight matrices defined in equation (4.34) with

$$W_t = \frac{1}{\sqrt{2}}Q_t^{-1}\,. \tag{4.63}$$

Teunissen and Amiri-Simkooei (2006) show that for elliptically contoured distributed data, the inverse of Q_{vh} is of the form

$$Q_{\mathrm{vh}}^{-1} = \frac{1}{2(\kappa+1)}D^T\left[Q_t^{-1} \otimes Q_t^{-1} - \frac{\kappa}{2(\kappa+1)+\kappa b}\mathrm{vec}(Q_t^{-1})\mathrm{vec}(Q_t^{-1})^T\right]D\,. \tag{4.64}$$

This means that for elliptical distributions the above matrix is an element of the weight matrix defined in equation (4.35) with

$$W_t = \frac{1}{\sqrt{2(\kappa+1)}}Q_t^{-1};\quad \text{and}\quad \alpha = \frac{-\kappa}{2(\kappa+1)+\kappa b}\,. \tag{4.65}$$

This is in fact an interesting result because we can now take the weight matrix as $W_{\mathrm{vh}} = Q_{\mathrm{vh}}^{-1}$ to obtain the minimum variance estimators of the (co)variance components (not only for the normal distribution but also for a larger class of elliptical distributions, see section 4.5). □

4.4 Properties of least-squares estimators

4.4.1 Optimal properties

The above (weighted) least-squares solutions to the (co)variance components are independent of the distribution of $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ which has been left unspecified so far. Since the approach is based on the least-squares principle, we know without any additional derivation that the estimators are *unbiased*. This property is also independent of the distribution (or even the covariance matrix) of the observable vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$.

Note that the weighted least-squares solution is however *not* necessarily of minimum variance as we can use any positive-definite symmetric matrix as the weight matrix in

equation (4.34). This problem can however be circumvented by taking the weight matrix to be the inverse of the covariance matrix, i.e. $W_{\mathrm{vh}} = Q_{\mathrm{vh}}^{-1}$ (see next section). Therefore, in this alternative and indeed the most efficient technique, we can use the 'probabilistic' least-squares technique to obtain the BLUE estimators (or better say, BIQUE estimators, since the observables are in *quadratic form* of the original ones, i.e. $\mathrm{vh}(\underline{t}\,\underline{t}^T)$) of the (co)variance components. In the previous section, we specified the covariance matrix Q_{vh} for the normal distribution. Therefore, in this thesis, the property of minimum variance of the (co)variance component estimators is restricted to the class of normally distributed data. As a generalization, Teunissen and Amiri-Simkooei (2006) give the minimum variance estimators for a larger class of distributions, namely elliptical distributions.

4.4.2 Covariance matrix of estimators

Since the (weighted) least-squares estimators are in a linear form of the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$, applying the error propagation law to equation (4.17) automatically gives us the covariance matrix of the estimated (co)variance components, namely $Q_{\hat{\sigma}}$, provided that the proper Q_{vh}, or equivalently the proper Q_t, is input as the covariance matrix of the vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. Applying the error propagation law to equation (4.17) gives (cf. equation (2.9))

$$Q_{\hat{\sigma}} = N^{-1} A_{\mathrm{vh}}^T W_{\mathrm{vh}} Q_{\mathrm{vh}} W_{\mathrm{vh}} A_{\mathrm{vh}} N^{-1}\,, \tag{4.66}$$

or simply

$$\boxed{Q_{\hat{\sigma}} = N^{-1} M N^{-1},} \tag{4.67}$$

where the $p \times p$ matrix M is given as

$$M = A_{\mathrm{vh}}^T W_{\mathrm{vh}} Q_{\mathrm{vh}} W_{\mathrm{vh}} A_{\mathrm{vh}}\,. \tag{4.68}$$

It is also possible to simplify the preceding equation. Substituting for W_{vh} from equation (4.34) and for Q_{vh} from equation (4.61) gives

$$\boxed{m_{kl} = 2\mathrm{tr}(B^T Q_k B W_t Q_t W_t B^T Q_l B W_t Q_t W_t),} \tag{4.69}$$

where m_{kl} represents the $p \times p$ matrix M in index notation.

Proof. Matrix M is given as $M = A_{\mathrm{vh}}^T W_{\mathrm{vh}} Q_{\mathrm{vh}} W_{\mathrm{vh}} A_{\mathrm{vh}} = m_{kl} = \mathrm{vh}(B^T Q_k B)^T W_{\mathrm{vh}} Q_{\mathrm{vh}} W_{\mathrm{vh}} \mathrm{vh}(B^T Q_l B)$ where $W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D$ and for normal distribution $Q_{\mathrm{vh}} = 2D^+(Q_t \otimes Q_t)D^{+T}$. Substitution of these terms, with $\mathrm{vh}(.) = D^+\mathrm{vec}(.)$ and $DD^+\mathrm{vec}(.) = \mathrm{vec}(.)$, gives the above expression for M. ▼

Equation (4.67) can therefore provide us with the precision of the estimators. This is in fact an important feature of the least-squares variance component estimation. Since the covariance matrix Q_{vh} depends on the distribution of the original observations, our results are again restricted to the class of normal distributions.

Example 4.4 (Example 4.1 continued) Let us assume that the misclosures $\underline{t}$ are normally distributed. It is now possible to compute the covariance matrix of the estimated (co)variance elements of Q_t, namely the covariance matrix of $\hat{\underline{\sigma}} = \mathrm{vh}(\hat{\underline{Q}}_t) = A_{\mathrm{vh}}^{-1}\mathrm{vh}(\underline{t}\,\underline{t}^T) = \mathrm{vh}(\underline{t}\,\underline{t}^T)$. This is

in fact a very simple example as $A_{\text{vh}} = I$. Application of the error propagation law to $\underline{\hat{\sigma}} = \text{vh}(\underline{t}\,\underline{t}^T)$ gives

$$Q_{\hat{\sigma}} = Q_{\text{vh}} = 2\,D^+(Q_t \otimes Q_t)D^{+T}\,. \tag{4.70}$$

In the preceding equation Q_t is unknown apriori. One will thus have to be satisfied with an estimate of Q_t, namely $\hat{Q}_t = \underline{t}\,\underline{t}^T$. This leads to an estimate of $Q_{\hat{\sigma}}$ as

$$\hat{Q}_{\hat{\sigma}} = 2\,D^+(t\,t^T \otimes t\,t^T)D^{+T}\,. \tag{4.71}$$

Strictly speaking, the (co)variance elements of Q_t are estimated as $\underline{\hat{\sigma}}_{ij} = \underline{t}_i\,\underline{t}_j$, with the variance of $\hat{\sigma}^2_{\hat{\sigma}_{ij}} = 2t_i^2\,t_j^2$ (follow similar way to Theorem 4.2 when $k = i$ and $l = j$). Note that the estimator $\underline{\hat{\sigma}}_{ij}$ is of the same order as its standard deviation $\hat{\sigma}_{\hat{\sigma}_{ij}}$. This indicates that the estimated (co)variance elements $\hat{\sigma}_{ij}$ may not be precise enough. In practice one may not have all the elements of Q_t to be unknown but only $p << \frac{b(b+1)}{2}$ (co)variance components which leads to a better precision of their related estimators. In the sequel, just for the sake of convenience, we will drop the hat notation located on top and simply denote $\hat{Q}$ as Q. ■

Example 4.5 (Example 4.3 continued) To determine the precision of the empirical autocovariance function, we will assume that y is normally distributed. The entries of matrix N in the expression of the covariance matrix $Q_{\hat{\sigma}} = N^{-1}MN^{-1}$ of $\hat{\sigma} = (\hat{\sigma}_0, \ldots, \hat{\sigma}_{m-1})^T$, are given in equation (4.44). To determine the entries of matrix M, we substitute $W_t = W_y = I$ and $B = I$ into equation (4.69). This gives

$$m_{\tau\kappa} = 2\text{tr}(Q_\tau Q_y Q_\kappa Q_y). \tag{4.72}$$

Therefore, the entries of the covariance matrix are given as

$$(Q_{\hat{\sigma}})_{\tau\kappa} = \frac{\text{tr}(Q_\tau Q_y Q_\kappa Q_y)}{2a_\tau^2(m-\tau)^2}. \tag{4.73}$$

This expression is easily evaluated numerically. A simple, but approximate, analytical expression can be obtained, if we approximate the covariance matrix of y by $Q_y \approx \sigma_0 I$. This is a good approximation, the closer the randomness of the time-series resembles that of white noise. With this approximation, the covariance matrix of equation (4.73) becomes a diagonal matrix, with entries

$$\sigma^2_{\hat{\sigma}_\tau} = \frac{\sigma^4}{a_\tau(m-\tau)},\ \tau = 0, 1, ..., m-1, \tag{4.74}$$

with $\sigma^4 = \sigma_0^2$. To obtain the variance of the autocorrelation function one can apply the error propagation law to the linearized form of equation (4.46)

$$\sigma^2_{\hat{\rho}_\tau} = \frac{1}{m-\tau} + \frac{2\rho_\tau^2}{m},\ \tau = 1, ..., m-1. \tag{4.75}$$

This shows that the precision of the autocorrelation function gets poorer with increasing time-lag τ. This also makes sense, since less data are used when τ increases. ■

4.4.3 Quadratic form of residuals

Since the approach is based on the least-squares principle, parts of the standard quality control theory can be applied to the model in equation (4.36) and the result in equation (4.18). One can in particular apply the idea of hypotheses testing to the stochastic

model. Chapter 5 deals with the w-test statistic and the quadratic form of the residuals in the stochastic model.

As an important measure of any least-squares solution, one can compute the quadratic form of the residuals. This quadratic form is given by equation (2.11) for the functional model. This can in fact be generalized for the least-squares (co)variance component estimation, i.e. to the stochastic model. Therefore, one will obtain

$$\underline{\hat{e}}_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{\hat{e}}_{\mathrm{vh}} = \mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)^T [W_{\mathrm{vh}} - W_{\mathrm{vh}} A_{\mathrm{vh}} N^{-1} A_{\mathrm{vh}}^T W_{\mathrm{vh}}] \mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)\,. \tag{4.76}$$

In a similar manner to the previous derivations, substitution $W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D$ results in the following formula:

$$\underline{\hat{e}}_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{\hat{e}}_{\mathrm{vh}} = (\underline{t}^T W_t \underline{t})^2 - 2\underline{t}^T W_t B^T Q_0 B W_t \underline{t} + \mathrm{tr}(B^T Q_0 B W_t B^T Q_0 B W_t) - \underline{l}^T N^{-1} \underline{l}\,, \tag{4.77}$$

with N and $\underline{l}$ introduced in equations (4.39) and (4.41), respectively. If $Q_0 = 0$, then it follows that

$$\underline{\hat{e}}_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{\hat{e}}_{\mathrm{vh}} = (\underline{t}^T W_t \underline{t})^2 - \underline{l}^T N^{-1} \underline{l}\,, \tag{4.78}$$

with $\underline{l}_k = \underline{t}^T W_t B^T Q_k B W_t \underline{t}$ obtained from equation (4.41). Note that the quadratic form is also independent of the (unspecified) distribution of the observable vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$.

4.4.4 Prior information

In some cases, we may have prior information about the (co)variance components. Such information can be provided by equipment manufacturers or from a previous process. Let us assume that this information can be expressed as $\underline{\sigma}_0$ in the following way

$$\mathsf{E}\{\underline{\sigma}_0\} = C\sigma\,; \quad W_{\sigma_0}\,, \tag{4.79}$$

where W_{σ_0} is the weight matrix of $\underline{\sigma}_0$ and C is a known matrix. One important feature of the LS-VCE is the possibility of incorporating such prior information with the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. Without additional derivations, we can obtain the weighted least-squares (co)variance estimators as (Teunissen, 2000a)

$$\underline{\hat{\sigma}} = N^{-1}\underline{l} + N^{-1}C^T(CN^{-1}C^T + W_{\sigma_0}^{-1})(\underline{\sigma}_0 - CN^{-1}\underline{l}). \tag{4.80}$$

The method uses all of the information in an optimal way. Concerning the preceding formulation, we can at least mention that

- the method is considered as a competitor to the Bayesian method. The prior information, here for the LS-VCE, is in the form of the weight or covariance matrix rather than the probability density function,
- the method can be used as a regularization technique for ill-posed (co)variance component models (see ill-posed problems in section 4.8.3),

- the method may also be used for incorporating several estimators obtained from various data sets in a sequential way.
- if we choose $W_{\sigma_0}^{-1} = 0$, then the solution obtained corresponds to using the hard constraints $\sigma_0 = C\sigma$.
- if we choose $W_{\rm vh} = Q_{\rm vh}^{-1}$ and $W_{\sigma_0} = Q_{\sigma_0}^{-1}$, we then obtain the minimum variance estimators. In this case, the covariance matrix of estimators is simply given as $Q_{\hat{\sigma}} = (N + C^T Q_{\sigma_0}^{-1} C)^{-1} = N^{-1} - N^{-1}C^T(CN^{-1}C^T + Q_{\sigma_0})^{-1}CN^{-1}$ (cf. later on equation (4.88)),

4.4.5 Robust estimation

Since we estimated the (co)variance components on the basis of a linear model of observation equations (see equation (4.36)), we can think of *robust estimation* methods rather than the least-squares. One can in particular think of an L_1 norm minimization problem. The usual method for implementation of the L_1 norm adjustment leads to the solving of a linear programming problem. For more information we refer to e.g. Amiri-Simkooei (2003). This may be an important alternative if one wants to be guarded against misspecifications in the functional part of the model.

Example 4.6 In example 4.2 we can apply the idea of L_1 norm minimization problem. The stochastic model expressed in terms of a model of observation equations is as follows:

$$\mathrm{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = \mathrm{vh}(I)\sigma^2, \quad W_{\rm vh} = I\,. \tag{4.81}$$

The L_1 norm principle states that the sum of the absolute values of the 'residual' vector $\mathrm{vh}(\underline{t}\,\underline{t}^T) - \mathrm{vh}(I)\sigma^2$ should be minimized. Because $\frac{b(b-1)}{2}$ elements of $\mathrm{vh}(I)$ are zeros, they do not effect the minimization problem. Therefore, the L_1 norm minimization problem is reduced to the following problem:

$$\sum_{i=1}^{m} |\underline{t}_i^2 - \sigma^2| \rightarrow \min\,. \tag{4.82}$$

This is in fact very similar to the problem of measuring one unknown quantity, where the observables are $\underline{t}_i^2,\ i = 1, ..., m$. It is well known that the L_1 norm minimization will lead to the sample *median* as an estimator of the population mean for repeated measurements of an unknown quantity. Therefore, the variance component is estimated as the median of the $\underline{t}_1^2, \underline{t}_2^2, ..., \underline{t}_m^2$, namely

$$\breve{\underline{\sigma}}^2 = \underline{t}^2_{\frac{m+1}{2}}\,, \tag{4.83}$$

whereas the least-squares solution gives the mean of $\underline{t}_1^2, \underline{t}_2^2, ..., \underline{t}_m^2$ (see example 4.2). ■

4.5 Minimum variance estimators

As with the BLUE estimator introduced in chapter 2, the (co)variance components can be estimated as the best linear unbiased estimator of the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. One can obtain such estimators by taking the weight matrix $W_{\rm vh}$ as the inverse of the covariance matrix of the observables, $Q_{\rm vh}^{-1}$. Then this linear form of the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ can be written as the best (minimum variance) quadratic unbiased estimator of the misclosures

$\underline{t}$. This can therefore be considered as an alternative derivation of the BIQUE estimator introduced in chapter 3. In section 4.3 we derived the covariance matrix Q_{vh} for normally distributed data. Therefore, the 'best' (minimum variance) property is restricted to the normal distribution (or even to a larger class of elliptical distributions).

Estimator To obtain the minimum variance estimators, one needs to substitute $W_{\mathrm{vh}} = Q_{\mathrm{vh}}^{-1} = \frac{1}{2}D^T(Q_t^{-1} \otimes Q_t^{-1})D$ in equation (4.17)

$$\underline{\hat{\sigma}} \;=\; N^{-1}\underline{l} = (A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} A_{\mathrm{vh}})^{-1} A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)\,. \tag{4.84}$$

After simplification, the $p \times p$ normal matrix N, and the $p \times 1$ vector $\underline{l}$ read, respectively (follow the same way as the previous derivations)

$$\boxed{n_{kl} = \frac{1}{2}\mathrm{tr}(B^T Q_k B Q_t^{-1} B^T Q_l B Q_t^{-1}),} \tag{4.85}$$

and

$$\boxed{\underline{l}_k = \frac{1}{2}\underline{t}^T Q_t^{-1} B^T Q_k B Q_t^{-1}\underline{t} - \frac{1}{2}\mathrm{tr}(B^T Q_k B Q_t^{-1} B^T Q_0 B Q_t^{-1}).} \tag{4.86}$$

Note that the preceding equations can simply be obtained from equations (4.39) and (4.41) with $W_t = \frac{1}{\sqrt{2}}Q_t^{-1}$ (see equation (4.63)). In the case that $Q_0 = 0$, the preceding equation reads

$$\underline{l}_k = \frac{1}{2}\underline{t}^T Q_t^{-1} B^T Q_k B Q_t^{-1}\underline{t}\,. \tag{4.87}$$

NOTE 4.5 For elliptically contoured distributed data one can obtain the above estimators by choosing the inverse of covariance matrix Q_{vh} given in equation (4.64) as the weight matrix. The normal matrix N and the vector $\underline{l}$ are respectively given as (follow the same way as the previous derivations)

$$\begin{aligned} n_{kl} &= \tfrac{1}{2(\kappa+1)}[\mathrm{tr}(B^T Q_k B Q_t^{-1} B^T Q_l B Q_t^{-1}) + \alpha\,\mathrm{tr}(B^T Q_k B Q_t^{-1})\mathrm{tr}(B^T Q_l B Q_t^{-1})]; \\ \underline{l}_k &= \tfrac{1}{2(\kappa+1)}[\underline{t}^T Q_t^{-1} B^T Q_k B Q_t^{-1}\underline{t} + \alpha(\underline{t}^T Q_t^{-1}\underline{t})\mathrm{tr}(B^T Q_k B Q_t^{-1})], \end{aligned}$$

with κ the kurtosis parameter and α defined in equation (4.65). Note again that for normal distribution $\kappa = \alpha = 0$. $\boxdot$

Covariance matrix Since we have formulated our variance component estimation problem in the least-squares sense, we know without any additional derivation that the inverse of the normal matrix N automatically gives us the covariance matrix of the estimated (co)variance components

$$\boxed{Q_{\hat{\sigma}} = N^{-1},} \tag{4.88}$$

which provides us with the precision of the (co)variance estimators. The preceding equation can also be obtained from equation (4.67) because in this case $M = N$.

Implementation of LS-VCE (B model)

Input:
1. matrix B of condition equations;
2. observation vector y;
3. cofactor matrices $Q_k,\ k = 0, ..., p$;
4. initial (co)variance components $\sigma = \sigma^0 = [\sigma_1^0, ..., \sigma_p^0]^T$;
5. small value for ϵ;

begin
check for presence of gross errors in observations;
compute misclosures $t = B^T y$;
set iteration counter $i = 0$;
begin
evaluate matrix $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$;
compute matrix $Q_t = B^T Q_y B$;
invert covariance matrix Q_t;
calculate matrix N and vector l from equations (4.85) and (4.86);
solve for a new $\hat{\sigma}$ from normal equations $N\hat{\sigma} = l$;
increase i by one step;
update vector $\sigma^i \leftarrow \hat{\sigma}$;
while $\|\sigma^i - \sigma^{i-1}\|_{Q_{\hat{\sigma}}^{-1}} > \epsilon$ **repeat**;

end
obtain $\hat{\sigma}$ and its covariance matrix $Q_{\hat{\sigma}} = N^{-1}$.
end

Figure 4.1: Symbolic algorithm for implementation of least-squares variance component estimation in terms of linear model of condition equations (B-model); σ^i contains the (co)variance components estimated in iteration i.

Quadratic form of residuals In a similar manner, from equation (4.77), the quadratic form of the residuals is given as

$$\begin{aligned} \hat{\underline{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \hat{\underline{e}}_{\mathrm{vh}} &= \tfrac{1}{2}(\underline{t}^T Q_t^{-1} \underline{t})^2 - \underline{t}^T Q_t^{-1} B^T Q_0 B Q_t^{-1} \underline{t} \\ &+ \tfrac{1}{2}\mathrm{tr}(B^T Q_0 B Q_t^{-1} B^T Q_0 B Q_t^{-1}) - \underline{l}^T N^{-1} \underline{l} \ , \end{aligned} \tag{4.89}$$

with N and $\underline{l}$ defined in equations (4.85) and (4.86), respectively. When $Q_0 = 0$, one obtains

$$\hat{\underline{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \hat{\underline{e}}_{\mathrm{vh}} = \frac{1}{2}(\underline{t}^T Q_t^{-1} \underline{t})^2 - \underline{l}^T N^{-1} \underline{l} \, . \tag{4.90}$$

Implementation In this thesis we mainly deal with *linear* models; the observations and the unknown parameters are related to each other by the linear model of observation equations, namely $\mathsf{E}\{\underline{y}\} = Ax$. An equivalent representation of this model is expressed

in terms of model of condition equations, i.e. $B^T\mathsf{E}\{\underline{y}\} = 0$. We will now concentrate on the implementation of this formulation. Before we are able to use the observations, we need to check for the presence of some (large) gross errors in observations, because outliers in the data will lead to biases in the (co)variance components. Since the functional model is generally known, obtaining some rough measures for verifying the correctness of observations is not difficult. For instance, GPS pseudo-range observations can be roughly checked on exceeding the interval 18000-26000 km.

From equations (4.85) and (4.86) with equation (4.18) we see that we need $Q_t = B^TQ_0B + \sum_{k=1}^{p}\sigma_k B^TQ_kB$ in order to compute the estimators $\hat{\underline{\sigma}}_k$. But the (co)variance components σ_k are unknown apriori. One way out of this dilemma is to perform *iterations*. One starts with an initial guess for the σ_k. Using these values, one computes with equation (4.18) estimates for the σ_k, which in the next cycle are considered the improved guess for σ_k. Figure 4.1 shows an iterative algorithm for performing a least-squares (co)variance component estimation in terms of the model of condition equations. This can only be a straightforward and symbolic procedure. In real problems, specially when the redundancy of the functional model is large, one may use more simplified procedures (see e.g. section 3.4).

NOTE 4.6 It should be perceived, from equation (4.14), that this model is intrinsically linear in terms of the observable vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. But when applying the (probabilistic) least squares technique, one needs to know the covariance matrix Q_{vh} of the observables which is unknown and is to be estimated (because Q_t is unknown). Therefore, the usual strategy is to estimate the stochastic model through iteration. For this purpose the apriori values of the (co)variance components are needed. Such values may be obtainable all the time since the nominal precision of the instruments is usually known. One may also employ approximate methods to obtain these values.

4.6 Maximum likelihood estimators

The maximum likelihood (ML) method to (co)variance component estimation is only applicable when the general structure of the probability density function (PDF) is known (see section 2.1.5). In this section we will restrict ourselves to the normal distribution for which we obtain

$$\underline{y} \sim \mathsf{N}(Ax, Q_y)\,, \quad \text{with} \quad Q_y = Q_0 + \sum_{k=1}^{p}\sigma_k Q_k\,, \tag{4.91}$$

where the first two moments of the observables $\underline{y}$, namely $\mathsf{E}\{\underline{y}\}$ and $\mathsf{D}\{\underline{y}\}$, are partially unknown. As mentioned in section 2.1.6 an equivalent expression for the model of observation equations $\mathsf{E}\{\underline{y}\} = Ax$ is in terms of the model of condition equations. With this, the preceding equation is reformulated as

$$\underline{t} = B^T\underline{y} \sim \mathsf{N}(0, Q_t)\,, \quad \text{with} \quad Q_t = B^TQ_0B + \sum_{k=1}^{p}\sigma_k B^TQ_kB\,, \tag{4.92}$$

in which we have left out the unknown parameters x of the functional model. The (restricted) likelihood function is therefore of the form (see also section 3.3.4)

$$L(t;\sigma) = \frac{1}{(2\pi)^{\frac{b}{2}}\det(Q_t)^{\frac{1}{2}}}\exp(-\frac{1}{2}t^TQ_t^{-1}t)\,, \tag{4.93}$$

with Q_t expressed in equation (4.92). In order to solve for the unknown (co)variance components, one needs to maximize the likelihood function over these unknowns

$$\hat{\sigma} = \arg\max_{\sigma} \frac{1}{(2\pi)^{\frac{b}{2}} \det(Q_t)^{\frac{1}{2}}} \exp(-\frac{1}{2} t^T Q_t^{-1} t) \tag{4.94}$$

In practice, it is more convenient to work with the log likelihood function, namely

$$\hat{\sigma} = \arg\max_{\sigma} \ln L(t;\sigma) = \arg\max_{\sigma} -\frac{b}{2}\ln 2\pi - \frac{1}{2}\ln\det(Q_t) - \frac{1}{2} t^T Q_t^{-1} t \tag{4.95}$$

where $\ln$ is the natural logarithm. The preceding maximization problem is equivalent to the following minimization problem:

$$\hat{\sigma} = \arg\max_{\sigma} \ln L(t;\sigma) = \arg\min_{\sigma} \frac{1}{2}\ln\det(Q_t) + \frac{1}{2} t^T Q_t^{-1} t \tag{4.96}$$

To minimize the preceding equation (or maximize log likelihood) one needs to take the partial derivatives with respect to σ and set them to zero. With doing so, the following set of nonlinear equations is obtained

$$\boxed{\frac{1}{2}\mathrm{tr}(Q_t^{-1} B^T Q_k B) - \frac{1}{2} t^T Q_t^{-1} B^T Q_k B Q_t^{-1} t = 0,\ k = 1, ..., p\,.} \tag{4.97}$$

There are standard procedures for solving a nonlinear problem. We use here a very simple technique to solve for the unknown parameters which is equivalent to the well-known Newton-Raphson method. The term $\mathrm{tr}(Q_t^{-1} B^T Q_k B)$ in the above equation can be rewritten as $\mathrm{tr}(Q_t^{-1} B^T Q_k B\, Q_t^{-1} Q_t)$. If we substitute Q_t from equation (4.92) we obtain

$$\frac{1}{2}\mathrm{tr}(Q_t^{-1} B^T Q_k B Q_t^{-1}[B^T Q_0 B + \sum_{l=1}^{p} \sigma_l B^T Q_l B]) = \frac{1}{2} t^T Q_t^{-1} B^T Q_k B Q_t^{-1} t\,. \tag{4.98}$$

If we now take the known term $\frac{1}{2}\mathrm{tr}(Q_t^{-1} B^T Q_k B Q_t^{-1} B^T Q_0 B)$ to the right-hand side, the preceding equation reads

$$\sum_{l=1}^{p} n_{kl}\hat{\sigma}_l = l_k,\ k = 1, ..., p\,, \tag{4.99}$$

with n_{kl} and l_k defined in equations (4.85) and (4.86), respectively. This is an interesting result which confirms that the least-squares (co)variance estimators are identical to the restricted maximum likelihood (REML) estimators if the observables are normally distributed. These estimators are therefore unbiased, of minimum variance (best), and restrictedly of maximum likelihood (cf. NOTE 2.3).

4.7 In terms of model of observation equations

The least-squares technique to (co)variance component estimation can directly be used, through equation (4.18) with equations (4.85) and (4.86), if the matrix B is available (model of condition equations). In practice however one will usually have the design matrix A available (model of observation equations), instead of B. In this section, we will extend the least-squares method for estimation of (co)variance components to the model of observation equations. We will again restrict ourselves to the class of normally distributed data.

4.7.1 Weighted least-squares estimators

Formulation We consider again the case that the covariance matrix can be split into a known part Q_0 and an unknown (co)variance component model, namely $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$. To apply the deterministic (weighted) least-squares variance component estimation to the model of observation equations we shall therefore have to rewrite equations (4.39) and (4.41) in terms of the design matrix A. Using the *trace* property, equation (4.39) can be rewritten as

$$n_{kl} = \mathrm{tr}(Q_k B W_t B^T Q_l B W_t B^T)\,. \tag{4.100}$$

If W is the weight matrix in the model of observation equations and correspondingly W_t is the weight matrix in the model of condition equations, to obtain identical estimators, it follows with identity (2.20) that

$$B W_t B^T = W(I - A(A^T W A)^{-1} A^T W) = W P_A^{\perp}\,. \tag{4.101}$$

The preceding identity with equation (4.100) gives

$$\boxed{n_{kl} = \mathrm{tr}(Q_k W P_A^{\perp} Q_l W P_A^{\perp})\,.} \tag{4.102}$$

Similarly, it follows with

$$\underline{\hat{e}} = P_A^{\perp}\underline{y} = W^{-1} B W_t B^T \underline{y} = W^{-1} B W_t \underline{t}\,, \tag{4.103}$$

from equation (4.41) that

$$\underline{l}_k = \underline{y}^T P_A^{\perp T} W Q_k W P_A^{\perp} \underline{y} - \mathrm{tr}(Q_k W P_A^{\perp} Q_0 W P_A^{\perp})\,. \tag{4.104}$$

In terms of the least-squares residuals $\underline{\hat{e}}$ the preceding equation reads

$$\boxed{\underline{l}_k = \underline{\hat{e}}^T W Q_k W \underline{\hat{e}} - \mathrm{tr}(Q_k W P_A^{\perp} Q_0 W P_A^{\perp})\,.} \tag{4.105}$$

Strictly speaking, the weighted least-squares estimator is given as $\underline{\hat{\sigma}} = N^{-1}\underline{l}$ with N and $\underline{l}$ given by equations (4.102) and (4.105), respectively.

Covariance matrix As with any weighted least-squares problem, one can compute the covariance matrix of estimators $\underline{\hat{\sigma}}$, namely $Q_{\hat{\sigma}} = N^{-1} M N^{-1}$, where the $p \times p$ matrix M obtained from equation (4.69) simplifies to

$$m_{kl} = 2\mathrm{tr}(Q_k W P_A^{\perp} Q_y W P_A^{\perp} Q_l W P_A^{\perp} Q_y W P_A^{\perp})\,. \tag{4.106}$$

Quadratic form of residuals In a similar manner, in terms of the design matrix A, the quadratic form of the residuals in the stochastic model reads (see equation (4.77))

$$\underline{\hat{e}}_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{\hat{e}}_{\mathrm{vh}} = (\underline{\hat{e}}^T W \underline{\hat{e}})^2 - 2\underline{\hat{e}}^T W Q_0 W \underline{\hat{e}} + \mathrm{tr}(Q_0 W P_A^{\perp} Q_0 W P_A^{\perp}) - \underline{l}^T N^{-1} \underline{l}\,, \tag{4.107}$$

with N and $\underline{l}$ given by equations (4.102) and (4.105), respectively. If $Q_0 = 0$, it follows that

$$\underline{\hat{e}}_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{\hat{e}}_{\mathrm{vh}} = (\underline{\hat{e}}^T W \underline{\hat{e}})^2 - \underline{l}^T N^{-1} \underline{l}\,, \tag{4.108}$$

where $\underline{l}_k = \underline{\hat{e}}^T W Q_k W \underline{\hat{e}}$ is obtained from equation (4.105).

4.7.2 Minimum variance estimators

Formulation To obtain the minimum variance estimators, we should choose the weight matrix as the inverse of the covariance matrix. In an analogous way to equation (4.63), one can take the symmetric and positive-definite matrix W as

$$W = \frac{1}{\sqrt{2}} Q_y^{-1} \,. \tag{4.109}$$

If we now substitute $W = \frac{1}{\sqrt{2}} Q_y^{-1}$ into the formulas of the previous subsection, equations (4.100) and (4.106) read then

$$\boxed{n_{kl} = m_{kl} = \frac{1}{2} \mathrm{tr}(Q_k Q_y^{-1} P_A^{\perp} Q_l Q_y^{-1} P_A^{\perp}),} \tag{4.110}$$

with the orthogonal projector $P_A^{\perp} = I - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$ (see chapter 2). Similarly, from equation (4.104) it follows that

$$\underline{l}_k = \frac{1}{2} \underline{y}^T P_A^{\perp T} Q_y^{-1} Q_k Q_y^{-1} P_A^{\perp} \underline{y} - \frac{1}{2} \mathrm{tr}(Q_k Q_y^{-1} P_A^{\perp} Q_0 Q_y^{-1} P_A^{\perp}) \,. \tag{4.111}$$

In terms of the least-squares (BLUE) residuals, one obtains

$$\boxed{\underline{l}_k = \frac{1}{2} \underline{\hat{e}}^T Q_y^{-1} Q_k Q_y^{-1} \underline{\hat{e}} - \frac{1}{2} \mathrm{tr}(Q_k Q_y^{-1} P_A^{\perp} Q_0 Q_y^{-1} P_A^{\perp}).} \tag{4.112}$$

The minimum variance (co)variance component estimation in terms of the model of observation equations is therefore formulated as follows:

$$\underline{\hat{\sigma}} = N^{-1} \underline{l} \,, \tag{4.113}$$

with N and $\underline{l}$ specified in equation (4.110) and (4.112), respectively. The inverse of the normal matrix N gives automatically the covariance matrix of estimators

$$\boxed{Q_{\hat{\sigma}} = N^{-1}.} \tag{4.114}$$

Note that while the inverse of the normal matrix gives the covariance matrix of the (co)variance components, i.e. $Q_{\hat{\sigma}} = N^{-1}$, the normal matrix itself is the covariance matrix of the p-vector $\underline{l}$. From equation (4.107), the quadratic form of the residuals in the stochastic model reads

$$\underline{\hat{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \underline{\hat{e}}_{\mathrm{vh}} = \frac{1}{2} (\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2 - \underline{\hat{e}}^T Q_y^{-1} Q_0 Q_y^{-1} \underline{\hat{e}} + \frac{1}{2} \mathrm{tr}(Q_0 Q_y^{-1} P_A^{\perp} Q_0 Q_y^{-1} P_A^{\perp}) - \underline{l}^T N^{-1} \underline{l} \,. \tag{4.115}$$

When the known part is zero, i.e. $Q_0 = 0$, the preceding equation simplifies to

$$\underline{\hat{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \underline{\hat{e}}_{\mathrm{vh}} = \frac{1}{2} (\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2 - \underline{l}^T N^{-1} \underline{l} \,. \tag{4.116}$$

Implementation Equations (4.110) and (4.112) with equation (4.113) show that we need $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$ in order to compute the estimators $\hat{\underline{\sigma}}_k$. But the (co)variance components σ_k are unknown apriori! We can therefore perform *iterations*. We can start with an initial guess for the σ_k. Based on these values, we compute with equation (4.113) estimates for the σ_k, which in the next iteration are considered the improved values for σ_k. The procedure is repeated until the estimated components do not change by further iteration. That is, convergence is achieved if the guess for σ_k equals the computed estimate $\hat{\sigma}_k$. Figure 4.2 gives a straightforward and symbolic iterative algorithm for implementing a least-squares variance component estimation in terms of the model of observation equations. Note that when the number of observations m is large, one may employ more efficient procedures (see e.g. section 3.4).

One point which may be useful to notice is the issue of computing (co)variance components. Actually, there are two ways of estimating such unknown parameters. The *first* way is to consider the cofactor matrices as a whole and try to estimate unknown unit *factors* (scale factors). That is, in each iteration we modify the cofactor matrices by multiplying them with the estimated (co)variance factors. After a few iterations we expect the factors to converge to ones. In the *second* way, we consider the cofactor matrices to be fixed (constant). In each iteration, the (co)variance *components* rather than the cofactor matrices are modified. After a few iterations, the modified components converge so that the maximum absolute difference with their modified previous iteration values does not exceed a small upper bound. For example consider the covariance matrix as $Q_y = \sigma_1 Q_1 + \sigma_2 Q_2$. At the point of convergence, the above strategies look as follows: In the first way, we obtain the factors f_1 and f_2, therefore $Q_y = f_1 \times \hat{\sigma}_1 Q_1 + f_2 \times \hat{\sigma}_2 Q_2$ where $f_1 = f_2 = 1$ and in the second way we estimate the components σ_1 and σ_2, therefore $Q_y = \hat{\sigma}_1 \times Q_1 + \hat{\sigma}_2 \times Q_2$. The same situation is also valid for the covariance matrix (or precision) of the estimators. The latter way is used throughout this thesis.

NOTE 4.7 We need to take good care of the a-priori as well as the estimated values of the (co)variance components to make sure that the resulting initial and estimated variance component model Q_y is positive definite. For example, consider the following cofactor matrices for two types of observations:

$$Q_1 = \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix}; \quad Q_2 = \begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix}; \quad Q_3 = \begin{pmatrix} 0 & Q \\ Q & 0 \end{pmatrix}, \tag{4.117}$$

with Q a symmetric and positive-definite matrix of appropriate size. Assuming $\sigma_1 = \sigma_2 = \sigma_3 = \sigma > 0$, one obtains

$$Q_y = \sigma \begin{pmatrix} Q & Q \\ Q & Q \end{pmatrix}, \tag{4.118}$$

which is a singular matrix, i.e. $\mathrm{rank}(Q_y) = \mathrm{rank}(Q) \neq \mathrm{size}(Q_y)$. In addition when $\sigma_1 = \sigma_2 = \sigma$ and $\sigma_3 = \sigma' > \sigma$, Q_y is not positive definite. One may for instance initiate the variance components as $\sigma_1 = \sigma_2 = \sigma$ and the covariance component as $\sigma_3 = 0$. □

Optimal properties The estimators obtained in each cycle are unbiased estimators of the σ_k. But they are not strictly of minimum variance, not even after convergence of the iterations. Since the computed estimate $\hat{\sigma}_k$ is not necessarily equal to σ_k, the property

Implementation of LS-VCE (A model)

Input:
1. design matrix A of observation equations;
2. observation vector y;
3. cofactor matrices $Q_k,\ k = 0, ..., p$;
4. initial (co)variance components $\sigma = \sigma^0 = [\sigma_1^0, ..., \sigma_p^0]^T$;
5. small value for ϵ;

begin
check for presence of gross errors in observations;
set iteration counter $i = 0$;
begin
evaluate matrix $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$;
invert covariance matrix Q_y;
calculate vector of least-squares residuals $\hat{e} = P_A^{\perp} y$;
compute matrix $Q_y^{-1} P_A^{\perp}$;
calculate matrix N and vector l from equations (4.110) and (4.112);
solve for a new $\hat{\sigma}$ from normal equations $N\hat{\sigma} = l$;
increase i by one step;
update vector $\sigma^i \leftarrow \hat{\sigma}$;
while $\|\sigma^i - \sigma^{i-1}\|_{Q_{\hat{\sigma}}^{-1}} > \epsilon$ **repeat**;

end
obtain $\hat{\sigma}$ and its covariance matrix $Q_{\hat{\sigma}} = N^{-1}$.
end

Figure 4.2: Symbolic algorithm for implementation of least-squares variance component estimation in terms of linear model of observation equations (A-model); σ^i is the (co)variance components estimated in iteration i.

of minimum variance may not necessarily be achieved. Therefore, in practice one usually will have to be satisfied with *almost* minimum variance unbiased estimators. The amount in which the computed estimates lack the property of minimum variance, depends on the initial guess and the number of iterations. There are however special cases where the σ_k are *not* needed apriori. One such case we will meet when discussing the estimator for the variance of unit weight (see Example 4.7). Another important case where the apriori values are not needed occurs when one wants to estimate the covariance matrix from a multivariate linear model (see chapter 6).

NOTE 4.8 Since we assumed $\underline{t}$ (or in fact $\underline{y}$) to have a normal distribution when deriving the covariance matrix Q_{vec}, the BLUE (or better say BIQUE) property of the estimator $\underline{\hat{\sigma}}$ is restricted to normally distributed data. It is possible to generalize the 'best' property to a larger class than normal, namely to elliptical distributions. For example, the multivariate Student t distribution is a special case in this class for which the kurtosis parameter is

not zero, $\kappa = \frac{2}{b-4}$. The minimum variance estimators can thus be obtained for the class of elliptically contoured distributed data (see Teunissen and Amiri-Simkooei, 2006). Therefore, if $\kappa \neq 0$, then $N\underline{\hat{\sigma}} = \underline{l}$ with equations (4.110) and (4.112) does not *strictly* provide minimum variance estimators but only *asymptotically*; i.e. when the redundancy in the functional model is large enough. The unbiasedness property of course still holds in general. □

Example 4.7 Let us as a simple application of the least-squares variance component estimation LS-VCE, assume that there is only one variance component in the stochastic model, namely $Q_y = \sigma^2 Q$. If our original observables $\underline{y}$ are normally distributed, it follows with equations (4.110) and (4.112) from equation (4.113) that

$$\underline{\hat{\sigma}}^2 = \frac{\underline{l}}{n} = \frac{\frac{1}{2}\underline{\hat{e}}^T Q_y^{-1} Q Q_y^{-1} \underline{\hat{e}}}{\frac{1}{2}\mathrm{tr}(Q Q_y^{-1} P_A^\perp Q Q_y^{-1} P_A^\perp)}\,, \tag{4.119}$$

with the expectation and the variance of

$$\mathsf{E}\{\underline{\hat{\sigma}}^2\} = \sigma^2 \text{ (unbiased)}\,; \quad \mathsf{D}\{\underline{\hat{\sigma}}^2\} = \frac{2}{\mathrm{tr}(Q Q_y^{-1} P_A^\perp Q Q_y^{-1} P_A^\perp)} \equiv \min \text{ (best)}\,, \tag{4.120}$$

respectively. With $Q_y = \sigma^2 Q$, $P_A^\perp P_A^\perp = P_A^\perp$, and $\mathrm{tr}(P_A^\perp) = \mathrm{rank}(P_A^\perp) = m - n = b$, the preceding equations simplify to

$$\underline{\hat{\sigma}}^2 = \frac{\underline{\hat{e}}^T Q^{-1} \underline{\hat{e}}}{m-n}\,; \quad \mathsf{E}\{\underline{\hat{\sigma}}^2\} = \sigma^2 \;\text{ and }\; \mathsf{D}\{\underline{\hat{\sigma}}^2\} = \frac{2\sigma^4}{m-n}\,. \tag{4.121}$$

These are the well-known results for the estimator of the variance of unit weight. This estimator can thus be obtained from the least-squares residuals without any iteration. The least-squares approach implies that the above estimator is optimal in the sense that it is *unbiased* and has *minimum variance*. With the least-squares approach we now also have a unified framework in which the well-known estimator of the variance of unit weight finds its logical place. That is, we now do not have to introduce the estimator of the variance of unit weight in an ad hoc way! In addition, the quadratic form of residuals, i.e. equation (4.116) for our model, simplifies to

$$\underline{\hat{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \underline{\hat{e}}_{\mathrm{vh}} = \frac{b-1}{2b} (\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2, \tag{4.122}$$

with $b = m - n$ the redundancy of the functional model. ■

4.7.3 Nonlinear covariance function

Let us now consider a nonlinear (co)variance component model, namely $Q_y = Q(\sigma)$. To overcome the nonlinearity, one can expand the stochastic model into a Taylor series. For this purpose we need to know the initial values for the unknown vector σ, namely σ^0. The linear term of the Taylor series expansion reads

$$Q_y = Q(\sigma^0) - \sum_{k=1}^{p} \frac{\partial Q(\sigma)}{\partial \sigma_k} \sigma_k \Big|_{\sigma^0} + \sum_{k=1}^{p} \frac{\partial Q(\sigma)}{\partial \sigma_k} \Big|_{\sigma^0} \sigma_k\,. \tag{4.123}$$

The above equation is in the form of the general (co)variance component model $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$ with

$$Q_0 = Q(\sigma^0) - \sum_{k=1}^{p} \frac{\partial Q(\sigma)}{\partial \sigma_k} \sigma_k \Big|_{\sigma^0}\,; \quad Q_k = \frac{\partial Q(\sigma)}{\partial \sigma_k} \Big|_{\sigma^0}\,. \tag{4.124}$$

Implementation of non linear LS-VCE (A model)

Input:
1. design matrix A of observation equations;
2. observation vector y;
3. function $Q(\sigma)$ and its partial derivatives $\partial_{\sigma_k} Q(\sigma),\ k = 1, ..., p$;
4. initial (co)variance components $\sigma = \sigma^0 = [\sigma_1^0, ..., \sigma_p^0]^T$;
5. small value for ϵ;

begin
check for presence of gross errors in observations;
set iteration counter $i = 0$;
begin
obtain (modify) matrices $Q_k^i,\ k = 0, 1, ..., p$, from equation (4.124);
evaluate matrix $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$;
invert covariance matrix Q_y;
calculate vector of least-squares residuals $\hat{e} = P_A^{\perp} y$;
compute matrix $Q_y^{-1} P_A^{\perp}$;
calculate matrix N and vector l from equations (4.110) and (4.112);
solve for a new $\hat{\sigma}$ from normal equations $N\hat{\sigma} = l$;
increase i by one step;
update vector $\sigma^i \leftarrow \hat{\sigma}$;
while $\|\sigma^i - \sigma^{i-1}\|_{Q_{\hat{\sigma}}^{-1}} > \epsilon$ **repeat**;

end
obtain $\hat{\sigma}$ and its covariance matrix $Q_{\hat{\sigma}} = N^{-1}$.
end

Figure 4.3: Symbolic algorithm for implementation of nonlinear least-squares variance component estimation in terms of linear model of observation equations.

We can now apply the least-squares method of the previous subsection to (co)variance components. What we will estimate is the least-squares estimator for σ . The estimated $\underline{\hat{\sigma}}$ can now be considered as a new update for σ_0 and the same procedure can be repeated. We can iterate until the estimated (co)variance components do not change by further iterations. The Gauss-Newton iteration has a linear rate of convergence. This convergence rate is dictated by the normal curvature of the nonlinear manifold $A_{\text{vh}}(\sigma)$. For more information we refer to Teunissen (1990).

NOTE 4.9 In the nonlinear LS-VCE, the cofactor matrices $Q_k,\ k = 1, ..., p$ are not fixed. They are modified from one iteration to the next. It is important to perceive that we have in fact two sorts of iterations for this model; one to come up with the nonlinearity and one due to the covariance matrix Q_y which is unknown apriori. Because of the Taylor series expansion, the initial values σ_0 should be close enough to their final values. Rough approximate values can lead to the divergence of the computed (co)variance components. The implementation algorithm is given in figure 4.3. ⊡

4.8 Remarks on stochastic model

4.8.1 Negative variance components

Since the LS-VCE method is based on the 'unrestricted' or 'unconstrained' least-squares principle, occurrence of *negative* variance components is not impossible. Estimated negative variance components may result in a non-positive definite covariance matrix Q_y, which can not be physically interpreted. Occurrence of negative variances makes sense when we compare the estimators with their precision (see equation (4.114)); they may not be precise enough because of insufficient redundancy in the functional model. Though non-interpretable, a significantly negative variance component can be an important indication of defects in our variance component model. We can enumerate at least

- an improperly designed (co)variance component model;
  The problem occurs when the (co)variance component model is not correct for the data at hand. In this case one has to look for an appropriate stochastic model. In chapter 5 we introduce and employ hypotheses testing (e.g. w-test statistic) in conjunction with the least-squares variance component estimation. The method provides a powerful tool to look for a realistic stochastic model. The problem of an improperly designed stochastic model is thus not subject of discussion in this section.

- a low redundancy of our functional model;
  Low *redundancy* b of the functional model (and thereby df of stochastic model) can give rise to some negative variance components. This also makes sense because the precision of all parameters of interest directly depends on the number of observations involved in the model. The more observations we have, the more precision for our (co)variance components we will obtain, and the less chance for estimators being inadmissible (e.g. negative) we have.

  NOTE 4.10 If the redundancy is too low, at times, the variance component model becomes singular. That is, we do not have enough observations for unique determination of the unknowns in the stochastic model with that special structure. There might be different ways to handle the problem. One can for instance add more observations to handle the problem (see next subsection). The other way, if allowed, is to modify the stochastic model. The observations might be sufficient for estimating the parameters of a modified stochastic model. Low redundancy may also be interpreted as a weak functional model. Given the stochastic model, a weak functional model may lead to a high correlation between (co)variance components. Such estimators are usually of poor precision and occurrence of negative variances is more probable (see section 4.8.3). ⊡

- a badly chosen set of apriori (co)variance components (initial weight matrix W).
  To avoid the non-negativity, one may for instance think of a proper set of a priori values for (co)variance components (a realistic weight matrix). That is, we try to choose the initial components somewhat close to their actual values. The application of the so-called *almost unbiased estimator* must be useful. Since this estimator, as long as the cofactor matrix Q_k is semi-positive-definite, always results in a non-negative variance component, we can apply this method in the first iteration and

then apply the least squares method. The almost unbiased estimator for the variance component is given by Förstner (1979) (see also section 3.4).

Example 4.8 Consider one unknown quantity that is measured with two different instruments twice, i.e., y_1, y_2 and s_1, s_2. In the following, the design matrix, vector of observations, and two cofactor matrices are given as

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \; y = \begin{bmatrix} y_1 \\ y_2 \\ s_1 \\ s_2 \end{bmatrix}, \; Q_1 = \begin{bmatrix} I_2 & 0 \\ 0 & 0 \end{bmatrix}, \; Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & I_2 \end{bmatrix}. \tag{4.125}$$

Assuming $\sigma_1^0 = 1$ and $\sigma_2^0 = 1$, i.e. $W = I_4$ in equations (4.102) and (4.104), after a few simple mathematical operations, the estimated variance components read

$$\begin{cases} \hat{\sigma}_1 = \frac{1}{6}\{3(y_1^2 + y_2^2) + 2(s_1 s_2 - 2y_1 y_2) - (s_1 + s_2)(y_1 + y_2)\}\,; \\ \hat{\sigma}_2 = \frac{1}{6}\{3(s_1^2 + s_2^2) + 2(y_1 y_2 - 2s_1 s_2) - (s_1 + s_2)(y_1 + y_2)\}\,, \end{cases} \tag{4.126}$$

which are not guaranteed to be positive all the time. Assume for instance observation values $y_1 = 10.0016\,\mathrm{m}$, $y_2 = 10.0009\,\mathrm{m}$, $s_1 = 9.9991\,\mathrm{m}$, and $s_2 = 10.0036\,\mathrm{m}$, which lead to a negative variance component

$$\begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \end{bmatrix} = \begin{bmatrix} -1.48 \\ 8.40 \end{bmatrix} \mathrm{mm}^2\,. \tag{4.127}$$

Note that this is an example which leads to a negative variance component. One can find lots of examples (different y in the above) with the resulting variances both being positive. Also if one increases the number of observations, e.g. y_3, s_3, ..., one will rarely encounter the negative variance problem.

To see the role of the a-priori (co)variance components let us refer to the preceding results. For instance, if we choose the a-priori variance components as $\sigma_1^0 = 1$ and $\sigma_2^0 = 10$, the following variance components can be estimated for the aforementioned sample (first iteration):

$$\begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \end{bmatrix} = \begin{bmatrix} 0.198 \\ 5.463 \end{bmatrix} \mathrm{mm}^2. \tag{4.128}$$

After a few iterations, the converged solution is given as

$$\begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \end{bmatrix} = \begin{bmatrix} 0.235 \\ 5.184 \end{bmatrix} \mathrm{mm}^2, \tag{4.129}$$

which are both positive. Note that the above structure for our model is known as a *disjunctive group* model with one variance component for each group. These components are expected to be positive as the cofactor matrices Q_1 and Q_2 are both positive semi-definite. ■

Non-negativity The (co)variance components are usually estimated in the absence of non-negative variance constraints. But in order to ensure non negative variance components, one may also incorporate non-negativity constraints $\sigma \geq 0$ in equation (4.36). This leads to a non-negative least squares problem. The standard form of the model of observation equations with non-negativity constraints is as follows: $\mathrm{E}\{\underline{y}\} = Ax,\; x \geq 0,\; \mathrm{D}\{\underline{y}\} = Q_y$ (cf. equation (2.4)). There are standard procedures for solving this problem (see e.g. Lawson and Hanson, 1974; Haskell and Hanson, 1981). One simple method to restore

the non-negativeness is to reparameterize the variance components as follows (Teunissen, 1988):

$$\sigma_i = e^{a_i},\ i = 1, \cdots, p', \tag{4.130}$$

where p' is the number of variance components. If one solves for a_i, namely $\hat{a}_i$, the estimated variance components then read

$$\hat{\sigma}_i = e^{\hat{a}_i},\ i = 1, \cdots, p', \tag{4.131}$$

which are forcefully and restrictively positive. The above reparameterization will lead us to a nonlinear (co)variance component model which was explained in section 4.7.3. Note that estimation of the variance components, subject to non-negativity constraints, conflicts with unbiasedness property. But such estimators are definitely of lower minimum variance than those obtained in the absence of the non-negativity constraints because we are using the additional information in a logical way.

4.8.2 Singular stochastic model

As previously mentioned, the (co)variance component model has a unique solution *if and only if* the matrices $B^T Q_1 B, ..., B^T Q_p B$ are linearly independent. If the cofactor matrices $Q_1, ..., Q_p$ are linearly dependent, so are matrices $B^T Q_1 B, ..., B^T Q_p B$ (i.e. rank deficient normal matrix N). But if the cofactor matrices $Q_1, ..., Q_p$ are linearly independent, there is no guarantee for the stochastic model to have a unique solution. Therefore, linear independence of the cofactor matrices $Q_1, ..., Q_p$ is a necessary condition to have a full rank normal matrix in the stochastic model. This is not however the sufficient condition, meaning that matrices $B^T Q_1 B, ..., B^T Q_p B$ can be linearly dependent whereas $Q_1, ..., Q_p$ are linearly independent. To see this, we will consider two examples.

Example 4.9 Assume that we have observed an unknown quantity m times, i.e. a repeated observation. In this case we have the following model of observation equation:

$$\mathsf{E}\{\underline{y}\} = u\,x;\quad \mathsf{D}\{\underline{y}\} = \sigma_1 I + \sigma_2 u u^T, \tag{4.132}$$

where u is the $m \times 1$ *summation vector* containing only ones. It is clear that the cofactor matrices $Q_1 = I$ and $Q_2 = uu^T$ are linearly independent. One can however show that the unknown covariance component σ_2 is not estimable. Given the $m \times 1$ design matrix $A = u$, it follows that the $m \times (m-1)$ matrix B in the model of condition equations reads $B = [u_{m-1}\ -I_{m-1}]^T$. Then $B^T Q_1 B = u_{m-1} u_{m-1}^T + I_{m-1}$ and $B^T Q_2 B = 0$. They are not linearly independent and therefore σ_2 remains inestimable. Note that increasing m can not overcome the singularity. Statistically this makes sense since strictly speaking a mutually constant covariance between observations, namely $Q_2 = uu^T$, can be interpreted as a systematic effect (constant bias) in the data. Such a bias can not be estimated from the functional model of an unknown quantity because $A = [u\ u]$ is also not of full rank (in fact the estimator $\hat{x}$ is biased by the systematic effect). Therefore, if we now try to estimate such a bias in the stochastic model, this also does not help. ■

Let us now consider another simple example that the (co)variance components are not simultaneously estimable.

Example 4.10 Consider that we are measuring the same set of unknown quantities (e.g. distances) with two different instruments which are in general of different precision. If we measure m different unknowns using these two instruments, then the number of unknowns is $n = m$, and the total number of observations would be $2m$. The functional model which relates the observations to the unknown parameters reads

$$E\{\underline{y}\} = Ax, \quad \text{with} \quad \underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}; \quad A = \begin{bmatrix} I \\ I \end{bmatrix}. \tag{4.133}$$

The redundancy of the functional model is $b = 2m - m = m$. In this example we have 2 types of observations for which one needs to estimate a variance component (σ_1 and σ_2), and also a covariance component between $\underline{y}_1$ and $\underline{y}_2$ (σ_3). The cofactor matrices may have the following forms:

$$Q_1 = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}; \quad Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}; \quad Q_3 = \begin{bmatrix} 0 & Q \\ Q & 0 \end{bmatrix}, \tag{4.134}$$

where Q is an arbitrary $m \times m$ symmetric and positive-definite matrix. The cofactor matrices Q_1, Q_2, and Q_3 are linearly independent, but σ_1, σ_2, and σ_3 are not simultaneously estimable. To show this, we need to obtain the matrix B of the model of condition equations. It is not difficult to see that $B = [I \ -I]^T$. After a few simple mathematical operations one gets, $B^T Q_1 B = Q$, $B^T Q_2 B = Q$, and $B^T Q_3 B = -2Q$. This means that the design matrix of the stochastic model, namely A_{vh}, is not of full rank. In fact the 3×3 normal matrix N suffers from a rank deficiency of two. Therefore, two out of three (co)variance components are not estimable; only one (co)variance component is estimable and not three. To handle the singularity of the above, at least, two ways are recommended.

1. assuming $\sigma_3 = 0$ and the relative scale between Q_1 and Q_2 is known (e.g., they are equal $\sigma_1 = \sigma_2$), one can try to estimate a common variance component, namely $Q_y = \sigma_1(Q_1 + Q_2)$;

2. assuming σ_1 and σ_2 are known, we can try to estimate the only covariance component σ_3. If this is the case, one obtains $Q_y = Q_0 + \sigma_3 Q_3$.

Generalization A third way can be to modify the functional and stochastic model by extending the observations plan. Let us consider 3 observation types instead of two kinds of observations, namely y_1, y_2 and y_3. In this case, the VCE model consists of 3 cofactor matrices incorporated with 3 variance components, namely

$$A = \begin{bmatrix} I \\ I \\ I \end{bmatrix}; \ Q_1 = \begin{bmatrix} Q & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ Q_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ Q_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Q \end{bmatrix}. \tag{4.135}$$

The matrix B of the condition equations reads

$$B = \begin{bmatrix} I & I \\ -I & 0 \\ 0 & -I \end{bmatrix}, \tag{4.136}$$

which leads to the matrices

$$B^T Q_1 B = \begin{bmatrix} Q & Q \\ Q & Q \end{bmatrix}; \quad B^T Q_2 B = \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix}; \quad B^T Q_3 B = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}. \tag{4.137}$$

The above listed cofactor matrices are linearly independent and therefore the variance components σ_1, σ_2 and σ_3 are simultaneously estimable. One can simply show that if a covariance component σ_4 is in addition included between any two observation types, the stochastic model becomes singular. This shows that only three components out of six (co)variance components are estimable. To summarize, the above results can be generalized as follows: with r observation types and the above structure for the functional and stochastic model, we have $\frac{r(r+1)}{2}$ unknown (co)variance components in total (r variance and $\frac{r(r-1)}{2}$ covariance components). In this case at maximum only $\frac{r(r-1)}{2}$ out of $\frac{r(r+1)}{2}$ (co)variance components are estimable and the other r components remain to be inestimable. ■

NOTE 4.11 The aforementioned singular examples have nothing to do with the least-squares variance component estimation. This is in fact an intrinsic behavior (property) of the problem itself and is independent of the estimation principle which one applies to the (co)variance component model. To handle the problem of the rank-deficient system of normal equations $N\underline{\hat{\sigma}} = \underline{l}$ some remedies were proposed. To handle the singularity one can also either apply the theory of generalized inverses (see e.g. Rao and Mitra (1971)), or apply the theory of S-transformations (see Baarda, 1973; Teunissen, 1985). ⊡

4.8.3 Ill-posedness of stochastic model

As mentioned, the normal matrix N of the VCE formulas will become singular if and only if the matrices $B^TQ_1B, ..., B^TQ_pB$ are linearly dependent. If the matrix N is strictly rank deficient then the cofactor matrices are linearly dependent. But practically, N, sometimes, is nearly rank-deficient (ill-conditioned) rather than strictly rank deficient. This means, theoretically, that N is of full rank but practically the solution is severely sensitive to round-off and random errors; so the solution is not stable. In this case, usually, one or more functions of the estimators may have very poor precision. To handle the problem, we suggest to compute the *condition number* of N.

For the positive-definite matrix N, the condition number is defined as the ratio $c = \lambda_{\max}/\lambda_{\min}$, with $\lambda_{\max}$ and $\lambda_{\min}$ the maximum and minimum eigenvalues of the matrix, respectively. For such a matrix, the solution $\sigma = N^{-1}l$ and the error $\delta\sigma = N^{-1}\delta l$ always satisfy the inequality (Strang, 1988)

$$\|\delta\sigma\| \leq c\frac{\|\sigma\|}{\|l\|}\|\delta l\|, \quad \text{or} \quad \frac{\|\delta\sigma\|}{\|\sigma\|} \leq c\frac{\|\delta l\|}{\|l\|}, \tag{4.138}$$

where $\|.\|$ denotes the norm of a vector. The above equation, in a relative sense, shows that any error $\|\delta l\|$ in the observations, both rounding and random errors, is amplified into the solution (at maximum) by the condition number. Ideally, for a well-conditioned matrix, the condition number is one, e.g., an identity matrix. This implies that the condition number c is not directly affected by the size of the matrix; so it is a better measure of ill-conditioning than the *determinant*. If c is large, the solution is ill-conditioned and thus the cofactor matrices $B^TQ_1B, ..., B^TQ_pB$ are nearly linearly dependent. It is important to perceive, at times, that the cofactor matrices $Q_1, ..., Q_p$ are strictly independent but the problem is ill-posed. This can be the case when combining non-precise observations with precise ones (e.g. when combining GPS code and phase observations; see chapter 7).

Since the condition number depends on the type of the formulation of the problem, and, in addition, it is an overall measure, meaning that it does not employ all information in an

optimal way, it is suggested to evaluate the *degree of dependency* of the columns of the design matrix $A_{\rm vh}$ in equation (4.15). For this purpose we consider the following partition of the design matrix of the stochastic model:

$$(A_J \vdots a_j);\quad j = 2, 3, \ldots, p,\quad J = \{1, 2, \ldots, j-1\}\,, \tag{4.139}$$

where A_J contains the first $j-1$ columns of the VCE design matrix $A_{\rm vh}$ and a_j is the $j^{\rm th}$ column of the matrix $A_{\rm vh}$ and its dependency from A_J is to be evaluated. Let θ_j be the angle between a_j and $\mathcal{R}(A_J)$. Then, the degree of dependence θ_j between a_j and $\mathcal{R}(A_J)$ can be obtained from the following equation:

$$\sin^2_{\theta_j} = \frac{\overline{a}_j^T Q_{\rm vh}^{-1} \overline{a}_j}{a_j^T Q_{\rm vh}^{-1} a_j},\ j = 2, \ldots, p\,, \tag{4.140}$$

where

$$\overline{a}_j^T Q_{\rm vh}^{-1} \overline{a}_j = a_j^T Q_{\rm vh}^{-1} a_j - a_j^T Q_{\rm vh}^{-1} A_J (A_J^T Q_{\rm vh}^{-1} A_J)^{-1} A_J^T Q_{\rm vh}^{-1} a_j\,. \tag{4.141}$$

The above measure can be simply obtained by some elements of the normal matrix as follows:

$$\sin^2_{\theta_j} = \frac{n_{jj} - N_{jJ} N_{JJ}^{-1} N_{Jj}}{n_{jj}};\quad j = 2, \ldots, p\,. \tag{4.142}$$

The term $N_{JJ} = A_J^T Q_{\rm vh}^{-1} A_J$ is the $(j-1) \times (j-1)$ normal matrix part corresponding to A_J, $N_{Jj} = A_J^T Q_{\rm vh}^{-1} a_j$ is a $j-1$ vector containing the first $j-1$ elements of the $j^{\rm th}$ column of N, and $n_{jj} = a_j^T Q_{\rm vh}^{-1} a_j$ is the $j^{\rm th}$ diagonal element of N. The above equation can be simplified to the following:

$$\cos^2_{\theta_j} = \frac{N_{jJ} N_{JJ}^{-1} N_{Jj}}{n_{jj}},\ j = 2, \ldots, p\,. \tag{4.143}$$

When dealing with two unknown components, $p = 2$, the above equation is reduced to the following simple formula:

$$\cos_\theta = \frac{\pm n_{12}}{\sqrt{n_{11}\, n_{22}}}\,, \tag{4.144}$$

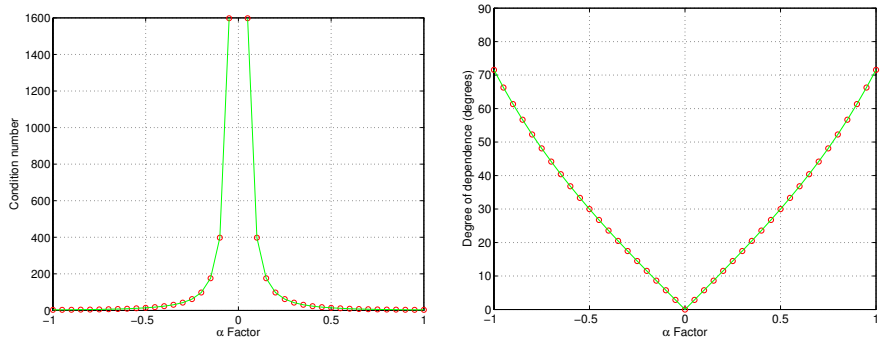
which represents a correlation coefficient ρ. If $\theta = \pi/2$, then a_j is orthogonal to (fully independent of) $\mathcal{R}(A_J)$; i.e. $\rho = 0$. On the other hand, if $\theta = 0$, then a_j belongs to the range space of A_J (fully dependent); i.e. $\rho = \pm 1$. If θ is small, then a_j nearly belongs to $\mathcal{R}(A_J)$ and therefore σ_j is poorly estimable.

Example 4.11 To see how the preceding measures work, in example 4.10 consider $m = 2$ and $\sigma_3 = 0$. If we now assume that the cofactor matrices Q_1 and Q_2 have the form

$$Q_1 = \begin{bmatrix} I_2 & 0 \\ 0 & 0 \end{bmatrix};\quad Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \text{ with } Q = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}, \tag{4.145}$$

with $-1 < \alpha < 1,\ \alpha \neq 0$. Taking the above cofactor matrices into account and assuming the weight matrix as $W = (Q_1 + Q_2)^{-1}$ (i.e. in the first iteration $\sigma_1^0 = \sigma_2^0 = 1$) and after not so short mathematical and matrix operations, the normal matrix of the stochastic model reads

$$N = \frac{1}{(4-\alpha^2)^2} \begin{bmatrix} 4+\alpha^2 & 4-3\alpha^2 \\ 4-3\alpha^2 & 4-3\alpha^2+\alpha^4 \end{bmatrix};\quad -1 < \alpha < 1\,, \tag{4.146}$$



(a) (b)

Figure 4.4: Condition number of the normal matrix (4.146), (a), and degree of dependence between two columns of design matrix $A_{\rm vh}$, (b), versus modifier factor α.

which is of full rank when $\alpha \neq 0$. When $\alpha = 0$, it reduces to example 4.10. Since the non-singular matrix N in equation (4.146) is expressed in terms of parameter α, with $-1 < \alpha < 1,\ \alpha \neq 0$, an analytical expression for the condition number of the normal matrix, after simple but not so short mathematical operations, can be obtained. Figure 4.4a shows the condition number versus α. As seen, the smaller value for α is, the larger value for condition number would be, and then the more ill-posed the problem would be. It is important to notice that when $\alpha = 0$, the condition number would become infinity and the problem is strictly rank deficient. Another way of evaluating the ill-posedness of the problem is based on the other criterion, i.e., the degree of dependency (DoD) of the columns of the design matrix. Applying equation (4.144) to the elements of the normal matrix, the DoD reads

$$\cos\theta = \frac{4 - 3\alpha^2}{\sqrt{(4+\alpha^2)(4-3\alpha^2+\alpha^4)}}\,. \tag{4.147}$$

Figure 4.4b illustrates the degree of dependency of the columns of design matrix versus parameter α, where $-1 < \alpha < 1$. When $\alpha = 0$, θ would be zero as well. This means that the columns of the design matrix of the stochastic model are strictly dependent which leaves the problem to be rank deficient. Also, when α is small, θ would be small too, meaning that the columns of the design matrix are not seriously independent. This will give poor results. When $\alpha = \pm 1$, though both the cofactor matrix Q_2 and the covariance matrix Q_y are rank deficient, the columns of the design matrix $A_{\rm vh}$ are as independent as possible. ■

4.9 Summary and concluding remarks

There are various VCE formulas based on optimality properties as unbiasedness, best (minimum variance), minimum norm, and maximum likelihood. In chapters 2 and 3 we reviewed the principle of all these estimation methods. In this chapter we generalized the method of least-squares for estimating the stochastic model. In this method any symmetric and positive-definite weight matrix can be used (weighted least-squares). The derivation given of the least squares VCE formula is based on the model of condition equations rather than the model of observation equations. The method is easily understood and very flexible indeed. It can be used for estimation of both variance and covariance components

Functional model: $\mathsf{E}\{\underline{y}\} = Ax,\ Q_y$			Stochastic model: $\mathsf{E}\{\underline{y}_{\mathrm{vh}}\} = A_{\mathrm{vh}}\sigma,\ Q_{\mathrm{vh}}$		
$\underline{\hat{x}}$	$=$	$(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y}$	$\underline{\hat{\sigma}}$	$=$	$(A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} A_{\mathrm{vh}})^{-1} A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \underline{y}_{\mathrm{vh}}$
$Q_{\hat{x}}$	$=$	$(A^T Q_y^{-1} A)^{-1}$	$Q_{\hat{\sigma}}$	$=$	$(A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} A_{\mathrm{vh}})^{-1}$
P_A	$=$	$A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$	$P_{A_{\mathrm{vh}}}$	$=$	$A_{\mathrm{vh}}(A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} A_{\mathrm{vh}})^{-1} A_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1}$
$P_A^\perp$	$=$	$I - P_A$	$P_{A_{\mathrm{vh}}}^\perp$	$=$	$I - P_{A_{\mathrm{vh}}}$
$\underline{\hat{y}}$	$=$	$P_A \underline{y}$	$\underline{\hat{y}}_{\mathrm{vh}}$	$=$	$P_{A_{\mathrm{vh}}} \underline{y}_{\mathrm{vh}}$
$Q_{\hat{y}}$	$=$	$P_A Q_y = Q_y P_A^T$	$Q_{\hat{y}_{\mathrm{vh}}}$	$=$	$P_{A_{\mathrm{vh}}} Q_{\mathrm{vh}} = Q_{\mathrm{vh}} P_{A_{\mathrm{vh}}}^T$
$\underline{\hat{e}}$	$=$	$P_A^\perp \underline{y}$	$\underline{\hat{e}}_{\mathrm{vh}}$	$=$	$P_{A_{\mathrm{vh}}}^\perp \underline{y}_{\mathrm{vh}}$
$Q_{\hat{e}}$	$=$	$P_A^\perp Q_y = Q_y P_A^{\perp T}$	$Q_{\hat{e}_{\mathrm{vh}}}$	$=$	$P_{A_{\mathrm{vh}}}^\perp Q_{\mathrm{vh}} = Q_{\mathrm{vh}} P_{A_{\mathrm{vh}}}^{\perp T}$
$\underline{T}_b$	$=$	$\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}}$	$\underline{T}_{\mathrm{df}}$	$=$	$\underline{\hat{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \underline{\hat{e}}_{\mathrm{vh}}$

with:
$\underline{y}_{\mathrm{vh}} = \mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B),\ A_{\mathrm{vh}} = [\mathrm{vh}(B^T Q_1 B), ..., \mathrm{vh}(B^T Q_p B)],\ Q_{\mathrm{vh}} = 2D^+(Q_t \otimes Q_t) D^{+T}$

Table 4.1: Comparison of least-squares estimators as well as their covariance matrices in functional and stochastic model; T is the quadratic form of least-squares residuals.

in the Gauss-Markov (A-model), Gauss-Helmert (mixed model) and condition model (B-model); both for linear and nonlinear stochastic models. Since the approach is based on the least-squares principle we know without any additional derivation that the estimators are *unbiased*. One advantage of this technique over the previous methods of (co)variance component estimation is that the least-squares estimators are *independent* of the distribution of the data.

We derived the minimum variance estimators (BIQUE) by taking the weight matrix as the inverse of covariance matrix of observables. These estimators are however restricted to the class of the multivariate *normal* distribution. The estimators obtained by this technique are therefore of *minimum variance* (in addition to the unbiasedness property). As it turns out, our results for the (co)variance component estimation are identical to the MINQUE, BIQUE, RELM (restricted maximum likelihood) results if a normal distribution is assumed. We have now therefore a more satisfactory derivation of the MINQUE and BIQUE theories. Though, the solution vector of this method is identical to the results of minimum variance solution proposed by Koch (1999) and the minimum norm solution proposed by Rao and Kleffe (1988), the least squares estimators have special and unique properties that we pointed out. For example, the inverse of the normal matrix gives automatically the *covariance* matrix of the estimators. Also, our existing body of knowledge on least-squares theory is directly applicable to LS-VCE. For instance, LS-VCE has a similar insightful (linear and nonlinear) geometric interpretation as standard least-squares. Properties of the normal equations and projector properties are therefore easily established. Table 4.1 summarizes the results of least-squares both in functional model for estimating the parameter vector x and in the stochastic model for estimating the (co)variance components σ.

Since the method is based on the least-squares principle, measures of inconsistency, such as the quadratic form of residuals and the w-test statistic can directly be given. We can also apply testing hypotheses with the stochastic model. This, as an important feature of LS-VCE, is the subject of discussion in the next chapter.

5 Detection and Validation in Stochastic Model

5.1 Introduction

In this chapter we generalize the idea of detection and validation introduced in chapter 2 into the stochastic model. The goal is to find misspecifications in the stochastic model, or to improve an existing covariance matrix description by including more (or excluding) (co)variance components. This can be done by introducing the w-test, the v-test, and the overall model test with the stochastic model. We show that employing the w-test and v-test in the stochastic model is in fact equivalent to the testing of (co)variance components.

We will point out the limitations of using the proposed formulations. Note that because the observables, here in the stochastic model, are in a quadratic form of the original observables $\underline{y}$ or $\underline{t}$, i.e. $\mathrm{vh}(\underline{y}\,\underline{y}^T)$ or $\mathrm{vh}(\underline{t}\,\underline{t}^T)$, the distribution of the test statistics becomes (much) more complicated when compared to the functional model. For example, the overall model test has the most complicated form of the distribution. This chapter is therefore considered as a first attempt for the detection and validation in the stochastic model.

5.2 The w-test statistic

5.2.1 Introduction

The goal of this section is to generalize the square-root of the 1-dimensional T-test statistic or the w-test statistic introduced in chapter 2 to the stochastic model. In the functional model, the w-test statistic is given by equations (2.39) and (2.42). In a similar way to the functional model, it is also possible, for the stochastic model, to obtain the w-test statistic and to generalize the idea of *datasnooping* on the vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$.

The hypotheses that will be considered in this section are all hypotheses on the *mean* of the observable vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ in the stochastic model. Such hypotheses are in fact equivalent to those considered on the (co)variance components of covariance matrix Q_t. From the explicit form of the observation equations of the VCE model, the following two hypotheses are put forward:

$$\mathrm{H}_o\colon \mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = A_{\mathrm{vh}}\sigma \quad \text{versus} \quad \mathrm{H}_a\colon \mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = A_{\mathrm{vh}}\sigma + c_{\mathrm{vh}}\nabla, \quad \nabla \neq 0\,, \tag{5.1}$$

The design matrix A_{vh} is supposed to be of full column rank, i.e. $\mathrm{rank}(A_{\mathrm{vh}}) = p$. Furthermore, vector c_{vh} is independent of the range space of the matrix A_{vh}, i.e. $\mathrm{rank}(A_{\mathrm{vh}} : c_{\mathrm{vh}}) = p + 1$. The scalar ∇ is unknown under H_a. Because both vh and E are linear operators, the above null and alternative hypotheses can be reformulated as

follows:

$$\mathrm{H}_o\colon \mathsf{E}\{\underline{t}\,\underline{t}^T\} = B^T Q_0 B + \sum_{k=1}^{p} \sigma_k B^T Q_k B, \tag{5.2}$$

versus

$$\mathrm{H}_a\colon \mathsf{E}\{\underline{t}\,\underline{t}^T\} = B^T Q_0 B + \sum_{k=1}^{p} \sigma_k B^T Q_k B + C_t \nabla, \quad \nabla \neq 0\,, \tag{5.3}$$

where the $b \times b$ symmetric matrix C_t satisfies $c_{\mathrm{vh}} = \mathrm{vh}(C_t)$. The constant term $B^T Q_0 B$ has been included to keep the generality of the formulation. When there is no misspecification in the functional part of the model, namely when $\mathsf{E}\{\underline{y}\} = Ax$ or $\mathsf{E}\{\underline{t}\} = 0$, it follows that $\mathsf{E}\{\underline{t}\,\underline{t}^T\} = Q_t$ which yields

$$\mathrm{H}_o\colon Q_t = B^T Q_0 B + \sum_{k=1}^{p} \sigma_k B^T Q_k B\,, \tag{5.4}$$

versus

$$\mathrm{H}_a\colon Q_t = B^T Q_0 B + \sum_{k=1}^{p} \sigma_k B^T Q_k B + C_t \nabla, \quad \nabla \neq 0\,. \tag{5.5}$$

The above formulation can be used for finding misspecifications in the covariance matrix of the misclosure vector $\underline{t}$, i.e. Q_t. The preceding hypotheses, in fact, provides us with a powerful tool to judge whether or not additional (co)variance components are likely to be included in the stochastic model. Note, however, that in most Geodesy and GPS applications, one needs to evaluate the misspecification in the covariance matrix of the original observables $\underline{y}$, namely Q_y, rather than Q_t (see section 5.2.5). From now on, for the sake of brevity, we will ignore the constant term $Q_{t_0} = B^T Q_0 B$ in our derivations, but when needed, we will make a few comments on that. Also, the matrices $B^T Q_1 B$, ..., $B^T Q_p B$ are simply denoted as Q_{t_1}, ..., Q_{t_p}. In addition, the statement $\nabla \neq 0$ in the alternative hypothesis will be disregarded (simply note that ∇ is always non-zero).

Carrying through the similarity with the functional model, we will just try to obtain equation (2.42) for the stochastic model and to determine its distribution under H_o assuming the original observables $\underline{y}$ or misclosures $\underline{t}$ to be *normally distributed*. The w-test statistic for the stochastic model reads then (cf. equation (2.42))

$$\underline{w} = \frac{\langle P^{\perp}_{A_{\mathrm{vh}}} \mathrm{vh}(C_t), P^{\perp}_{A_{\mathrm{vh}}} \mathrm{vh}(\underline{t}\,\underline{t}^T)\rangle}{\|P^{\perp}_{A_{\mathrm{vh}}} \mathrm{vh}(C_t)\|_{Q^{-1}_{\mathrm{vh}}}} = \|\underline{\hat{e}}_{\mathrm{vh}}\|_{Q^{-1}_{\mathrm{vh}}} \cos\varphi\,, \tag{5.6}$$

where most of the terms are given in table 4.1. Note that φ is the angle between two vectors $P^{\perp}_{A_{\mathrm{vh}}} \mathrm{vh}(C_t)$ and $P^{\perp}_{A_{\mathrm{vh}}} \mathrm{vh}(\underline{t}\,\underline{t}^T)$. In the next subsection, we will give a simple expression for the preceding formula. The above test statistic has a similar geometrical interpretation to that introduced in section 2.2.4 with the functional model.

5.2.2 Formulation in terms of B-model

To obtain a simple expression for the w-test, one needs to reach a simplified formula for the norm and the inner product (in metric $Q_{\rm vh}^{-1}$) in equation (5.6). For this purpose, let us consider the following theorem:

Theorem 5.1 *Let the $b \times b$ matrices Z_1 and Z_2 be symmetric. Then*

$$\left\langle P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_1)\,,\, P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_2)\right\rangle = \frac{1}{2}{\rm tr}(Z_1 Q_t^{-1} Z_2 Q_t^{-1}) - g^{(1)T} N^{-1} g^{(2)}\,, \tag{5.7}$$

where $\langle .\,,\,.\rangle$ denotes the inner (scalar) product of two vectors in the metric $Q_{\rm vh}^{-1}$ and the p-vector $g^{(1)}$ is given as

$$g_k^{(1)} = \frac{1}{2}{\rm tr}(Z_1 Q_t^{-1} Q_{t_k} Q_t^{-1}), \quad k = 1, 2, ..., p\,, \tag{5.8}$$

and so is $g^{(2)}$ with Z_2. When $Z_2 = Z_1$, the inner product becomes the squared norm (L_2 norm) of $P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_1)$. Therefore

$$\|P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_1)\|_{Q_{\rm vh}^{-1}}^2 = \frac{1}{2}{\rm tr}(Z_1 Q_t^{-1} Z_1 Q_t^{-1}) - g^{(1)T} N^{-1} g^{(1)}\,, \tag{5.9}$$

where $\|.\|_{Q_{\rm vh}^{-1}}$ denotes the norm of a vector in the metric $Q_{\rm vh}^{-1}$. ⊠

Proof. The inner product of vectors $P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_1)$ and $P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_2)$, in the metric $Q_{\rm vh}^{-1}$, is given as

$$\left\langle P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_1)\,,\, P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_2)\right\rangle = {\rm vh}(Z_1)^T P_{A_{\rm vh}}^{\perp T} Q_{\rm vh}^{-1} P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_2)\,. \tag{5.10}$$

Since $P_{A_{\rm vh}}^{\perp T} Q_{\rm vh}^{-1} P_{A_{\rm vh}}^{\perp} = Q_{\rm vh}^{-1} P_{A_{\rm vh}}^{\perp}$, the preceding equation reads

$$\left\langle P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_1)\,,\, P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_2)\right\rangle = {\rm vh}(Z_1)^T Q_{\rm vh}^{-1} P_{A_{\rm vh}}^{\perp}\,{\rm vh}(Z_2)\,. \tag{5.11}$$

If we now substitute the terms ${\rm vh}(.) = D^{+}{\rm vec}(.)$, $Q_{\rm vh}^{-1} = \frac{1}{2} D^T (Q_t^{-1} \otimes Q_t^{-1}) D$, and $P_{A_{\rm vh}}^{\perp} = I - A_{\rm vh} N^{-1} A_{\rm vh}^T Q_{\rm vh}^{-1}$ in the preceding equation, one can leave out, similar to previous derivations of chapter 4, the duplication matrices D and simply prove the theorem. The second part of the theorem is trivial. ▼

Because of the symmetry of the $b \times b$ matrices C_t and $\underline{t}\,\underline{t}^T$, if we now substitute $Z_1 = C_t$ and $Z_2 = \underline{t}\,\underline{t}^T$, it follows from equation (5.6) with theorem 5.1 that

$$\underline{w} = \frac{\underline{w}^n}{w^d} = \frac{\frac{1}{2}{\rm tr}(C_t Q_t^{-1} \underline{t}\,\underline{t}^T Q_t^{-1}) - g^T N^{-1} \underline{l}}{\left[\frac{1}{2}{\rm tr}(C_t Q_t^{-1} C_t Q_t^{-1}) - g^T N^{-1} g\right]^{1/2}}\,, \tag{5.12}$$

with g and $\underline{l}$ respectively as

$$g_k = \frac{1}{2}{\rm tr}(C_t Q_t^{-1} Q_{t_k} Q_t^{-1}); \quad \underline{l}_k = \frac{1}{2}\underline{t}^T Q_t^{-1} Q_{t_k} Q_t^{-1} \underline{t}\,. \tag{5.13}$$

NOTE 5.1 (SUMMATION CONVENTION) In the sequel, summation $\sum$, for the sake of brevity, is sometimes disregarded. This is based on the *summation convention* which states that whenever there arises an expression in which an index occurs twice on the same side of any equation, or term within an equation, it is understood to represent a summation on these repeated indices. A repeated index is called a *summation index*, while an unrepeated index is called a *free index*. For example, the *trace* of a matrix S can simply be denoted as $\mathrm{tr}(S) = s_{ii}$ while the matrix itself is symbolized as $S = s_{ij}$. $\boxdot$

With the summation convention and substituting $\underline{l}_l = \frac{1}{2}\underline{t}^T Q_t^{-1} Q_{t_l} Q_t^{-1}\underline{t}$, the numerator of the w-test statistic in equation (5.12) is rewritten as

$$\begin{aligned}
\underline{w}^n &= \tfrac{1}{2}\underline{t}^T Q_t^{-1} C_t Q_t^{-1}\underline{t} - \tfrac{1}{2} g_k n_{kl}^{-1}\underline{t}^T Q_t^{-1} Q_{t_l} Q_t^{-1}\underline{t} \\
&= \tfrac{1}{2}\underline{t}^T Q_t^{-1} C_t Q_t^{-1}\underline{t} - \tfrac{1}{2}\underline{t}^T Q_t^{-1}[g_k n_{kl}^{-1} Q_{t_l}] Q_t^{-1}\underline{t} \qquad (5.14)\\
&= \underline{t}^T Q_t^{-1}\left(\tfrac{1}{2} C_t - \tfrac{1}{2}[g_k n_{kl}^{-1} Q_{t_l}]\right) Q_t^{-1}\underline{t}\,.
\end{aligned}$$

Substitution of the preceding equation into equation (5.12) yields

$$\boxed{\underline{w} = \underline{t}^T M \underline{t} = \frac{\underline{t}^T Q_t^{-1}\left(\frac{1}{2} C_t - \frac{1}{2}[g_k n_{kl}^{-1} Q_{t_l}]\right) Q_t^{-1}\underline{t}}{\left[\frac{1}{2}\mathrm{tr}(C_t Q_t^{-1} C_t Q_t^{-1}) - g^T N^{-1} g\right]^{1/2}}\,,} \qquad (5.15)$$

with the symmetric matrix M as

$$M = \frac{Q_t^{-1}\left(\frac{1}{2} C_t - \frac{1}{2}[g_k n_{kl}^{-1} Q_{t_l}]\right) Q_t^{-1}}{\left[\frac{1}{2}\mathrm{tr}(C_t Q_t^{-1} C_t Q_t^{-1}) - g^T N^{-1} g\right]^{1/2}}\,, \qquad (5.16)$$

Therefore, the w-test statistic of the stochastic model can be written as a *quadratic form* in the vector $\underline{t}$ which is supposed to be normally distributed. This is an important result since the distribution of quadratic forms can be derived. This can be considered an important issue if one wants to apply hypotheses testing and to compute the critical values of the test. Note that, if a constant term $Q_{t_0} = B^T Q_0 B$ is also included in both null and alternative hypotheses, the numerator of equation (5.15) should be added up with the constant term $-\frac{1}{2}\mathrm{tr}(C_t Q_t^{-1} Q_{t_0} Q_t^{-1})$.

NOTE 5.2 The symmetric matrix C_t should be linearly independent from the matrices Q_{t_1}, ..., Q_{t_p}. Otherwise, the w-test can not be defined since *zero* over *zero* is undefined. To see this, let us assume that C_t is a linear combination of Q_{t_1}, ..., Q_{t_p}

$$C_t = \sum_{i=1}^{p} \alpha_i B^T Q_i B = \sum_{i=1}^{p} \alpha_i Q_{t_i} = \alpha_i Q_{t_i}\,, \qquad (5.17)$$

with α_i some non-zero real numbers. Substituting the preceding equation in equation (5.13) gives $g_k = \frac{1}{2}\mathrm{tr}(\alpha_i Q_{t_i} Q_t^{-1} Q_{t_k} Q_t^{-1}) = \alpha_i n_{ik}$. Substituting $C_t = \alpha_i Q_{t_i}$ and $g_k = \alpha_i n_{ik}$ into the numerator of equation (5.15), namely $\underline{w}^n$, gives

$$\underline{w}^n = \frac{1}{2}\alpha_i \underline{t}^T Q_t^{-1}\left(Q_{t_i} - [n_{ik} n_{kl}^{-1} Q_{t_l}]\right) Q_t^{-1}\underline{t}, \qquad (5.18)$$

in which $n_{ik}n_{kl}^{-1}$, multiplication of the $i^{\text{th}},\ i=1,...,p$ row of N with the $l^{\text{th}},\ l=1,...,p$ column of N^{-1}, is the well-known Kronecker delta, i.e.

$$n_{ik}n_{kl}^{-1}=\delta_{il}=\begin{cases}1 & \text{if} \quad i=l\,;\\ 0 & \text{if} \quad i\neq l\,.\end{cases} \tag{5.19}$$

Substituting $n_{ik}n_{kl}^{-1}=\delta_{il}$ in equation (5.18) yields

$$\underline{w}^n=\frac{1}{2}\alpha_i\underline{t}^TQ_t^{-1}\left(Q_{t_i}-Q_{t_i}\right)Q_t^{-1}\underline{t}=0\,. \tag{5.20}$$

In an analogous way, one can show that the denominator of equation (5.15) simplifies to

$$w^d=\left[\alpha^TN\alpha-\alpha^TN\alpha\right]^{1/2}=0\,, \tag{5.21}$$

which proves the claim. In the next subsections, we show how to formulate matrix C_t. ⊡

Special case 5.1 (1) In some applications the covariance matrix is known under the null hypothesis, i.e. $\mathrm{H}_o: Q_t=Q_{t_0}$ versus $\mathrm{H}_o: Q_t=Q_{t_0}+C_t\nabla$. Therefore, there is no unknown in the stochastic model to estimate. In this case, the normal matrix N is empty. The w-test statistic can therefore be given from equation (5.15) when the constant term Q_{t_0} is included in the formulation

$$\underline{w}=\frac{\frac{1}{2}\left[\underline{t}^TQ_t^{-1}C_tQ_t^{-1}\underline{t}-\mathrm{tr}(C_tQ_t^{-1})\right]}{\left[\frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}C_tQ_t^{-1})\right]^{1/2}}\,, \tag{5.22}$$

which has in fact a very simple form. The above w-test may be written as $\underline{w}=\underline{t}^TM\underline{t}-m_0$ with

$$M=\frac{\frac{1}{2}Q_t^{-1}C_tQ_t^{-1}}{\left[\frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}C_tQ_t^{-1})\right]^{1/2}}\,;\quad m_0=\frac{\frac{1}{2}\mathrm{tr}(C_tQ_t^{-1})}{\left[\frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}C_tQ_t^{-1})\right]^{1/2}}\,. \tag{5.23}$$

The constant m_0 only changes the mean of the statistic $\underline{t}^TM\underline{t}$ (just shifts it into the origin). ■

Special case 5.2 (2) Another special case of the w-test occurs when only one variance component is involved in the stochastic model, $p=1$. The null and alternative hypotheses are given as

$$\mathrm{H}_o\colon Q_t=\sigma_1Q_{t_1}\ \text{ versus }\ \mathrm{H}_a\colon Q_t=\sigma_1Q_{t_1}+C_t\nabla\,, \tag{5.24}$$

with $\sigma_1=\sigma^2$ the variance of unit weight. Under the null hypothesis one obtains

$$Q_y=\sigma^2Q_1\ \text{ and }\ Q_t^{-1}=\frac{1}{\sigma^2}(B^TQ_1B)^{-1}=\frac{1}{\sigma^2}Q_{t_1}^{-1}\,. \tag{5.25}$$

The scalars $\underline{l}$, g, and N read then (note that $p=1$)

$$\begin{aligned}\underline{l}\ &=\ \underline{l}_1\ &=\ \tfrac{1}{2}\underline{t}^TQ_t^{-1}Q_{t_1}Q_t^{-1}\underline{t}\ &=\ \tfrac{1}{2\sigma^2}\underline{t}^TQ_t^{-1}\underline{t}\,;\\ g\ &=\ g_1\ &=\ \tfrac{1}{2}\mathrm{tr}(C_tQ_t^{-1}Q_{t_1}Q_t^{-1})\ &=\ \tfrac{1}{2\sigma^2}\mathrm{tr}(C_tQ_t^{-1})\,;\\ N\ &=\ n_{11}\ &=\ \tfrac{1}{2}\mathrm{tr}(Q_t^{-1}Q_{t_1}Q_t^{-1}Q_{t_1})\ &=\ \tfrac{b}{2\sigma^4}\,.\end{aligned} \tag{5.26}$$

Substituting g and N in equation (5.15) yields the w-test as

$$\underline{w}=\underline{t}^TM\underline{t}=\frac{\underline{t}^TQ_t^{-1}\left[\frac{1}{2}C_t-\frac{\mathrm{tr}(C_tQ_t^{-1})}{2b}Q_t\right]Q_t^{-1}\underline{t}}{\left[\frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}C_tQ_t^{-1})-\frac{1}{2b}\mathrm{tr}(C_tQ_t^{-1})\mathrm{tr}(C_tQ_t^{-1})\right]^{1/2}}\,. \tag{5.27}$$

In the sequel, we will refer to these two special cases several times. ■

5.2.3 Distribution of w-test statistic

In order to derive the distribution of the w-test statistic for the stochastic model, we need the distribution of $\underline{w} = \underline{t}^T M \underline{t}$ given in the previous subsection. It is based on the following theorem, the distribution of *quadratic forms* in normal variables. We closely but not exactly follow Teunissen (1988).

Theorem 5.2 *Let the $b \times 1$ vector $\underline{t}$ be normally distributed with mean $\mathsf{E}\{\underline{t}\} = 0$ and positive definite covariance matrix Q_t. Let M be a symmetric matrix of order b. Then there exists a diagonal matrix $\Lambda_r = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r)$ such that*

$$\underline{t}^T M \underline{t} = \underline{z}^T \Lambda_r \underline{z} = \sum_{i=1}^{r} \lambda_i \underline{z}_i^2 \,, \tag{5.28}$$

where $\underline{z}$ has the standard normal distribution, i.e. $\underline{z} \sim \mathsf{N}(0, I_r)$. The number r is the rank of M or MQ_t. The diagonal elements of Λ_r are the r eigenvalues of MQ_t or Q_tM. ⊠

Proof. If we define the random vector $\underline{x} = Q_t^{-1/2}\underline{t}$, then clearly $\underline{x}$ has a standard normal distribution, i.e. $\underline{x} \sim \mathsf{N}(0, I_b)$. Substitution of $\underline{t} = Q_t^{1/2}\underline{x}$ in $\underline{t}^T M \underline{t}$ gives

$$\underline{t}^T M \underline{t} = \underline{x}^T Q_t^{1/2} M Q_t^{1/2} \underline{x} \,. \tag{5.29}$$

Since the matrix $Q_t^{1/2} M Q_t^{1/2}$ is *symmetric*, it has real-valued eigenvalues (not necessarily positive) and corresponding orthonormal eigenvectors. In case M is of full rank ($b = r$), if we collect the b-number of non-zero eigenvalues in the $b \times b$ diagonal matrix Λ and the corresponding orthonormal eigenvectors as columns in the $b \times b$ matrix U then

$$Q_t^{1/2} M Q_t^{1/2} = U \Lambda U^T \,, \tag{5.30}$$

with

$$U^T U = U U^T = I_b \,. \tag{5.31}$$

If $\mathrm{rank}(Q_t^{1/2} M Q_t^{1/2}) = r < b$, then r-number of eigenvalues are non-zero and $(b-r)$-number of eigenvalues are zero. We may therefore partition equation (5.30) as

$$Q_t^{1/2} M Q_t^{1/2} = [\, U_r \, U_{b-r}] \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^T \\ U_{b-r}^T \end{bmatrix} = U_r \Lambda_r U_r^T \,, \tag{5.32}$$

with

$$U_r^T U_r = I_r \tag{5.33}$$

Substitution of equation (5.32) into equation (5.29) gives

$$\underline{t}^T M \underline{t} = \underline{x}^T U_r \Lambda_r U_r^T \underline{x} = \underline{z}^T \Lambda_r \underline{z} \,, \tag{5.34}$$

with

$$\underline{z} = U_r^T \underline{x} = U_r^T Q_t^{-1/2} \underline{t} \,. \tag{5.35}$$

Since $\underline{x}$ is distributed as $\underline{x} \sim \mathsf{N}(0, I_b)$, and $U_r^T U_r = I_r$, it follows that $\underline{z} = U_r^T \underline{x}$ is distributed as $\underline{z} \sim \mathsf{N}(0, I_r)$. Note that since $|Q_t^{1/2} M Q_t^{1/2} - \lambda I_b| = |MQ_t - \lambda I_b| = |Q_t M - \lambda I_b|$, the eigenvalues of $Q_t^{1/2} M Q_t^{1/2}$, MQ_t and $Q_t M$ are the same. ▼

COROLLARY 5.1 The above theorem says that $\underline{t}^T M \underline{t}$ is distributed as a linear combination of r independent central χ^2-distributions each with 1 degree of freedom, i.e.

$$\underline{t}^T M \underline{t} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0)\,. \tag{5.36}$$

The mean and variance of a *central* χ^2-distribution with 1 degree of freedom is 1 and 2, respectively. This means that the mean and variance of $\underline{t}^T M \underline{t}$ follow as

$$\mathsf{E}\{\underline{t}^T M \underline{t}\} = \sum_{i=1}^{r} \lambda_i = \mathrm{tr}(MQ_t); \quad \mathsf{D}\{\underline{t}^T M \underline{t}\} = 2\sum_{i=1}^{r} \lambda_i^2 = 2\mathrm{tr}(MQ_tMQ_t)\,, \tag{5.37}$$

respectively. If all the eigenvalues of MQ_t *equal* 1, i.e. if $M = Q_t^{-1}$, then the quadratic form is distributed as a central χ^2-distribution with b degrees of freedom

$$\underline{t}^T M \underline{t} \sim \chi^2(b,0) \;\text{ if }\; \Lambda_r = I_b\,, \tag{5.38}$$

with the mean and variance of $\mathsf{E}\{\underline{t}^T M \underline{t}\} = b$ and $\mathsf{D}\{\underline{t}^T M \underline{t}\} = 2b$, respectively. ⊠

NOTE 5.3 It is well-known, in the functional model and under the null hypothesis, that the mean and variance of the w-test statistic are *zero* and *one*, respectively. Under the alternative hypothesis, however, only the mean changes and not the variance. We will show that the mean and the variance of the w-test statistic in the stochastic model, under the null hypothesis and with the normality assumption, is also zero and one, respectively. This conclusion can however not be drawn when either of these assumptions is violated, i.e. when the original observables are not normally distributed or when, for instance, the alternative hypothesis is correct. In the following, for the sake of verification and certainty, we will derive the first two moments of this statistic under the null hypothesis. In the sequel, it is assumed that the observable vector $\underline{y}$ and therefore the residual vector $\underline{\hat{e}}$ as well as the misclosure vector $\underline{t}$ are *normally distributed*. ⊡

Mean For computing the expectation of the w-test statistic in the stochastic model, one needs to compute the expectation of the numerator in equation (5.15). Using equation (5.37) in which M is the numerator of equation (5.16), one obtains (see NOTE 5.1)

$$\begin{array}{rcl} \mathsf{E}\{\underline{w}^n\} &=& \frac{1}{2}\mathrm{tr}(Q_t^{-1}(C_t - [g_k n_{kl}^{-1} Q_{t_l}])) \\ &=& \frac{1}{2}\mathrm{tr}(Q_t^{-1}C_t) - \frac{1}{2}\mathrm{tr}(Q_t^{-1}[g_k n_{kl}^{-1} Q_{t_l}]) \\ &=& \frac{1}{2}\mathrm{tr}(Q_t^{-1}C_t) - \frac{1}{2}[g_k n_{kl}^{-1}]\mathrm{tr}(Q_t^{-1}Q_{t_l}) \\ &=& \frac{1}{2}\mathrm{tr}(Q_t^{-1}C_t) - g_k n_{kl}^{-1} l_l\,, \end{array} \tag{5.39}$$

where

$$l_l = \mathsf{E}\{\underline{l}_l\} = \frac{1}{2}\mathrm{tr}(Q_t^{-1}Q_{t_l})\,. \tag{5.40}$$

Since $\underline{\hat{\sigma}}$ is as an *unbiased* estimator of σ, it follows that $\mathsf{E}\{\underline{\hat{\sigma}}\} = \mathsf{E}\{N^{-1}\underline{l}\} = N^{-1}l = \sigma$. This with equation (5.39) yields

$$\begin{array}{rcl} \mathsf{E}\{\underline{w}^n\} &=& \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}) - g_k\sigma_k \\ &=& \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}) - \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}Q_{t_k}Q_t^{-1})\sigma_k \\ &=& \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}) - \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}[\sigma_k Q_{t_k}]Q_t^{-1}) \\ &=& \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}) - \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}Q_tQ_t^{-1}) \\ &=& \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}) - \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}) = 0\,. \end{array} \tag{5.41}$$

Therefore, from equations (5.41) and (5.15), it can simply be concluded that the expectation of the w-test statistic is zero, namely

$$\boxed{\mathrm{E}\{\underline{w}\} = 0 \;\; \text{under} \;\; \mathrm{H}_o\,.} \tag{5.42}$$

Variance The variance of the w-test statistic in the stochastic model is given as

$$\mathrm{D}\{\underline{w}\} = \frac{\mathrm{D}\{\underline{w}^n\}}{(w^d)^2}\,. \tag{5.43}$$

Using equation (5.37), the variance of $\underline{w}^n$ reads (note that M here is again the numerator of equation (5.16))

$$\begin{array}{rcl} \mathrm{D}\{\underline{w}^n\} & = & \frac{1}{2}\mathrm{tr}(Q_t^{-1}C_tQ_t^{-1}C_t) - \mathrm{tr}(Q_t^{-1}C_tQ_t^{-1}[g_in_{ij}^{-1}Q_{t_j}]) \\ & + & \frac{1}{2}\mathrm{tr}(Q_t^{-1}[g_in_{ij}^{-1}Q_{t_j}]Q_t^{-1}[g_kn_{kl}^{-1}Q_{t_j}])\,. \end{array} \tag{5.44}$$

The second term of equation (5.44) can be rewritten as

$$\mathrm{tr}(Q_t^{-1}C_tQ_t^{-1}[g_in_{ij}^{-1}Q_{t_j}]) = g_in_{ij}^{-1}\mathrm{tr}(Q_t^{-1}C_tQ_t^{-1}Q_{t_j}) = 2g_in_{ij}^{-1}g_j = 2g^TN^{-1}g\,. \tag{5.45}$$

The third term in equation (5.44) reads

$$\tfrac{1}{2}\mathrm{tr}(Q_t^{-1}[g_in_{ij}^{-1}Q_{t_j}]Q_t^{-1}[g_kn_{kl}^{-1}Q_{t_l}]) = \tfrac{1}{2}g_in_{ij}^{-1}g_kn_{kl}^{-1}\mathrm{tr}(Q_t^{-1}Q_{t_j}Q_t^{-1}Q_{t_l})\,, \tag{5.46}$$

or

$$\begin{array}{rcl} \frac{1}{2}\mathrm{tr}(Q_t^{-1}[g_in_{ij}^{-1}Q_{t_j}]Q_t^{-1}[g_kn_{kl}^{-1}Q_{t_l}]) & = & g_in_{ij}^{-1}g_kn_{kl}^{-1}n_{jl} \;=\; g_i\delta_{il}g_kn_{kl}^{-1} \\ & = & g_ig_kn_{ki}^{-1} \;=\; g^TN^{-1}g\,. \end{array} \tag{5.47}$$

Substitution of equations (5.47) and (5.45) into equation (5.44) gives

$$\mathrm{D}\{\underline{w}^n\} = \frac{1}{2}\mathrm{tr}(C_tQ_t^{-1}C_tQ_t^{-1}) - g^TN^{-1}g\,. \tag{5.48}$$

From equations (5.48) and (5.15), the equation (5.43) yields

$$\boxed{\mathrm{D}\{\underline{w}\} = 1 \;\; \text{under} \;\; \mathrm{H}_o\,.} \tag{5.49}$$

It is important to perceive that the preceding result on the mean and variance of the w-test statistic is only valid as long as the true covariance matrix of the observations, namely Q_y or Q_t (and therefore the true Q_{vh}) has been used in equation (5.15).

Distribution From theorem 5.2, it can be concluded that the w-test statistic (5.15), under the null hypothesis H_o, i.e. $\mathrm{E}\{\underline{t}\} = 0$ and $Q_t = \sum_{k=1}^{p}\sigma_kB^TQ_kB$, is distributed as[1]

$$\boxed{\underline{w} \sim \sum_{i=1}^{r}\lambda_i\chi_i^2(1,0) \;\; \text{under} \;\; \mathrm{H}_o\,,} \tag{5.50}$$

[1] just for notational convenience; one may also write $\underline{w} = \sum_{i=1}^{r}\lambda_i\underline{\chi}_i^2$ where $\underline{\chi}_i^2 \sim \chi^2(1,0)$.

where the $\chi_i^2(1,0)$ are mutually independent chi-squared distributions with a degree of freedom of one. The non-zero eigenvalues λ_i of $Q_t M$, with M expressed in equation (5.16), can be obtained as

$$\left| \frac{\left(\frac{1}{2}C_t - \frac{1}{2}[g_k n_{kl}^{-1} Q_{t_l}]\right) Q_t^{-1}}{\left[\frac{1}{2}\mathrm{tr}(C_t Q_t^{-1} C_t Q_t^{-1}) - g^T N^{-1} g\right]^{1/2}} - \lambda I_b \right| = 0\,, \tag{5.51}$$

or

$$\boxed{\left| C_t Q_t^{-1} - [g_k n_{kl}^{-1} Q_{t_l}] Q_t^{-1} - 2\lambda w^d I_b \right| = 0\,.} \tag{5.52}$$

where w^d is the denominator of equation (5.16). One can show in fact that $\sum_k \lambda_k = 0$ and $\sum_k = 2\lambda_k^2 = 1$ (see equations (5.42) and (5.49), respectively).

Special case 5.3 (1) Let us refer to the case of known covariance matrix under the null hypothesis, i.e. $Q_t = Q_{t_0}$. The w-test, in this case, is given by equation (5.22): $\underline{w} = \underline{t}^T M \underline{t} - m_0$. From theorem 5.2, one obtains

$$\underline{w} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0) - m_0\,, \tag{5.53}$$

with λ_i the non-zero eigenvalues of $Q_t M$, where M is given by equation (5.23). They can therefore be obtained from the following equation:

$$\left| C_t Q_t^{-1} - 2\lambda w^d I_b \right| = 0\,, \tag{5.54}$$

with w^d the denominator of equation (5.23), namely $w^d = \left[\frac{1}{2}\mathrm{tr}(C_t Q_t^{-1} C_t Q_t^{-1})\right]^{1/2}$. ■

Special case 5.4 (2) When dealing only with the variance of unit weight, i.e. when $p = 1$ or $Q_t = \sigma^2 Q_{t_1}$ (under H_o), the w-test statistic is given by equation (5.27). The distribution again is given by $\underline{w} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0)$ with λ_i the non-zero eigenvalues of the $Q_t M$, where M is obtained from equation (5.27). After a few simple mathematical operations it follows that

$$\left| b C_t Q_t^{-1} - [2\lambda b w^d + \mathrm{tr}(C_t Q_t^{-1})] I_b \right| = 0\,. \tag{5.55}$$

It is important to note that the eigenvalues λ of equation (5.55) are independent of the unknown variance component σ^2 (it can be canceled out from both terms in equation (5.55)). However, this is not the case for computing the value of the w-test statistic itself. ■

5.2.4 Datasnooping on Q_t

In this section, we generalize the idea of data snooping for the stochastic model, i.e. data snooping on vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ corresponding to the elements of Q_t, the (co)variances, since $\mathrm{E}\{\underline{t}_i \underline{t}_j\} = \sigma_{t_i t_j}$. For this purpose, the $b \times b$ symmetric matrix C_t takes the following form:

$$C_t = C_{ij} = c_i c_j^T + c_j c_i^T, \quad i \neq j\,, \tag{5.56}$$

with $c_i = (0 \ldots 1 \ldots 0)^T$, as before, the canonical unit vector. The preceding equation refers to a single couple of entries in vector $\mathrm{vec}(\underline{t}\,\underline{t}^T)$ or equivalently just to a single entry in

vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. For the sake of unification, the above structure can also be used for $i = j$, i.e. $C_{ii} = 2c_i c_i^T$. With these in mind, the null and alternative hypotheses are formulated as

$$\mathrm{H}_o\colon Q_t = \sum_{k=1}^{p} \sigma_k Q_{t_k} \quad \text{versus} \quad \mathrm{H}_a\colon Q_t = \sum_{k=1}^{p} \sigma_k Q_{t_k} + C_{ij}\nabla, \quad \nabla \neq 0\,. \tag{5.57}$$

Substituting $C_t = C_{ij} = c_i c_j^T + c_j c_i^T$ into equation (5.15), after a few mathematical operations, yields

$$\boxed{\underline{w}_{ij} = \frac{\underline{t}^T Q_t^{-1}\left(\frac{1}{2}C_{ij} - \frac{1}{2}[g_k^{(ij)} n_{kl}^{-1} Q_{t_l}]\right) Q_t^{-1}\underline{t}}{\left[(q_{ij}^{-1})^2 + q_{ii}^{-1}\, q_{jj}^{-1} - g^{(ij)T} N^{-1} g^{(ij)}\right]^{1/2}}}\,, \tag{5.58}$$

where $q_{ij}^{-1} = q_{t_i t_j}^{-1} = c_i^T Q_t^{-1} c_j$ and vector $g^{(ij)}$ is obtained from equation (5.13) as

$$g_k^{(ij)} = \frac{1}{2}\mathrm{tr}(C_{ij} Q_t^{-1} Q_{t_k} Q_t^{-1}) = c_i^T Q_t^{-1} Q_{t_k} Q_t^{-1} c_j\,. \tag{5.59}$$

In this case, the non-zero eigenvalues λ associated with the distribution of the w-test statistic in equation (5.50) follow as

$$\boxed{\left|C_{ij} Q_t^{-1} - [g_k^{(ij)} n_{kl}^{-1} Q_{t_l}] Q_t^{-1} - 2\lambda w^d I_b\right| = 0\,,} \tag{5.60}$$

where w^d is the denominator of w-test in equation (5.58).

5.2.5 Formulation in terms of A-model

In most Geodesy and GPS applications, one needs to evaluate the misspecification in terms of the covariance matrix of the original observables $\underline{y}$, namely Q_y, rather than Q_t. One may thus retrieve the formulation in Q_y from the one in Q_t. For this purpose, $\underline{t} = B^T\underline{y}$ implies that $C_t = B^T C_y B$. With this, the hypotheses (5.4) and (5.5) then read (note that the constant term $Q_{t_0} = B^T Q_0 B$ has been left out)

$$\mathrm{H}_o\colon B^T Q_y B = \sum_{k=1}^{p} \sigma_k B^T Q_k B, \text{ versus } \mathrm{H}_a\colon B^T Q_y B = \sum_{k=1}^{p} B^T[\sigma_k Q_k + C_y\nabla]B, \tag{5.61}$$

The above hypotheses cannot directly be rewritten in terms of Q_y. In other words, it is not allowed to just delete the matrix B from both sides of the above hypotheses, as we deal with an *expansion* in dimension of the system in which we need some more information for that to be uniquely specified. As an example, if Q_y is a solution of one of the above equations, then $Q_y + AQ_xA^T$ would be a solution as well, since $B^TA = 0$, with A the design matrix of the model of observation equations and Q_x an arbitrary symmetric matrix of size n. In the simplest case, i.e. when $Q_x = 0$, one obtains

$$\mathrm{H}_o\colon Q_y = \sum_{k=1}^{p} \sigma_k Q_k, \quad \text{versus} \quad \mathrm{H}_a\colon Q_y = \sum_{k=1}^{p} \sigma_k Q_k + C_y\nabla. \tag{5.62}$$

One can also argue that such a formulation for the above null and alternative hypotheses could directly be assumed from beginning.

As previously mentioned, the matrix C_t through $c_{\rm vh} = {\rm vh}(C_t)$ plays the role of c_y of the functional model, in testing the stochastic model (see equations (5.1) and (5.3)). It is assumed that this matrix is symmetric and its *vector*, i.e. ${\rm vh}(C_t)$, is independent of the columns of the design matrix $A_{\rm vh}$ in the stochastic model. In the previous sections, the problem was formulated in terms of the model of *condition* equations (B-model). This formulation is not always advantageous, since the misspecification in the matrix Q_t may not be properly interpretable. However, for better interpretation of the testing results, it is possible to formulate the w-test statistic in terms of the model of *observation* equations (A-model). To see this, let us rewrite the matrix C_t in equation (5.15) as follows:

$$C_t = B^T C_y B\,, \tag{5.63}$$

in which the matrix C_y is of order m, the size of observations. A necessary (and not sufficient) condition that the symmetric matrix C_y should fulfill is its independence from any linear combination of the cofactor matrices Q_1, ..., Q_p (cf. NOTE 4.2)

$$C_y \neq \sum_{k=1}^{p} \alpha_k Q_k = \alpha_k Q_k\,. \tag{5.64}$$

Substituting $\underline{t} = B^T \underline{y}$ and C_t from equation (5.63) into equation (5.15) yields

$$\underline{w} = \frac{\underline{y}^T B Q_t^{-1} B^T \left(\frac{1}{2} C_y - \frac{1}{2}[g_k n_{kl}^{-1} Q_l]\right) B Q_t^{-1} B^T \underline{y}}{\left[\frac{1}{2}{\rm tr}(B^T C_y B Q_t^{-1} B^T C_y B Q_t^{-1}) - g^T N^{-1} g\right]^{1/2}}\,. \tag{5.65}$$

Using the identity ${\rm tr}(EF) = {\rm tr}(FE)$ and the orthogonal projector $P_A^{\perp} = P_{Q_y B} = Q_y B Q_t^{-1} B^T$ (see equation (2.21)) or equivalently $B Q_t^{-1} B^T = Q_y^{-1} P_A^{\perp}$, one obtains

$$\underline{w} = \frac{\underline{y}^T Q_y^{-1} P_A^{\perp} \left(\frac{1}{2} C_y - \frac{1}{2}[g_k n_{kl}^{-1} Q_l]\right) Q_y^{-1} P_A^{\perp} \underline{y}}{\left[\frac{1}{2}{\rm tr}(C_y Q_y^{-1} P_A^{\perp} C_y Q_y^{-1} P_A^{\perp}) - g^T N^{-1} g\right]^{1/2}}\,. \tag{5.66}$$

Denoting $Q_y^{-1} P_A^{\perp} = P_A^{\perp T} Q_y^{-1} = P_A^{\perp T} Q_y^{-1} P_A^{\perp} = Q_y^{-1} Q_{\hat{e}} Q_y^{-1}$ as $Q_{\hat{e}}^{-}$, the preceding equation reads

$$\boxed{\underline{w} = \frac{\underline{y}^T Q_{\hat{e}}^{-} \left(\frac{1}{2} C_y - \frac{1}{2}[g_k n_{kl}^{-1} Q_l]\right) Q_{\hat{e}}^{-} \underline{y}}{\left[\frac{1}{2}{\rm tr}(C_y Q_{\hat{e}}^{-} C_y Q_{\hat{e}}^{-}) - g^T N^{-1} g\right]^{1/2}}\,,} \tag{5.67}$$

expressed as a *quadratic form* of the *observation* vector $\underline{y}$. Note that $Q_{\hat{e}}^{-}$ is in fact the reflexive inverse of $Q_{\hat{e}} = P_A^{\perp} Q_y$. Using the least-squares (BLUE) residual vector $\underline{\hat{e}} = P_A^{\perp} \underline{y}$, equation (5.66) yields

$$\boxed{\underline{w} = \frac{\underline{\hat{e}}^T Q_y^{-1} \left(\frac{1}{2} C_y - \frac{1}{2}[g_k n_{kl}^{-1} Q_l]\right) Q_y^{-1} \underline{\hat{e}}}{\left[\frac{1}{2}{\rm tr}(C_y Q_{\hat{e}}^{-} C_y Q_{\hat{e}}^{-}) - g^T N^{-1} g\right]^{1/2}}\,,} \tag{5.68}$$

expressed as a *quadratic form* of the *residual* vector $\underline{\hat{e}}$. In a similar way, the elements of the vector g, in terms of the model of observation equations, read

$$g_k = \frac{1}{2}{\rm tr}(C_y Q_{\hat{e}}^{-} Q_k Q_{\hat{e}}^{-})\,. \tag{5.69}$$

As before, the w-test statistic is distributed as $\underline{w} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0)$. From equation (5.52), the eigenvalues λ then read

$$\left| B^T C_y B Q_t^{-1} - B^T [g_k n_{kl}^{-1} Q_l] B Q_t^{-1} - 2\lambda w^d I_b \right| = 0\,. \tag{5.70}$$

The preceding equation can be rewritten as

$$(-\lambda)^{b-m} \left| C_y B Q_t^{-1} B^T - [g_k n_{kl}^{-1} Q_l] B Q_t^{-1} B^T - 2\lambda w^d I_m \right| = 0\,. \tag{5.71}$$

For $\lambda \neq 0$, one obtains

$$\left| C_y B Q_t^{-1} B^T - [g_k n_{kl}^{-1} Q_l] B Q_t^{-1} B^T - 2\lambda w^d I_m \right| = 0\,, \tag{5.72}$$

or

$$\boxed{\left| C_y Q_{\hat{e}}^{-} - [g_k n_{kl}^{-1} Q_l] Q_{\hat{e}}^{-} - 2\lambda w^d I_m \right| = 0\,,} \tag{5.73}$$

where

$$w^d = \left[\frac{1}{2} \mathrm{tr}(C_y Q_{\hat{e}}^{-} C_y Q_{\hat{e}}^{-}) - g^T N^{-1} g \right]^{1/2}, \tag{5.74}$$

is the denominator of the w-test statistic.

Special case 5.5 (1) When the covariance matrix is known under the null hypothesis, i.e. H_o: $Q_y = Q_0$ versus H_a: $Q_y = Q_0 + C_y \nabla$, the w-test statistic simplifies to

$$\underline{w} = \frac{\frac{1}{2} \underline{y}^T Q_{\hat{e}}^{-} C_y Q_{\hat{e}}^{-} \underline{y} - \frac{1}{2} \mathrm{tr}(C_y Q_{\hat{e}}^{-})}{\left[\frac{1}{2} \mathrm{tr}(C_y Q_{\hat{e}}^{-} C_y Q_{\hat{e}}^{-}) \right]^{1/2}} = \frac{\frac{1}{2} \underline{\hat{e}}^T Q_y^{-1} C_y Q_y^{-1} \underline{\hat{e}} - \frac{1}{2} \mathrm{tr}(C_y Q_{\hat{e}}^{-})}{\left[\frac{1}{2} \mathrm{tr}(C_y Q_{\hat{e}}^{-} C_y Q_{\hat{e}}^{-}) \right]^{1/2}} = \underline{\hat{e}}^T M \underline{\hat{e}} - m_0. \tag{5.75}$$

The preceding w-test is distributed as $\underline{w} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0) - m_0$, where the eigenvalues λ associated with the distribution are obtained as

$$\left| C_y Q_{\hat{e}}^{-} - 2\lambda w^d I_m \right| = 0\,, \tag{5.76}$$

where w^d is the denominator of equation (5.75). ■

Example 5.1 (Known variance) As a simple application of the above special case, let us assume that we want to test the only variance component of unit weight. For such a simple case, the following null and alternative hypotheses are put forward:

$$\mathrm{H}_o\colon Q_y = \sigma_0^2 Q_1 \quad \text{versus} \quad \mathrm{H}_a\colon Q_y = (\sigma_0^2 + \nabla) Q_1, \tag{5.77}$$

with the known variance component σ_0^2. This structure is in fact the same as the above special case when $Q_0 = \sigma_0^2 Q_1$ and $C_y = Q_1$. After a few simple mathematical operations, the w-test statistic given by equation (5.75) and its distribution given as $\underline{w} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0) - m_0$ simplify to (note that $\mathrm{tr}(P_A^{\perp}) = m - n = b$ and b-number of eigenvalues of $P_A^{\perp}$ are equal to one)

$$\underline{w} = \frac{1}{\sqrt{2b}} \left(\frac{\underline{\hat{e}}^T Q_1^{-1} \underline{\hat{e}}}{\sigma_0^2} - b \right) \quad \text{and} \quad \underline{w} \sim \frac{1}{\sqrt{2b}} \left(\chi^2(b,0) - b \right). \tag{5.78}$$

Denoting $\underline{\hat{e}}^T Q_1^{-1} \underline{\hat{e}} = b\, \underline{\hat{\sigma}}^2$, the above equation reads

$$\frac{\underline{\hat{\sigma}}^2}{\sigma_0^2} \sim \frac{\chi^2(m-n,0)}{m-n}; \quad m - n = b\,. \tag{5.79}$$

This is the well-known distribution of the famous variance of unit weight. We will also derive this expression after introducing the v-test statistic. ■

Special case 5.6 (2) When, under H_o, there is only one variance component in the stochastic model, i.e. $Q_y = \sigma^2 Q_1$, equations (5.67) and (5.68) can be simplified to (cf. equation (5.27))

$$\underline{w} = \frac{\underline{y}^T Q_{\hat{e}}^- \left[\frac{1}{2} C_y - \frac{\mathrm{tr}(C_y Q_{\hat{e}}^-)}{2b} Q_y\right] Q_{\hat{e}}^- \underline{y}}{\left[\frac{1}{2}\mathrm{tr}(C_y Q_{\hat{e}}^- C_y Q_{\hat{e}}^-) - \frac{1}{2b}\mathrm{tr}(C_y Q_{\hat{e}}^-)\mathrm{tr}(C_y Q_{\hat{e}}^-)\right]^{1/2}}\,, \tag{5.80}$$

and

$$\underline{w} = \frac{\underline{\hat{e}}^T Q_y^{-1} \left[\frac{1}{2} C_y - \frac{\mathrm{tr}(C_y Q_{\hat{e}}^-)}{2b} Q_y\right] Q_y^{-1} \underline{\hat{e}}}{\left[\frac{1}{2}\mathrm{tr}(C_y Q_{\hat{e}}^- C_y Q_{\hat{e}}^-) - \frac{1}{2b}\mathrm{tr}(C_y Q_{\hat{e}}^-)\mathrm{tr}(C_y Q_{\hat{e}}^-)\right]^{1/2}}\,, \tag{5.81}$$

respectively. The eigenvalues λ corresponding to the distribution of test statistics (5.80) and (5.81) follows as (cf. equation (5.55))

$$\left| bC_y Q_{\hat{e}}^- - [2\lambda b w^d + \mathrm{tr}(C_y Q_{\hat{e}}^-)] I_m \right| = 0\,. \tag{5.82}$$

It is important to note that the eigenvalues λ of equation (5.82) can be obtained independently from the unknown parameter σ^2. ■

Example 5.2 (Known mean) If the actual value of a quantity is known, one may assess the noise characteristic of the measurements of this quantity (see NOTE 4.1). For example, we can assess the noise of a GPS coordinate time series by measuring a zero baseline (two receivers connected to one antenna). Under H_o, if we assume that the time series has only white noise, we can test this hypothesis against some alternative hypotheses. For this purpose, one can write

$$\mathrm{H}_o\colon Q_y = \sigma^2 I \quad \text{versus} \quad \mathrm{H}_a\colon Q_y = \sigma^2 I + C_y \nabla\,. \tag{5.83}$$

One can simply show that $Q_{\hat{e}}^- = \sigma^{-2} I$. After a few mathematical operations, one can obtain the w-test statistic as

$$\underline{w} = \frac{m\underline{\hat{e}}^T C_y \underline{\hat{e}} - \mathrm{tr}(C_y)\underline{\hat{e}}^T \underline{\hat{e}}}{\sigma^2 \left[2m^2 \mathrm{tr}(C_y C_y) - 2m\mathrm{tr}(C_y)\mathrm{tr}(C_y)\right]^{1/2}} \tag{5.84}$$

Let us now assume that $C_y = uu^T$, with u the summation vector of size m. Matrix C_y contains all ones and describes a random constant noise process. A systematic effect (bias) in the observations has a similar behavior. Note that a random constant noise is assessed in the stochastic model, but a systematic bias is usually estimated in the functional model. The preceding statistic then reads

$$\underline{w} = \frac{(u^T \underline{\hat{e}})^2 - \underline{\hat{e}}^T \underline{\hat{e}}}{\sigma^2 \left[2m(m-1)\right]^{1/2}}\,. \tag{5.85}$$

The above statistic is distributed as

$$\underline{w} \sim \frac{1}{\left[2m(m-1)\right]^{1/2}} \left[(m-1)\chi_1^2(1) - \chi_2^2(m-1)\right]\,. \tag{5.86}$$

If m is large enough, the above statistic and its distribution are approximated by

$$\underline{w} = \frac{1}{\sqrt{2}} \left[\frac{(u^T \underline{\hat{e}})^2}{m\sigma^2} - 1\right] \quad \text{and} \quad \underline{w} \sim \frac{1}{\sqrt{2}} [\chi_1^2(1) - 1]\,, \tag{5.87}$$

which yields

$$\frac{(u^T\underline{\hat{e}})^2}{m\sigma^2} \sim \chi^2_1(1) \;\text{ or }\; \frac{u^T\underline{\hat{e}}}{\sqrt{m}\sigma} \sim N(0,1)\,. \tag{5.88}$$

The preceding formula could directly be obtained from the functional model when we include an unknown *bias* term, namely b_x, in the functional model. In this case, its least-squares solution reads $\hat{b}_x = u^T\underline{\hat{e}}/m$. When testing the hypothesis $\mathrm{H}_o:\ b_x = 0$ versus $\mathrm{H}_a:\ b_x \neq 0$ (parameter significance test, see section 2.2.5), the same formula as above can be obtained. ■

5.2.6 Datasnooping on Q_y

As with the datasnooping on the elements of the vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ or correspondingly on the elements of Q_t, suggested in section 5.2.4, one can also apply this idea to the elements of Q_y, the variances and covariances of the observables. In this case, the $m \times m$ symmetric matrix C_y takes the form

$$C_y = C_{ij} = c_i c_j^T + c_j c_i^T,\ i,j = 1,2,...,m,, \tag{5.89}$$

with c_i and c_j the canonical *unit vectors* of length m, referring to a single couple of entries in matrix Q_y, the ij^{th} and ji^{th} element, the (co)variances. Taking the above form for the matrix C_y, the null and alternative hypotheses read

$$\mathrm{H}_o\colon Q_y = \sum_{k=1}^{p} \sigma_k Q_k \text{ versus } \mathrm{H}_a\colon Q_y = \sum_{k=1}^{p} \sigma_k Q_k + C_{ij}\nabla,\ \nabla \neq 0\,. \tag{5.90}$$

Substituting C_y from equation (5.89) into equation (5.67) gives

$$\underline{w} = \frac{\underline{y}^T Q_{\hat{e}}^- \left(\frac{1}{2}C_{ij} - \frac{1}{2}[g_k n_{kl}^{-1} Q_l]\right) Q_{\hat{e}}^- \underline{y}}{\left[(q_{ij}^-)^2 + q_{ii}^- q_{jj}^- - g^{(ij)T} N^{-1} g^{(ij)}\right]^{1/2}}\,, \tag{5.91}$$

with $q_{ij}^- = c_i Q_{\hat{e}}^- c_j$. The p-vector $g^{(ij)}$ follows from equation (5.69): $g_k^{(ij)} = c_i^T Q_{\hat{e}}^- Q_k Q_{\hat{e}}^- c_j$ (substitute C_y from equation (5.89)). The eigenvalues associated with the distribution of the preceding w-test statistic follow from equation (5.73)

$$\left|C_{ij}Q_{\hat{e}}^- - [g_k n_{kl}^{-1} Q_l] Q_{\hat{e}}^- - 2\lambda w^d I_m\right| = 0 \tag{5.92}$$

with w^d the denominator of equation (5.91). In terms of the least-squares residuals $\underline{\hat{e}} = P_A^{\perp}\underline{y}$, a similar formulation can be given for the w-test statistic.

Special case 5.7 (1) In the case that the covariance matrix is known, i.e. $\mathrm{H}_o\colon Q_y = Q_0$ versus $\mathrm{H}_a\colon Q_y = Q_0 + C_{ij}\nabla$, the w-test statistic (5.75) reads

$$\underline{w} = \frac{\frac{1}{2}\underline{y}^T Q_{\hat{e}}^- C_{ij} Q_{\hat{e}}^- \underline{y} - q_{ij}^-}{\left[(q_{ij}^-)^2 + q_{ii}^- q_{jj}^-\right]^{1/2}} = \frac{\frac{1}{2}\underline{\hat{e}}^T Q_y^{-1} C_{ij} Q_y^{-1} \underline{\hat{e}} - q_{ij}^-}{\left[(q_{ij}^-)^2 + q_{ii}^- q_{jj}^-\right]^{1/2}} = \underline{\hat{e}}^T M \underline{\hat{e}} - m_0. \tag{5.93}$$

This test statistic is again distributed as $\underline{w} \sim \lambda_k \chi^2_k(1,0) - m_0$, where the r number of eigenvalues λ follow from

$$\left|C_{ij}Q_{\hat{e}}^- - 2\lambda w^d I_m\right| = 0\,, \tag{5.94}$$

where w^d is the denominator of equation (5.93). ■

Special case 5.8 (2) As another special case, let us consider the following. When there is only one variance component in the stochastic model, i.e. when $p = 1$, the null and alternative hypotheses are given as

$$\mathrm{H}_o\colon Q_y = \sigma^2 Q_1 \quad \text{versus} \quad \mathrm{H}_a\colon Q_y = \sigma^2 Q_1 + C_{ij}\nabla, \tag{5.95}$$

Substituting for C_y from equation (5.89) into equation (5.80), after a few operations, gives

$$\underline{w}_{ij} = \frac{\underline{y}^T Q_{\hat{e}}^- \left(\frac{1}{2}C_{ij} - \frac{1}{b}q_{ij}^- Q_y\right) Q_{\hat{e}}^- \underline{y}}{\left[\frac{b-2}{b}(q_{ij}^-)^2 + q_{ii}^- q_{jj}^-\right]^{1/2}}. \tag{5.96}$$

For the purpose of the datasnooping on the elements of Q_y (testing the elements of the covariance matrix Q_y), one needs the distribution of $\underline{w}_{ij},\ i, j = 1, 2, ..., m$. Here, the $\underline{w}_{ij}$ is again distributed as $\underline{w}_{ij} \sim \lambda_k \chi_k^2(1, 0)$ with the eigenvalues λ following from

$$\left|bC_{ij}Q_{\hat{e}}^- - [2\lambda b w^d + 2q_{ij}^-]I_m\right| = 0\,, \tag{5.97}$$

where w^d is the denominator of equation (5.96). ■

NOTE 5.4 Since C_{ij} has a simple form, namely $C_{ij} = c_i c_j^T + c_j c_i^T$, the eigenvalues $\lambda' = 2\lambda w^d$ in equation (5.94) and $\lambda' = 2\lambda w^d + 2q_{ij}^-/b$ in equation (5.97) (and therefore λ) can be analytically computed. For this purpose, one can write

$$C_{ij}Q_{\hat{e}}^- = c_i c_j^T Q_{\hat{e}}^- + c_j c_i^T Q_{\hat{e}}^-\,. \tag{5.98}$$

This is a matrix of order m with zero elements except for its i^{th} and j^{th} row in which, respectively, the j^{th} and i^{th} row of the $Q_{\hat{e}}^-$ have been housed (i^{th} and j^{th} row have been interchanged). The rank of this matrix is then 2. It has therefore two non-zero eigenvalues which are given as the eigenvalues of the following matrix:

$$[C_{ij}Q_{\hat{e}}^-]_{2\times 2} = \begin{pmatrix} q_{ij}^- & q_{jj}^- \\ q_{ii}^- & q_{ij}^- \end{pmatrix}. \tag{5.99}$$

One can simply show that the eigenvalues of the preceding matrix read

$$\lambda_1' = q_{ij}^- + \sqrt{q_{ii}^- q_{jj}^-} \quad \text{and} \quad \lambda_2' = q_{ij}^- - \sqrt{q_{ii}^- q_{jj}^-}\,. \tag{5.100}$$

The remaining eigenvalues of $C_{ij}Q_{\hat{e}}^-$ are zero, i.e. $\lambda_3' = ... = \lambda_b' = 0$. Note that when $i = j$, the only eigenvalue is given as $\lambda_1' = 2q_{ii}^-$. The eigenvalues λ associated with the distribution of w-test in equations (5.93) and (5.96) then read

$$\lambda = \frac{\lambda'}{2w^d}; \quad \lambda = \frac{b\lambda' - 2q_{ij}^-}{2bw^d}\,, \tag{5.101}$$

respectively. If we now substitute λ' from equation (5.100) and w^d from the denominators of equations (5.93) and (5.96) we can obtain analytical expressions for $\lambda_1, ..., \lambda_b$. For example, the eigenvalues associated with the distribution of w-test in equation (5.93) are given as

$$\lambda_1 = \frac{q_{ij}^- + \sqrt{q_{ii}^- q_{jj}^-}}{2\left[(q_{ij}^-)^2 + q_{ii}^- q_{jj}^-\right]^{1/2}}; \quad \lambda_2 = \frac{q_{ij}^- - \sqrt{q_{ii}^- q_{jj}^-}}{2\left[(q_{ij}^-)^2 + q_{ii}^- q_{jj}^-\right]^{1/2}}; \quad \lambda_3 = ... = \lambda_b = 0. \tag{5.102}$$

Note that the eigenvalues can, in general, be both positive and negative. From the preceding equations, it can simply be verified that

$$\mathsf{E}\{\underline{w}_{ij}\} = \sum_{k=1}^{b} \lambda_k - m_0 = 0 \quad \text{and} \quad \mathsf{D}\{\underline{w}_{ij}\} = \sum_{k=1}^{b} 2\lambda_k^2 = 1\,. \tag{5.103}$$

Because a $b-2$ number of the eigenvalues are zero, the distribution of $\underline{w}_{ij}$ is given as

$$\underline{w}_{ij} \sim \lambda_1 \chi_1^2(1,0) + \lambda_2 \chi_2^2(1,0) - m_0. \tag{5.104}$$

In case $i = j$, when testing the diagonal elements of Q_y, the preceding distribution can, to a greater degree, be simplified to

$$\underline{w}_{ii} \sim \frac{1}{\sqrt{2}} \left[\chi_1^2(1,0) - 1\right]\,, \tag{5.105}$$

which is independent of the elements of the matrix $Q_{\hat{e}}^-$. ⊡

Example 5.3 (Known mean and variance) As a very simple example of the above, let one functionally known quantity be measured independently and with known variance σ_0^2 (under the null hypothesis) m times (see NOTE 4.1). In this case, the model of observation equations reads

$$\mathsf{E}\{\underline{y}\} = \mu_y \;\text{ with }\; b = m \;\text{ and }\; \mathsf{D}\{\underline{y}\} = Q_y = \sigma_0^2 I_m. \tag{5.106}$$

It is not difficult to show that $Q_{\hat{e}}^-$ reads $q_{ij}^- = \frac{1}{\sigma_0^2}$ if $i = j$ and $q_{ij}^- = 0$ if $i \neq j$. Taking the above circumstances into account, the w-test statistic reads

$$\underline{w}_{ij} = \frac{\hat{e}_i \hat{e}_j}{\sigma_0^2},\ \text{if } i \neq j; \quad \underline{w}_{ii} = \frac{\hat{e}_i^2 - \sigma_0^2}{\sigma_0^2}\,. \tag{5.107}$$

Accordingly, the eigenvalues (5.102) can be simplified to $\lambda_1 = \frac{1}{2}$ and $\lambda_2 = -\frac{1}{2}$. Therefore, the distribution of $\underline{w}_{ij}$, $i \neq j$ is given as

$$\underline{w}_{ij} \sim \frac{1}{2} \left[\chi_1^2(1,0) - \chi_2^2(1,0)\right]\,. \tag{5.108}$$

The above distribution can be rewritten as

$$\underline{w}_{ij} \sim \frac{1}{2} \left[\chi_1(1) - \chi_2(1))(\chi_1(1) + \chi_2(1)\right] = \frac{\chi_1(1) - \chi_2(1)}{\sqrt{2}} \cdot \frac{\chi_1(1) + \chi_2(1)}{\sqrt{2}} = z_1 z_2, \tag{5.109}$$

where $\chi_1(1) = \sqrt{\chi_1^2(1)}$, and $z_1 = \frac{\chi_1(1) - \chi_2(1)}{\sqrt{2}}$ and $z_2 = \frac{\chi_1(1) + \chi_2(1)}{\sqrt{2}}$ are two independent standard normal variables. One can show that the distribution of $\underline{w}_{ij}$ reads (see e.g. Ware and Lad, 2003):

$$\underline{w}_{ij} \sim f(w) = \int_0^\infty \frac{1}{\pi x} \exp\left\{-\frac{w^2 + x^4}{2x^2}\right\} dx\,. \tag{5.110}$$

This integration can be undertaken numerically (e.g. in mathematical computer package MATLAB, see `quadl.m`). In case $i = j$, the distribution of $\underline{w}_{ii}$ is given by equation (5.105). Figure 5.1 (a and b) show the histogram of $\underline{w}_{ij}$ and $\underline{w}_{ii}$ on the basis of simulated data according to equations (5.108) and (5.105), respectively; the standard normal distribution is also included in the figures, since the w-test statistic has a mean *zero* and a variance *one*! Also, in figure 5.1 (a), the distribution of $\underline{w}_{ij}$, obtained from numerical evaluation of (5.110), has been illustrated. ■

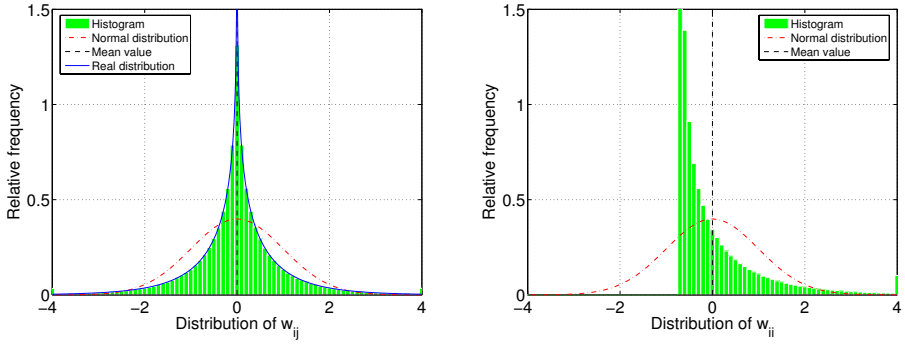


Figure 5.1: Histogrammed distribution of $\underline{w}_{ij}$ (a) and $\underline{w}_{ii}$ (b) on the basis of a simulated data set of size 1000 000, according to equations (5.108) and (5.105), respectively, compared with the standard normal distribution. Real distribution of $\underline{w}_{ij}$, obtained from numerical evaluation of equation (5.110) is also included in figure (a).

5.2.7 Two illustrative examples

In the previous section, the idea of data snooping on the elements of the matrix Q_y was introduced. However, it should be noted that this technique (elementwise method) is neither computationally efficient, since a $\frac{m(m+1)}{2}$ number of elements should be tested, nor practically useful, since they can not simply be interpreted. On the other hand, in most Geodesy and GPS applications, it is possible to divide the observations into different *groups* or *categories*. This may cross our mind to generalize the idea of data snooping *groupwise* rather than *elementwise*. To do so and to illustrate, at the moment, we will present two simple examples.

Example 5.4 Let us assume that we are measuring one unknown parameter simultaneously with two different instruments m times. Now we have two different sorts of observations, namely $\underline{y}_1$ and $\underline{y}_2$ each of size m. In this case we have only one unknown parameter x and the total number of observations is $2m$. Assuming the observations are uncorrelated and with the same precision (under $\mathrm{H}_o;\ p = 1$), one obtains

$$\mathrm{E}\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}\} = Ax = \begin{bmatrix} u \\ u \end{bmatrix} x; \quad Q_y = \sigma^2 I_{2m} = \sigma^2 \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \tag{5.111}$$

with u, as before, the summation vector of size m. If we want now to test whether the covariance matrix chosen Q_y is a realistic one, we may put forward the following two hypotheses:

$$\mathrm{H}_o\colon Q_y = \sigma^2 \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \quad \text{versus} \quad \mathrm{H}_a\colon Q_y = \sigma^2 \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \nabla, \tag{5.112}$$

where the extra $2m \times 2m$ matrix, under the alternative hypothesis, is the matrix C_y. The unknown parameter ∇ implies that the second group observations, has a precision different from the first group. In the following, we derive analytical expressions for the w-test statistic and for its distribution. For this purpose, strictly speaking, we need to simplify equations (5.81) and

(5.82), respectively. The w-test statistic simplifies to

$$\underline{w} = \frac{\underline{\hat{e}}_2^T \underline{\hat{e}}_2 - \underline{\hat{e}}_1^T \underline{\hat{e}}_1}{2\sigma^2\sqrt{m-1}}, \tag{5.113}$$

with $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ the least squares residual vectors of the first and second group of observations, under H_o, respectively.

Proof. The entries of matrix $Q_{\hat{e}}^- = Q_y^{-1} P_A^\perp$, of order $2m$, are given as $q_{ij}^- = \frac{2m-1}{2m\sigma^2}$ if $i = j$ and $q_{ij}^- = \frac{-1}{2m\sigma^2}$ if $i \neq j$. After a few mathematical operations, one obtains

$$\mathrm{tr}(C_y Q_{\hat{e}}^-) = \frac{2m-1}{2\sigma^2}, \quad \mathrm{tr}(C_y Q_{\hat{e}}^- C_y Q_{\hat{e}}^-) = \frac{4m-3}{4\sigma^4} \tag{5.114}$$

from which w^d in equation (5.81) follows as $w^d = \frac{\sqrt{m-1}}{2\sigma^2}$. The terms in the numerator of equation (5.81) simplify to

$$\underline{\hat{e}}^T Q_y^{-1} C_y Q_y^{-1} \underline{\hat{e}} = \frac{\underline{\hat{e}}_2^T \underline{\hat{e}}_2}{\sigma^4}, \quad \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} = \frac{\underline{\hat{e}}_1^T \underline{\hat{e}}_1 + \underline{\hat{e}}_2^T \underline{\hat{e}}_2}{\sigma^2} \tag{5.115}$$

Substituting the preceding terms in equation (5.81) completes the proof. ▼

Note that the σ^2 is unknown even under the null hypothesis. In practice, one has to be satisfied with a least-squares estimate of this variance component, namely $\hat{\sigma}^2 = \frac{\hat{e}_1^T \hat{e}_1 + \hat{e}_2^T \hat{e}_2}{2m-1}$. If $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ are of similar magnitude, the preceding test statistic intuitively says that the w-test value is not very large, and therefore the null hypothesis is likely to be accepted. One will then obtain $\hat{\nabla} \approx 0$ and therefore $\hat{\sigma}_1^2 \approx \hat{\sigma}_2^2$.

The w-test statistic is distributed as $\underline{w} \sim \lambda_k \chi_k^2(1,0)$, where the eigenvalues λ are obtained from equation (5.82). After not very long derivations, the eigenvalues λ are given, independent of the unknown variance component σ^2, as (see appendix B.1)

$$\lambda_1 = 0; \quad \lambda_2 = ... = \lambda_m = \frac{1}{2\sqrt{m-1}}; \quad \lambda_{m+1} = ... = \lambda_{2m-1} = \frac{-1}{2\sqrt{m-1}}, \tag{5.116}$$

from which it can simply be concluded that the w-test statistic (5.113) is distributed as

$$\underline{w} \sim \frac{1}{2\sqrt{m-1}} \left[\chi_1^2(m-1,0) - \chi_2^2(m-1,0)\right] \tag{5.117}$$

where $\chi_1^2(m-1,0)$ and $\chi_2^2(m-1,0)$ are two independent chi-squared distributions with $m-1$ degrees of freedom. One can simply show that the above distribution has a mean of *zero* and a variance of *one*. Based on the *central limit theorem* for independent and identically distributed data, when m is large enough, one can approximate the chi-squared distribution by a normal one, i.e. $\chi_1^2(m-1,0) \approx \mathsf{N}(m-1, 2(m-1))$. Taking these into account, one can simply show that the preceding distribution is approximated by the standard normal distribution, namely $\underline{w} \simeq N(0,1)$. To show this, one needs to use the following identity:

$$\alpha_1 \mathsf{N}_1(\mu_1, \sigma_1^2) + \alpha_2 \mathsf{N}_2(\mu_2, \sigma_2^2) = \mathsf{N}(\alpha_1\mu_1 + \alpha_2\mu_2, \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2), \tag{5.118}$$

for two independent normal distributions N_1 and N_2, and α_1 and α_2 two real numbers. ■

Example 5.5 In previous example, if one wants to test whether the first group of observations is independent of the second group, one may write the null and alternative hypotheses as follows:

$$\mathrm{H}_o\colon Q_y = \sigma^2 \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \quad \text{versus} \quad \mathrm{H}_a\colon Q_y = \sigma^2 \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}}_{C_y} \nabla, \tag{5.119}$$

where ∇ here is a covariance component. In the following, we give formulas for the w-test statistic and its distribution. After a few mathematical and matrix operations, one can show that (see appendix B)

$$\underline{w} = \frac{(\underline{\hat{e}}_2 - \underline{\hat{e}}_1)^T(\underline{\hat{e}}_2 - \underline{\hat{e}}_1) + 4m\underline{\hat{e}}_1^T\underline{\hat{e}}_2}{2\sigma^2\sqrt{2m(2m-1)(m-1)}}\,, \tag{5.120}$$

where $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ are, again, the least squares residuals of the first and second group of observations.

Proof. After a few mathematical and matrix operations, one can show that

$$\mathrm{tr}(C_y Q_{\hat{e}}^-) = \frac{-1}{\sigma^2}, \quad \mathrm{tr}(C_y Q_{\hat{e}}^- C_y Q_{\hat{e}}^-) = \frac{2m-1}{\sigma_1^4} \tag{5.121}$$

which yields $w^d = \frac{1}{\sigma^2}\sqrt{\frac{2m(m-1)}{2m-1}}$ in equation (5.81). The terms in the numerator of equation (5.81) then read

$$\underline{\hat{e}}^T Q_y^{-1} C_y Q_y^{-1}\underline{\hat{e}} = \frac{2\,\underline{\hat{e}}_1^T\underline{\hat{e}}_2}{\sigma_1^4}, \quad \underline{\hat{e}}^T Q_y^{-1}\underline{\hat{e}} = \frac{\underline{\hat{e}}_1^T\underline{\hat{e}}_1 + \underline{\hat{e}}_2^T\underline{\hat{e}}_2}{\sigma^2} \tag{5.122}$$

Substituting the preceding terms in equation (5.81) completes the proof. ▼

If $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ have similar corresponding values, then the term $\underline{\hat{e}}_1^T\underline{\hat{e}}_2$ in the preceding equation becomes large, and therefore the null hypothesis is likely to be rejected. This implies that the first and the second observation groups are likely to be correlated.

To see how the distribution looks like, we need the eigenvalues λ from equation (5.82). We can show that such non-zero eigenvalues (a $2m-1$ number) can be obtained as (see appendix B.2)

$$\begin{aligned} \lambda_1 \quad &= \quad ... \quad = \quad \lambda_m \quad &= \quad -\tfrac{m-1}{\sqrt{2m(2m-1)(m-1)}}\,; \\ \lambda_{m+1} \quad &= \quad ... \quad = \quad \lambda_{2m-1} \quad &= \quad \tfrac{m}{\sqrt{2m(2m-1)(m-1)}}\,. \end{aligned} \tag{5.123}$$

With these in mind, the distribution of the w-test statistic, namely $\lambda_k\chi_k^2(1,0)$, simplifies to

$$\underline{w} \sim \frac{1}{\sqrt{2m(2m-1)(m-1)}}\left[m\chi_1^2(m-1,0) - (m-1)\chi_2^2(m,0)\right]\,, \tag{5.124}$$

where $\chi_1^2(m-1,0)$ and $\chi_2^2(m,0)$ are two independent chi-squared distributions with $m-1$ and m degrees of freedom, respectively. One can again show that the preceding distribution has mean *zero* and variance *one*. If m is large enough, again, the chi-squared distributions can be approximated by normal ones. Therefore, using identity (5.118), the above distribution can be approximated by the standard normal distribution, i.e. $\underline{w} \simeq \mathsf{N}(0,1)$. ■

Figure (5.2), a and b, respectively, show the histogram distribution of w-test statistic presented in formulae (5.117) and (5.124), in case of $m = 50$ and on the basis of a simulated data set of size 100,000. For a *visual comparison with normality*, the standard normal distribution has been plotted on the histograms. In addition, figure 5.2, c and d, illustrate the *normal probability plot* for the aforementioned (simulated) data sets. The normal probability plot is a graphical technique to assess whether or not a data set is approximately normally distributed. The data is plotted against a theoretical normal distribution in such a way that the points should form a straight line. Departures from this straight line indicate departures from normality. These plots may seem curved at the ends (tails) of the distribution function. However, this minor curvature seems to be not significant at $\alpha = 0.05$ or even $\alpha = 0.01$ significance level tests.

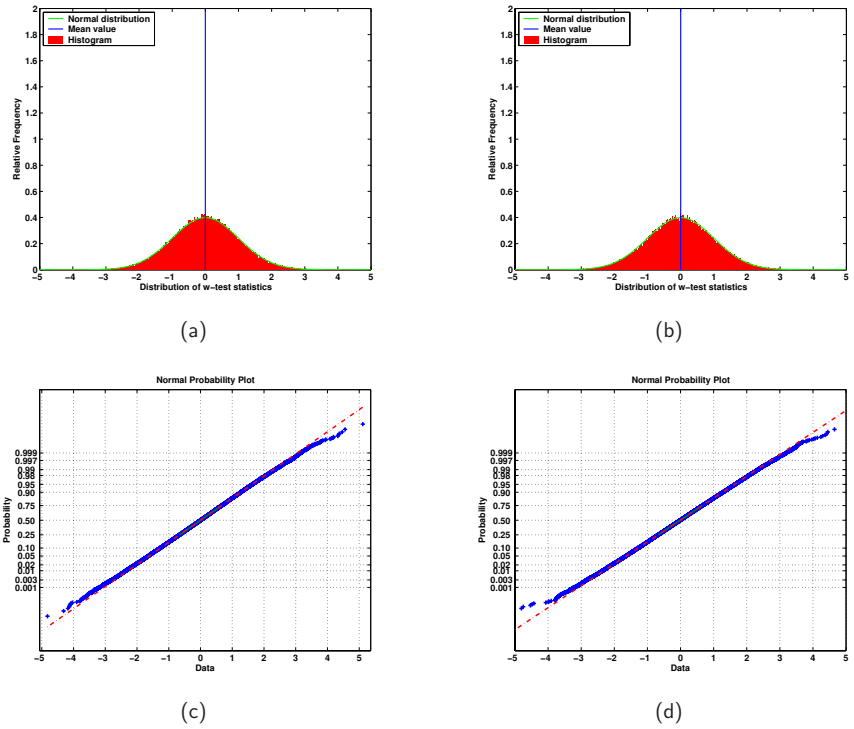


Figure 5.2: Histogram distribution of w-test statistic given by equation (5.117), (a), and by equation (5.124), (b), in case of $m = 50$ and on the basis of a simulated data set of size 100 000 (compared with the standard normal distribution); plots (c) and (d) show normal probability plot of simulated data sets corresponding to plots (a) and (b), respectively.

5.3 The v-test statistic

5.3.1 Formulation in terms of B-model

In the following we consider a testing problem that, though from a mathematical aspect equivalent to the above formulation, occurs when we want to test the *significance* of the (co)variance components. For this purpose, the following two hypotheses are considered:

$$\mathrm{H}_o\colon \mathrm{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = A_{\mathrm{vh}}\sigma,\ d^T\sigma = c_o \text{ versus } \mathrm{H}_a\colon \mathrm{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = A_{\mathrm{vh}}\sigma,\ d^T\sigma = c_a\,. \quad (5.125)$$

The two hypotheses H_o and H_a differ in the sense that under H_o it is assumed that the linear function $d^T\sigma$ is identical to the known scalar c_o, whereas under H_a, this function is identical to the unknown scalar $c_a \neq c_o$. In terms of the model of condition equations, the preceding hypotheses read

$$\mathrm{H}_o\colon Q_t = \sum_{k=1}^{p} \sigma_k Q_{t_k},\ d^T\sigma = c_o \text{ versus } \mathrm{H}_a\colon Q_t = \sum_{k=1}^{p} \sigma_k Q_{t_k},\ d^T\sigma = c_a\,. \quad (5.126)$$

In a similar way to the w-test statistic, we can here also obtain (generalize) the v-test statistic, introduced in chapter 2, for the stochastic model. From equation (2.44), it can simply be concluded that the v-test statistic reads

$$\underline{v} = \frac{d^T N^{-1} \underline{l} - c_0}{\sqrt{d^T N^{-1} d}} = \frac{d_k n_{kl}^{-1} \underline{l}_l - c_0}{\sqrt{d^T N^{-1} d}} = \frac{\frac{1}{2} d_k n_{kl}^{-1} \, \underline{t}^T Q_t^{-1} Q_{t_l} Q_t^{-1} \underline{t} - c_0}{\sqrt{d^T N^{-1} d}} \,. \tag{5.127}$$

The preceding formula can be rewritten as follows:

$$\boxed{\underline{v} = \frac{\underline{t}^T Q_t^{-1} [\frac{1}{2} d_k n_{kl}^{-1} Q_{t_l}] Q_t^{-1} \underline{t} - c_0}{\sqrt{d^T N^{-1} d}} = \underline{t}^T M \underline{t} - m_0,} \tag{5.128}$$

where M is a symmetric matrix and m_0 a constant as

$$M = \frac{Q_t^{-1} [\frac{1}{2} d_k n_{kl}^{-1} Q_{t_l}] Q_t^{-1}}{\sqrt{d^T N^{-1} d}} \,; \quad m_0 = \frac{c_0}{\sqrt{d^T N^{-1} d}} \,, \tag{5.129}$$

respectively. Based on the theorem 5.2, the preceding v-test statistic is distributed as (cf. equation (5.50))

$$\boxed{\underline{v} \sim \sum_{i=1}^{r} \lambda_i \chi_i^2(1,0) - m_0,} \tag{5.130}$$

where the χ_i^2 are mutually independent and the λ_i are the non-zero eigenvalues of $Q_t M$

$$\boxed{\left| [d_k n_{kl}^{-1} Q_{t_l}] Q_t^{-1} - 2\lambda v^d I_b \right| = 0 \,,} \tag{5.131}$$

where $v^d = \sqrt{d^T N^{-1} d}$ is the denominator of equation (5.128).

5.3.2 Formulation in terms of A-model

In terms of the model of observation equations, the null and alternative hypotheses in equation (5.126) may be reformulated as

$$\mathrm{H}_o \colon Q_y = \sum_{k=1}^{p} \sigma_k Q_k; \;\; d^T \sigma = c_o \;\; \text{versus} \;\; \mathrm{H}_a \colon Q_y = \sum_{k=1}^{p} \sigma_k Q_k; \;\; d^T \sigma = c_a. \tag{5.132}$$

The test statistic $\underline{v} = \underline{t}^T M \underline{t} - m_0$ is reformulated in terms of the model of observation equations as

$$\boxed{\underline{v} = \frac{\underline{\hat{e}}^T Q_y^{-1} [\frac{1}{2} d_k n_{kl}^{-1} Q_l] Q_y^{-1} \underline{\hat{e}} - c_0}{\sqrt{d^T N^{-1} d}} = \underline{\hat{e}}^T M \underline{\hat{e}} - m_0,} \tag{5.133}$$

which is again distributed as $\underline{v} \sim \lambda_k \chi_k^2(1,0) - m_0$, and M different from equation (5.129) obviously. The eigenvalues associated with this distribution are given as

$$\boxed{\left| [d_k n_{kl}^{-1} Q_l] Q_{\hat{e}}^{-} - 2\lambda v^d I_b \right| = 0 \,,} \tag{5.134}$$

where $v^d = \sqrt{d^T N^{-1} d}$ is the denominator of equation (5.133).

Example 5.6 As a simple application, let us assume that we want to test the only variance component of the stochastic model, namely the variance of unit weight ($p = 1$). For such a simple case, the null and alternative hypotheses read

$$\mathrm{H}_o\colon Q_y = \sigma^2 Q_1;\quad \sigma^2 = \sigma_0^2,\quad \text{versus}\quad \mathrm{H}_a\colon Q_y = \sigma^2 Q_1;\quad \sigma^2 = \sigma_a^2 \neq \sigma_0^2. \tag{5.135}$$

We can simply see that $d = 1$. From equation (5.26) follows that $v^d = \sqrt{d^T N^{-1} d} = \sqrt{\frac{2}{b}}\sigma^2$. Also, the numerator of M in equation (5.133) reads $\underline{v}^n = \underline{\hat{e}}^T Q_1^{-1}\underline{\hat{e}}/b - \sigma_0^2$ (note that the solution under H_o has to read: $\sigma^2 = \sigma_0^2$). The non-zero eigenvalues associated with the distribution of the v-test read then $\lambda_1 = \lambda_2 = ... = \lambda_b = \frac{1}{\sqrt{2b}}$. Therefore, one can simply show that the v-test statistic and its distribution read

$$\underline{v} = \frac{1}{\sqrt{2b}}\left[\frac{b\,\underline{\hat{\sigma}}^2}{\sigma_0^2} - b\right];\quad \underline{v} \sim \frac{1}{\sqrt{2b}}\left[\chi^2(b,0) - b\right], \tag{5.136}$$

where $b\,\underline{\hat{\sigma}}^2 = \underline{\hat{e}}^T Q_1^{-1}\underline{\hat{e}}$. Mind that this is the solution of the model without constraints. Comparing the formulas in equation (5.136) gives the well-known distribution of the famous *variance of unit weight*, i.e.

$$\frac{b\,\underline{\hat{\sigma}}^2}{\sigma_0^2} \sim \chi^2(b,0);\ \rightarrow\ \frac{\underline{\hat{\sigma}}^2}{\sigma_0^2} \sim \frac{\chi^2(b,0)}{b};\quad b = m - n. \tag{5.137}$$

Note that this result has also been obtained with the w-test statistic (see example 5.1); see also the equivalence of the w-test and v-test statistic in appendix D.3. ■

5.4 The overall model test

5.4.1 Quadratic form of residuals

The goal of the present section is, as a first attempt, to generalize, or better say to introduce, the *overall model test* for the stochastic model. We are testing the overall validity of the assumed stochastic model. For this goal, assuming $\mathsf{E}\{\underline{t}\} = 0$, the null and alternative hypotheses are formulated as

$$\mathrm{H}_o\colon \mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = A_{\mathrm{vh}}\sigma \ \text{ versus }\ \mathrm{H}_a\colon \mathsf{E}\{\mathrm{vh}(\underline{t}\,\underline{t}^T)\} = [A_{\mathrm{vh}}\, C_{\mathrm{vh}}]\begin{bmatrix}\sigma\\ \nabla\end{bmatrix}, \tag{5.138}$$

with C_{vh} a matrix of order $\frac{b(b+1)}{2} \times df$ to imply that in the alternative hypothesis one leaves the observable vector $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ completely free, and ∇ a vector of order df (in fact $[A_{\mathrm{vh}}\, C_{\mathrm{vh}}]$ is a square and invertible matrix). The above null and alternative hypotheses can be formulated in terms of the elements of the model of condition equations as

$$\mathrm{H}_o\colon Q_t = \sum_{k=1}^{p}\sigma_k Q_{t_k} \quad \text{versus} \quad \mathrm{H}_a\colon Q_t = \sum_{k=1}^{\frac{b(b+1)}{2}}\sigma_k C_{t_k}, \tag{5.139}$$

with C_{t_k} and σ_k under the alternative hypothesis as

$$C_{t_k} = \begin{cases} c_i c_j^T & \text{if}\quad \sigma_k = \sigma_{ii},\quad i = j\,; \\ c_i c_j^T + c_j c_i^T & \text{if}\quad \sigma_k = \sigma_{ij},\quad i \leq j\,, \end{cases} \tag{5.140}$$

and i and j have to run from 1 through b. We now apply the idea of section 2.2.3 to the above hypotheses. This will lead us to the the quadratic form of the residuals in the stochastic model, namely to $\underline{T}_{\rm df} = \underline{\hat{e}}_{\rm vh}^T Q_{\rm vh}^{-1} \underline{\hat{e}}_{\rm vh}$. As we saw in chapter 4, this test statistic can be obtained as (see equation (4.89))

$$\underline{T}_{\rm df} = \frac{1}{2}(\underline{t}^T Q_t^{-1} \underline{t})^2 - \underline{t}^T Q_t^{-1} Q_{t_0} Q_t^{-1} \underline{t} + \frac{1}{2}{\rm tr}(Q_{t_0} Q_t^{-1} Q_{t_0} Q_t^{-1}) - \underline{l}^T N^{-1} \underline{l}\,, \tag{5.141}$$

where N and $\underline{l}$ are defined in equations (4.85) and (4.86), respectively. When $Q_0 = 0$, the preceding equation reads

$$\underline{T}_{\rm df} = \frac{1}{2}(\underline{t}^T Q_t^{-1} \underline{t})^2 - \underline{l}^T N^{-1} \underline{l}\,. \tag{5.142}$$

Without loss of generality, we consider the constant term Q_0 to be absent. When needed, we will, however, give a few comments for special cases. It is clear that the quadratic form is a *fourth-degree* function of $\underline{t}$; multiplication of two quadratic forms in $\underline{t}$. In terms of the A-model, one obtains

$$\underline{T}_{\rm df} = \frac{1}{2}(\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2 - \underline{\hat{e}}^T Q_y^{-1} Q_0 Q_y^{-1} \underline{\hat{e}} + \frac{1}{2}{\rm tr}(Q_0 Q_{\hat{e}}^- Q_0 Q_{\hat{e}}^-) - \underline{l}^T N^{-1} \underline{l}\,, \tag{5.143}$$

where N and $\underline{l}$ are evaluated from equations (4.110) and (4.112), respectively. If $Q_0 = 0$, the preceding equation reads

$$\underline{T}_{\rm df} = \frac{1}{2}(\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2 - \underline{l}^T N^{-1} \underline{l}\,. \tag{5.144}$$

The distribution of the preceding test statistic is not a trivial one, because of the complicated nature in which it depends on $\underline{\hat{e}}$ or $\underline{t}$. For some special cases, it is possible to determine the exact distribution of equation (5.143). However, in general, one will have to rely on alternative computer-based techniques such as Monte-Carlo simulation or bootstrapping, see e.g. Efron and Tibshirani (1993). In the sequel, just the first two moments of statistic (5.144), namely its *expectation* and its *dispersion* are derived. At the end, we will consider an approximation for this quadratic form. A function of this approximation will have a well-known distribution, namely the chi-squared distribution.

Example 5.7 Let the covariance matrix $Q_t = Q_{t_0}$ be known under the null hypothesis. In this case, equation (5.143) simplifies to, with $N = 0$,

$$\begin{aligned} \underline{T}_{\rm df} &= \tfrac{1}{2}(\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2 - \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} + \tfrac{b}{2}; \\ &= \tfrac{1}{2}(\underline{t}^T Q_t^{-1} \underline{t})^2 - \underline{t}^T Q_t^{-1} \underline{t} + \tfrac{b}{2}. \end{aligned} \tag{5.145}$$

The distribution of this statistic is nontrivial. However, if we define $\underline{T}_b = \sqrt{2\underline{T}_{\rm df} - b + 1} + 1$ it follows that

$$\underline{T}_b = \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} = \underline{t}^T Q_t^{-1} \underline{t} \tag{5.146}$$

which is distributed as $\underline{T}_b \sim \chi^2(b, 0)$. This result can also be obtained from the overall model test in the functional model. ■

5.4.2 Expectation of quadratic form

The usual assumption, from now on, is that the $b \times 1$ misclosure vector $\underline{t}$ is *normally* distributed, i.e. $\underline{t} \sim \mathsf{N}(0, Q_t)$. From equation (C.37), one obtains

$$\mathsf{E}\{\frac{1}{2}(\underline{t}^T Q_t^{-1} \underline{t})^2\} = \frac{1}{2}\mathsf{E}\{(\underline{t}^T Q_t^{-1} \underline{t})(\underline{t}^T Q_t^{-1} \underline{t})\} = \frac{1}{2} b(b+2)\,. \tag{5.147}$$

The expectation of the second term in equation (5.142) reads

$$\mathsf{E}\{\underline{l}^T N^{-1} \underline{l}\} = \mathsf{E}\{\mathrm{tr}(\underline{l}^T N^{-1} \underline{l})\} = \mathsf{E}\{\mathrm{tr}(N^{-1} \underline{l}\,\underline{l}^T)\} = \mathrm{tr}(N^{-1} \mathsf{E}\{\underline{l}\,\underline{l}^T\})\,, \tag{5.148}$$

where matrix $\mathsf{E}\{\underline{l}\,\underline{l}^T\}$ is given as (see equation (4.86))

$$\mathsf{E}\{\underline{l}\,\underline{l}^T\} = \mathsf{E}\{\underline{l}_i \underline{l}_j\} = \frac{1}{4}\mathsf{E}\{(\underline{t}^T \overbrace{Q_t^{-1} Q_{t_i} Q_t^{-1}}^{A} \underline{t})(\underline{t}^T \overbrace{Q_t^{-1} Q_{t_j} Q_t^{-1}}^{B} \underline{t})\}\,, \tag{5.149}$$

with A and B temporary matrices as in equation (C.33). One then obtains

$$\begin{aligned} \mathsf{E}\{\underline{l}\,\underline{l}^T\} &= \tfrac{1}{4}\mathrm{tr}(Q_t^{-1} Q_{t_i})\mathrm{tr}(Q_t^{-1} Q_{t_j}) + \tfrac{1}{2}\mathrm{tr}(Q_t^{-1} Q_{t_i} Q_t^{-1} Q_{t_j}) \\ &= l_i l_j + n_{ij} = l\,l^T + N\,, \end{aligned} \tag{5.150}$$

with $l_i = \frac{1}{2}\mathrm{tr}(Q_t^{-1} Q_{t_i})$, the expectation of $\underline{l}_i$. Substitution of the preceding equation in equation (5.148) yields

$$\mathsf{E}\{\underline{l}^T N^{-1} \underline{l}\} = \mathrm{tr}(N^{-1}[l l^T + N]) = p + l^T N^{-1} l\,. \tag{5.151}$$

The term $l^T N^{-1} l$ in the preceding formulas can be simplified to

$$l^T N^{-1} l = l^T \sigma = l_i \sigma_i = \frac{1}{2}\mathrm{tr}(Q_t^{-1} Q_{t_i})\sigma_i = \frac{1}{2}\mathrm{tr}(Q_t^{-1}[\sigma_i Q_{t_i}]) = \frac{b}{2}\,. \tag{5.152}$$

Taking the expectation of equation (5.142) and substituting the first and second terms from equations (5.147) and (5.151), respectively, under null-hypothesis H_o, result in the expectation of the quadratic form as the redundancy of the stochastic model, namely

$$\boxed{\mathsf{E}\{\underline{T}_{\mathrm{df}}\} = df\,,} \tag{5.153}$$

with $df = \frac{b(b+1)}{2} - p$, the redundancy of the stochastic model. This is an interesting result, when compared with the functional model, as the expectation of the quadratic form in the *functional model* leads to the redundancy of the functional model b, i.e. $\mathsf{E}\{\underline{T}_b\} = \mathsf{E}\{\underline{t}^T Q_t^{-1} \underline{t}\} = b$. Note however that the result obtained for the functional model is independent of the distribution of $\underline{t}$, but that obtained for the stochastic model takes the normality assumption into account.

5.4.3 Dispersion of quadratic form

The dispersion of the quadratic form of the residuals reads

$$\mathsf{D}\{\underline{T}_{\mathrm{df}}\} = \mathsf{E}\{\underline{T}_{\mathrm{df}}^2\} - (\mathsf{E}\{\underline{T}_{\mathrm{df}}\})^2\,. \tag{5.154}$$

From equation (5.153), the second term of the preceding equation reads

$$(\mathrm{E}\{\underline{T}_{\mathrm{df}}\})^2 = \frac{b^2(b+1)^2}{4} + p^2 - p\,b(b+1)\,. \tag{5.155}$$

From equation (5.142), the expectation of $\underline{T}_{\mathrm{df}}^2$ is given as

$$\mathrm{E}\{\underline{T}_{\mathrm{df}}^2\} = \mathrm{E}\left\{\frac{1}{4}(\underline{t}^T Q_t^{-1}\underline{t})^4 - (\underline{t}^T Q_t^{-1}\underline{t})^2(\underline{l}^T N^{-1}\underline{l}) + (\underline{l}^T N^{-1}\underline{l})^2\right\}, \tag{5.156}$$

which consists of three terms. In the following, we will evaluate each of the three terms in equation (5.156). The results will be given in a compact form. However, more explanation can be found in appendix C.

First term: From equation (C.51), the expectation of the first term in equation (5.156) reads:

$$\mathrm{E}\{\frac{1}{4}(\underline{t}^T Q_t^{-1}\underline{t})^4\} = b(b+2)(b+4)(b+6)/4\,. \tag{5.157}$$

Second term: The expectation of the second term in equation (5.156) simplifies to (see appendix B.3)

$$\mathrm{E}\{(\underline{t}^T Q_t^{-1}\underline{t})^2(\underline{l}^T N^{-1}\underline{l})\} = \frac{b^3}{2} + 5b^2 + b^2p + 12b + 10bp + 24p\,. \tag{5.158}$$

Third term: The expectation of the third term in equation (5.156) simplifies to (see appendix B.4)

$$\mathrm{E}\{(\underline{l}N^{-1}\underline{l})^2\} = \frac{b^2}{4} + p\,b + 2b + 10p + p^2 + 4N^{-1} : Z : N^{-1} + 2N^{-1} : Z^{23} : N^{-1}\,. \tag{5.159}$$

where : denotes the *double inner (dot) product* of two tensors and the *fourth-order* tensor Z is given as $z_{ijkl} = \frac{1}{2}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j}Q_t^{-1}Q_{t_k}Q_t^{-1}Q_{t_l})$, $i,j,k,l = 1,2,...,p$. This tensor satisfies $z_{ijkl} = z_{jkli} = z_{klij} = z_{lijk}$, $i,j,k,l = 1,2,...,p$. Note that here $N^{-1} : Z$ is a second order tensor (a matrix), and $N^{-1} : Z : N^{-1}$ is a scalar.

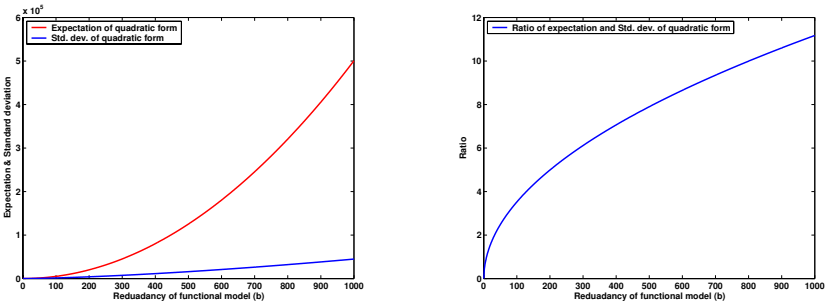
Substitution of equations (5.157), (5.158) and (5.159) into equation (5.156) and then substitution of equations (5.156) and (5.155) into equation (5.154), under the null-hypothesis H_o, gives the dispersion of the quadratic form as

$$\boxed{\mathrm{D}\{\underline{T}_{\mathrm{df}}\} = 2b(b^2+3b+1) - (8b+14)p + 4N^{-1} : Z : N^{-1} + 2N^{-1} : Z^{23} : N^{-1}\,.} \tag{5.160}$$

5.4.4 Quadratic form in case $p = 1$

In a simple case that only one variance component is involved in the stochastic model, i.e. when $Q_y = \sigma^2 Q_1$ or $Q_t = \sigma^2 Q_{t_1}$, one obtains $\underline{l}$ and N by equation (5.26). In this case, the quadratic form expressed in equation (5.142) simplifies to (cf. equation (4.116))

$$\underline{T}_{\mathrm{df}} = \frac{b-1}{2b}(\underline{t}^T Q_t^{-1}\underline{t})^2. \tag{5.161}$$



(a) (b)

Figure 5.3: Expectation and standard deviation of quadratic form of residuals in stochastic model, i.e. $\mathsf{E}\{\underline{T}_{\mathrm{df}}\}$ and $\sqrt{\mathsf{D}\{\underline{T}_{\mathrm{df}}\}}$, (a), and their ratio, (b), in case $p = 1$.

Using equation (C.37), one can obtain the expectation of the preceding quadratic form as (cf. equation (5.153))

$$\mathsf{E}\{\underline{T}_{\mathrm{df}}\} = \frac{(b-1)(b+2)}{2} = \frac{b(b+1)}{2} - 1 = df\,. \tag{5.162}$$

The variance of the quadratic form is, by definition, given as $\mathsf{D}\{\underline{T}_{\mathrm{df}}\} = \mathsf{E}\{\underline{T}^2_{\mathrm{df}}\} - (\mathsf{E}\{\underline{T}_{\mathrm{df}}\})^2$. The second part can be obtained from equation (5.162). Using equation (C.51), one can also simplify the first part. The variance of the quadratic form then reads

$$\mathsf{D}\{\underline{T}_{\mathrm{df}}\} = \frac{2(b-1)^2(b+2)(b+3)}{b}\,. \tag{5.163}$$

This result can also be obtained from the general formula (5.160). Therefore, the results show that the expectation of the quadratic form of residuals, i.e. $\underline{T}_{\mathrm{df}}$ test statistic, in case $p = 1$, is a second-order polynomial of b, the redundancy of the functional model, while its variance, more or less, looks like a third-order polynomial. Figure 5.3 shows the expectation and the standard deviation of the quadratic form of the residuals versus the redundancy of the functional model, for the special case $p = 1$.

It should be noted that the quadratic form (5.161) is exactly (without any approximation) a quadratic function of $\underline{t}^T Q_t^{-1} \underline{t}$, namely $\frac{b-1}{2b}(\underline{t}^T Q_t^{-1} \underline{t})^2$, from which the following function can be derived explicitly:

$$\underline{T}_b = \sqrt{\frac{2b}{b-1}\underline{T}_{\mathrm{df}}} = \underline{t}^T Q_t^{-1} \underline{t}\,, \tag{5.164}$$

which, provided that the vector $\underline{t}$ is normally distributed, i.e. $\underline{t} \sim \mathsf{N}(0, Q_t)$, has a distribution of chi-squared with b degrees of freedom. It is possible to construct an *exact* confidence interval for the true value of $\underline{T}_b$, namely T_b, meaning that it has exactly the stated confidence level, as opposed to so-called an *asympotic* or a *large-sample* confidence interval which only has approximately the stated confidence level and is valid when the sample size is large. An exact confidence interval is valid for any sample size. An exact

two-sided confidence interval $(1-\alpha)$ for T_b reads then

$$\chi^2_{b,1-\frac{\alpha}{2}} \leq T_b \leq \chi^2_{b,\frac{\alpha}{2}}\,, \tag{5.165}$$

with $\chi^2_{b,\frac{\alpha}{2}}$, the critical value obtained from the inverse of the chi-square cumulative distribution function. Substituting for T_b from equation (5.164) and solving for $T_{\rm df}$, i.e. transforming back from T_b into $T_{\rm df}$, gives a confidence interval for $T_{\rm df}$ as follows:

$$\frac{b-1}{2b}\chi^4_{b,1-\frac{\alpha}{2}} \leq T_{\rm df} \leq \frac{b-1}{2b}\chi^4_{b,\frac{\alpha}{2}}\,, \tag{5.166}$$

with notation $\chi^4_{b,\frac{\alpha}{2}} = (\chi^2_{b,\frac{\alpha}{2}})^2$. In the case that b is sufficiently large, the chi-square distribution may be approximated by a normal one, i.e. $\chi^2(b,0) \approx \mathsf{N}(b,2b)$. In this case, the critical value reads $\chi^2_{b,\frac{\alpha}{2}} \approx X_{\frac{\alpha}{2}}$ with

$$X_{\frac{\alpha}{2}} = \sqrt{2b}Z_{\frac{\alpha}{2}} + b\,. \tag{5.167}$$

Since the standard normal distribution is a symmetric one about zero, it follows $Z_{1-\frac{\alpha}{2}} = -Z_{\frac{\alpha}{2}}$. Substitution of the preceding terms in equation (5.166) gives an asymptotic confidence interval for $T_{\rm df}$ as

$$\frac{b^2}{2} - \sqrt{2}b^{\frac{3}{2}}Z_{\frac{\alpha}{2}} + bZ^2_{\frac{\alpha}{2}} \leq T_{\rm df} \leq \frac{b^2}{2} + \sqrt{2}b^{\frac{3}{2}}Z_{\frac{\alpha}{2}} + bZ^2_{\frac{\alpha}{2}}. \tag{5.168}$$

5.4.5 Approximation of quadratic form

What we want to mention here is the complexity of the distribution of the quadratic form $\underline{T}_{\rm df}$. In this subsection an approximation for this quadratic form is derived, and a function of it for which the distribution is completely known. To reach this goal, let us rewrite the quadratic form of the residuals for the stochastic model, i.e. equation (5.142) in terms of $\underline{t}$ vector, as follows:

$$\begin{aligned}
\underline{T}_{\rm df} &= \tfrac{1}{2}(\underline{t}^T Q_t^{-1}\underline{t})^2 - \underline{l}^T\underline{\hat{\sigma}} \\
&= \tfrac{1}{2}(\underline{t}^T Q_t^{-1}\underline{t})^2 - \underline{l}_i\underline{\hat{\sigma}}_i \\
&= \tfrac{1}{2}(\underline{t}^T Q_t^{-1}\underline{t})^2 - \tfrac{1}{2}\underline{t}^T Q_t^{-1}[\underline{\hat{\sigma}}_i Q_{t_i}]Q_t^{-1}\underline{t} \\
&= \tfrac{1}{2}(\underline{t}^T Q_t^{-1}\underline{t})^2 - \tfrac{1}{2}\underline{t}^T Q_t^{-1}\hat{Q}_t Q_t^{-1}\underline{t}\,.
\end{aligned} \tag{5.169}$$

Using the approximation $\hat{Q}_t = Q_t$, one obtains

$$\underline{T}_{\rm df} \approx \frac{1}{2}(\underline{t}^T Q_t^{-1}\underline{t})^2 - \frac{1}{2}\underline{t}^T Q_t^{-1}\underline{t} = \underline{T}'_{\rm df}. \tag{5.170}$$

This approximation is denoted as $\underline{T}'_{\rm df}$. Therefore, $\underline{T}_{\rm df} \approx \underline{T}'_{\rm df}$. The distribution of this approximation is still complicated. It is not however difficult to show that the expectation and the dispersion of this approximated statistic read

$$\mathsf{E}\{\underline{T}'_{\rm df}\} = \frac{b(b+1)}{2} = df + p\,; \quad \mathsf{D}\{\underline{T}'_{\rm df}\} = 2b^3 + 8b^2 + 8.5b\,, \tag{5.171}$$

respectively. It is now possible to derive a function of this new statistic whose distribution is completely known. To follow this, equation (5.169) can be rewritten as

$$\underline{T}_{\rm df} = \frac{1}{2}(\underline{t}^T Q_t^{-1}\underline{t})^2 - \frac{1}{2}\underline{t}^T Q_t^{-1}\underline{t} - \underline{\epsilon} = \underline{T}'_{\rm df} - \underline{\epsilon}\,, \tag{5.172}$$

with

$$\underline{\epsilon} = \frac{1}{2}\left(\underline{t}^T Q_t^{-1}\hat{Q}_t Q_t^{-1}\underline{t} - \underline{t}^T Q_t^{-1}\underline{t}\right). \tag{5.173}$$

If we now define $\underline{T}_b = \underline{t}^T Q_t^{-1}\underline{t}$ it then follows that (see appendix B.5)

$$\underline{T}'_b + \underline{\epsilon}' \approx \frac{1}{2}\left(\sqrt{8(\underline{T}_{\mathrm{df}} + \underline{\epsilon}) + 1} + 1\right) \approx \underline{T}_b = \underline{t}^T Q_t^{-1}\underline{t} \tag{5.174}$$

where the bias statistic $\underline{\epsilon}'$ is approximated as

$$\underline{\epsilon}' \approx \frac{1}{2}\left[\frac{\underline{t}^T Q_t^{-1}\hat{Q}_t Q_t^{-1}\underline{t}}{\underline{t}^T Q_t^{-1}\underline{t}} - 1\right]. \tag{5.175}$$

Neglecting the bias terms $\underline{\epsilon}$ and $\underline{\epsilon}'$ in equation (5.174) yields

$$\underline{T}'_b \approx \underline{T}_b = \frac{1}{2}\left(\sqrt{8\underline{T}'_{\mathrm{df}} + 1} + 1\right) = \underline{t}^T Q_t^{-1}\underline{t} \sim \chi^2(b, 0) \tag{5.176}$$

Note that this approximation gets better and better when b, the redundancy of the functional model, becomes larger and larger (see example 5.8). In fact, the estimator $\underline{T}'_b$ is a *consistent* estimator of T_b. This is an interesting result as the preceding function (statistic) is the *quadratic form* of the residuals for the *functional model*, denoted as $\underline{T}_b$. Therefore, the statistic of the *overall model test*, both for the functional and the stochastic model, is identical. That is, the $\underline{T}_b$ test statistic used in the functional model is a good general indicator of the presence of some blunders (gross errors) in the observations. On the other hand, with a large enough b, it can be interpreted as a general indication of detecting misspecifications in the stochastic model as well.

An exact two-sided confidence interval $(1-\alpha)$ for T_b (or accordingly an asymptotic confidence interval for T'_b) reads then

$$\chi^2_{b,1-\frac{\alpha}{2}} \le T'_b \le \chi^2_{b,\frac{\alpha}{2}} \tag{5.177}$$

Substituting T_b and solving for T'_{df}, i.e. transforming back to the original function, from T_b to T'_{df}, gives an approximate confidence interval for T'_{df} (or an asymptotic approximate confidence interval for T_{df}) as follows:

$$\frac{1}{2}\left(\chi^4_{b,1-\frac{\alpha}{2}} - \chi^2_{b,1-\frac{\alpha}{2}}\right) \le T'_{\mathrm{df}} \le \frac{1}{2}\left(\chi^4_{b,\frac{\alpha}{2}} - \chi^2_{b,\frac{\alpha}{2}}\right) \tag{5.178}$$

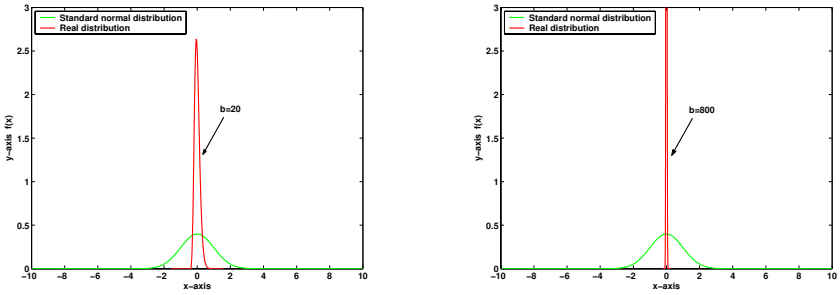
Again, if b is sufficiently large, the chi-square distribution can be approximated by a normal one for which the critical value reads $\chi^2_{b,\frac{\alpha}{2}} \approx \sqrt{2b}Z_{\frac{\alpha}{2}} + b$.

Example 5.8 Let us consider the simplest case with only one variance component in the stochastic model, namely $p = 1$. In this case, $Q_y = \sigma^2 Q$, which gives

$$\hat{Q}_t = \hat{\underline{\sigma}}^2 B^T Q_1 B, \quad Q_t = \sigma^2 B^T Q_1 B\,. \tag{5.179}$$

Substitution of the preceding terms into equation (5.175) yields

$$\underline{\epsilon}' \approx \frac{1}{2}\left[\frac{\hat{\underline{\sigma}}_1^2}{\sigma^2} - 1\right]. \tag{5.180}$$



(a) (b)

Figure 5.4: Distribution of $\underline{\epsilon}$ introduced in (5.181), i.e. $p = 1$, for case $b = 20$, (a), and $b = 800$, (b), compared with the standard normal distribution

This (bias) statistic is distributed as

$$\underline{\epsilon}' \simeq \frac{1}{2}\left[\frac{\chi^2(b,0)}{b} - 1\right], \tag{5.181}$$

with the expectation and the dispersion of

$$\mathsf{E}\{\underline{\epsilon}'\} = 0; \quad \mathsf{D}\{\underline{\epsilon}'\} = \frac{1}{2b}, \tag{5.182}$$

respectively. This is an important result since, in general, the statistic $\underline{\epsilon}'$ is not a *real bias term* as $\mathsf{E}\{\underline{\epsilon}'\} = 0$, and the variance gets smaller when b becomes larger. However, for a large enough redundancy of the functional model b, this variance when compared with the variance of the statistic $\underline{t}^T Q_t^{-1} \underline{t}$, i.e. $2b$, can then be considered negligible. To make it clearer, we have considered an example. Figure (5.4)-a illustrates the distribution of the error statistic $\underline{\epsilon}'$, for case $b = 20$ and again $p = 1$, compared to the standard normal distribution. However, this comparison may not be very much relevant. That is, the distribution of $\underline{\epsilon}'$ should be compared with the distribution of the quadratic form $\underline{t}^T Q_t^{-1} \underline{t}$, with the expectation of b and variance of $2b$. The ratio of variances is then

$$\frac{\mathsf{D}\{\underline{\epsilon}'\}}{\mathsf{D}\{\underline{t}^T Q_t^{-1} \underline{t}\}} = \frac{1}{4b^2}. \tag{5.183}$$

In other words, the comparison of $\underline{\epsilon}'$ and $\underline{t}^T Q_t^{-1} \underline{t}$, with b degrees of freedom, is approximately equivalent to the comparison of $\underline{\epsilon}'$, with $2b^2$ degrees of freedom, and the standard normal distribution. If $b = 20$, then $2b^2$ is set to 800. This is what we have deliberately considered in Fig. (5.4)-b, $b = 800$. Therefore,

$$\underline{\epsilon}' \xrightarrow{\mathsf{D}} 0 \quad \text{as} \quad b \to \infty, \tag{5.184}$$

meaning that its distribution is converged to zero. However, as previously mentioned, for the above example we need not use this approximation as the quadratic form (5.161) is strictly a quadratic function of $\underline{t}^T Q_t^{-1} \underline{t}$. ■

5.5 Summary and concluding remarks

We generalized the idea of detection and validation into the stochastic model. It was aimed to find misspecifications in the stochastic model, to improve an existing covariance matrix, and to test the general validity of the stochastic model. This was done by introducing the w-test, the v-test, and the overall model test with the stochastic model in systematic analogy with testing in the functional model. We gave expressions for these test statistics and derived the distributions for the w-test and the v-test statistic based on the normality assumptions of the original observables $\underline{y}$ or $\underline{t}$. For the examples given, when the redundancy b of the functional model is large enough, the distributions can be approximated by the (standard) normal distribution.

The overall model test statistic has in general a complicated form of distribution because of the complicated nature in which it depends on $\underline{\hat{y}}$ or $\underline{\hat{e}}$. For some special cases, it might be possible to determine the exact distribution of equation (5.143). We considered an approximation for this test statistic of which a function of this approximation has the well-known chi-squared distribution. But, in general, one will have to rely on alternative computer-based techniques such as Monte-Carlo simulation or bootstrapping to make for instance bootstrap confidence intervals. One may consider the simplest technique, the percentile method. This method works well when the statistic used has a symmetric distribution. The percentile method requires a large number of Monte Carlo replications for the intervals to be both accurate (i.e. be as small as possible for the given confidence level) and nearly exact (i.e. if the procedure were repeated many times the percentage of intervals that would actually include the 'true' parameter value is approximately the stated confidence level). For more information we refer to Efron and Tibshirani (1993)

Multivariate Variance-Covariance Analysis

6

6.1 Introduction

The primary purpose of this chapter is to introduce basic concepts of a multivariate statistical analysis which occurs frequently in practice. It may be helpful to think of the 'analysis' as involving multiple dependent models. Experiments produce data where measurements were obtained on one repeated model. In repeated model analysis, the multiple dependent variables are the same measures (e.g. they are repeated over time). In multivariate models, the multiple dependent variables are measures of multiple outcomes, usually measured at the same point in time. For example, a repeated model might be used to analyze the vertical components measured at three successive months while a multivariate analysis might be used to model three coordinate components (north, east, and vertical) at a single point in time. The analysis of (co)variance is often applied in a multivariate model, if the effects of factors can be explained by several characteristics. This is called a multivariate analysis of (co)variance. Multivariate models can also be set up if observations are repeated at different times. In this case we can record temporal changes of a phenomenon.

A multivariate linear model, also known as a repeated linear model, is in fact an extension of the univariate linear model. If for one design matrix in a Gauss-Markov model instead of one observation vector several observation vectors with identical covariance matrices are given and the corresponding parameter vectors have to be determined, the model is referred to as a multivariate Gauss-Markov model. We consider a generalization of the univariate linear model $\mathrm{E}\{\underline{y}\} = Ax,\ \mathrm{D}\{\underline{y}\} = \sigma^2 Q$, when the variance σ^2 of unit weight is unknown. This is in fact the simplest form of an unknown covariance matrix in a linear model. One can for instance show that the expressions for the BLUE of x, y, and e are invariant against a change in the variance σ^2 (see chapter 3). We will show in fact that such an independence of the variance holds true also for a multivariate linear model.

In order to properly understand the estimation and validation of these model parameters, a foundation in multivariate statistics is needed. In particular, we may think of concepts such as variances and covariances (or correlations) between different groups (models). We show how the estimation and validation of (co)variance components can be expressed in terms of the least-squares residuals. An advantage of the multivariate approach is to allow for designs of experiments where the resulting parameter estimators will be uncorrelated, thus making it easier to interpret results. One might approach multivariate outcomes by examining each outcome separately. There might be dangers with this, however. With univariate model analysis we may not use the whole observables in an optimal way. One problem is that we are not doing multiple comparisons on the same data. More importantly, we usually expect the multiple outcomes to be correlated, and we may wish to determine if the outcomes are affected by treatment factors independent of other outcomes. We

can however show, for this special model, that the univariate models can be treated individually. The mutual correlation coefficients between different models, obtained from the least-squares residuals, propagate directly into the correlation between the corresponding parameters in the functional model.

6.2 Multivariate parameter estimation

The least squares approach can be used to estimate the *covariance matrix* from repeated measurements. We can think of having r groups (categories) where each has m observations. Consider the following linear model consisting of r groups of observations:

$$\mathsf{E}\{\underline{y}_i\} = Ax_i; \quad \mathsf{D}\{\underline{y}_i, \underline{y}_j\} = \sigma_{ij}Q; \quad i,j = 1,2,\ldots,r\,, \tag{6.1}$$

with $\underline{y}_i$ the normally distributed m-vector of observables for group i, and correspondingly x_i the n-vector of unknowns; the $m \times n$ design matrix A and the $m \times m$ cofactor matrix Q is supposed to be identical for all groups and Q is assumed to be symmetric and positive-definite. For example, we may have measurements of daily position estimates of permanent GPS stations ($r = 3$, north, east, and vertical) for a whole year ($m = 365$). If for each time series a simple linear regression model is used then the number of parameters becomes $n = 2$ (intercept and slope). Written out in full, the preceding linear models read

$$\underbrace{\mathsf{E}\left\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_r \end{bmatrix}\right\}}_{mr\times 1} = \underbrace{\begin{bmatrix} A & & & \\ & A & & \\ & & \ddots & \\ & & & A \end{bmatrix}}_{mr\times nr} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}}_{nr\times 1}, \tag{6.2}$$

with the $mr \times mr$ covariance matrix of the form

$$\underbrace{\mathsf{D}\left\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_r \end{bmatrix}\right\}}_{mr\times 1} = \underbrace{\begin{bmatrix} \sigma_1^2\,Q & \sigma_{12}Q & \cdots & \sigma_{1r}Q \\ \sigma_{12}Q & \sigma_2^2\,Q & \cdots & \sigma_{2r}Q \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r}Q & \sigma_{2r}Q & \cdots & \sigma_r^2\,Q \end{bmatrix}}_{mr\times mr}. \tag{6.3}$$

The unknowns in this model are the $n\,r$-number of elements of the vector x (those in the functional part of the model)

$$x = \begin{bmatrix} x_1^T & x_2^T & \cdots & x_r^T \end{bmatrix}^T, \tag{6.4}$$

and the $\frac{r(r+1)}{2}$ number of elements $\sigma_{ii} = \sigma_i^2$ (variances) and σ_{ij} (covariances) of the symmetric matrix (those in the stochastic part of the model)

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r} & \sigma_{2r} & \cdots & \sigma_{rr} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1r} & \sigma_{2r} & \cdots & \sigma_r^2 \end{bmatrix}. \tag{6.5}$$

If we collect all unknows x_i in the $n \times r$ matrix X, all observables $\underline{y}_i$ in the $m \times r$ matrix $\underline{Y}$, and correspondingly all residuals $\underline{e}_i$ in the $m \times r$ matrix $\underline{E}$, we will then obtain

$$X = [x_1\, x_2\, \ldots\, x_r];\quad \underline{Y} = [\underline{y}_1\, \underline{y}_2\, \ldots\, \underline{y}_r];\quad \underline{E} = [\underline{e}_1\, \underline{e}_2\, \ldots\, \underline{e}_r]\,. \tag{6.6}$$

With these notations, the functional part of the model in equation (6.2) can simply be rewritten as $\mathsf{E}\{\underline{Y}\} = AX$. Using the properties of vec operator and Kronecker product, we can write equations (6.2) and (6.3) using equations (6.5) and (6.6) more compactly as

$$\mathsf{E}\{\mathrm{vec}(\underline{Y})\} = (I \otimes A)\mathrm{vec}(X)\,;\quad Q_{\mathrm{vec}(Y)} = \Sigma \otimes Q\,. \tag{6.7}$$

We can now apply the least-squares method to estimate both X and Σ. To derive simplified formulas for the BLUE of X, Y, and E, we need simple expressions for $P^{\perp}_{I\otimes A}$ and $P_{I\otimes A}$. They follow as

$$P^{\perp}_{I\otimes A} = I \otimes P^{\perp}_{A}\,;\quad P_{I\otimes A} = I \otimes P_A\,, \tag{6.8}$$

where $P_A = A(A^TQ^{-1}A)^{-1}A^TQ^{-1}$ and $P^{\perp}_A = I - P_A$.

Proof. By definition the orthogonal projector is given as $P^{\perp}_{I\otimes A} = I \otimes I_m - P_{I\otimes A}$ where $P_{I\otimes A} = (I\otimes A)[(I\otimes A^T)(\Sigma^{-1}\otimes Q^{-1})(I\otimes A)]^{-1}(I\otimes A^T)(\Sigma^{-1}\otimes Q^{-1})$. Simplification of the equation gives $P^{\perp}_{I\otimes A} = I\otimes I_m - I\otimes A(A^TQ^{-1}A)^{-1}A^TQ^{-1} = I\otimes P^{\perp}_A$, where $P^{\perp}_A = I_m - A(A^TQ^{-1}A)^{-1}A^TQ^{-1}$. ▼

One can now simply show that the BLUE of X, Y, and E follows as

$$\hat{\underline{X}} = (A^TQ^{-1}A)^{-1}A^TQ^{-1}\underline{Y}\,;\quad \hat{\underline{Y}} = P_A\underline{Y}\,;\quad \hat{\underline{E}} = \underline{Y} - \hat{\underline{Y}} = P^{\perp}_A\underline{Y}\,, \tag{6.9}$$

respectively. Note that the preceding expressions are independent of the unknown matrix Σ. This can be considered as a generalization of the univariate linear model $\mathsf{E}\{\underline{y}\} = Ax$, $\mathsf{D}\{\underline{y}\} = \sigma^2 Q$, when the variance σ^2 of unit weight is unknown. We can therefore determine the outcomes of individual models separately.

One can simply show that the covariance matrices of the estimators $\mathrm{vec}(\hat{\underline{X}})$, $\mathrm{vec}(\hat{\underline{Y}})$, and $\mathrm{vec}(\hat{\underline{E}})$ are given as

$$Q_{\mathrm{vec}(\hat{X})} = \Sigma \otimes (A^TQ^{-1}A)^{-1};\quad Q_{\mathrm{vec}(\hat{Y})} = \Sigma \otimes P_AQ;\quad Q_{\mathrm{vec}(\hat{E})} = \Sigma \otimes P^{\perp}_AQ\,. \tag{6.10}$$

The preceding equations are also very similar to the univariate model. Note that to obtain the covariance matrix of the estimators, the matrix Σ should be known. If Σ is unknown, one can rely on an estimate $\hat{\Sigma}$ instead. We have, in fact, shown that mutual correlation between different models propagates directly into the correlation between the corresponding parameters in the functional model.

6.3 Multivariate variance-covariance estimation

We shall now apply equation (4.113), with equations (4.110) and (4.112), to the model in equation (6.7) in order to find *unbiased* and *minimum variance* estimators for the elements of the matrix Σ of equation (6.5). The least-squares estimator of $\sigma = \mathrm{vh}(\Sigma)$ reads

$$\hat{\underline{\sigma}} = \mathrm{vh}(\hat{\underline{\Sigma}}) = N^{-1}\underline{l}. \tag{6.11}$$

where the $\frac{r(r+1)}{2}$-vector $\underline{l}$ and the $\frac{r(r+1)}{2} \times \frac{r(r+1)}{2}$ matrix N^{-1} are given as

$$\underline{l} = \frac{1}{2} D^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathrm{vec}(\hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}), \tag{6.12}$$

and

$$N^{-1} = \frac{2}{m-n} D^+ (\Sigma \otimes \Sigma) D^{+T}, \tag{6.13}$$

respectively, with D the duplication matrix and the $m \times r$ matrix of least-squares residuals $\hat{\underline{E}}$ given in equation (6.9).

Proof. With appropriate matrices C_α, matrix Σ can be written as $\Sigma = \sum_{k=1}^{r(r+1)/2} \sigma_\alpha C_\alpha$ where σ_α is respectively $\sigma_1^2,\ \sigma_{12}, \ldots, \sigma_r^2$ and $C_\alpha = c_i c_i^T$, for $\sigma_\alpha := \sigma_i^2$, and $C_\alpha = c_i c_j^T + c_j c_i^T$, for $\sigma_\alpha := \sigma_{ij}$ $(i \neq j)$. Equation (6.7) then reads

$$\mathsf{E}\{\mathrm{vec}(\underline{Y})\} = (I \otimes A)\mathrm{vec}(X)\,; \quad Q_{\mathrm{vec}(Y)} = \sum_{k=1}^{r(r+1)/2} \sigma_\alpha (C_\alpha \otimes Q) \tag{6.14}$$

We can now apply LS-VCE to find the minimum variance estimator of $\sigma = [\sigma_1, \cdots, \sigma_{r(r+1)/2}]^T = \mathrm{vh}(\Sigma)$. We have for multivariate model $Q_{\mathrm{vec}(Y)}^{-1} = \Sigma^{-1} \otimes Q^{-1}$, $Q_\alpha = C_\alpha \otimes Q$, $P_{I\otimes A}^\perp = I \otimes P_A^\perp$, and $\mathrm{vec}(\hat{\underline{E}}) = \mathrm{vec}(P_A^\perp \underline{Y})$, with $P_A^\perp = I_m - A(A^T Q^{-1} A)^{-1} A^T Q^{-1}$. Substitution into $\underline{l}_\alpha = \frac{1}{2}\mathrm{vec}(\hat{\underline{E}})^T Q_{\mathrm{vec}(Y)}^{-1} Q_\alpha Q_{\mathrm{vec}(Y)}^{-1} \mathrm{vec}(\hat{\underline{E}})$ gives the entries of the right-hand side vector as

$$\begin{aligned} \underline{l}_\alpha &= \tfrac{1}{2}\mathrm{vec}(\hat{\underline{E}})^T \Sigma^{-1} C_\alpha \Sigma^{-1} \otimes Q^{-1} \mathrm{vec}(\hat{\underline{E}}) \\ &= \tfrac{1}{2}\mathrm{tr}(\Sigma^{-1} C_\alpha \Sigma^{-1} \hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}) \\ &= \tfrac{1}{2}\mathrm{vec}(C_\alpha)^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathrm{vec}(\hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}) \\ &= \tfrac{1}{2}\mathrm{vh}(C_\alpha)^T D^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathrm{vec}(\hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}) \end{aligned} \tag{6.15}$$

Since $\mathrm{vh}(C_\alpha)$ is a canonical unit vector, when α runs from 1 to $\frac{r(r+1)}{2}$, $\mathrm{vh}(C_\alpha)$ will make an identity matrix and thus the preceding equation reads

$$\underline{l} = \frac{1}{2} D^T (\Sigma^{-1} \otimes \Sigma^{-1}) D \mathrm{vh}(\hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}) \tag{6.16}$$

In a similar manner the entries of the normal matrix N are obtained as

$$\begin{aligned} n_{\alpha\beta} &= \tfrac{1}{2}\mathrm{tr}(C_\alpha \Sigma^{-1} C_\beta \Sigma^{-1} \otimes P_A^\perp) \\ &= \tfrac{1}{2}\mathrm{tr}(C_\alpha \Sigma^{-1} C_\beta \Sigma^{-1}) \mathrm{tr}(P_A^\perp) \\ &= \tfrac{1}{2}\mathrm{vh}(C_\alpha)^T D^T (\Sigma^{-1} \otimes \Sigma^{-1}) D \mathrm{vh}(C_\beta) \mathrm{tr}(P_A^\perp) \\ &= \tfrac{m-n}{2}\mathrm{vh}(C_\alpha)^T D^T (\Sigma^{-1} \otimes \Sigma^{-1}) D \mathrm{vh}(C_\beta) \end{aligned} \tag{6.17}$$

since $\mathrm{tr}(P_A^\perp) = \mathrm{rank}(P_A^\perp) = m-n$. Since vectors $\mathrm{vh}(C_\alpha)$ and $\mathrm{vh}(C_\beta)$ are canonical unit vectors, it follows that $N = \frac{m-n}{2} D^T (\Sigma^{-1} \otimes \Sigma^{-1}) D$ and thus

$$N^{-1} = \frac{2}{m-n} D^+ (\Sigma \otimes \Sigma) D^{+T} \quad \blacktriangledown \tag{6.18}$$

Substitution of N^{-1} and $\underline{l}$ into equation (6.11) gives, with $D^+ \mathrm{vec}(.) = \mathrm{vh}(.)$ and leaving out the term $D^{+T} D^T$

$$\hat{\underline{\sigma}} = \frac{1}{m-n} \mathrm{vh}(\hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}), \tag{6.19}$$

or finally (cf. equation (4.121))

$$\boxed{\hat{\underline{\Sigma}} = \frac{\hat{\underline{E}}^T Q^{-1} \hat{\underline{E}}}{m-n}.} \tag{6.20}$$

The above formula can therefore be thought of as a generalization of the estimation of the variance of unit weight. Here, instead of a single variance, we estimate a full matrix consisting of variance-covariance components. With the above formulation we now have a straightforward derivation of the estimates for the variance and covariance components in a multivariate linear model given by e.g. Schaffrin (1981).

The covariance matrix of estimators $\underline{\hat{\sigma}} = \mathrm{vh}(\hat{\underline{\Sigma}})$ is given as the inverse of the normal matrix N expressed by equation (6.13) (see equation (4.114))

$$Q_{\hat{\sigma}} = Q_{\mathrm{vh}(\hat{\Sigma})} = N^{-1} = \frac{2}{m-n} D^{+}(\Sigma \otimes \Sigma) D^{+T}. \tag{6.21}$$

In the preceding equation Σ is unknown a-priori and one has to be satisfied with an estimate $\hat{\Sigma}$. This yields an estimate of $Q_{\hat{\sigma}}$, namely $\hat{Q}_{\hat{\sigma}}$. From the preceding covariance matrix, we can now evaluate the (co)variance between element α and β (i.e. between $\sigma_\alpha = \sigma_{ij}$ and $\sigma_\beta = \sigma_{kl}$). In a similar (but in fact reverse) way to what we saw in theorem 4.2, one can easily show that the (co)variances between the estimated variance-covariance components are given as

$$\sigma_{\sigma_\alpha \sigma_\beta} = \sigma_{\sigma_{ij} \sigma_{kl}} = \frac{\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}}{m-n}. \tag{6.22}$$

In the case that $\alpha = \beta$, i.e. $i = k$ and $j = l$, the variances of the variance-covariance components read

$$\sigma^2_{\sigma_\alpha} = \sigma^2_{\sigma_{ij}} = \frac{\sigma_i^2 \sigma_j^2 + \sigma_{ij}^2}{m-n}. \tag{6.23}$$

In a special case that $i = j$, the variances of the variance components read then (cf. equation (4.121))

$$\sigma^2_{\sigma_\alpha} = \sigma^2_{\sigma_i^2} = \frac{2\sigma_i^4}{m-n}. \tag{6.24}$$

From the estimated matrix Σ, one can also compute the mutual correlation coefficient between different observation groups

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i \hat{\sigma}_j}\,. \tag{6.25}$$

This is a nonlinear function of the variables $\hat{\sigma}_{ij}$, $\hat{\sigma}_{ii}$, and $\hat{\sigma}_{jj}$. Based on equations (6.22), (6.23) and (6.24), the covariance matrix of these three estimators is given by

$$Q^{ij}_{\hat{\sigma}} = \mathrm{D}\{\begin{bmatrix} \hat{\sigma}_{ij} \\ \hat{\sigma}_{ii} \\ \hat{\sigma}_{jj} \end{bmatrix}\} = \frac{1}{m-n} \begin{bmatrix} \sigma_{ii}\sigma_{jj} + \sigma_{ij}^2 & 2\sigma_{ii}\sigma_{ij} & 2\sigma_{jj}\sigma_{ij} \\ 2\sigma_{ii}\sigma_{ij} & 2\sigma_{ii}^2 & 2\sigma_{ij}^2 \\ 2\sigma_{jj}\sigma_{ij} & 2\sigma_{ij}^2 & 2\sigma_{jj}^2 \end{bmatrix} \tag{6.26}$$

To obtain the variance of the correlation coefficient $\hat{\rho}_{ij}$, one needs to apply the error propagation law to the linearized form of equation (6.25). The Jacobian (vector) of $\hat{\rho}_{ij}$ is given as (take partial derivatives of equation (6.25) with respect to σ_{ij}, σ_{ii}, and σ_{jj})

$$J = \rho_{ij} \begin{bmatrix} \frac{1}{\sigma_{ij}} & \frac{-1}{2\sigma_{ii}} & \frac{-1}{2\sigma_{jj}} \end{bmatrix} \tag{6.27}$$

Application of the error propagation law yields $\sigma^2_{\hat{\rho}_{ij}} = JQ^{ij}_{\hat{\sigma}} J^T$. Substituting $Q^{ij}_{\hat{\sigma}}$ and J, respectively, from equations (6.26) and (6.27) yields

$$\sigma^2_{\hat{\rho}_{ij}} = \frac{(1-\rho_{ij})^2}{m-n} \tag{6.28}$$

If $\rho_{ij} = 0$, it follows that $\sigma^2_{\hat{\rho}_{ij}} = \frac{1}{m-n}$.

6.4 Multivariate variance-covariance validation

6.4.1 The w-test statistic

Let us consider again the multivariate linear model given in equation (6.7). We are now going to apply the idea of hypothesis testing to the stochastic part of this model. For the w-test statistic, the following null and alternative hypotheses are put forward:

$$\mathrm{H}_o : \mathrm{D}\{\mathrm{vec}(\underline{Y})\} = \sigma_0^2 I \otimes Q \quad \text{versus} \quad \mathrm{H}_a : \mathrm{D}\{\mathrm{vec}(\underline{Y})\} = (\sigma_0^2 I + \nabla C_\alpha) \otimes Q\,, \tag{6.29}$$

where ∇ is an unknown parameter and C_α is an arbitrary $r \times r$ matrix. In this section we will restrict ourselves to the case that σ_0^2 is known. This means, under the null hypothesis, that all groups of observables have the same known variance component σ_0^2 and that the observable groups are uncorrelated. We want now to test this hypothesis using the w-test statistic. For this purpose we can use the special case introduced in section 5.2.5 (see equation (5.75)). We just need to substitute the terms from the multivariate model as $\hat{\underline{e}} \leftarrow \mathrm{vec}(\underline{\hat{E}})$, $Q_y \leftarrow \sigma_0^2 I \otimes Q$, $C_y \leftarrow C_\alpha \otimes Q$, and $P_A^\perp \leftarrow I \otimes P_A^\perp$. By doing so and after a few matrix operations, the following w-test statistic can be given:

$$\underline{w} = \frac{\frac{1}{2\sigma_0^4}\mathrm{tr}(C_\alpha \underline{\hat{E}}^T Q^{-1} \underline{\hat{E}}) - \frac{b}{2\sigma_0^2}\mathrm{tr}(C_\alpha)}{\left[\frac{b}{2\sigma_0^4}\mathrm{tr}(C_\alpha C_\alpha)\right]^{1/2}}. \tag{6.30}$$

Under H_o, the preceding test statistic is distributed as $\underline{w} \sim \sum_k \lambda_k \chi^2_k(1,0) - m_0$, where m_0 is the constant term on the right-hand side of equation (6.30) divided by the denominator, and the eigenvalues λ_k follow from

$$\left| C_\alpha \otimes P_A^\perp - \lambda [2b\,\mathrm{tr}(C_\alpha C_\alpha)]^{1/2} I_{rm} \right| = 0\,. \tag{6.31}$$

We will now consider two important cases which can frequently be used in practice. They are related to the testing of variances and covariances.

Testing of variances

In order to test whether or not the predefined variance components σ_0^2 are appropriate, in equation (6.29), one can take $C_\alpha = c_i c_i^T$. Therefore, $\mathrm{tr}(C_\alpha) = \mathrm{tr}(C_\alpha C_\alpha) = 1$. One can simply show that the w-test statistic (6.30) simplifies to

$$\underline{w}_{ii} = \frac{1}{\sqrt{2b}} \left(\frac{\underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_i}{\sigma_0^2} - b \right). \tag{6.32}$$

In order to obtain the distribution of the preceding test statistic, we need to compute λ_k from equation (6.31). It goes without saying that the only non-zero eigenvalue of $C_\alpha = c_i c_i^T$ is one. Therefore, the b-number of non-zero eigenvalues of $C_\alpha \otimes P_A^\perp$ are one. The eigenvalues λ_k then read $\lambda_1 = \lambda_2 = ... = \lambda_b = \frac{1}{\sqrt{2b}}$. Therefore, under the null hypothesis, the w-test statistic is distributed as

$$\underline{w}_{ii} \sim \frac{1}{\sqrt{2b}} \left(\chi^2(b,0) - b \right). \tag{6.33}$$

Denoting $\underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_i = b \times \underline{\hat{\sigma}}_i^2$, the preceding equations yield

$$\boxed{\frac{\underline{\hat{\sigma}}_i^2}{\sigma_0^2} \sim \frac{\chi^2(b,0)}{b}.} \tag{6.34}$$

This is very similar to and in fact a generalization of the distribution of the famous variance of unit weight (see example 5.1). Here we were concerned about the variance component of only one group of observables. We may also want to consider a few groups of observables together. For example one can define $C_\alpha = c_i c_i^T + c_j c_j^T$ (if two groups are assumed to have an identical variance component) or $C_\alpha = I$ (if we consider all groups together).

NOTE 6.1 In equation (6.29) we assumed that σ_0^2 is known. It is still possible to perform hypotheses testing even though the variance component σ^2 is unknown. In this case equivalent expressions for equations (6.32) and (6.33) can be given as (employ equations (5.81) and (5.82))

$$\underline{w}_{ii} = \frac{1}{\sqrt{2br(r-1)}} \left(\frac{r \underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_i - \sum_{k=1}^r \underline{\hat{e}}_k^T Q^{-1} \underline{\hat{e}}_k}{\sigma^2} \right) ; \quad r \geq 2, \tag{6.35}$$

and

$$\underline{w}_{ii} \sim \frac{1}{\sqrt{2br(r-1)}} \Big((r-1)\chi_1^2(b,0) - \chi_2^2(b(r-1),0) \Big), \tag{6.36}$$

respectively. Note that the value of the test statistic (but not its distribution) depends on the unknown variance component σ^2 for which an estimate can be used as $\hat{\sigma}^2 = (\sum_{k=1}^r \hat{e}_k^T Q^{-1} \hat{e}_k)/br$. ⊡

Testing of covariances

We can also test whether or not additional covariance components are needed to be introduced in the stochastic model. For this purpose we may take $C_\alpha = c_i c_j^T + c_j c_i^T$ to see

if the covariance component between groups i and j is in fact zero. With this structure we have $\mathrm{tr}(C_\alpha) = 0$ and $\mathrm{tr}(C_\alpha C_\alpha) = 2$. One can show that the w-test statistic (6.30) simplifies to

$$\underline{w}_{ij} = \frac{1}{\sqrt{b}} \left(\frac{\hat{\underline{e}}_i^T Q^{-1} \hat{\underline{e}}_j}{\sigma_0^2} \right). \tag{6.37}$$

In order to determine the distribution of the above test statistic, we need to compute λ_k from equation (6.31). One can simply show that the two non-zero eigenvalues of C_α are ± 1. Therefore, only $2b$-number of eigenvalues of $C_\alpha \otimes P_A^\perp$ are non zero. They follow as $\lambda_1 = \lambda_2 = ... = \lambda_b = \frac{1}{2\sqrt{b}}$ and $\lambda_{b+1} = \lambda_{b+2} = ... = \lambda_{2b} = \frac{-1}{2\sqrt{b}}$. The preceding w-test statistic is then distributed as

$$\underline{w}_{ij} \sim \frac{1}{2\sqrt{b}} \left(\chi_1^2(b,0) - \chi_2^2(b,0) \right), \tag{6.38}$$

where $\chi_1^2(b,0)$ and $\chi_2^2(b,0)$ are two independent chi-squared distributions with b degrees of freedom. Denoting $\hat{\underline{e}}_i^T Q^{-1} \hat{\underline{e}}_j = b \times \hat{\underline{\sigma}}_{ij}$, the preceding equations yield

$$\hat{\underline{\rho}}_{ij} = \frac{\hat{\underline{\sigma}}_{ij}}{\sigma_0^2} \sim \frac{1}{2b} \left(\chi_1^2(b,0) - \chi_2^2(b,0) \right), \tag{6.39}$$

with $\hat{\underline{\rho}}_{ij}$ the correlation coefficient between groups i and j. There are standard procedures for computing the distribution of quadratic forms in normal variables. The reader is referred to e.g. Imhof (1961). Based on the *central limit theorem* for independent and identically distributed data, when the redundancy b is large enough, one can approximate the chi-squared distribution by a normal one. Taking these approximations into account, the preceding equation reads

$$\hat{\underline{\rho}}_{ij} = \frac{\hat{\underline{\sigma}}_{ij}}{\sigma_0^2} \simeq \mathsf{N}(0, 1/b). \tag{6.40}$$

Here we were only concerned about the covariance component between two groups of observables. We may also want to consider a few groups together. For example we can define $C_\alpha = c_i c_j^T + c_j c_i^T + c_i c_k^T + c_k c_i^T$ (three groups).

6.4.2 The v-test statistic

The idea here is to test whether or not it is possible to reduce the number of covariance components. To perform the *(co)variance significance test* (v-test), we can, for instance, test the following null hypotheses versus the alternative ones:

1. $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ii} = \sigma_{ii}^0$ versus $\mathrm{H}_a : Q_y = \Sigma \otimes Q$,
2. $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ij} = \sigma_{ij}^0$ versus $\mathrm{H}_a : Q_y = \Sigma \otimes Q$,
3. $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ii} = \sigma_{kk}$ versus $\mathrm{H}_a : Q_y = \Sigma \otimes Q$,
4. $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ij} = \sigma_{ik}$ versus $\mathrm{H}_a : Q_y = \Sigma \otimes Q$,
5. $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ij} = \sigma_{kl}$ versus $\mathrm{H}_a : Q_y = \Sigma \otimes Q$.

To do the testing we can make use of equations (5.133) and (5.134) for v-test statistic and the corresponding eigenvalues associated with its distribution, respectively. However, since the constant terms c_0 and $d^T N^{-1} d$ do not have any effect on testing results, we will ignore them both in the test statistic and its distribution. Therefore, equation (5.133) can now be reformulated as

$$\underline{v} = d_\alpha \underline{\hat{\sigma}}_\alpha = \underline{\hat{e}} Q_y^{-1} [\frac{1}{2} d_\alpha n_{\alpha\beta}^{-1} Q_\beta] Q_y^{-1} \underline{\hat{e}}\,, \tag{6.41}$$

which is distributed as $\underline{v} \sim \sum_k \lambda_k \chi_k^2$. The eigenvalues associated with this distribution can be given by the following equation:

$$\left| [d_\alpha n_{\alpha\beta}^{-1} Q_\beta] Q_y^{-1} P_A^\perp - 2\lambda I \right| = 0. \tag{6.42}$$

Note that this new test statistic, in general, is no longer of mean *zero* and variance *one*.

To obtain the v-test statistics for the multivariate linear model we just need to substitute the preceding terms as follows: $\hat{e} \leftarrow \mathrm{vec}(\underline{\hat{E}})$, $Q_y \leftarrow \Sigma \otimes Q$, $Q_\beta \leftarrow C_\beta \otimes Q$, and $P_A^\perp \leftarrow I \otimes P_A^\perp$. In the following we will make use of the canonical unit vector for d_α, e.g. $d_\alpha = \mathrm{vh}(c_i c_i^T)$ or $d_\alpha = \mathrm{vh}(c_i c_j^T + c_j c_i^T)$ if $i \neq j$. When we specify values for α or correspondingly for i and j, the term $d_\alpha n_{\alpha\beta}^{-1}$ simplifies to $n_{\alpha\beta}^{-1}$ where the only free index is β (it is in fact the α^{th} row of N^{-1}). Note that β becomes a summation index when multiplied with Q_β. We will first evaluate the term $d_\alpha n_{\alpha,\beta}^{-1} Q_\beta$. For this purpose we use the fixed index as $\alpha \equiv i, j$ and the summation index as $\beta \equiv k, l$. Therefore, one obtains

$$[d_\alpha n_{\alpha,\beta}^{-1} Q_\beta] = [n_{ij,kl}^{-1} Q_{kl}] = [n_{ij,kl}^{-1} c_k c_l^T] \otimes Q = \frac{1}{b} [(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T] \otimes Q\,. \tag{6.43}$$

Equation (6.42) can therefore be reformulated for the multivariate model as

$$\left| [(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T] \Sigma^{-1} \otimes P_A^\perp - 2b\lambda I_{rm} \right| = 0\,, \tag{6.44}$$

in which the us^{th} entry of the matrix $[(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T]$, i.e. the element of the u^{th} row and s^{th} column, is given as

$$[(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T] = \sigma_{ui}\sigma_{js} + \sigma_{uj}\sigma_{is}\,. \tag{6.45}$$

When post-multiplied with Σ^{-1}, it follows that

$$[(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T] \Sigma^{-1} = \sigma_{ui}\sigma_{js}\sigma_{sv}^{-1} + \sigma_{uj}\sigma_{is}\sigma_{sv}^{-1} = \sigma_{ui}\delta_{jv} + \sigma_{uj}\delta_{iv}\,, \tag{6.46}$$

which simplifies to

$$[(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T] \Sigma^{-1} = \begin{cases} \sigma_{uj} & \text{if} \quad v = i; \\ \sigma_{ui} & \text{if} \quad v = j; \\ 0 & \text{if} \quad v \neq i, j. \end{cases} \tag{6.47}$$

Therefore, the matrix $[(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) c_k c_l^T] \Sigma^{-1}$ is a zero matrix except for two columns, namely i and j, with entries of the j^{th} and i^{th} columns of Σ, respectively (the i^{th} and j^{th} columns have been interchanged). The two non-zero eigenvalues of this matrix are given by the eigenvalues of the following matrix:

$$\Lambda = \begin{bmatrix} \sigma_{ij} & \sigma_{ii} \\ \sigma_{jj} & \sigma_{ij} \end{bmatrix} = \begin{bmatrix} \sigma_{ij} & \sigma_i^2 \\ \sigma_j^2 & \sigma_{ij} \end{bmatrix}. \tag{6.48}$$

The eigenvalues of the preceding matrix can easily be computed as

$$\lambda_{\max} = \sigma_{ij} + \sigma_i\sigma_j; \quad \lambda_{\min} = \sigma_{ij} - \sigma_i\sigma_j. \tag{6.49}$$

Since the matrix $P_A^{\perp}$ is a projector it also has b-number of non-zero eigenvalues i.e. $\lambda'_1 = ... = \lambda'_b = 1$. Equation (6.44) has therefore $2b$-number of non-zero eigenvalues. In the following we will consider different testing hypotheses (different possibilities for i and j) for which we present test statistics and their corresponding distributions.

Testing of $\sigma_{ii} = \sigma_{ii}^0$

We consider here the null hypothesis $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ii} = \sigma_{ii}^0$ versus the alternative one $\mathrm{H}_a : Q_y = \Sigma \otimes Q$. Under the null hypothesis it is assumed that the value of variance component σ_{ii} is to be equal to the known value σ_{ii}^0. To see how the v-test statistic and its distribution of the above testing hypotheses look like, one can take $d_\alpha = \mathrm{vh}(c_i c_i^T)$. It goes without saying that the test statistic (6.41) can be written as

$$\underline{v} = \mathrm{vh}(c_i c_i^T)\mathrm{vh}(\hat{\underline{\Sigma}}) = \hat{\underline{\sigma}}_{ii} = \frac{\hat{\underline{e}}_i^T Q^{-1} \hat{\underline{e}}_i}{b}. \tag{6.50}$$

In order to determine the distribution of the aforementioned statistic, i.e. $\underline{v} = \hat{\underline{\sigma}}_{ii}$, we need to compute the eigenvalues λ_k from equation (6.44) as (note that here $d_\alpha = \mathrm{vh}(c_i c_i^T)$)

$$\left|[\sigma_{ik}\sigma_{il} c_k c_l^T]\Sigma^{-1} \otimes P_A^{\perp} - b\lambda I_{rm}\right| = 0. \tag{6.51}$$

In a similar way to equation (6.47), one can show that

$$[\sigma_{ik}\sigma_{il} c_k c_l^T]\Sigma^{-1} = \begin{cases} \sigma_{ui} & \text{if} \quad v = i; \\ 0 & \text{if} \quad v \neq i. \end{cases} \tag{6.52}$$

Therefore, the matrix $[\sigma_{ik}\sigma_{il} c_k c_l^T]\Sigma^{-1}$ is a zero matrix except for its i^{th} column with entries of the i^{th} column of Σ. The only non-zero eigenvalue of this matrix is σ_{ii}. The matrix $P_A^{\perp}$ is a projector with non-zero eigenvalues $\lambda'_1 = ... = \lambda'_b = 1$. With these in mind, the non-zero eigenvalues λ of equation (6.51) are given as

$$\lambda_1 = \lambda_2 = ... = \lambda_b = \frac{\sigma_{ii}}{b}. \tag{6.53}$$

Under the null hypothesis, the distribution of the test statistic $\hat{\underline{\sigma}}_{ii}$ then reads

$$\boxed{\hat{\underline{\sigma}}_{ii} \sim \frac{\sigma_{ii}^0}{b}\chi^2(b,0) \Rightarrow \frac{\hat{\sigma}_{ii}}{\sigma_{ii}^0} \sim \frac{\chi^2(b,0)}{b}.} \tag{6.54}$$

The above result can be thought of as being a generalization of the testing of the *variance of unit weight*. This is an interesting result and very simple indeed as the $\chi^2(b,0)$ is a well-known central chi-squared distribution. Note that we have obtained the same results using the w-test statistic (see equation (6.34)).

Testing of $\sigma_{ij} = \sigma_{ij}^0$

We consider here the null hypothesis $\mathrm{H}_o : Q_y = \Sigma \otimes Q,\ \sigma_{ij} = \sigma_{ij}^0$ versus the alternative one $\mathrm{H}_a : Q_y = \Sigma \otimes Q$. Under the null hypothesis the value of covariance component σ_{ij} is equal to the known value σ_{ij}^0. To see how the v-test statistic and its distribution of the above testing hypotheses look like, one can take $d_\alpha = \mathrm{vh}(c_i c_j^T + c_j c_i^T)$. One can simply show that the v-test statistic is written as follows:

$$\boxed{\underline{v} = \underline{\hat{\sigma}}_{ij} = \frac{\underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_j}{b}.} \tag{6.55}$$

In order to determine the distribution of the aforementioned statistic, we need to compute the eigenvalues of equation (6.44). We have, in fact, obtained these values as those presented in equation (6.49). When taking into account the projector $P_A^\perp$, one obtains

$$\lambda_1 = \lambda_2 = ... = \lambda_b = \frac{\sigma_{ij} + \sigma_i \sigma_j}{2b};\quad \lambda_{b+1} = \lambda_{b+2} = ... = \lambda_{2b} = \frac{\sigma_{ij} - \sigma_i \sigma_j}{2b}. \tag{6.56}$$

Under the null hypothesis, the test statistic $\underline{\hat{\sigma}}_{ij}$ is thus distributed as

$$\underline{\hat{\sigma}}_{ij} \sim \frac{1}{2b}\left[(\sigma_{ij}^0 + \sigma_i \sigma_j)\chi_1^2(b,0) + (\sigma_{ij}^0 - \sigma_i \sigma_j)\chi_2^2(b,0)\right], \tag{6.57}$$

where $\chi_1^2(b,0)$ and $\chi_2^2(b,0)$ are two independent chi-squared distributions with b degrees of freedom. If the redundancy b is large enough, we may approximate the chi-squared distributions with normal ones. In this case the preceding distribution approximately yields

$$\underline{\hat{\sigma}}_{ij} \simeq \sigma_i \sigma_j \mathsf{N}\left(\rho_{ij}^0, \frac{1 + {\rho_{ij}^0}^2}{b}\right), \tag{6.58}$$

or

$$\boxed{\underline{\hat{\rho}}_{ij} \simeq \mathsf{N}\left(\rho_{ij}^0, \frac{1 + {\rho_{ij}^0}^2}{b}\right),} \tag{6.59}$$

where $\rho_{ij}^0 = \frac{\sigma_{ij}^0}{\sigma_i \sigma_j}$ and $\underline{\hat{\rho}}_{ij} = \frac{\underline{\hat{\sigma}}_{ij}}{\sigma_i \sigma_j}$ are the correlation coefficients. If we now want to test the significance of the correlation coefficient, i.e. under null hypothesis if $\rho_{ij}^0 = 0$, it then follows that

$$\underline{\hat{\rho}}_{ij} \simeq \mathsf{N}(0, 1/b), \tag{6.60}$$

which is identical to what we found in equation (6.40).

Testing of $\sigma_{ii} = \sigma_{kk}$

The goal here is to test whether or not the variance component of group i is the same as that of group k. In a similar way to what was done previously, one can define the vector d as $d_\alpha = \mathrm{vh}(c_i c_i^T - c_k c_k^T)$. The test statistic simplifies to

$$\boxed{\underline{v} = \underline{\hat{\sigma}}_{ii} - \underline{\hat{\sigma}}_{kk} = \frac{1}{b}[\underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_i - \underline{\hat{e}}_k^T Q^{-1} \underline{\hat{e}}_k].} \tag{6.61}$$

The eigenvalues associated with the distribution, apart from those of $P_A^{\perp}$, i.e. b-number of *ones*, of $\underline{v}$ can be obtained from

$$\left| \begin{bmatrix} \sigma_{ii} & -\sigma_{ik} \\ \sigma_{ik} & -\sigma_{kk} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0, \tag{6.62}$$

which gives

$$\lambda_{\max} = \frac{1}{2}\left[\sigma_{ii} - \sigma_{kk} + \sqrt{(\sigma_{ii} + \sigma_{kk})^2 - 4\sigma_{ik}^2}\right], \tag{6.63}$$

and

$$\lambda_{\min} = \frac{1}{2}\left[\sigma_{ii} - \sigma_{kk} - \sqrt{(\sigma_{ii} + \sigma_{kk})^2 - 4\sigma_{ik}^2}\right], \tag{6.64}$$

Under the null hypothesis we have $\sigma_{ii} = \sigma_{kk}$. Therefore, the v-test statistic is distributed as (take into account the projector $P_A^{\perp}$)

$$\underline{v} \sim \frac{\sqrt{\sigma_{ii}^2 - \sigma_{ik}^2}}{b}\left[\chi_1^2(b,0) + \chi_2^2(b,0)\right]. \tag{6.65}$$

If we approximate $\chi_1^2(b,0)$ and $\chi_2^2(b,0)$ by two independent normal distributions $\mathsf{N}_1(b,2b)$ and $\mathsf{N}_2(b,2b)$, we obtain

$$\boxed{\underline{v} \simeq \mathsf{N}\left(0, \frac{4(\sigma_{ii}^2 - \sigma_{ik}^2)}{b}\right).} \tag{6.66}$$

The maximum possible variance of the v-test statistic occurs when the covariance component between groups i and k is not significant, i.e. when $\sigma_{ik} = 0$. If σ_{ii} and σ_{ik} are unknown a-priori, one can have recourse to estimates of those as $\sigma_{ii} = \frac{1}{2}(\hat{\sigma}_{ii} + \hat{\sigma}_{kk})$ and $\sigma_{ik} = \hat{\sigma}_{ik}$, respectively.

Testing of $\sigma_{ij} = \sigma_{ik}$

Here we want to test whether or not the covariance component σ_{ij} equals σ_{ik} (i is the common index). To perform this test, one can define $d_\alpha = \mathrm{vh}(c_i c_j^T + c_j c_i^T - c_i c_k^T - c_k c_i^T)$ from which it can be concluded that

$$\boxed{\underline{v} = \underline{\hat{\sigma}}_{ij} - \underline{\hat{\sigma}}_{ik} = \frac{1}{b}[\underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_j - \underline{\hat{e}}_i^T Q^{-1} \underline{\hat{e}}_k].} \tag{6.67}$$

The matrix Λ, a matrix whose eigenvalues are associated with the distribution of the v-test statistic, is then given by

$$\Lambda = \begin{bmatrix} \sigma_{ij} - \sigma_{ik} & \sigma_{ii} & -\sigma_{ii} \\ \sigma_{jj} - \sigma_{jk} & \sigma_{ij} & -\sigma_{ij} \\ \sigma_{jk} - \sigma_{kk} & \sigma_{ik} & -\sigma_{ik} \end{bmatrix}, \tag{6.68}$$

in which the third column is a multiple of the second column. Therefore, the preceding matrix is of rank 2. The eigenvalues of Λ then can be given as

$$\lambda_{\max} = \sigma_{ij} - \sigma_{ik} + \sigma_i \sqrt{\sigma_{jj} + \sigma_{kk} - 2\sigma_{jk}}, \tag{6.69}$$

and

$$\lambda_{\min} = \sigma_{ij} - \sigma_{ik} - \sigma_i\sqrt{\sigma_{jj} + \sigma_{kk} - 2\sigma_{jk}}; \quad \lambda_3 = 0. \tag{6.70}$$

Therefore, under the null hypothesis, the v-test statistic is distributed according to

$$\underline{v} \sim \frac{\sigma_i\sqrt{\sigma_{jj} + \sigma_{kk} - 2\sigma_{jk}}}{2b}\left[\chi_1^2(b,0) - \chi_2^2(b,0)\right]. \tag{6.71}$$

which can be approximated by a normal distribution as

$$\boxed{\underline{v} \simeq \mathsf{N}\left(0, \frac{\sigma_{ii}(\sigma_{jj} + \sigma_{kk} - 2\sigma_{jk})}{b}\right).} \tag{6.72}$$

If the (co)variance components σ_{ii}, σ_{jj}, σ_{kk}, and σ_{jk} are unknown a-priori, one can rely on the estimates $\hat{\sigma}_{ii}$, $\hat{\sigma}_{jj}$, $\hat{\sigma}_{kk}$, and $\hat{\sigma}_{jk}$, respectively.

Testing of $\sigma_{ij} = \sigma_{kl}$

The goal here is to see whether or not the covariance component σ_{ij} equals σ_{kl}. To perform this test, one can define $d_\alpha = \mathrm{vh}(c_i c_j^T + c_j c_i^T - c_k c_l^T - c_l c_k^T)$ from which it can be concluded that

$$\boxed{\underline{v} = \underline{\hat{\sigma}}_{ij} - \underline{\hat{\sigma}}_{kl} = \frac{1}{b}[\underline{\hat{e}}_i^T Q^{-1}\underline{\hat{e}}_j - \underline{\hat{e}}_k^T Q^{-1}\underline{\hat{e}}_l].} \tag{6.73}$$

The matrix Λ, a matrix whose eigenvalues are associated with the distribution of the preceding test statistic reads then

$$\Lambda = \begin{bmatrix} \sigma_{ij} & \sigma_{ii} & -\sigma_{il} & -\sigma_{ik} \\ \sigma_{jj} & \sigma_{ij} & -\sigma_{jl} & -\sigma_{jk} \\ \sigma_{jk} & \sigma_{ik} & -\sigma_{kl} & -\sigma_{kk} \\ \sigma_{jl} & \sigma_{il} & -\sigma_{ll} & -\sigma_{kl} \end{bmatrix}, \tag{6.74}$$

which, in general, is of rank 4. Denoting the eigenvalues of Λ as λ_1, λ_2, λ_3 and λ_4, the test statistic $\underline{v}$ is distributed as

$$\underline{v} \sim \frac{1}{2b}\left[\lambda_1\chi_1^2(b,0) + \lambda_2\chi_2^2(b,0) + \lambda_3\chi_3^2(b,0) + \lambda_4\chi_4^2(b,0)\right]. \tag{6.75}$$

Under the null hypothesis and when $\sigma_{il} = \sigma_{jk} = \sigma_{jl} = \sigma_{ik} = 0$ it follows that

$$\begin{aligned} \lambda_1 &= \quad (\sigma_{ij} + \sigma_i\sigma_j); \qquad \lambda_2 = \quad (\sigma_{ij} - \sigma_i\sigma_j); \\ \lambda_3 &= -(\sigma_{ij} + \sigma_k\sigma_l); \qquad \lambda_4 = -(\sigma_{ij} - \sigma_k\sigma_l). \end{aligned} \tag{6.76}$$

Again, if we approximate the chi-squared distributions with normal ones, equation (6.75) yields

$$\boxed{\underline{v} \simeq \mathsf{N}\left(0, \frac{\sum_{u=1}^4 \lambda_u^2}{2b}\right).} \tag{6.77}$$

If in addition $\sigma_i = \sigma_j = \sigma_k = \sigma_l$, the preceding distribution reads

$$\underline{v} \simeq \mathsf{N}\left(0, \frac{2(\sigma_{ii} + \sigma_{ij})^2}{b}\right). \tag{6.78}$$

6.4.3 Overall model test

Let us now obtain the overall model test for the multivariate model. Under the null hypothesis the stochastic model reads $Q_{\mathrm{vec}(Y)} = \Sigma \otimes Q$, in which we only consider that Σ is known. Under the alternative hypothesis the covariance matrix is fully unknown (see e.g. equation (5.139)).

$\Sigma = \Sigma_0$ **known** In multivariate model we need to obtain $\underline{T}_{\mathrm{df}}$ where df is defined as $df = \frac{rb(rb+1)}{2}$. We can extend example 5.7 and equation (5.145) for a multivariate model. We should substitute $\hat{\underline{e}} \leftarrow \mathrm{vec}(\hat{\underline{E}})$, $Q_y^{-1} \leftarrow \Sigma^{-1} \otimes Q^{-1}$, and $b \leftarrow rb$. One will then obtain

$$\underline{T}_{\mathrm{df}} = \frac{1}{2}\mathrm{tr}(\Sigma^{-1}\hat{\underline{E}}^T Q^{-1}\hat{\underline{E}})\,\mathrm{tr}(\Sigma^{-1}\hat{\underline{E}}^T Q^{-1}\hat{\underline{E}}) - \mathrm{tr}(\Sigma^{-1}\hat{\underline{E}}^T Q^{-1}\hat{\underline{E}}) + \frac{rb}{2}, \tag{6.79}$$

which can have a complicated distribution. However, if we use the function $\underline{T}_{rb} = \sqrt{2\underline{T}_{\mathrm{df}} - rb + 1} + 1$, it then follows that

$$\underline{T}_{rb} = \mathrm{tr}(\Sigma^{-1}\hat{\underline{E}}^T Q^{-1}\hat{\underline{E}}), \tag{6.80}$$

which is distributed as $\underline{T}_{rb} \sim \chi^2(rb, 0)$.

7 GPS Geometry-Free Model

7.1 Introduction

The geometry-free observation model is one of the simplest approaches to integer GPS double differenced ambiguity estimation. This linear model dispenses with geometric a priori information and parameterizes the GPS observations simply in carrier-phase ambiguities and satellite-receiver ranges. Although simple, this parameterization does imply that an adjustment according to the geometry-free observation model does not allow a direct estimation of the baseline coordinates. It is therefore not the common mode of operation for surveying applications.

Of the two basic GPS observation types, namely carrier-phase and pseudorange observations, phase observations are greatly the more precise. Although more precise, carrier-phase observations are inherently biased by unknown integer numbers of carrier wavelengths which are called carrier-phase *ambiguities*. Therefore, the favorable precision of the carrier-phase observations can only be fully exploited for relative positioning if the ambiguities are estimated and kept *fixed* at their integer values. In this case, the carrier-phase observations are corrected on the basis of these estimates (see later on).

In spite of the fact that the geometry-free approach is less well suited for surveying applications, its use does offer some advantages (see next section). Foremost of these advantages is the ease with which the geometry-free observation model can be implemented in computer code. This ease stems from the linearity of the observation model and its independence of satellite orbit information. In addition, the geometry-free observation model is also very versatile, as it will allow the estimation of ambiguities even if observations to only two satellites are available and even if both receivers are moving. Moreover, ambiguities estimated with the geometry-free model are also known to be free from residual tropospheric delay in the observations, as these delays are automatically lumped with the satellite-receiver ranges. Therefore, for some applications, like data analysis and ionosphere monitoring, the geometry-free approach does have its appeal. Because of these advantages, the GPS geometry-free model will be used in this chapter as a favorable model for estimation of (co)variance components via the LS-VCE method. In the following section we consider the geometry-free approach in more detail. We start with a description of the geometry-free observation model.

7.2 Functional and stochastic models

The geometry-free observation model consists of two parts: the functional model and the stochastic model. The functional model relates the observations to the parameters of interest whereas the stochastic model describes the precision and the mutual correlation be-

tween the observations. The geometry-free functional model is based on the non-linearized double differenced (DD) dual frequency pseudorange and carrier-phase observation equations. Consider two receivers r and j simultaneously observing the same satellites s and k. Neglecting the dispersive DD ionospheric delay, the double differenced dual frequency pseudorange observation equations read (Jonkman, 1998)

$$\underline{p}^{sk}_{rj,L}(t_i) = \rho^{sk}_{rj,L}(t_i) + \underline{e}^{sk}_{rj,L}(t_i)\,, \tag{7.1}$$

where $(.)^{sk}_{rj}$ is an abbreviation for $(.)^{k}_{rj} - (.)^{s}_{rj} = (.)^{k}_{j} - (.)^{k}_{r} - ((.)^{s}_{j} - (.)^{s}_{r})$, p denotes the 'observed' DD pseudoranges on the L1 or L2 frequency, e.g. C1 and P2, ρ denotes the combination of all non-dispersive effects, L is either L1 or L2 frequency, e denotes the pseudorange measurement errors on the L1 or L2 frequency, and t_i indicates the time instant or epoch to which the observations refer. The most important difference between the phase and the code observations is the entry of an integer DD carrier-phase ambiguity in the phase observation equations. Expressed in units of distance (m) rather then in cycles, the DD carrier-phase observation equations read

$$\underline{\phi}^{sk}_{rj,L}(t_i) = \rho^{sk}_{rj,L}(t_i) + \lambda_L a^{sk}_{rj,L} + \underline{\epsilon}^{sk}_{rj,L}(t_i)\,, \tag{7.2}$$

where ϕ denotes the 'observed' DD carrier-phase on the L1 or L2 frequency, a denotes the integer carrier-phase ambiguities on the L1 or L2 frequency, and ϵ denotes the carrier-phase measurement errors on the L1 or L2 frequency. The dual frequency DD pseudorange and carrier-phase observation equations can be summarized in the following linear system of equations:

$$\mathsf{E}\{\begin{bmatrix} \underline{p}^{sk}_{rj,1}(t_i) \\ \underline{p}^{sk}_{rj,2}(t_i) \\ \underline{\phi}^{sk}_{rj,1}(t_i) \\ \underline{\phi}^{sk}_{rj,2}(t_i) \end{bmatrix}\} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & \lambda_1 & 0 \\ 1 & 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \rho^{sk}_{rj}(t_i) \\ a^{sk}_{rj,1} \\ a^{sk}_{rj,2} \end{bmatrix}, \tag{7.3}$$

where E denotes the expectation operator; the expectation of the measurement errors is assumed to be zero. This system of observation equations is referred to as the single epoch geometry-free functional model. It refers to two receivers simultaneously observing two satellites. It is of course possible to include, in the functional model, the observations to more than two satellites or even the observations of more than two receivers. In this study, however, we will restrict ourselves to the observations of two receivers and to more than two satellites. The above functional model over K epochs can be summarized in a convenient vector-matrix notation as (still for two satellites)

$$E\{\begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \\ \underline{\phi}_1 \\ \underline{\phi}_2 \end{bmatrix}\} = \begin{bmatrix} I & 0 & 0 \\ I & 0 & 0 \\ I & u\lambda_1 & 0 \\ I & 0 & u\lambda_2 \end{bmatrix} \begin{bmatrix} \rho \\ a_1 \\ a_2 \end{bmatrix} \tag{7.4}$$

with $u = [1,\ 1, \cdots, 1]^T$. If one, in one step, obtains a least squares solution of the above, fixes the double differenced carrier-phase ambiguities to their integer values by using the least-squares ambiguity decorrelation adjustment (LAMBDA) method (see Teunissen (1993, 1995); Verhagen (2005); Odijk (2002)) and ignores the stochastic behavior

of the integer valued ambiguities, one, in the next step, will be able to simplify the above observation equations by subtracting from the observation vector a constant vector, namely $\left[\begin{array}{cccc} 0 & 0 & \lambda_1 \breve{a}_1 u^T & \lambda_2 \breve{a}_2 u^T \end{array}\right]^T$, with $\breve{a}_1$ and $\breve{a}_2$ the least squares integer valued ambiguities. In this case, only the identity matrices of the design matrix will remain on the right hand side of the preceding equation.

For a single observation epoch and dual frequency undifferenced pseudorange and carrier-phase observations, a very simple stochastic model reads

$$\mathrm{D}\{\underline{y}\} = \left[\begin{array}{cccc} \sigma^2_{p_1} & 0 & 0 & 0 \\ 0 & \sigma^2_{p_2} & 0 & 0 \\ 0 & 0 & \sigma^2_{\phi_1} & 0 \\ 0 & 0 & 0 & \sigma^2_{\phi_2} \end{array}\right], \tag{7.5}$$

where $\sigma^2_{p_i}$ and $\sigma^2_{\phi_i},\ i = 1, 2$ denote the variances of the undifferenced pseudorange and carrier-phase observables. At this moment, mutual correlation between the observations is assumed to be absent. Moreover, for the sake of simplicity, in some applications, the observations on the L1 and L2 frequency are assumed to have a constant precision, i.e. $\sigma^2_{p_1} = \sigma^2_{p_2} = \sigma^2_p$ and $\sigma^2_{\phi_1} = \sigma^2_{\phi_2} = \sigma^2_\phi$. This however may not be the case in general (see next section); with these assumptions the stochastic model is very poor and immature. The purpose of this chapter is to come up with a realistic and adequate covariance matrix of GPS observables in case of a linear and simple geometry-free observation model. We apply the least squares variance component estimation (LS-VCE) method to the stochastic part of the geometry-free observation model. The construction of the covariance matrix (for undifferenced observables) starts from a scaled unit matrix per observation type and takes place in different steps. In this application, the following six suppositions regarding the noise of the undifferenced observables are considered:

1. to handle the ill-posedness of the VCE model (specifically for geometry-free model), the precision of the carrier-phase observables on L1 and L2 is assumed to be identical,

2. the precision of the pseudorange observations is assumed to be different for different observation types (on different frequencies),

3. the observables on the L1 and L2 frequency may or may not be correlated,

4. satellite elevation dependence of the observables precision will in part be considered,

5. time correlation of the observables is assumed to be absent (this may not be true for some observation types for particular receivers), and

6. the correlation between different channels/satellites is assumed to be absent for all observation types.

For detailed discussions on the noise characteristics of the GPS observables we refer to Langley (1997); Tiberius et al. (1999).

7.3 Sophisticated stochastic models

7.3.1 Satellite elevation dependence

If, in addition to the satellites s and k, we now assume that the receivers r and j are also simultaneously tracking the satellite l, the other DD pseudorange reads

$$\underline{p}^{sl}_{rj,L}(t_i) = \rho^{sl}_{rj,L}(t_i) + \underline{e}^{sl}_{rj,L}(t_i)\,, \tag{7.6}$$

in which satellite s has been considered as reference. Assuming that the correlation between channels is absent for each observation type, one can consider different precisions for different channels. The covariance matrix of the undifferenced pseudorange observables then reads (two receivers)

$$Q_y = \mathrm{diag}(\sigma^2_{p_s}\ \sigma^2_{p_s}\ \sigma^2_{p_k}\ \sigma^2_{p_k}\ \sigma^2_{p_l}\ \sigma^2_{p_l})\,. \tag{7.7}$$

Applying the error propagation law to the DD pseudoranges yields

$$Q^D_y = JQ_yJ^T = 2\begin{bmatrix} \sigma^2_{p_s}+\sigma^2_{p_k} & \sigma^2_{p_s} \\ \sigma^2_{p_s} & \sigma^2_{p_s}+\sigma^2_{p_l} \end{bmatrix}, \tag{7.8}$$

where

$$J = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix}. \tag{7.9}$$

One can simplify the above covariance matrix if one assumes the same precision for different channels, i.e. $\sigma^2_{p_s} = \sigma^2_{p_k} = \sigma^2_{p_l} = \sigma^2_p$. Therefore, if one neglects the satellite elevation dependence of GPS observation precision, one can estimate only one variance component σ_1 (a scale) for each observation type

$$Q^D_y = \begin{bmatrix} 4\sigma^2_p & 2\sigma^2_p \\ 2\sigma^2_p & 4\sigma^2_p \end{bmatrix}\sigma_1\,, \tag{7.10}$$

when the initial value of the pseudorange precision, $\sigma_p = 30\,\mathrm{cm}$, has been included in the cofactor matrix. In general, however, one needs to estimate 3 variance components for each observation type

$$Q^D_y = \begin{bmatrix} 2\sigma^2_p & 2\sigma^2_p \\ 2\sigma^2_p & 2\sigma^2_p \end{bmatrix}\sigma_s + \begin{bmatrix} 2\sigma^2_p & 0 \\ 0 & 0 \end{bmatrix}\sigma_k + \begin{bmatrix} 0 & 0 \\ 0 & 2\sigma^2_p \end{bmatrix}\sigma_l \tag{7.11}$$

In other words, instead of estimating just one variance component, one can estimate different variance components for different satellites (or channels) observed at different elevation angles. The above structure can be employed for all observation types (e.g. for code and phase observations on L1 and L2).

7.3.2 Correlation between observation types

We now want to consider the correlation between different observation types. For the sake of simplicity, one can ignore the satellite elevation dependence of the observables precision. When dealing with four observation types, one needs to estimate 10 (co)variance

components, namely 4 variances and 6 covariances. As an example, consider the following covariance matrix for two types of undifferenced observations (e.g. single frequency), namely pseudorange and carrier phase (as before, 2 receivers simultaneously tracking 3 satellites)

$$Q_y = \begin{bmatrix} \sigma_p^2\, I_6 & \sigma_{p\phi} I_6 \\ \sigma_{p\phi} I_6 & \sigma_\phi^2\, I_6 \end{bmatrix}. \tag{7.12}$$

From the above 12 observations one can create 4 DD equations; 2 DD pseudorange and 2 DD carrier-phase observation equations. Applying the error propagation law, will give rise to the following covariance matrix for DD observations:

$$Q_y^D = \begin{bmatrix} J & 0 \\ 0 & J \end{bmatrix} \begin{bmatrix} \sigma_p^2\, I_6 & \sigma_{p\phi} I_6 \\ \sigma_{p\phi} I_6 & \sigma_\phi^2\, I_6 \end{bmatrix} \begin{bmatrix} J^T & 0 \\ 0 & J^T \end{bmatrix} = \begin{bmatrix} \sigma_p^2\, Q & \sigma_{p\phi} Q \\ \sigma_{p\phi} Q & \sigma_\phi^2\, Q \end{bmatrix}, \tag{7.13}$$

where

$$Q = J\, J^T = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}. \tag{7.14}$$

When we have more satellites, namely k, Q can be generalized as well; $q_{ii} = 4,\ q_{ij} = 2,\ i \neq j$ and $i, j = 1, 2, \ldots, k-1$. Therefore, in general, the following covariance matrix should be estimated by the LS-VCE method (dual frequency):

$$Q_y^D = \begin{bmatrix} \sigma_{p_1}^2\, Q & \sigma_{p_1 p_2} Q & \sigma_{p_1 \phi_1} Q & \sigma_{p_1 \phi_2} Q \\ \sigma_{p_1 p_2} Q & \sigma_{p_2}^2\, Q & \sigma_{p_2 \phi_1} Q & \sigma_{p_2 \phi_2} Q \\ \sigma_{p_1 \phi_1} Q & \sigma_{p_2 \phi_1} Q & \sigma_{\phi_1}^2\, Q & \sigma_{\phi_1 \phi_2} Q \\ \sigma_{p_1 \phi_2} Q & \sigma_{p_2 \phi_2} Q & \sigma_{\phi_1 \phi_2} Q & \sigma_{\phi_2}^2\, Q \end{bmatrix}, \tag{7.15}$$

or

$$Q_y^D = \begin{bmatrix} \sigma_{p_1}^2 & \sigma_{p_1 p_2} & \sigma_{p_1 \phi_1} & \sigma_{p_1 \phi_2} \\ \sigma_{p_1 p_2} & \sigma_{p_2}^2 & \sigma_{p_2 \phi_1} & \sigma_{p_2 \phi_2} \\ \sigma_{p_1 \phi_1} & \sigma_{p_2 \phi_1} & \sigma_{\phi_1}^2 & \sigma_{\phi_1 \phi_2} \\ \sigma_{p_1 \phi_2} & \sigma_{p_2 \phi_2} & \sigma_{\phi_1 \phi_2} & \sigma_{\phi_2}^2 \end{bmatrix} \otimes Q = \Sigma \otimes Q\,, \tag{7.16}$$

where $\otimes$ denotes the Kronecker product of two matrices. In our investigations, the initial values (undifferenced) $\sigma_{\phi_i} = 0.003\,\mathrm{m}$ and $\sigma_{p_i} = 0.3\,\mathrm{m},\ i = 1, 2$ have also been included in Q matrices. In other words, the Q matrices above have the same structure but, in general, different scales. Therefore, the above equation is just symbolically correct.

With four observation types, one needs to estimate 10 (co)variance components, namely 4 variances and 6 covariances. Strictly speaking not all unknown components are estimable. The maximum number of estimable parameters is 6, e.g. 4 variances and 2 covariances or just 6 covariances, etc (see example 4.10). The other problem as to the estimability is due to the fact that two observation types are much more precise than the others, namely carrier phase when compared to the pseudoranges. This will cause the problem to be ill-posed for the geometry-free model. That is, the estimated components of the precise observations are highly correlated and poorly estimable. To handle the problem, one can assume the same variance for both carrier-phase observations on L1 and L2. It should be noted, when dealing with 4 observation types as with GPS, only 3 components are precisely estimable,

namely one (co)variance component for carrier-phase observations and two (co)variance components for pseudorange observations. When we deal with 5 observation types, e.g. $\phi 1$, $\phi 2$, C1, P1, and P2, 7 components are precisely estimable, namely 1 (co)variance component for carrier-phase observations and all 6 (co)variance components of the code observations.

7.3.3 Time correlation of observables

We concentrate here on the off-diagonal elements (covariances) in the $K\,m \times K\,m$ covariance matrix of one GPS observation type over K epochs, i.e. the covariance between *epochs* or *time correlation*, instead of between observation types. For the reason of convenience, assume only two epochs of observations for one observation type (as before, 2 receivers simultaneously tracking 3 satellites). The covariance matrix of undifferenced observables then reads

$$Q_y = \begin{bmatrix} \sigma^2_{(1)}\, I_6 & \sigma_{(12)} I_6 \\ \sigma_{(12)} I_6 & \sigma^2_{(2)}\, I_6 \end{bmatrix}, \tag{7.17}$$

in which $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$ are the variance components of the first and second epoch, respectively and $\sigma_{(12)}$ is the covariance between the two epochs. Applying the error propagation law will bring about the following covariance matrix for DD observations:

$$Q_y^D = \begin{bmatrix} J & 0 \\ 0 & J \end{bmatrix} \begin{bmatrix} \sigma^2_{(1)}\, I_6 & \sigma_{(12)} I_6 \\ \sigma_{(12)} I_6 & \sigma^2_{(2)}\, I_6 \end{bmatrix} \begin{bmatrix} J^T & 0 \\ 0 & J^T \end{bmatrix} = \begin{bmatrix} \sigma^2_{(1)}\, Q & \sigma_{(12)} Q \\ \sigma_{(12)} Q & \sigma^2_{(2)}\, Q \end{bmatrix}, \tag{7.18}$$

where again

$$Q = J\,J^T = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}. \tag{7.19}$$

If we have k satellites, Q can be generalized as $q_{ii} = 4,\ q_{ij} = 2,\ i \neq j$ and $i, j = 1, 2, \ldots, k-1$. Therefore, in general, for K epochs, the following covariance matrix should be estimated by the LS-VCE method:

$$Q_y^D = \begin{bmatrix} \sigma^2_{(1)}\, Q & \sigma_{(12)} Q & \cdots & \sigma_{(1K)} Q \\ \sigma_{(12)} Q & \sigma^2_{(2)}\, Q & \cdots & \sigma_{(2K)} Q \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{(1K)} Q & \sigma_{(2K)} Q & \cdots & \sigma^2_{(K)}\, Q \end{bmatrix}, \tag{7.20}$$

or

$$Q_y^D = \begin{bmatrix} \sigma^2_{(1)} & \sigma_{(12)} & \cdots & \sigma_{(1K)} \\ \sigma_{(12)} & \sigma^2_{(2)} & \cdots & \sigma_{(2K)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{(1K)} & \sigma_{(2K)} & \cdots & \sigma^2_{(K)} \end{bmatrix} \otimes Q\,. \tag{7.21}$$

The above structure can be employed for all different observation types. The preceding covariance matrix contains $K(K+1)/2$ (co)variance components, K variances and

$K(K-1)/2$ covariances. Since the observations are usually equidistant in time, assuming stationarity of the GPS observations implies that the elements on the negative-sloping diagonals of the above equation should be equal (a symmetric Toeplitz matrix). In this case the (co)variance component $\sigma_{(ij)}$ is a function of the time difference $\tau = |j-i|$, namely $\sigma_{(ij)} = \sigma_\tau = \sigma_{|j-i|}$. Therefore, the number of auto-(co)variance components will get reduced to K. The reader is referred to Tiberius and Kenselaar (2003).

7.4 Multivariate linear model

The LS-VCE model is a powerful method for the estimation of the stochastic model parameters. However, the computational burden of this method is still a challenging problem. The procedure proposed in this section aims to reduce the computational load and memory usage. For the GPS geometry-free model one can estimate the (co)variance components by using only a few epochs of observations. However, to improve the precision of the estimates one usually likes using all available observations. Because of the special structure of the covariance matrix of observations and the design matrix, the final estimates can be easily obtained. To this end, consider the following model of observation equations with unknown (co)variance components:

$$E\{\underline{y}\} = Ax,\ Q = \sum_{k=1}^{p} \sigma_k Q_k \tag{7.22}$$

with $\underline{y}$ the $m \times 1$ vector of observables, A the $m \times n$ design matrix, x the $n \times 1$ vector of unknowns, Q the $m \times m$ covariance matrix of observables, Q_k the $m \times m$ cofactor matrices, and σ_k the unknown (co)variance components to be estimated. Now this model is repeated r times $\alpha = 1, \ldots, r$, where the unknown parameters are allowed to vary between the groups. Moreover, the measurements of different groups are assumed to be uncorrelated. This model can, for example, be applied to the geometry-free observation model when first the observations are split into r groups or categories, second the time correlation of observables is absent, and third the statistical behavior of the integer DD carrier-phase ambiguities is ignored and they are kept fixed. Such a multivariate model (repeated model) is of the form

$$\mathsf{E}\left\{\begin{bmatrix} \underline{y}^{(1)} \\ \underline{y}^{(2)} \\ \vdots \\ \underline{y}^{(r)} \end{bmatrix}\right\} = \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(r)} \end{bmatrix}, \quad Q_y = \begin{bmatrix} Q & 0 & \cdots & 0 \\ 0 & Q & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q \end{bmatrix} \tag{7.23}$$

with $\underline{y}^{(i)}$ and $x^{(i)}$ the observable vector and the unknown vector of group i, respectively. The same design matrix A appears for each group, but the unknowns in x are allowed to vary from group to group. One can prove that the final estimates are obtained by averaging the groupwise estimates

$$\hat{\sigma} = \frac{1}{r}\sum_{i=1}^{r} \hat{\sigma}^{(i)}, \quad \text{or} \quad \hat{\sigma}_k = \frac{1}{r}\sum_{i=1}^{r} \hat{\sigma}_k^{(i)},\ k = 1, \ldots, p \tag{7.24}$$

with the covariance matrix

$$Q_{\hat{\sigma}} = \frac{1}{r} N^{-1} \tag{7.25}$$

where N is the normal matrix of a single group. Note that if certain (co)variance components can not be estimated in a single model (which is sometimes the case for the geometry-free model), they can never be estimated if the model is repeated r times.

When a model repeats every 1 sec, collecting 10 (10 is arbitrary) secs of these data in one group (a 10-epoch group) is repeated every 10 secs. For the multivariate model introduced above, we now assume that the covariance matrix Q is the same for all 10-epoch groups as in equation (7.23). This is our strategy for handling a full hour of GPS data.

7.5 Presentation and interpretation of results

The estimates obtained should be presented as clearly as possible. For this purpose, usually visualizing techniques are very appealing. Apart from that, when dealing with *numbers*, it is more convenient to demonstrate the numerical estimates in such a way that they are readily understandable. For example, if our original observations are expressed in unit of meter (m), then the (co)variance components estimated by LS-VCE will be expressed in unit of m^2 and the variance of these estimators in m^4. It may not be convenient to deal with m^2 and m^4. In this section it is intended to derive simple formulas for standard deviation estimators and correlation coefficients as well as their precision only for the reason of *presentation*. Let $\hat{\sigma}_i^2 = \hat{\sigma}_{ii}$ and $\sigma_{\hat{\sigma}_i^2} = \sigma_{\hat{\sigma}_{ii}}$ be the variance estimator and its standard deviation, respectively. They are both expressed in units of m^2. To extract the more convenient indicators, let us apply the square root to the variance estimator which gives the standard deviation estimator expressed in units of meters. To obtain the precision of this estimator, consider the following equation:

$$\hat{\sigma}_i = \sqrt{\hat{\sigma}_i^2} = \sqrt{\hat{\sigma}_{ii}}. \tag{7.26}$$

It is now possible to derive the precision of the variable $\hat{\sigma}_i$, namely $\sigma_{\hat{\sigma}_i}$, using the error propagation law for non-linear functions. One can simply show that the precision of the standard deviation estimate can be approximated using the following equation:

$$\sigma_{\hat{\sigma}_i} \approx \frac{\sigma_{\hat{\sigma}_i^2}}{2\hat{\sigma}_i} = \frac{\sigma_{\hat{\sigma}_{ii}}}{2\hat{\sigma}_i} \tag{7.27}$$

where both $\hat{\sigma}_i$ and $\sigma_{\hat{\sigma}_{ii}}$ are given.

For correlation coefficients, assume that we are given the covariance estimate $\hat{\sigma}_{ij}$ (m^2) and its precision $\sigma_{\hat{\sigma}_{ij}}$ (m^2) and two variance estimates $\hat{\sigma}_{ii}$ (m^2) and $\hat{\sigma}_{jj}$ (m^2) with their precision $\sigma_{\hat{\sigma}_{ii}}$ (m^2) and $\sigma_{\hat{\sigma}_{jj}}$ (m^2). In addition to the standard deviations of the estimates, there are also covariances between estimates. The 3×3 matrix $Q_{\hat{\sigma}}^{ij}$ denotes the covariance matrix of the estimates. In practice, it is more convenient to present the correlation coefficient instead of the covariance estimate

$$\hat{\rho} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i\,\hat{\sigma}_j} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}}\,\sqrt{\hat{\sigma}_{jj}}} \tag{7.28}$$

To obtain the variance of the correlation coefficient $\hat{\rho}_{ij}$, one needs to apply the error propagation law to the linearized form of the preceding equation which yields $\sigma_{\hat{\rho}}^2 = JQ_{\hat{\sigma}}^{ij}J^T$, where J is the Jacobian vector given in equation (6.27). If we assume that the (co)variance

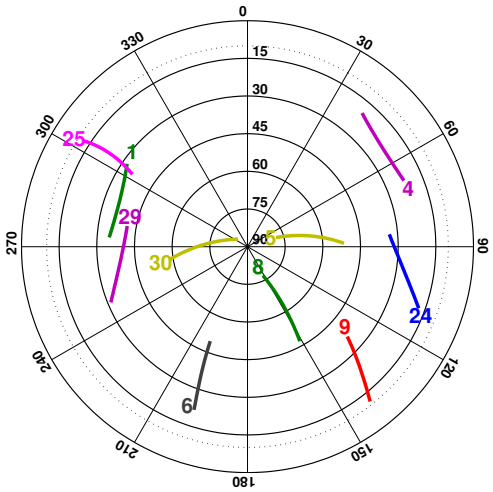


Figure 7.1: Sky plot (azimuth versus elevation) of ten satellites, indicated by their PRN numbers, on April 9, 1999 for Delfland, the Netherlands.

estimates are statistically uncorrelated ($Q_{\hat{\sigma}}^{ij}$ is diagonal), the precision of the correlation coefficient simplifies to

$$\sigma_{\hat{\rho}} \approx \hat{\rho} \left[\left(\frac{\sigma_{\hat{\sigma}_i}}{\hat{\sigma}_i} \right)^2 + \left(\frac{\sigma_{\hat{\sigma}_j}}{\hat{\sigma}_j} \right)^2 + \left(\frac{\sigma_{\hat{\sigma}_{ij}}}{\hat{\sigma}_{ij}} \right)^2 \right]^{1/2} \tag{7.29}$$

where $\sigma_{\hat{\sigma}_i}$ is given in equation (7.27), and so is $\sigma_{\hat{\sigma}_j}$.

7.6 Numerical results

7.6.1 Experiment description

The goal here is not to measure the performance of a particular GPS receiver but instead the sole aim is to apply the LS-VCE method to real data and interpret the results. A data set was obtained from the Delfland 99 campaign, carried out by the MGP group of TU Delft in an open meadow area near Delft, the Netherlands. A 1 hour Ashtech Z-XII3 and Trimble 4000 SSi zero baseline data set from April 9, 1999 was considered (08:00:00-08:59:59 UTC), with 10 and 8 satellites, five and four observation types, respectively, i.e. C1-P1-P2-L1-L2 and C1-P2-L1-L2, and a 1 sec interval. In the sequel, for the reason of convenience, the receivers are simply called 'Ashtech receiver' and 'Trimble receiver'. The total number of epochs is $K = 3600$. The sky plot for this campaign is given in figure 7.1. For the Trimble receiver, the satellites PRN 30 and 25 have not been considered in our evaluations. It is mainly because this receiver has only 9 channels to track the satellites, and the data for satellite PRN 30 was not complete in the above time span.

The results presented here concern zero baselines. Two receivers were connected to the same antenna. The antenna amplifier is responsible for a (large) part of the noise,

and with only one amplifier this noise is common to both receivers, which may cause too optimistic outcomes concerning the level of measurement noise. In general, when measuring an ordinary baseline with two receivers and antennas at separate locations, other effects like atmospheric and multipath effects will also play an important role. For a brief but comprehensive review see Bona and Tiberius (2000).

The estimated (co)variance components for both receivers over 3600 epochs divided into 360 10-epoch groups will be presented. Five observation types have been processed for Ashtech receiver, i.e. carriers L1 and L2 and codes C1, P1 and P2. Also, four observation types have been considered for Trimble receiver, i.e. carriers L1 and L2 and codes C1 and P2. Our conclusion regarding implementation of LS-VCE is that only 2, 3 or 4 iterations are needed to obtain converged variance components. The results presented here, for all observation types, are given for the last iteration (after convergence). In all graphs, the estimates are given for each of the 360 groups. The mean values are also presented based on equation (7.24). The final variances can be obtained by multiplying the mean estimates with their a-priori values $30^2\,\mathrm{cm}^2$ and $3^2\,\mathrm{mm}^2$ for undifferenced code and phase observations, respectively. These a-priori values turned out to be too large for the data of this zero baseline, resulting in small unit variance components (less then one). The estimates that include also the a-priori values are given in the tables.

Because of the special structure of the model, the estimation of two individual variance components for L1 and L2 carrier-phase data would cause the problem to be ill-posed. The ill-posedness will be removed if we estimate one single variance component instead (e.g. if we assume that the precision of L1 and L2 phase observables is the same).

7.6.2 Variances of observation types

Float ambiguity

For each group, in the functional model, the ambiguities have been considered as float unknown parameters. That is, based on the observation of that group, the float ambiguities are used to estimate the variance components (the ambiguities have not been kept fixed but float as estimated). Actually, in this case, the simple averaging of the estimates over 360 groups may not be correct because the structure of the functional model is not identical to equation (7.23). The design matrix A has a structure similar to equation (7.3). The goal is to estimate one variance component for each observation type in each group neglecting

Obs. Type	$\hat{\sigma}^2\ (\mathrm{mm}^2)$	$\sigma_{\hat{\sigma}^2}\ (\mathrm{mm}^2)$	$\hat{\sigma}\ (\mathrm{mm})$	$\sigma_{\hat{\sigma}}\ (\mathrm{mm})$
L1/L2	0.10	0.001	0.32	0.001
C1	1336.39	10.07	36.56	0.140
P1	52229.84	374.28	228.54	0.820
P2	56690.70	405.38	238.10	0.850
L1/L2	0.12	0.001	0.34	0.001
C1	4475.55	27.77	66.90	0.210
P1	49637.10	359.94	222.79	0.810
P2	54232.85	395.03	232.88	0.840

Table 7.1: Variance and standard deviation estimates of phase and code observations (all satellites) as well as their precision for Ashtech receiver; float ambiguities (top); fixed ambiguities (bottom).

Obs. Type	$\hat{\sigma}^2$ (mm^2)	$\sigma_{\hat{\sigma}^2}$ (mm^2)	$\hat{\sigma}$ (mm)	$\sigma_{\hat{\sigma}}$ (mm)
L1/L2	0.09	0.0006	0.29	0.001
C1	8882.24	77.46	94.25	0.410
P2	24054.40	158.74	155.09	0.510

Table 7.2: Variance and standard deviation estimates of phase and code observations (all satellites) as well as their precision for Trimble receiver (fixed ambiguities).

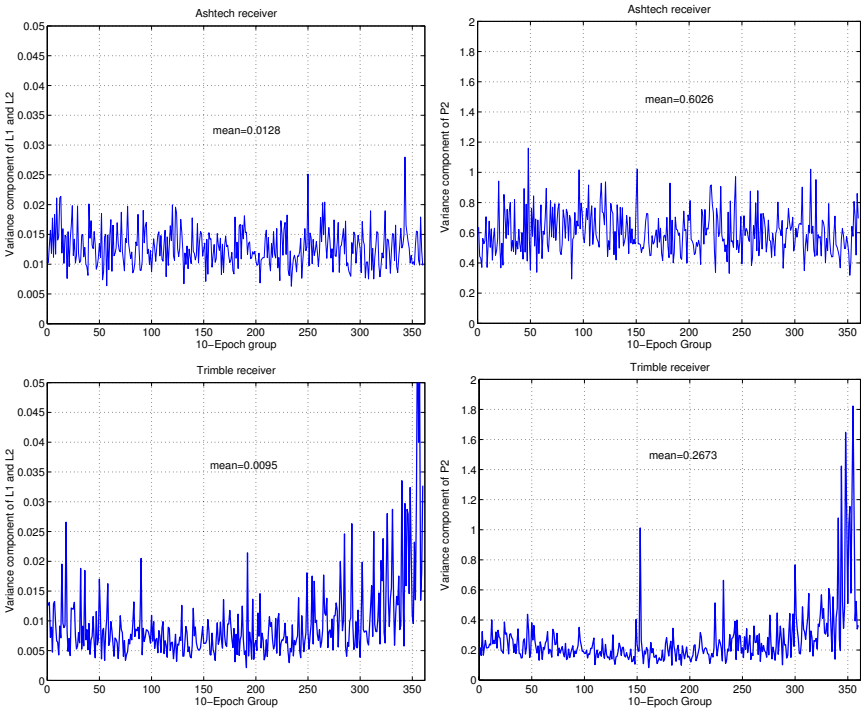


Figure 7.2: Groupwise variance components estimated for two types of GPS observations with fixed ambiguities (all satellites); Ashtech receiver (top); Trimble receiver (bottom); L1 and L2 phase (left); and P2 code (right).

the satellite elevation dependence of the observables precision, the time correlation, the covariance between channels, and the covariance between different observation types. To overcome the ill-posedness problem, we just estimate one single variance component for the L1 and L2 carrier-phase data.

Table 7.1 (top) gives the variance $\hat{\sigma}^2$ and the standard deviation $\hat{\sigma}$ estimates as well as their precision. The results indicate that the noise of GPS observations is at sub-millimeter level (0.3 mm), centimeter level (3.6 cm) and decimeter level (2.3-2.4 dm) for phase, C1, and P1 and P2, respectively. The precision of these estimates are at micrometer level, sub-millimeter level, and millimeter level, respectively. The estimated standard deviation of phase observations (combined L1 and L2) is about 0.3 mm which may look very optimistic.

This is only because we are here dealing with a zero baseline in which mainly the receiver noise is present.

Fixed ambiguity

In the sequel, the ambiguities are kept fixed using the LAMBDA method. The LAMBDA method is described by Teunissen (1995). Therefore, all our subsequent results will be presented on the basis of fixed ambiguities. The success rate of the integer least squares ambiguity estimation is good enough to ignore the randomness of the integer estimators. Using 300 epochs of observations and taking advantage of LAMBDA method, the DD carrier-phase ambiguities are kept fixed. Now the same procedure is employed but this time with fixed ambiguities. The vector of observations has been therefore corrected for the integer ambiguities. The goal is to estimate one variance component for each observation type neglecting the satellite elevation dependence of the observables precision, the time correlation, the covariance between channels, and the covariance between different observation types.

Figure 7.2 (top) shows the groupwise estimate of variance components using the full hour of the data for the L1/L2 phase and P2 code observations of the Ashtech receiver; the estimated components, when multiplied with their initial values in the cofactor matrices, will give the final estimates. Table 7.1 (bottom) gives the variance and standard deviation estimates as well as their precision. Again, the results indicate that the noise of GPS observations is at sub-millimeter level, centimeter level, and decimeter level for phase, C1, and P1 and P2, respectively. The precision of these estimates are at micrometer level, sub-millimeter level, and millimeter level, respectively. The only difference with the results of the float ambiguities is related to the precision of C1 code observables which increases from 37 mm to 67 mm . The other estimates remain the same approximately.

Figure 7.2 (bottom) shows the groupwise estimate of variance components of the L1/L2 phase and P2 code observations for the Trimble receiver. Note that the estimated variance components grow up at the end of the graphs. It is likely because satellite PRN 09 is setting and has a low elevation angle (nearly 10°). This cannot, however, be seen for the Ashtech receiver. It is likely due to the antenna pattern, which has a significant influence on the noise level (e.g. at low elevation angles). Table 7.2 gives the variance and standard deviation estimates as well as their precision for the Trimble receiver. The noise of observables is at sub-millimeter level (0.3 mm), centimeter level (9.4 cm) and decimeter level (1.6 dm) for phase, C1 and P2, respectively. The results show that the P2-code observations are more precise than those of Ashtech receiver; but this is not the case for the C1-code observations. The precision of the estimates are at micrometer and millimeter level for phase and code observations, respectively.

7.6.3 Covariances between code observations

In this subsection, in addition to the variances, the covariances between code observation types have been estimated, namely 4 variance components and 3 covariance components. The satellite elevation dependence of the observables precision, the time correlation and the covariance between channels are disregarded.

Figure 7.3 shows the groupwise estimates of the covariance components between C1 and P2 using the data from Ashtech receiver. The graph at right shows the groupwise estimate of correlation coefficients. As can be seen, these coefficients are estimated around zero

Obs. Type	$\hat{\sigma}^2$ (mm^2)	$\sigma_{\hat{\sigma}^2}$ (mm^2)	$\hat{\sigma}$ (mm)	$\sigma_{\hat{\sigma}}$ (mm)
L1/L2	0.12	0.0008	0.34	0.001
C1	4475.55	26.45	66.90	0.200
P1	49637.11	341.56	222.79	0.770
P2	54232.85	373.62	232.88	0.800
L1/L2	0.11	0.0027	0.34	0.004
C1	4636.20	107.06	68.09	0.790
P1	50109.20	1133.06	223.85	2.530
P2	54547.06	1255.66	233.55	2.690

Table 7.3: Variance and standard deviation estimates as well as their precision for Ashtech receiver when considering all (co)variance components in stochastic model (all satellites); 1 second interval (top); 10 second interval (bottom).

Between	$\hat{\sigma}_{ij}$ (mm^2)	$\sigma_{\hat{\sigma}_{ij}}$ (mm^2)	$\hat{\rho_{ij}}$	$\sigma_{\hat{\rho_{ij}}}$
P1 – P2	324.79	255.31	0.006	0.005
P1 – C1	12.89	69.87	0.001	0.004
P2 – C1	114.84	72.89	0.007	0.005
P1 – P2	-327.77	849.99	-0.006	0.016
P1 – C1	176.32	248.44	0.012	0.016
P2 – C1	59.70	262.08	0.004	0.017

Table 7.4: Covariance and correlation ($\hat{\sigma}_{ij}$ and $\hat{\rho_{ij}}$) estimates of code observations with their precision for Ashtech receiver when considering all (co)variance components in stochastic model (all satellites); 1 second interval (top); 10 second interval (bottom).

and they will be averaged out. Therefore, the correlation between code observations C1 and P2 for the Ashtech receiver is not significant.

Having averaged over 360 groups, table 7.3 (top) gives the variance and standard deviation estimates as well as their precision. The variance components computed by this

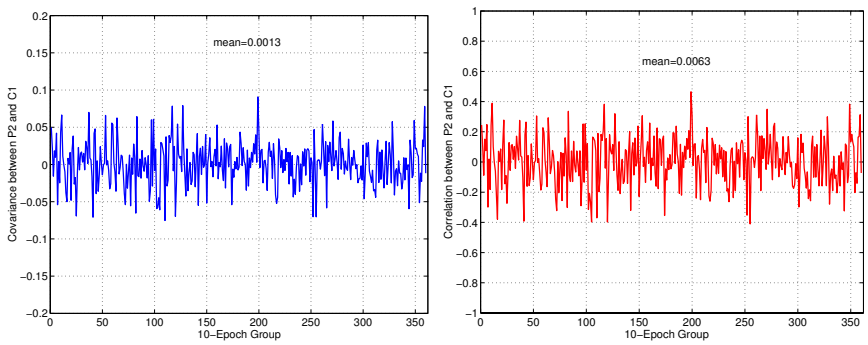


Figure 7.3: Estimated covariances (left) and correlation coefficients (right) between C1 and P2 code observations for Ashtech receiver.

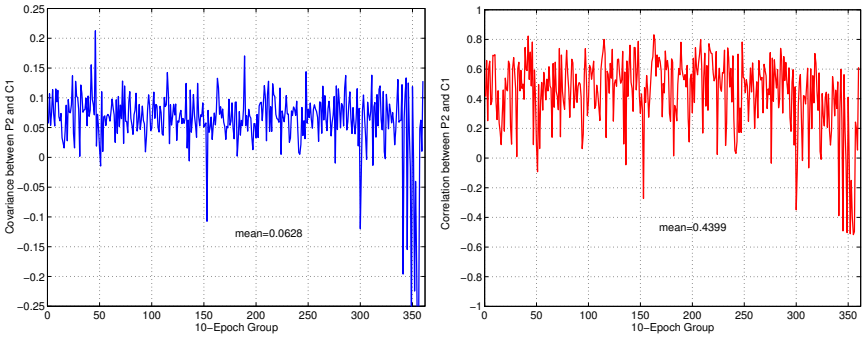


Figure 7.4: Estimated covariances (left) and correlation coefficients (right) between C1 and P2 code observations for Trimble receiver.

Between	$\hat{\sigma}_{ij}$ (mm^2)	$\sigma_{\hat{\sigma}_{ij}}$ (mm^2)	$\hat{\rho_{ij}}$	$\sigma_{\hat{\rho_{ij}}}$
P2 – C1	5649.11	90.53	0.387	0.007

Table 7.5: Covariance and correlation ($\hat{\sigma}_{ij}$ and $\hat{\rho_{ij}}$) estimates between code observations as well as their precision for Trimble receiver.

method coincide with the results when we do not include the covariances in the model (see table 7.1 (bottom)). This implies that the estimated correlation may not be significant. Table 7.4 (top) gives three estimated covariance components and correlation coefficients along with their precision for different code observation types of the Ashtech receiver. The estimated correlation coefficients are at the level of their precision and thus not significant.

We now concentrate on the effect of time correlation on the estimates for the Ashtech receiver. Instead of 1 second intervals, we used 10 second intervals. The 10-epoch groups are performed using these 10-second interval epochs. Therefore, the number of observations gets reduced by a factor of 10 (the number of groups is 36 instead of 360). The observations within these groups are assumed to be relatively uncorrelated in time. Again, satellite elevation dependence of the observables precision, time correlation and the covariance between channels are disregarded. The previous structure in which we estimated variance and covariance components simultaneously, has now been employed here, i.e. 4 variance components and 3 covariance components. Tables 7.3 and 7.4 (bottom) give the results. The (co)variance components obtained here nearly coincide with those when we took all observations into account. This implies that the one-second interval GPS observations for the Ashtech receiver are not severely correlated in time.

For the Trimble receiver, we have only 4 observation types and not 5. We have therefore some restrictions for estimation of the (co)variance components. For example, in case of C1 and P2 code observations, only 2 (co)variance components are precisely estimable and not 3. To handle this problem, one can consider only one variance component for C1 and P2 pseudorange observables and try to estimate it in addition to the covariance. Since the C1 and P2 code observations have different precision (see previous subsection), the aforementioned estimated components have been also included in the related cofactor matrix.

Figure 7.4 (left) shows the groupwise estimates $\hat{\sigma}$ of the covariances between C1 and P2

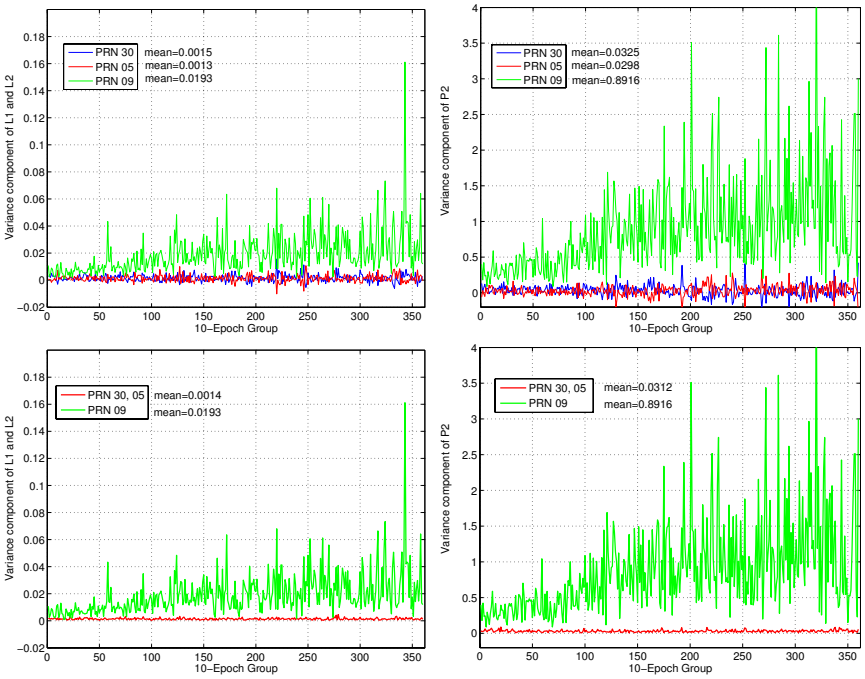


Figure 7.5: Groupwise variance components of satellites PRN 30, 05, and 09 estimated for L1/L2 phase (left) and P2 code (right) observables (Ashtech receiver); 3 individual components (top); 2 components (bottom).

for Trimble receiver. The graph at right shows the corresponding correlation coefficients. These coefficients are estimated around 0.4 and they will not be averaged out. Therefore, the correlation between C1 and P2 code observations for the Trimble receiver is significant. Table 7.5 gives the estimated covariance component and correlation coefficient, as well as their precision, between C1 and P2 code observations. The correlation coefficient $\hat{\rho} = 0.39$ is significant when compared to its precision.

7.6.4 Single channel variances

To evaluate the satellite elevation dependence of the GPS observable precision, we used the structure introduced in equation (7.11). The observations from 3 satellites have been employed, namely satellites PRN 30, 05 and 09 for Ashtech receiver and satellites PRN 05, 29 and 09 for Trimble receiver.

The graphs in figure 7.5 (top) show the groupwise estimates of variance components using the data (L1/L2 phase and P2 code) of the satellites PRN 30, 05 and 09 for Ashtech receiver. The variance components estimated for satellite PRN 09, which is descending and has the lowest elevation angle, are larger than those estimated for satellites PRN 30 and 05. Also, when the elevation angle decreases, a positive trend is observed (for the last groups, on average, the estimated variance components are larger than those of the first groups). Another point is that the variance components estimated for satellites PRN

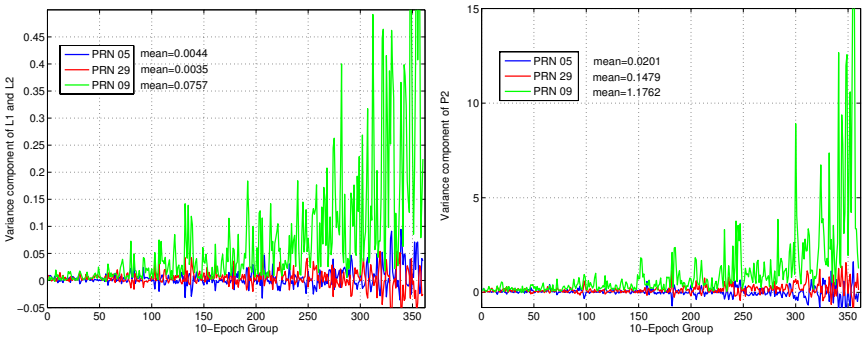


Figure 7.6: Groupwise variance components of satellites PRN 05, 29 and 09 estimated for L1/L2 phase (left) and P2 code (right) observables (Trimble receiver).

Obs. Type	Sat. No.	$\hat{\sigma}^2$ (mm^2)	$\sigma_{\hat{\sigma}^2}$ (mm^2)	$\hat{\sigma}$ (mm)	$\sigma_{\hat{\sigma}}$ (mm)
L1/L2	PRN 30	0.01	0.0006	0.12	0.003
L1/L2	PRN 05	0.01	0.0006	0.11	0.003
L1/L2	PRN 09	0.17	0.0017	0.42	0.002
C1	PRN 30	644.89	10.66	25.39	0.210
C1	PRN 05	1044.91	11.52	32.32	0.180
C1	PRN 09	6045.99	27.57	77.76	0.180
P1	PRN 30	2485.71	179.61	49.86	1.800
P1	PRN 05	2389.67	175.77	48.88	1.800
P1	PRN 09	72991.71	706.83	270.17	1.310
P2	PRN 30	2923.95	196.79	54.07	1.820
P2	PRN 05	2683.99	195.72	51.81	1.890
P2	PRN 09	80241.34	798.95	283.27	1.410

Table 7.6: Variance and standard deviation estimates along with their precision for different observation types of satellites PRN 30, 05, and 09 (Ashtech receiver).

30 and 05 are (highly) negatively correlated. This can also be seen from the correlation matrix of the estimates since the correlation coefficients of variance components between satellites PRN 30 and 05 are $\hat{\rho}_{\phi} = -0.85$, $\hat{\rho}_{p1} = -0.93$, $\hat{\rho}_{p2} = -0.93$ and $\hat{\rho}_{c1} = -0.79$. This implies that the precision of the data of the satellites PRN 30 and 05 is nearly the same. This makes sense since they both have high elevation angles.

Table 7.6 gives the numerical results. The results indicate that the noise of satellites PRN 30 and 05 observations is about 0.1 millimeter, 3 centimeters and 5 centimeters level for phase, C1, and P1 and P2, respectively. The precision of these estimators are at micrometer level, sub-millimeter level, and millimeter level, respectively. The noise of satellite PRN 09 observations, with the lowest elevation angle, is about 0.4 millimeter, 8 centimeters, and 27-28 centimeter level for phase, C1, and P1 and P2, respectively. The precision of these estimators is at micrometer level, sub-millimeter level, and millimeter level, respectively.

To get rid of the high correlation between the estimates of the variance components,

Obs. Type	Sat. No.	$\hat{\sigma}^2$ (mm^2)	$\sigma_{\hat{\sigma}^2}$ (mm^2)	$\hat{\sigma}$ (mm)	$\sigma_{\hat{\sigma}}$ (mm)
L1/L2	PRN 30, 05	0.01	0.0002	0.11	0.001
L1/L2	PRN 09	0.17	0.0019	0.42	0.002
C1	PRN 30, 05	844.90	5.82	29.07	0.100
C1	PRN 09	6045.99	40.79	77.76	0.260
P1	PRN 30, 05	2437.69	39.34	49.37	0.400
P1	PRN 09	72991.69	829.72	270.17	1.540
P2	PRN 30, 05	2803.97	43.39	52.95	0.410
P2	PRN 09	80241.33	914.21	283.27	1.610

Table 7.7: Variance and standard deviation estimates along with their precision for different observation types of satellites PRN 30, 05, and 09 (Ashtech receiver).

Obs. Type	Sat. No.	$\hat{\sigma}^2$ (mm^2)	$\sigma_{\hat{\sigma}^2}$ (mm^2)	$\hat{\sigma}$ (mm)	$\sigma_{\hat{\sigma}}$ (mm)
L1/L2	PRN 05	0.04	0.0008	0.20	0.002
L1/L2	PRN 29	0.03	0.0008	0.18	0.002
L1/L2	PRN 09	0.68	0.0012	0.83	0.001
C1	PRN 05	2365.62	121.62	48.64	1.250
C1	PRN 29	6181.53	144.74	78.62	0.920
C1	PRN 09	15803.42	206.17	125.71	0.820
P2	PRN 05	1809.15	201.77	42.53	2.370
P2	PRN 29	13308.06	196.41	115.36	0.850
P2	PRN 09	105857.73	394.76	326.36	0.610

Table 7.8: Variance and standard deviation estimates along with their precision for different observation types of satellites PRN 05, 29, and 09 (Trimble receiver).

one may think of using one variance component for each observation type for high elevation satellites, i.e. satellites PRN 30 and 05 with the Ashtech receiver. Again, the structure introduced in equation (7.11) was employed; but the first and second cofactor matrices are lumped together in one term and so are their variance components.

The graphs at bottom in figure 7.5 show the groupwise estimates of variance components using the full hour of data (L1/L2 phase and P2 code) of satellites PRN 30, 05 and 09 for the Ashtech receiver. For satellite PRN 09 the same results as before are obtained. Again, when the elevation angle decreases, a positive trend has been observed. Another point which can be clearly seen from the figures is that the combined variance components estimated for satellites PRN 30 and 05 are modestly smoothed. Actually, it seems that the combined variance component estimates are nothing else but the simple arithmetic mean of the two previous components (cf. graphs at top).

Table 7.7 gives the numerical results. Again, the same conclusion as above is logical (cf. table 7.6). The results indicate that the noise of satellites PRN 30 and 05 observations is about 0.1 millimeter, 3 centimeters and 5 centimeters level for phase, C1, and P1 and P2, respectively. The noise of satellite PRN 09 observations, with the lowest elevation angle, is about 0.4 millimeter, 8 centimeters and 27-28 centimeters for phase, C1, and P1 and P2, respectively.

The same procedure can be done for the Trimble receiver, but with the data of satellites

PRN 05, 29 and 09. Two graphs in figure 7.6 show the groupwise estimates of variance components using data (L1/L2 phase and P2 code) of satellites PRN 05, 29 and 09 for the Trimble receiver. It goes without saying that the variance components computed for satellite PRN 09, which has the lowest elevation angle, are larger than those estimated for satellites PRN 05 and 29. Also, when the elevation angle decreases, a positive trend can be observed. Another point which can be clearly seen from the figures is that the variance components estimated for satellites PRN 05 and 29 are negatively correlated. This can also be verified from the correlation matrix of the estimators since the correlation coefficients between satellites PRN 05 and 29 are $\hat{\rho}_{\phi} = -0.70$, $\hat{\rho}_{p2} = -0.67$ and $\hat{\rho}_{c1} = -0.63$. This implies that the precision of the data of the satellites PRN 30 and 05 is nearly the same. This also makes sense since they both have high elevation angles.

Table 7.8 gives the numerical results. The results indicate that the noise of satellites PRN 05-29-09 observations is about 0.2-0.2-0.8 millimeter, 5-8-13 centimeters and 4-12-33 centimeters for phase, C1, and P2, respectively. The precision of these estimators are at micrometer level and millimeter level, respectively.

7.7 Summary and conclusions

This chapter dealt with the application of least squares variance component estimation (LS-VCE) on real GPS data. The LS-VCE method was applied to the GPS geometry-free observation model. This model is linear and can be easily implemented in computer code. Mainly, the satellite elevation dependence of the GPS observables precision and the correlation between different observation types were considered. On the basis of the numerical results, it is concluded that the variance of a GPS observable generally depends on the elevation of the satellite. Also, significant correlation can occur between different observation types, e.g. between the C1 and P2 codes. This can be a good motivation to study the GPS stochastic model in more detail.

The huge number of observations in case of GPS is always an advantage when estimating the stochastic model parameters; the larger the redundancy of the functional model, the better the precision of the estimates in the stochastic model. Note that, however, when dealing with large amounts of GPS data, the computation burden will be drastically increased. One may think of some efficient techniques to do so. For this purpose, the multivariate observation equation model was used in which the observations are divided into small groups.

In this chapter we employed the GPS geometry-free observation model. The GPS *geometry-based* model rather than the geometry-free model is likely more fruitful. For example, the redundancy of the geometry-based model is larger than that of the geometry-free model.

8 GPS Coordinate Time Series

8.1 Introduction

Continuous GPS measurements have been used now nearly 15 years for estimation of crustal deformation. Station positions are determined with respect to an earth-fixed terrestrial reference system. Geophysical studies using geodetic measurements of surface displacement or strain require not only accurate estimates of these parameters but also accurate error estimates. The precision of these estimates is often assessed by their repeatability defined by the mean squared error of individual coordinate components (i.e. north, east, and vertical) about a linear trend. Except for the significant episodic deformation, such as large earthquakes, a linear trend can be a good representative of the deformation behavior. The site velocities are usually determined by linear regression of individual coordinate components. The least-squares technique is used to estimate the line parameters, i.e. the intercept and the slope (site velocity).

In the ideal case, it is desired that the time series possess only white noise and all functional effects are fully understood. The noise in GPS coordinate time series turns out not to be white. Several geodetic data sets have provided evidence for error sources that introduce large temporal correlations into the data. The ultimate goal of noise-studies is to come up with a stochastic model which allows one to process the coordinate time series such that the 'best' solution (most precise solution together with proper precision description) of the station positions and site velocities can be determined. An intermediate goal is therefore to better understand and identify the various noise components of the stochastic model.

Two techniques have generally been employed to assess the noise characteristics of geodetic time series, namely the power spectral method and the maximum likelihood estimation (MLE) method. The former is aimed to examine the data in the frequency domain while the latter is used to examine the data covariance matrix in the time (space) domain. The MLE can estimate the parameters of a noise model effectively in contrast to the classical power spectra techniques. MLE has generally been used to compute the amount of white noise, flicker noise and random walk noise in the time series (see e.g. Zhang et al., 1997; Langbein and Johnson, 1997; Mao et al., 1999; Williams et al., 2004; Langbein, 2004). In this chapter we employ the least-squares variance component estimation (LS-VCE) method introduced in chapters 4 and 5.

8.2 Review of previous work

Zhang et al. (1997) processed 19 months of continuous GPS coordinates from 10 sites in southern California. Using MLE with integer spectral indices, they found that the noise in

the GPS time series was best described as a combination of white noise and flicker noise. This combination suggested that the velocity uncertainties should be 3-6 times larger than those obtained from a pure white noise model. Using the power spectra, the noise was characterized by a fractal noise process with a spectral index of -0.4. Neglecting this fractal white noise model, site velocity uncertainties could be underestimated by a factor of 2-4. In an analogous way, Calais (1999); Mao et al. (1999) processed permanent GPS coordinates. They found that GPS position time series best fitted a noise model consisting of both white noise and flicker noise. In global GPS solutions, Williams et al. (2004) showed that a combination of white and flicker noise is appropriate for all three coordinate components. In regional GPS solutions, the noise was substantially lower than in global solutions. Higher-frequency (1–30 s) GPS position time series have also been shown to contain white plus flicker noise (Bock et al., 2000; Langbein and Bock, 2004).

Several studies have also recognized random walk noise in geodetic data. Random walk noise was detected in continuous measurements of strainmeters as well as very short baseline GPS data at Piñon Flat Observatory in southern California (Wyatt, 1982, 1989; Wyatt et al., 1989; Johnson and Agnew, 2000). Langbein and Johnson (1995, 1997) showed that the noise in the Electronic Distance Measuring (EDM) data is well characterized by a combination of white and random walk noise. Random walk noise can be mitigated by carefully designed monuments. GPS stations deployed in southern California for which the base is deeply seated and laterally braced can be mentioned as a good sample (Wyatt et al., 1989; Bock et al., 1997). The random walk amplitude for a very short baseline at Piñon Flat Observatory is only 0.4 $\mathrm{mm/yr}^{1/2}$ (Johnson and Agnew, 2000). Beavan (2005) shows that the noise properties of GPS time series for concrete pillar monuments are very similar to those of deeply drilled braced monuments. Using two-color EDM measurements in California, Langbein (2004) shows that the random walk noise model is valid for about 30% of the data. In some cases a combination of random walk and band-pass-filtered noise best characterizes the data.

Site-position time series obtained from continuous GPS arrays show significant seasonal variations with annual and semiannual periods. Such seasonal deformation is present in both global and regional GPS coordinate time series (see van Dam et al., 2001; Dong et al., 2002). The latter examine the nature of the observed annual and semiannual site position variations using 4.5 years of global continuous GPS time series. Joint contributions from seasonal surface mass redistribution are the primary causes for the observed annual vertical variations of site positions. Kusche and Schrama (2005) show that after removing the atmospheric pressure loading effect, estimated annual variations of continental-scale mass redistribution exhibit pattern similar to those obtained with GRACE. Ding et al. (2005) used time series of daily positions of eight co-located GPS and VLBI stations to assess the seasonal signals using the wavelet transform. Blewitt and Lavallée (2002) showed that annual signals can significantly bias the site velocity if they are not estimated in the model. Another important systematic error in GPS time series is the presence of offsets (jumps). Williams (2003b) discusses offset detection and estimation strategies. The effect offset estimation has on rate uncertainty depends on the noise characteristics in the series. Kenyeres and Bruyninx (2004) estimate offsets for coordinate time series in the EUREF permanent network. Perfetti (2006) applies the Detection Identification Adaptation (DIA) procedure, suggested by Teunissen (2000b), to permanent GPS time series to detect jumps and to reject outliers.

This study differs in several ways from previous work. We use the least-squares variance

component estimation (LS-VCE) method which has attractive and unique features that we pointed out earlier. First, LS-VCE is generally applicable and can cope with any type of noise (and with any number of noise components) in the data series. The method can be implemented in a relatively simple and straightforward manner. Second, using LS-VCE one can obtain the covariance matrix of the estimators describing the uncertainty of the (co)variance components. Third, the LS-VCE method is employed to estimate time correlation assuming that the time series are stationary in time. Fourth, we use the w-test statistic with which one can simply test the 'contribution' of single noise components. One can thus determine which noise combination is best describing the noise characteristics of GPS position time series. In the same framework as LS-VCE we then introduce the least-squares harmonic estimation (LS-HE) method. The goal is to introduce harmonic functions to capture unmodelled effects in the time series. It is then shown that practically only white noise remains which is very attractive from the data processing point of view. Such a duality between the stochastic and functional model is useful to correctly judge on the amount and behavior of noise.

8.3 Analysis of GPS coordinate time series

This section demonstrates how to estimate the time correlation of GPS coordinate time series using LS-VCE. We rely on a commonly accepted structure of the functional and stochastic model. To see an application of LS-VCE, time correlation of a time series is estimated using a simple expression.

8.3.1 Introduction to noise process

The power spectra, P_y, of many geophysical phenomena, including the noise in GPS position time series, are well approximated as a power-law process (Mandelbrot, 1983; Agnew, 1992; Mao et al., 1999; Williams, 2003a; Williams et al., 2004). The one dimensional time behavior of the stochastic process is such that its power spectrum has the form

$$P_y(f) = P_0 \left(\frac{f}{f_0} \right)^{\kappa} \tag{8.1}$$

where f is the temporal frequency, P_0 and f_0 are normalizing constants, and κ is the spectral index (see e.g. Mandelbrot and van Ness, 1968). Typical spectral index values lie within $[-3, 1]$; for stationary processes $-1 < \kappa < 1$ and for non-stationary processes $-3 < \kappa < -1$. A smaller spectral index implies a more correlated process and more relative power at lower frequencies. Special cases within this stochastic process occur at the integer values for κ. Classical white noise has a spectral index of 0, flicker noise has a spectral index of -1, and a random walk noise has a spectral index of -2. The power spectral method can be employed to assess the noise characteristic of GPS time series.

The second way is to use (co)variance component estimation (VCE) methods. The role of the data series covariance matrix is considered to be an important element with respect to the quality criteria of the unknown parameters. Therefore, VCE methods are of great importance. There are many different methods for VCE. For more information on VCE methods and their applications we refer to chapter 3. The noise components of GPS coordinate time-series, i.e. white noise, flicker noise and random walk noise, are

usually estimated by the MLE method which is a well-known estimation principle. The MLE problem can be solved in several ways (see Kubik, 1970; Koch, 1986; Ou, 1989; Yu, 1996; Grodecki, 1999, 2001). In time series analysis of GPS coordinates, the MLE is described by Langbein and Johnson (1997), Zhang et al. (1997), Mao et al. (1999), Calais (1999), and Williams et al. (2004). They all selected the *downhill simplex* method developed by Nelder and Mead (Press et al., 1992). Using expressions given by Sahin et al. (1992), some authors used a simplified algorithm for VCE (see Davies and Blewitt, 2000; Altamimi et al., 2002).

In contrast to MLE which gives biased estimators, LS-VCE provides *unbiased* and *minimum variance* estimators. The unbiasedness property is independent of the (unspecified) distribution of the data. LS-VCE is also faster than MLE since it iterates in a Newton-Raphson scheme towards a solution rather than using the downhill simplex which can be extremely slow (see Press et al., 1992). Another advantage of LS-VCE over downhill simplex method is the possibility of incorporating any number of noise components in the stochastic model. Using hypothesis testing one can also simply judge in an objective manner which noise components are likely to be present in the series (see sections 8.4.2 and 8.5.4.2).

8.3.2 Functional model

We restrict ourselves to the problem of time correlation estimation for an individual component of GPS coordinate time series. In functional model $\mathrm{E}\{\underline{y}\} = Ax$, y is the m-vector of time series observations, e.g. daily GPS positions of one component. Hereinafter it is denoted by $y(t)$ where t refers to the time instant. When a linear trend describes the deformation behavior, the functional model will read: $\mathrm{E}\{y(t)\} = y_0 + r\,t$. When there are in addition q periodic signals in the data series, the functional model is extended to

$$\mathrm{E}\{y(t)\} = y_0 + r\,t + \sum_{k=1}^{q} a_k \cos\omega_k t + b_k \sin\omega_k t \tag{8.2}$$

Two trigonometrical terms $\cos$ and $\sin$ together represent a sinusoidal wave with in general a non-zero initial phase. The structure introduced above has the advantage of being linear. The unknown vector x consists of the intercept y_0, the slope r and the coefficients a_k and b_k. In case of a linear trend and annual and semiannual signals ($q = 2$), the design matrix A is of size $m \times 6$. Its i^{th} row at time instant t_i is given as

$$\mathbf{a}_i = [1 \;\; t_i \;\; \cos 2\pi t_i \;\; \sin 2\pi t_i \;\; \cos 4\pi t_i \;\; \sin 4\pi t_i] \tag{8.3}$$

where t_i is expressed in terms of year (yr). In section 8.4.1 we show how to obtain an appropriate functional model.

8.3.3 Stochastic model

If the time series of GPS coordinates is composed of white noise, flicker noise, and random walk noise with variances σ_w^2, σ_f^2, and σ_{rw}^2, respectively, the covariance matrix of the time series can then be written as ($Q_0 = 0$)

$$Q_y = \sigma_w^2 I + \sigma_f^2 Q_f + \sigma_{rw}^2 Q_{rw} \tag{8.4}$$

where I is the $m \times m$ identity matrix, and Q_f and Q_{rw} are the cofactor matrices relating to flicker noise and random walk noise, respectively. The structure of Q_y matrix is known through I, Q_f, and Q_{rw}, but the contributions through σ_w, σ_f, and σ_{rw} are unknown. In section 8.4.2 we show how to improve an existing stochastic model.

The elements of the flicker noise cofactor matrix Q_f can be approximated by (Zhang et al., 1997)

$$q_{ij}^{(f)} = \begin{cases} \frac{9}{8} & \text{if} \quad \tau = 0 \\ \frac{9}{8}\left(1 - \frac{\log \tau / \log 2 + 2}{24}\right) & \text{if} \quad \tau \neq 0 \end{cases} \tag{8.5}$$

where $\tau = |t_j - t_i|$. For evenly spaced data, the matrix Q_f is a symmetric Toeplitz matrix which contains constant values along negative-sloping diagonals. It is important to note that the Hosking flicker noise covariance matrix which was introduced and used by Williams (2003a); Langbein (2004); Williams et al. (2004); Beavan (2005) can also be used. The main difference is a scaling of the amplitudes. Therefore, the flicker noise variances we use here are roughly one-half the size of those quoted in these papers.

A random walk process is derived by integrating white noise. Random walk noise is supposed to be zero at initial time t_0 . For evenly spaced data, the random walk cofactor matrix Q_{rw} is expressed as

$$Q_{rw} = f_s^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & m \end{bmatrix} ; \; f_s = \frac{m-1}{T} \tag{8.6}$$

where f_s is the sampling frequency in yr^{-1}, and T is the total observation span (Johnson and Wyatt, 1994; Zhang et al., 1997; Mao et al., 1999).

The variance components σ_w^2, σ_f^2 and σ_{rw}^2 can now be estimated using the LS-VCE method.

8.3.4 Misspecification in functional and stochastic model

A discussion on the misspecifications (errors) in the functional and stochastic model is in place as in practice man's models of real phenomena will always be misspecified. Concerning the number of parameters in a model two types of misspecifications can occur: overparametrization and underparameterization. Underparametrization in the functional model will generally lead to biases in the estimation of x and thus in the results of variance component estimation (aliasing) for noise assessment. The side effect of underparameterization in the functional model can also be a misinterpretation in terms of underparameterization in the stochastic model (see section 8.4.3). Overparametrization in general does not introduce biases provided the remaining redundancy is such that sufficiently precise estimates of x (and thus VCE-results) can be obtained. If however overparametrization in the functional model can seriously affect the noise amplitudes in the stochastic model, this simply means that the extended part of the functional model is closely related to the stochastic model. One can therefore expect to see a duality between the functional and stochastic model.

Misspecifications in the stochastic model will *not* lead to biases in the estimation of x (after all weighted least-squares is still unbiased). However, underparameterization in the

stochastic model leads to biases in (co)variance components. This will result in an incorrect precision description of the estimator x (also losing the 'best' property). To discuss the effect of misspecifications (in fact underparameterization) in the stochastic model, let Q_y be the correct covariance matrix and Q'_y be an incorrect one. If least-squares estimation of x is done with Q'_y, then $\underline{\hat{x}} = (A^T Q'^{-1}_y A)^{-1} A^T Q'^{-1}_y \underline{y}$ is still an unbiased estimator of x and

$$Q_{\hat{x}} = (A^T Q'^{-1}_y A)^{-1} A^T Q'^{-1}_y Q_y Q'^{-1}_y A (A^T Q'^{-1}_y A)^{-1} \tag{8.7}$$

is the correct covariance matrix of the estimator $\underline{\hat{x}}$. Therefore, if one uses $Q'_x = (A^T Q'^{-1}_y A)^{-1}$ as the matrix to describe the precision of $\hat{x}$ one will have an incorrect precision description which can be too optimistic if $Q_{\hat{x}} \geq Q'_x$, but also too pessimistic if $Q_{\hat{x}} \leq Q'_x$. Comparisons of precision description of estimates for different stochastic models are given in section 8.5.6.

8.4 Model identification

8.4.1 Least-squares harmonic estimation (LS-HE)

In this subsection it is aimed to determine an adequate design matrix A for the functional model through parameter significance testing. For a time series $\underline{y}^T = [\underline{y}_1, \underline{y}_2, ..., \underline{y}_m]$ defined on $\mathbb{R}^m$, we assume that it can be expressed as a linear trend plus a sum of q individual trigonometric terms, i.e. $\mathsf{E}\{\underline{y}(t)\} = y_0 + r\,t + \sum_{k=1}^{q} a_k \cos\omega_k t + b_k \sin\omega_k t$ (see equation (8.2)). In matrix notation, we may write

$$\mathsf{E}\{\underline{y}\} = Ax + \sum_{k=1}^{q} A_k x_k, \quad \mathsf{D}\{\underline{y}\} = Q_y \tag{8.8}$$

where the design matrix A contains for instance two columns of the linear regression terms and the matrix A_k consists of two columns corresponding to the frequency ω_k of the sinusoidal function

$$A_k = \begin{bmatrix} \cos\omega_k t_1 & \sin\omega_k t_1 \\ \cos\omega_k t_2 & \sin\omega_k t_2 \\ \vdots & \vdots \\ \cos\omega_k t_m & \sin\omega_k t_m \end{bmatrix} \text{ and } x_k = \begin{bmatrix} a_k \\ b_k \end{bmatrix} \tag{8.9}$$

with a_k, b_k and ω_k being (un)known real numbers. On the one hand, if the frequencies ω_k are known, one will deal with the most popular (linear) least-squares problem to estimate amplitudes a_k and b_k's. Petrov and Ma (2003) studied harmonic position variations of 40 VLBI stations at 32 *known* tidal frequencies. They found that the estimates of station displacements generally agree with the ocean loading computed on the basis of modern ocean tide models for the main diurnal and semidiurnal tides. On the other hand, if the frequencies ω_k are unknown, the problem of finding these unknown parameters is the task of least-squares harmonic estimation.

The problem now is to find the set of frequencies $\omega_1, \cdots, \omega_q$, and in particular the value q, in equation (8.8). The following null and alternative hypotheses are put forward (to start, set $i = 1$):

$$\mathrm{H}_o : \mathsf{E}\{\underline{y}\} = Ax + \sum_{k=1}^{i-1} A_k x_k \tag{8.10}$$

versus

$$\mathrm{H}_a : \mathsf{E}\{\underline{y}\} = Ax + \sum_{k=1}^{i} A_k x_k \tag{8.11}$$

The detection and validation of ω_i is completed through the following two steps:

Step I The goal is to find the frequency ω_i (and correspondingly A_i) by solving the following minimization problem:

$$\underline{\omega}_i = \arg\min_{\omega_j} ||P^{\perp}_{[\overline{A}\ A_j]}\underline{y}||^2_{Q_y^{-1}} = \arg\min_{\omega_j} ||\underline{\hat{e}}_a||^2_{Q_y^{-1}} \tag{8.12}$$

where $||.||^2_{Q_y^{-1}} = (.)^T Q_y^{-1}(.)$, $\overline{A} = [A\ A_1\ ...\ A_{i-1}]$ and $\underline{\hat{e}}_a$ is the least-squares residuals under the alternative hypothesis. The matrix A_j has the same structure as A_k in equation (8.9); the one which minimizes the preceding criterion is set to be A_i. The above minimization problem is equivalent to the following maximization problem (Teunissen, 2000a, p. 96):

$$\underline{\omega}_i = \arg\max_{\omega_j} ||P_{\overline{A}_j}\underline{y}||^2_{Q_y^{-1}}, \text{ with } \overline{A}_j = P^{\perp}_{\overline{A}} A_j \tag{8.13}$$

with $P_{\overline{A}_j} = \overline{A}_j(\overline{A}_j^T Q_y^{-1}\overline{A}_j)^{-1}\overline{A}_j^T Q_y^{-1}$. The preceding equation simplifies to

$$\underline{\omega}_i = \arg\max_{\omega_j} \underline{\hat{e}}_0^T Q_y^{-1} A_j (A_j^T Q_y^{-1} P^{\perp}_{\overline{A}} A_j)^{-1} A_j^T Q_y^{-1} \underline{\hat{e}}_0 \tag{8.14}$$

with $\underline{\hat{e}}_0 = P^{\perp}_{\overline{A}}\underline{y}$ the least-squares residuals under the null hypothesis. In the case that the time series contains only white noise, namely $Q_y = \sigma^2 I$, it follows that

$$\underline{\omega}_i = \arg\max_{\omega_j} \underline{\hat{e}}_0^T A_j (A_j^T P^{\perp}_{\overline{A}} A_j)^{-1} A_j^T \underline{\hat{e}}_0 \tag{8.15}$$

Analytical evaluation of the above maximization problem is complicated. In practice, one has to be satisfied with numerical evaluation. A plot of spectral values $||P_{\overline{A}_j}\underline{y}||^2_{Q_y^{-1}}$ versus a set of discrete values for ω_j can be used as a tool to investigate the contribution of different frequencies in the construction of the original time series. That is, we can compute the spectral values for different frequencies using equation (8.14) or (8.15). The frequency at which $||P_{\overline{A}_j}\underline{y}||^2_{Q_y^{-1}}$ achieves its maximum value is used to construct A_i.

Step II To test H_o against H_a in equation (8.11), we consider $Q_y = \sigma^2 I$, with σ^2 unknown. The following test statistic can be used (see Teunissen et al., 2005)

$$\underline{T}_2 = \frac{||P_{\overline{A}_i}\underline{y}||^2_{Q_y^{-1}}}{2\underline{\hat{\sigma}}_a^2} = \frac{\underline{\hat{e}}_0^T A_i (A_i^T P^{\perp}_{\overline{A}} A_i)^{-1} A_i^T \underline{\hat{e}}_0}{2\underline{\hat{\sigma}}_a^2} \tag{8.16}$$

where $\overline{A}_i = P^{\perp}_{\overline{A}} A_i$ and the estimator for the variance, $\underline{\hat{\sigma}}_a^2$, has to be computed under the alternative hypothesis. Under H_o, the test statistic has a central Fisher distribution

$$\underline{T}_2 \sim \mathsf{F}(2, m-n-2i) \tag{8.17}$$

The above hypothesis testing is in fact the parameter significance test because the test statistic $\underline{T}_2$ can also be expressed in terms of $\underline{\hat{x}}_i$ in equation (8.11) and its covariance matrix.

If the null hypothesis is rejected, we can increase i by one step and perform the same procedure for finding yet another frequency. As a generalization of the Fourier Spectral Analysis, the method is neither limited to evenly spaced data nor to integer frequencies.

When measuring a functionally known quantity, equation (8.15) reads $\underline{\omega}_i = \arg\max_{\omega_j} \ \underline{\hat{e}}_0^T A_j (A_j^T A_j)^{-1} A_j^T \underline{\hat{e}}_0$ which is identical to the least-squares spectral analysis (LSSA) method developed by Vanícek (1969). For some applications refer to Craymer (1998); Abbasi (1999); Asgari and Harmel (2005); Amiri-Simkooei (2005). Amiri-Simkooei and Tiberius (2007) assess the noise characteristics of GPS receivers using zero-baseline time series and arrive at the same conclusions using short-baseline time series when multipath effects were captured by a set of harmonic functions. With equations (8.8)–(8.11) and (8.14) we now have a formulation for LSSA even when an initial design matrix A is present in the model and the covariance matrix Q_y, in general, is not a scaled identity matrix.

Our application of harmonic estimation, in the first place, is to find any potential periodicities in the series. The remaining unmodelled effects (e.g. power-law noise) will also be interpreted and captured by a set of harmonic functions. Once we compensate for these effects in the functional model, the remaining noise characteristics of the series will be assessed. A nearly white noise combined with autoregressive noise can be shown to remain in the data series. This will bring us to see a duality between the stochastic and functional model; what is not captured in the functional model is captured in the stochastic model and vice versa.

8.4.2 The w-test statistic

Here we aim to determine the appropriate covariance matrix Q_y through significance testing on the stochastic model. One advantage of LS-VCE over other methods is that one can use statistical hypothesis testing in the stochastic model (similar as done with the functional model). When there is no misspecification in the functional part of the model $\mathrm{E}\{\underline{y}\} = Ax$, the following two hypotheses, as an example, are taken into account:

$$\mathrm{H}_o : Q_y = \sigma_w^2 I \ \text{ versus } \ \mathrm{H}_a : Q_y = \sigma_w^2 I + C_y \nabla \tag{8.18}$$

where C_y is a known cofactor matrix, for example Q_f or Q_{rw}, and ∇ is an unknown (co)variance parameter. We can use the generalized likelihood ratio test for testing H_o against H_a. The following w-test statistic can be obtained from equation (5.81), with $Q_y = \sigma_w^2 I$

$$\underline{w} = \frac{b\, \underline{\hat{e}}^T C_y \underline{\hat{e}} - \mathrm{tr}(C_y P_A^\perp) \underline{\hat{e}}^T \underline{\hat{e}}}{\sigma_w^2 \left[2b^2 \,\mathrm{tr}(C_y P_A^\perp C_y P_A^\perp) - 2b \,\mathrm{tr}(C_y P_A^\perp)^2 \right]^{1/2}} \tag{8.19}$$

with $b = m - n$ the redundancy of the functional model and $\underline{\hat{e}}$ the least-squares residuals under the null hypothesis.

The expectation and the variance of the w-test statistic are zero and one, respectively (see equations (5.42) and (5.49)). The distribution of this statistic, for large m, can be approximated by the standard normal distribution. The goal now is to compute the w-test statistic values for different alternative hypotheses, i.e. different C_y's in the preceding equation, and select the one that gives the maximum value for the w-test. In fact, equation (8.19) provides us with an objective measure to judge, whether or not (or which), additional

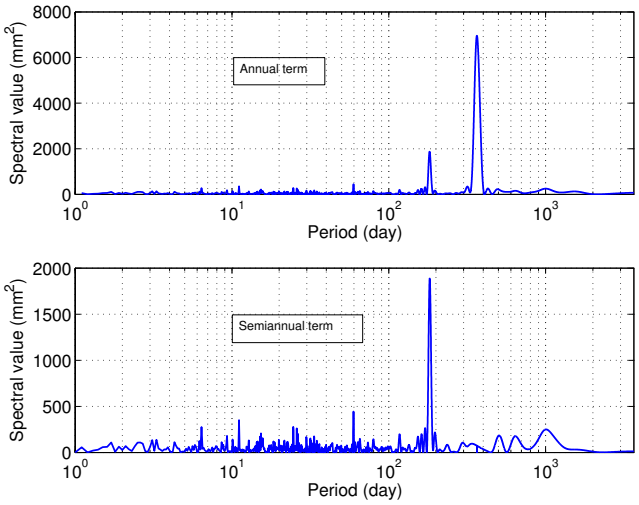


Figure 8.1: Estimated spectral values, equation (8.15), of simulated data-series to detect annual and semiannual signals; top gives annual term (364.7 days), bottom semiannual term (182.4 days).

noise processes are likely to be present in the data at hand. Because of the special structure of the above hypotheses, the numerical evaluation of the preceding test statistic is very simple. We do not need for instance to invert a full covariance matrix since it is diagonal under the (assumed) null hypothesis, namely $Q_y = \sigma_w^2 I$.

8.4.3 Demonstration using simulated data

To illustrate how the proposed harmonic estimation method and LS-VCE (focusing here on time correlation) work, we simulated a 10-year time-series (daily samples) containing only white noise with a standard deviation of 5 mm. Two sinusoidal functions with amplitude of 2 and 1 mm, respectively, for the annual and semiannual term, have then been added to the data (periods of 365.25 and 182.625 days, respectively). The goal now is to use the harmonic estimation method and find the frequencies (or periods) of these signals. The process has been repeated 100 times and it follows that the empirical standard deviation of the detected period is 1.4 and 0.7 days for the annual and semiannual term, respectively. Increasing the amplitude of the harmonic functions or increasing the time-series length gives more precise results for such periods. Figure 8.1 shows one typical example of application of the method to find the periods of harmonic functions. In the first step, the annual term is detected and in the second step the semiannual term.

A correlogram portrays the autocorrelation versus time-lag (the time interval between samples). The coefficient at lag zero equals one by definition. If a white noise process describes the stochastic characteristic of the series, then all other coefficients should be approximately zero. Figure 8.2 shows the typical example of the simulated data corresponding to Figure 8.1. In each graph, the top figure is the time series itself, the middle shows the running average of the series over one week, one month and one year, and the bottom gives the autocorrelation coefficients obtained from LS-VCE (see equation (4.46)).

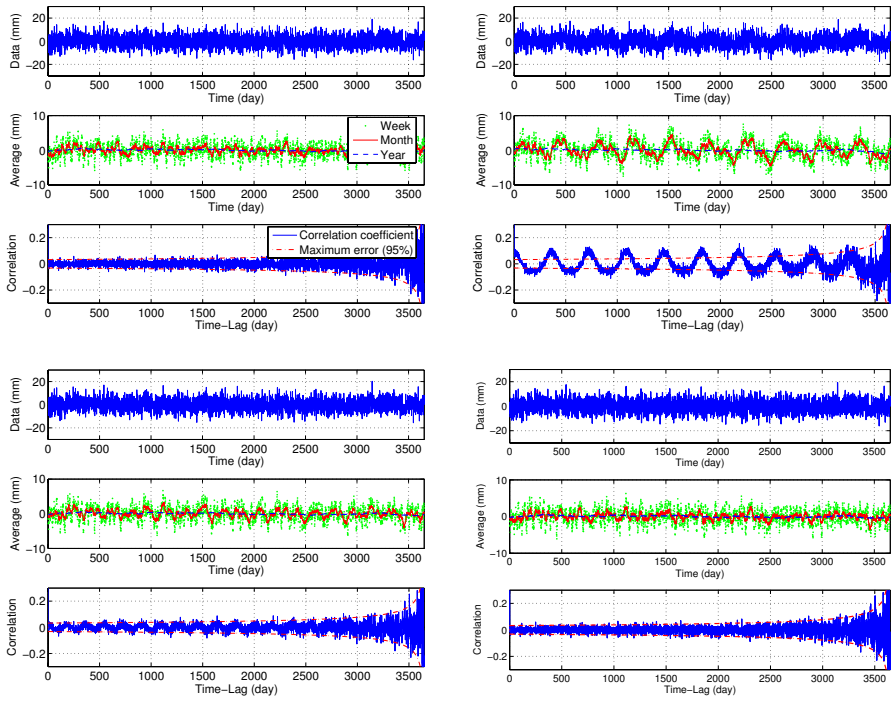


Figure 8.2: Simulated data-series with their running averages as well as their autocorrelation coefficients for white noise only (top-left); added two sinusoidal functions with annual and semiannual periods and amplitudes of 2 and 1 mm, respectively (top-right); removed annual term obtained from LS-HE method (bottom-left); and removed both annual and semiannual terms obtained from LS-HE method (bottom-right).

In case of pure white noise, the autocorrelation function behaves randomly around zero. When both annual and semiannual terms are added, the autocorrelation function shows a periodic behavior which resembles the periodicity of the annual signal. This makes sense because the autocorrelation function of a sinusoidal wave is again a sinusoidal wave with the same frequency but its amplitude is proportional to the square of the amplitude of the original signal (Priestley, 1981). If one includes the annual term in the functional model, the autocorrelation function will still show a periodicity which is due to the presence of the semiannual signal. When one also includes this signal in the functional model, the autocorrelation function becomes very similar to the case of pure white noise; we refer to duality of functional and stochastic model in section 8.3.4.

One can also compute the values of the w-test statistic using equation (8.19) for the different cases mentioned above. The cofactor matrix is chosen as that of flicker noise $C_y = Q_f$. Based on the simulation of 20 data sets, the w-test values on average become: in presence of annual and semiannual signals $w = 15.3$, removed annual signal $w = 1.9$, and removed both annual and semiannual signals $w = 0.3$. Using the LS-VCE method white and flicker noise amplitudes were estimated. The amplitudes on average are: in presence of both signals $\sigma_w = 4.85$ mm and $\sigma_f = 2.84$ mm, removed annual term $\sigma_w = 4.98$ mm

and $\sigma_f = 1.35$, mm and removed both terms $\sigma_w = 5.00$ mm and $\sigma_f = 0.07$ mm. These all together simply express that if there exist unmodelled effects in data, they can mistakenly be interpreted as time correlation (here flicker noise). One should therefore take care of these signals in the functional model.

8.5 Numerical results and discussions

8.5.1 Data and model description

Global time series of site positions are supposed to have more noise than those from a regional solution (Williams et al., 2004). Therefore, the daily GPS global solutions of different stations processed by the GPS Analysis Center at JPL are adopted. The data were processed using the precise point positioning method in the GIPSY software (Zumberge et al., 1997). The satellite orbits, satellite clocks and Earth rotation parameters (ERP) used for the daily solutions were estimated with data from 42 globally distributed IGS tracking stations (see Beutler et al., 1999). In addition, corrections for geophysical effects such as pole and ocean tide effects have been applied. The reader is referred to the JPL website [http://sideshow.jpl.nasa.gov/mbh/series.html].

The estimated coordinates of a site are uncorrelated with those of the other sites if the effects of the common errors in the satellite orbits and clocks and ERP on the estimated coordinates are insignificant. To make a proper statement, one can rely on multivariate time series analysis methods. The time series are processed component-by-component in this chapter.

Most of the results given are based on 5 stations, namely KOSG, WSRT, ONSA, GRAZ and ALGO. Four stations are in Europe of which KOSG and WSRT in the Netherlands, ONSA in Sweden, and GRAZ in Austria. ALGO is in Canada. We have used 10-years of daily solutions for all sites except WSRT which covers only 6.5 years. To justify some of the statements that we will make, 71 globally distributed GPS stations were also processed. Our point of departure is the original time series and its linear model of observation equations $y_t = y_0 + r\,t$. In some cases, the annual and semiannual signals have been considered as well. At times, we have included a set of harmonic functions to compensate for (parts of) unmodelled effects in the series.

8.5.2 Variance component analysis

Three stochastic models have been chosen to describe the noise characteristics of GPS coordinate time series. They include the pure white noise model (I), the white plus flicker noise model (IIa), and the white plus random walk noise model (III). We employed LS-VCE to estimate the white noise, flicker noise, and random walk noise amplitudes (see equation (8.4)). Williams investigations (pers. commun.; Williams (2006)) show that LS-VCE gives the same results as MLE. This holds in fact if $m >> n$ which is usually the case in time series analysis. Table 8.1 gives the noise amplitudes of different components for different stochastic models. The table also provides the precision (standard deviation) of the estimates which is an important feature of LS-VCE. We find, for different noise components, that the horizontal components are less noisy than the vertical components by a factor of 2-4. Compared to the white noise model only, the amplitude of white noise for the white plus flicker noise model is 30% smaller, while this reduction for white plus

Table 8.1: White noise, flicker noise and random walk noise amplitude estimates $\hat{\sigma}$ as well as their precision $\sigma_{\hat{\sigma}}$ for north (N), east (E), and vertical (V) components (C) of site time series in three different stochastic models (white noise only (model I); white noise plus flicker noise (model IIa); and white noise plus random walk noise (model III); functional model is linear regression model for individual time series; WN (mm), FN (mm), RW (mm/yr$^{1/2}$).

Site code			KOSG		WSRT		ONSA		GRAZ		ALGO	
Model		C	$\hat{\sigma}$	$\sigma_{\hat{\sigma}}$	$\hat{\sigma}$	$\sigma_{\hat{\sigma}}$	$\hat{\sigma}$	$\sigma_{\hat{\sigma}}$	$\hat{\sigma}$	$\sigma_{\hat{\sigma}}$	$\hat{\sigma}$	$\sigma_{\hat{\sigma}}$
		N	3.34	0.04	2.76	0.04	3.35	0.04	3.74	0.05	3.62	0.04
I	WN	E	3.44	0.04	2.82	0.04	3.65	0.04	4.75	0.06	3.60	0.04
		V	7.45	0.09	7.12	0.11	7.85	0.10	9.12	0.11	8.22	0.10
		N	2.45	0.05	2.12	0.05	2.54	0.05	2.81	0.06	2.32	0.05
IIa	WN	E	2.54	0.05	2.35	0.05	2.70	0.05	3.62	0.07	2.77	0.06
		V	5.21	0.12	5.08	0.14	5.28	0.13	6.53	0.15	5.20	0.13
		N	3.41	0.18	2.58	0.18	3.39	0.18	3.62	0.20	3.87	0.17
IIa	FN	E	3.67	0.18	2.37	0.19	3.45	0.19	4.16	0.25	3.67	0.19
		V	8.82	0.40	8.24	0.47	9.62	0.42	9.90	0.49	9.71	0.40
		N	2.79	0.04	2.32	0.04	2.85	0.04	3.14	0.04	2.74	0.04
III	WN	E	2.89	0.04	2.52	0.04	3.02	0.04	3.97	0.05	3.08	0.04
		V	6.17	0.08	5.95	0.10	6.29	0.09	7.51	0.10	6.29	0.09
		N	7.96	0.51	6.21	0.51	8.25	0.53	8.26	0.55	9.53	0.55
III	RW	E	9.06	0.56	5.11	0.48	7.50	0.52	9.12	0.65	9.38	0.58
		V	23.4	1.31	21.3	1.52	27.6	1.45	25.7	1.52	25.6	1.37

random walk noise model is about 20%.

In a similar manner to Zhang et al. (1997); Williams et al. (2004), we produced the difference in the log-likelihood values for each site, each component and each error model. The results are given in Table 8.2. The values given in this table are normalized such that the pure white noise model has a log-likelihood of zero. These results confirm that the white noise plus flicker noise model seems to be preferred over the pure white noise model or the white noise plus random walk noise model which coincide with Williams et al. (2004) findings for global solutions. We will give the results of the w-test statistic in section 8.5.4.2 and show how they can be different from those obtained by the MLE method.

Figure 8.3 gives typical examples of estimated variance components at each iteration step for two variance component models. The graphs show that the flicker and random walk noise variances systematically converge to their final estimates from one side. This in fact results in overestimated (biased upwards) flicker and random walk noise variances which coincides with Langbein (2004) findings (conversely white noise is biased downward). The statement can also be verified when we compute the position errors and compare them with the scatter of the time series themselves (see section 8.5.6). The overestimation of random walk noise is more significant than that of flicker noise. This also means that the white plus flicker noise model is the preferred model in these circumstances. The number of iterations in VCE methods is in fact an indication of (the lack of) appropriateness of the selected functional and/or stochastic model.

Table 8.2: Difference in log-likelihood values for different stochastic models; white noise plus flicker noise (WN+FN–left) versus white noise plus random walk noise (WN+RW–right), both compared with pure white noise model which is assumed to have Log=0. White noise plus flicker noise is the preferred model as it provides the largest values.

Model	IIa: WN+FN			III: WN+RW		
Site Code	North	East	Vertical	North	East	Vertical
KOSG	382	326	351	332	277	277
WSRT	229	138	227	213	129	182
ONSA	310	423	396	264	392	328
GRAZ	368	399	376	331	371	320
ALGO	682	290	598	620	253	535
Mean	394	315	390	352	284	328

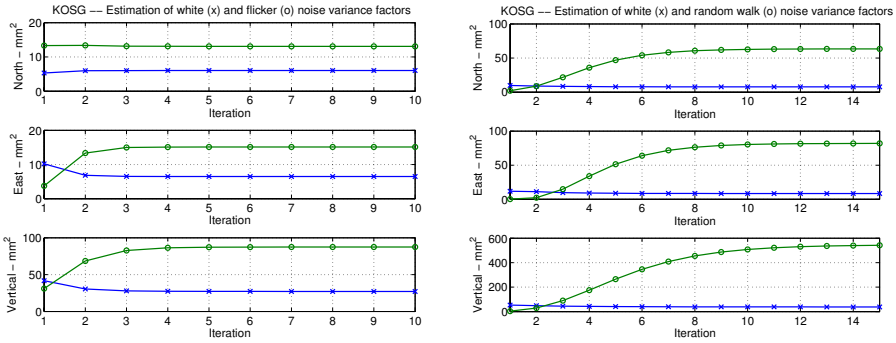


Figure 8.3: Estimated variance components at each iteration step for KOSG; combination of white plus flicker noise (left); and combination of white plus random walk noise (right).

The above discussion essentially means that there are still misspecifications (in fact underparameterization) in the model. In the functional model for instance one should take care of any potential periodicities in the series (see section 8.5.3). Also, a better stochastic model may include in addition to power-law noise other noise models like autoregressive noise (see section 8.5.4). Langbein (2004) proposed to use a combination of power-law noise and band-pass-filtered noise. The upward bias of power-law noise and downward bias of white noise can thus be circumvented by introducing a more sophisticated functional and stochastic model.

8.5.3 Functional model

8.5.3.1 Simple and intuitive technique

Seasonal variations in site positions consist of signals from various geophysical sources and systematic modeling errors (Dong et al., 2002). The weekly, monthly and yearly mean residuals calculated from averaging the daily residuals are shown in Figure 8.4 (left) for ALGO. Running averages naturally remove the high frequency noise and leave the lower frequency signals. Annual and seasonal variations can be observed in the running averages.

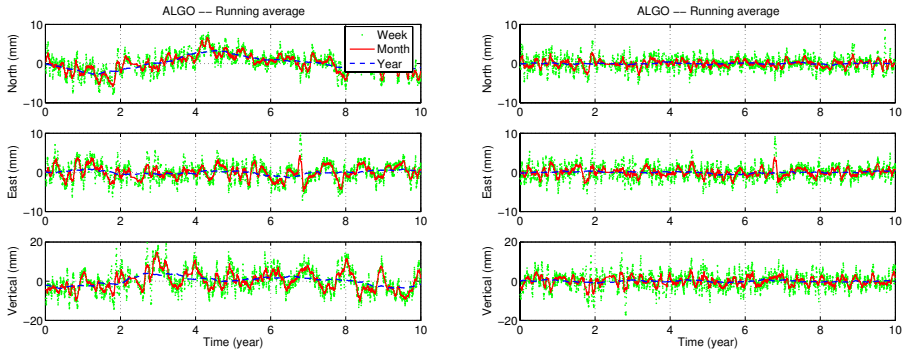


Figure 8.4: Weekly, monthly and yearly mean residual series of north, east and vertical components for different functional models (site ALGO); linear trend only (left); and linear trend plus 10 harmonic functions (right).

For example, the vertical component shows a clear annual signal. They should be eliminated from the time series in order to obtain a more realistic assessment of the noise behavior. We can also see some high and low frequency fluctuations which can likely be captured by flicker noise and random walk noise, respectively (see east and north components, respectively). When parts of these variations have a deterministic behavior, they should be compensated for in the functional rather than the stochastic model. Ding et al. (2005) tried to interpret this behavior as some inter-annual signals.

We now focus on time correlation in the series and estimate one covariance for each time-lag τ using equation (4.45). Figure 8.5 (top) shows the autocorrelation coefficients for the time series of the each component of two sites. The annual and seasonal variations as well as long term fluctuations can be seen in the correlograms. The variations are clearer here than those for the running averages in Figure 8.4. For example, the periodic behavior in the ALGO vertical correlogram shows the annual signal in the series. When the annual signal was included in the functional model, the annual periodicity of the correlogram disappeared. However, this was not the case for KOSG and ALGO east components which show an annual-like signal. This implies that there might still be some hidden periodicities in the data series.

8.5.3.2 Harmonic estimation

Figure 8.6 (left) shows the test statistic values given by equation (8.16) to find the first 15 frequencies. The step size used for $T_j = \frac{2\pi}{\omega_j}$ is taken small at high frequencies and gets larger at lower frequencies. We can see that the value for the test statistic levels off quickly. With 6–10 harmonic functions it gets close to the critical value. In all subsequent results the number of harmonic functions q in equation (8.8) was set to 10 starting with just the offset and slope model. The combination of all 10 harmonic functions included in the functional model of the series is given in Figure 8.6 (right). The periods of ten harmonic functions are given in Table 8.3. Our opinion is that the periodic functions detected by the LS-HE method are due to the following four reasons:

- Unmodelled periodic ground motion: The site is actually moving periodically in this

Table 8.3: Periods (days) of the first ten harmonic functions obtained from least-squares harmonic estimation for north, east and vertical components using equation (8.15), i.e. $Q_y = \sigma_w^2 I$; for all series initial functional model is linear regression model.

Number		1	2	3	4	5	6	7	8	9	10
	N	2620	1256	315	187	579	145	69.8	175	50.4	24.9
KOSG	E	354	384	862	176	116	248	97	70.4	264	14.2
	V	885	377	2782	334	94.5	1401	87.5	37.6	503	116
	N	1586	354	171	292	841	71	84.2	190	50.2	13.6
WSRT	E	349	168	265	533	115	59.3	188	70.2	14.2	1694
	V	2035	863	712	210	150	114	37.4	88.7	30.3	66.6
	N	3154	1307	235	50.4	145	278	185	173	372	57.4
ONSA	E	1418	343	2698	121	756	260	116	70.6	394	237
	V	337	1467	522	927	296	215	191	41.5	3627	94.7
	N	2349	175	187	1282	361	387	50.4	144	88.6	822
GRAZ	E	3651	1274	393	792	171	180	506	148	122	343
	V	910	2505	370	325	697	117	278	88	210	37.5
	N	2772	378	783	1459	122	13.6	200	86.2	73.2	182
ALGO	E	332	177	1483	86.9	108	74.5	762	527	93.4	43.9
	V	366	3651	1566	307	161	88.3	43.9	923	466	120

case. Annual and semiannual signals, for instance, can be specified into this category. Except for a few components, both annual and semiannual signals can be seen in the series. A good example is the first period obtained for ALGO vertical component (366 days) which reveals the annual signal.

- Periodic variation of the estimated time series: The site is *apparently* moving periodically. This is known as the aliasing effect. Unmodelled periodic systematic (e.g. tidal) errors present at a station will result in spurious longer periodic systematic effects in the resultant time series (see Penna and Stewart, 2003; Stewart et al., 2005). A harmonic function with a period of 13.63 days is detected in the north components of WSRT and ALGO. In the east component of KOSG and WSRT, a period of 14.2 days is seen.

- Aliased multipath effect (still a challenging problem): We observe periodic patterns with periods of roughly 350, 175, 117, 88, 70, 59, 50, and 44 days. To justify this, the time series of 71 GPS stations were processed. Figure 8.7 shows the stacked (weighted) power spectra for these stations after including the annual and semiannual signals. The peaks shown in the spectrum coincide well with the number given above. The set of stations was split into two parts and the same conclusion could again be drawn. The results also confirm Ray (2006) findings. Two possibilities which may lead to this effect are as follows: 1) Agnew and Larson (2007) show that the repeat time of the GPS constellation, through which multipath can repeat at permanent stations, averages at 247 s less than a solar day (24 h). However, daily GPS position estimates are based on a full solar day. The difference will alias to a frequency of

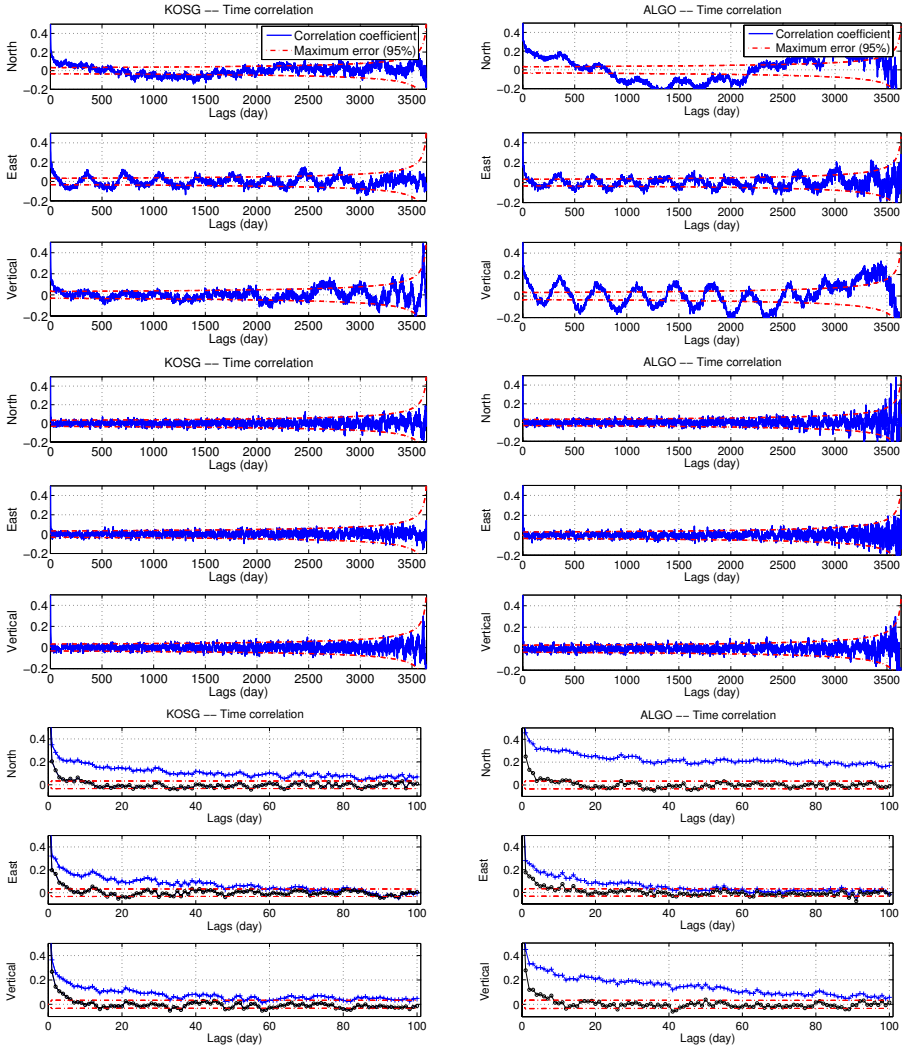


Figure 8.5: Autocorrelation coefficients for time series of north, east and vertical components before (top) and after (middle) removing 10 harmonic functions; graphs at bottom provide a 'zoom in' on the first 100 days for both before (+) and after (o) removing 10 harmonic functions.

0.0028565 cycles/day or 1.04333 cycles/yr (350 days period). The periods found fit with this frequency and its harmonics. 2) Periodic variations of the range residuals with maximum at the eclipse seasons indicate orbit modeling deficiencies for the GPS satellites (Urschl et al., 2005, 2006). The periods found above also coincide with the period of one draconitic GPS year (about $DJ = 351$ days) and fractions $DJ/n,\ n = 1, 2, ..., 8$ (see Beutler, 2006).

- Presence of power-law noise: There are still many numbers in the table (about 50%) which do not fit into the previous categories. The 10 harmonic functions uniformly

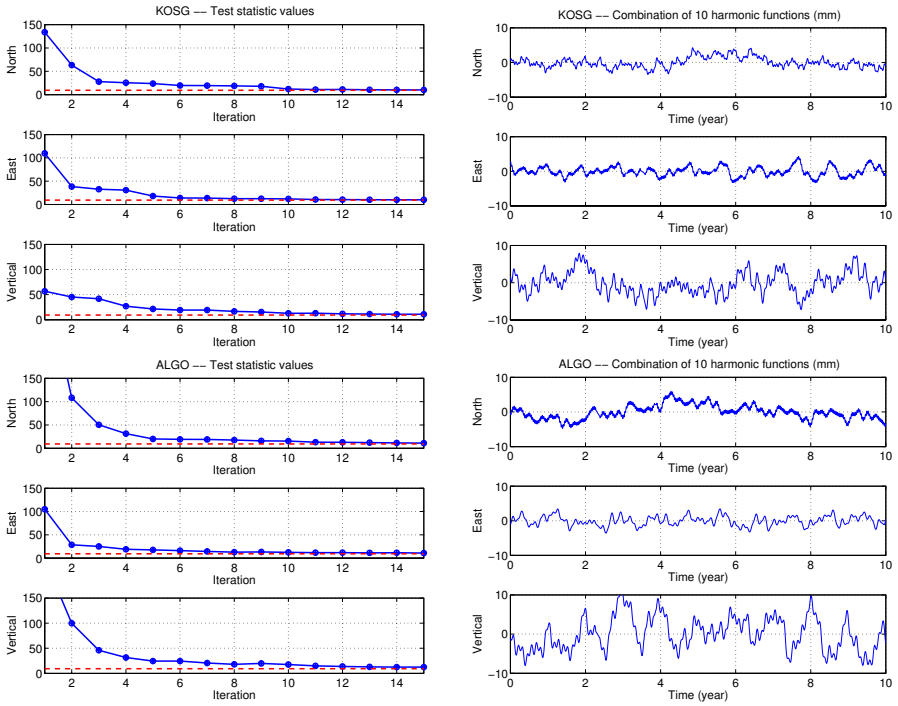


Figure 8.6: The first 15 test statistic values obtained from equation (8.16) as well as its critical value (dashed line) from equation (8.17) with type I error of $\alpha = 0.0001$ (left); combination of 10 harmonic functions (unmodelled effects) as removed from original coordinate component time series (right).

distributed in log-frequency space are sufficient to simulate power-law noise. The higher frequency effects are likely due to flicker noise. Long term period (e.g. larger than 500 days) effects are observed for most of the series. They can likely be considered as random walk noise. Note also that undetected offsets in the time series can mimic random walk noise (Williams, 2003b). We used so far equation (8.15) based on the pure white noise model to detect the frequencies. Therefore, some of the detected periods are most likely due to the presence of colored noise in the data which has been ignored in the results of Table 8.3 and Figure 8.7 (left). The graph shows that white noise is mainly present at high frequencies, flicker noise at medium frequencies, and random walk noise at low frequencies. To justify this, in the harmonic estimation, we used equation (8.14) with a more sophisticated noise model given by equation (8.20). Almost all the lower frequency effects that were detected in the white noise model could not be detected here. Figure 8.7 (right) shows the stacked power spectra of 71 stations using this new stochastic model. The spectrum looks more or less flat and thus does not contradict our statement.

Therefore, to avoid biases in the estimate of x and also the amplitude of noise components (see section 8.3.4) one should take good care of any potential periodicities in the GPS position time series. We can at least mention the annual and semiannual signals,

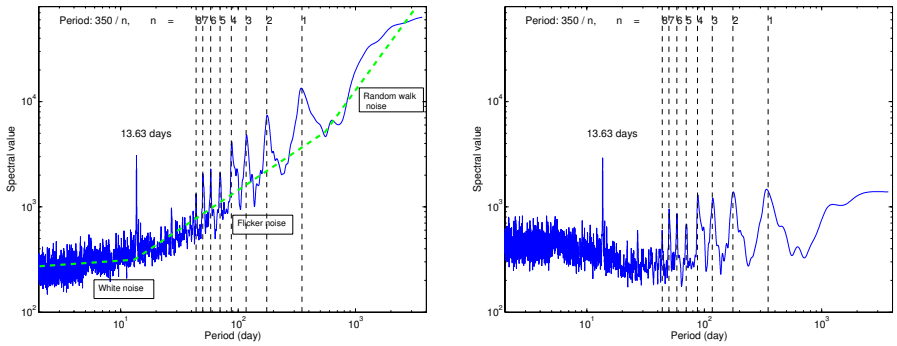


Figure 8.7: Stacked (weighted) power spectra of 71 permanent GPS stations (213 time series) for a linear regression model with annual and semiannual signals included; pure white noise model (left); short-memory plus flicker noise model (right).

periods of 13.66, 14.2, and 14.8 days, and most likely periods of 350 days and its fractions.

8.5.4 Stochastic model

8.5.4.1 Simple and intuitive technique

Figure 8.4 (right) shows the weekly, monthly and yearly running average of residuals when including 10 harmonic functions to describe unmodelled effects (site ALGO). Most of the signals and fluctuations have now been removed. One can also plot the autocorrelation coefficients for the corrected functional model. The correlograms of the time series are given in Figure 8.5 (middle). At first insight, they appear to represent white noise. The graphs at bottom provides a 'zoom-in' on the first part of the graphs at the top and in the middle, i.e. the correlograms over the first 100 days. Unlike the original data, the autocorrelation coefficients become small already after a few days (max 10 days). Our impression is that this remaining high-frequency correlation can be caused by common and well-known sources of errors like atmospheric effects and satellite orbit errors. Table 8.4 provides the numerical results over the first 5 days. The correlation coefficients reduce approximately exponentially, e.g. by $e^{-\alpha\tau}$ which is known as a first order autoregressive noise process AR(1). More results on this model are presented in the next subsection for cases $\alpha = 1$ and $\alpha = 0.25$.

8.5.4.2 The w-test statistic

The results of the w-test statistic are presented to find the most appropriate noise model for the global GPS coordinate time series. The larger (absolute value) the w-test statistic is, the more powerfully the null hypothesis tends to be rejected, and hence, the more likely the alternative model will be as a candidate for the noise in the time series. Five stochastic models were tested using the hypotheses as in equation (8.18). The results are given in Table 8.5. For original data (before removing harmonic functions), the maximum values are obtained for flicker noise and random walk noise models (columns w_2 and w_3). In addition, except for a few components (e.g. ALGO north), flicker noise is preferred to random walk noise. When the values are very close (e.g. ALGO vertical), both noise components (in addition to white noise) are likely to be present in the series.

Table 8.4: Estimated time correlation over the first five time-lags τ (days) for time series of north, east, and vertical components before (left) and after (right) removing harmonic functions; standard deviation of all estimates is 0.02; after removing 10 harmonic functions (right) time correlation presents only at very high frequencies.

		Correlation coefficient			Correlation coefficient		
Site Code	Lag τ (day)	N	E	V	N	E	V
KOSG	1	0.35	0.32	0.37	**0.20**	**0.20**	**0.26**
	2	0.28	0.29	0.26	**0.12**	**0.16**	**0.14**
	3	0.23	0.22	0.23	0.06	0.09	0.10
	4	0.22	0.21	0.21	0.05	0.08	0.08
	5	0.21	0.19	0.18	0.03	0.06	0.06
WSRT	1	0.31	0.21	0.36	**0.13**	**0.07**	**0.25**
	2	0.27	0.22	0.26	**0.09**	**0.08**	**0.13**
	3	0.26	0.17	0.23	0.08	0.03	0.09
	4	0.24	0.18	0.20	0.06	0.05	0.06
	5	0.24	0.18	0.17	0.07	0.06	0.03
ONSA	1	0.31	0.33	0.37	**0.20**	**0.16**	**0.27**
	2	0.25	0.29	0.29	**0.13**	**0.11**	**0.18**
	3	0.21	0.26	0.25	0.08	0.08	0.13
	4	0.20	0.25	0.22	0.07	0.06	0.11
	5	0.18	0.23	0.20	0.04	0.04	0.08
GRAZ	1	0.32	0.32	0.35	**0.16**	**0.14**	**0.23**
	2	0.27	0.27	0.29	**0.10**	**0.08**	**0.16**
	3	0.23	0.28	0.26	0.06	0.09	0.13
	4	0.21	0.25	0.22	0.03	0.06	0.08
	5	0.23	0.24	0.20	0.06	0.05	0.07
ALGO	1	0.46	0.28	0.45	**0.25**	**0.18**	**0.27**
	2	0.39	0.25	0.33	**0.13**	**0.14**	**0.12**
	3	0.36	0.23	0.33	0.11	0.12	0.12
	4	0.31	0.19	0.30	0.04	0.07	0.07
	5	0.32	0.17	0.30	0.07	0.05	0.08

Usually the white noise along with either flicker noise or random walk noise are estimated. To confirm the w-test results, we included all 3 variances in the stochastic model, namely white noise, flicker noise, and random walk noise as in equation (8.4). If a variance is negative, it is an indication that this noise model is most likely not the preferred model and can be excluded from the stochastic model. Table 8.6 shows the signs of the estimated variances using LS-VCE. In 53% of the cases the random walk noise variance is negative. They are correspondingly related to the cases that the w-test values for flicker noise given in Table 8.5 are significantly larger than those for random walk noise. In 44% of the cases both flicker and random walk noise variances are positive. They are related to the case that the w-test values of flicker noise are approximately identical to those of random walk noise. Only for the north component of ALGO, the flicker noise variance is negative which is also verified because the w-test value for flicker noise is smaller than that for the random

Table 8.5: The w-test statistic values for time series of position estimate, before (left) and after (right) removing 10 harmonic functions using equation (8.15), for different alternative hypotheses (different C_y's below). The larger (absolute value) the w-test statistic is, the more likely the alternative model will be as a candidate for stochastic model of series.

		Before removing 10 harmonics					After removing 10 harmonics				
Site	C	w_1	w_2	w_3	w_4	w_5	w_1	w_2	w_3	w_4	w_5
	N	-8.1	**124**	**126**	27.5	40.0	-10.6	11.1	4.9	**14.7**	**16.8**
KOSG	E	-1.8	**32.9**	**9.3**	26.0	37.3	-5.1	12.9	9.1	**15.6**	**18.8**
	V	0.3	**58.1**	**43.2**	27.7	37.6	0.6	10.4	1.1	**19.0**	**21.2**
	N	-11.0	**62.1**	**50.5**	20.4	31.7	-12.0	6.4	0.2	**8.4**	**11.0**
WSRT	E	-8.0	**25.4**	**10.9**	14.5	22.8	-9.8	6.4	2.4	**5.3**	**7.5**
	V	-8.3	**61.2**	**58.4**	22.4	30.8	-8.5	15.1	6.4	**14.7**	**16.6**
	N	-8.2	**71.0**	**66.5**	21.3	35.7	-11.1	13.5	6.1	**14.9**	**18.0**
ONSA	E	-8.4	**67.7**	**52.8**	17.1	41.4	-11.2	11.3	3.2	**12.8**	**15.5**
	V	-3.1	**43.1**	**32.7**	23.4	40.7	-4.5	16.9	4.5	**20.8**	**25.5**
	N	-8.8	**120**	**119**	25.5	38.5	-10.6	7.1	0.3	**12.2**	**13.9**
GRAZ	E	-4.5	**139**	**144**	26.0	41.0	-10.8	5.2	0.1	**10.7**	**13.5**
	V	-5.9	**70.9**	**50.7**	27.8	40.2	-7.0	14.9	3.9	**17.4**	**21.5**
	N	-7.9	**285**	**350**	37.6	57.1	- 8.7	13.1	3.3	**18.1**	**20.9**
ALGO	E	-5.4	**20.7**	**11.2**	14.4	34.4	-10.4	10.2	4.7	**14.2**	**18.2**
	V	-2.7	**99.2**	**99.3**	31.1	52.5	-5.7	14.3	4.6	**19.4**	**22.0**

w_1: $C_y = \mathrm{diag}(Q_{rw})$, only diagonal elements of Q_{rw} in equation (8.6);
w_2: $C_y = Q_f$, flicker noise structure introduced in equation (8.5);
w_3: $C_y = Q_{rw}$, random walk noise structure introduced in equation (8.6);
w_4: $C_y = c_{ij}$ full matrix extracted from an exponential function of the form $e^{-\tau}$;
w_5: $C_y = c_{ij}$ full matrix extracted from an exponential function of the form $e^{-0.25\tau}$.

walk noise.

In columns 4 and 5 we have respectively used the following matrices for C_y: $c_{ij} = e^{-\tau}$ and $c_{ij} = e^{-0.25\tau}$ with $i, j = 1, 2, ..., m$. The corresponding w-test statistic values are mostly significantly smaller than those for flicker and random walk noise. However, after removing 10 harmonic functions (individually per component) from the original data series, the largest values for w-test statistic are obtained for the $e^{-0.25\tau}$ stochastic model (column 5). Note also that the results given in columns 4 and 5 are not much different. This therefore confirms the existence of remaining correlation at very high-frequencies which is believed to be due to common sources of errors that last only a couple of successive days (see Figure 8.5 (bottom)). A relatively large value of the w-test statistic (values on the right) for flicker and random walk noise is most likely due to this correlation. A significant decrease in the w-test values for flicker and random walk noise implies that most parts of the power-law noise have now been captured by the harmonic functions. This statement

Table 8.6: Sign of white noise, flicker noise, and random walk noise variances for north, east, and vertical components. A negative variance component indicates that this noise model is most likely not an appropriate model and can be excluded from stochastic model.

		Sign of variance component		
Site code	Component	WN	FN	RW
KOSG	N	+	+	+
	E	+	+	-
	V	+	+	-
WSRT	N	+	+	-
	E	+	+	-
	V	+	+	+
ONSA	N	+	+	+
	E	+	+	-
	V	+	+	-
GRAZ	N	+	+	+
	E	+	+	+
	V	+	+	-
ALGO	N	+	-	+
	E	+	+	-
	V	+	+	+

was also justified when the white and flicker noise amplitudes were estimated using the extended functional model (with 10 harmonics) which led to small positive or negative values for flicker noise amplitudes.

Let us now turn our attention to the first column (w_1) in Table 8.5. The goal is to test stationarity of the white noise amplitude in the series. For this purpose we selected $C_y = \mathrm{diag}(Q_{rw})$. Most of the w-test values are negative implying that the white noise amplitude in the daily position estimates gets reduced towards the end of the series as $\sigma^2_{w_i} = \sigma^2_w + t_i \nabla$ (∇ negative). This reflects the improvements in analysis products (e.g. satellite orbits and Earth orientation parameters) which makes sense, of course, as equipment is improving and also our knowledge about error sources like atmosphere and orbit is continuously being improved. Note however that white noise is not the dominating source of error on the uncertainties of the parameters of interest. Williams et al. (2004) showed that such a reduction of noise with time holds true also for the flicker noise amplitude.

A large value for the w-test leads to the rejection of the null hypothesis. One can therefore obtain the w-test values for different alternative hypotheses. The one which gives the largest absolute value is considered as a superior candidate for describing the noise characteristics of the data. In our case, in general, the flicker noise model was preferred. After introducing 10 harmonic functions to the model, the largest values were obtained for the AR(1) noise process. The w-test statistic is considered to be a powerful tool to decide on the preferred noise model. Based on simulated data, Williams (pers. commun.) concludes that the w-test statistic and the difference in the log-likelihood values give very similar results (e.g. a correlation of 0.94). Note however that the w-test statistic can simply be used while the MLE method needs successive inversion of the covariance matrix.

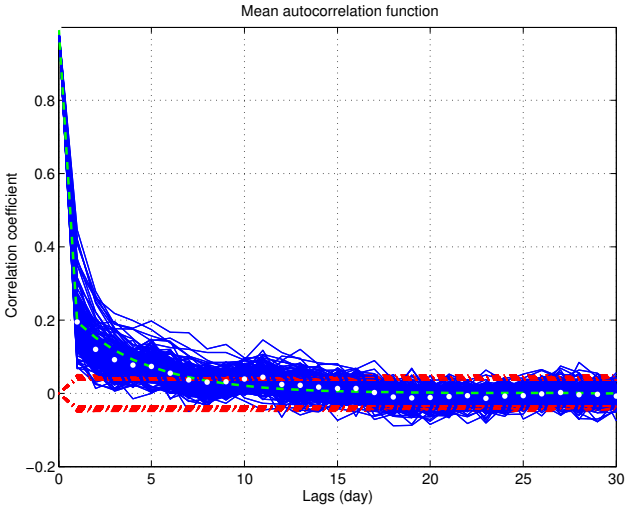


Figure 8.8: Individual as well as mean autocorrelation function (white circles) of 71 permanent GPS stations (213 time series) for a linear regression model with 10 harmonic functions; light dashed line represents the estimated short-memory WN+AR(1) noise process.

8.5.5 Remarks and discussions

Although not fully physically justified, we captured unmodelled effects by harmonic functions. We would like to point out here the duality between harmonic functions and the time correlated noise process. If we can compensate for all unmodelled effects in the functional model, this would be the best way to do so; refer to section 8.3.4. Otherwise, they will be interpreted as if the data were time correlated. On the other hand, if these variations can not be considered as deterministic signals to be compensated for in the functional model (e.g. by a set of harmonic functions), they can for instance be captured as power-law noise process (e.g. flicker noise or random walk noise) which need to be taken care of in the stochastic model.

When unmodelled effects were removed, practically only white noise is left in the series. The remaining minor time correlation as a short-memory process exponentially vanishes within a few days and can be expressed for instance as an AR(1) noise process. This essentially means that to avoid biases in the power-law noise amplitude due to underparameterization, one will have to include also the AR(1) noise process $\sigma_a^2 Q_a$ in equation (8.4), namely $Q_y = \sigma_w^2 I + \sigma_a^2 Q_a + \sigma_f^2 Q_f$ (when one ignores random walk noise). This holds indeed also for any potential periodicities in the functional part of the model.

Our investigations show that the time series are not yet long enough to separately estimate one variance for each noise component. Therefore, first the LS-HE method was employed to include a set of harmonic functions to compensate for power-law noise model. The remaining noise is now expressed as a combination of white and autoregressive noise. The unknowns in this case are the amplitudes of white and autoregressive noise (σ_w and σ_a), and the time-scale α of the noise process. In other words the short-memory process is expressed as: $Q_y = Q(\alpha; \sigma_w^2, \sigma_a^2) = \sigma_w^2 I + \sigma_a^2 Q_a$ where $q_{ij}^a = e^{-\alpha\tau}$ and $\tau = |t_j - t_i|$. This

Table 8.7: Short-memory (WN plus AR(1)) noise and flicker noise amplitude estimates for north, east, and vertical components; functional model is linear regression model plus annual and semi-annual signals as well as signals with period of 13.66 days and 350/n, n=1,...,8; stochastic model employs the form of equation (8.20).

		Short-memory and flicker noise (Model IIb)					
		WN+AR(1) (mm)			FN (mm)		
Site code	Estimate	N	E	V	N	E	V
KOSG	$\hat{\sigma}$	2.85	3.06	6.36	2.23	1.61	5.21
	$\sigma_{\hat{\sigma}}$	0.05	0.05	0.12	0.25	0.27	0.57
WSRT	$\hat{\sigma}$	2.43	2.67	6.03	1.45	1.03	5.66
	$\sigma_{\hat{\sigma}}$	0.05	0.05	0.15	0.28	0.31	0.69
ONSA	$\hat{\sigma}$	2.96	3.10	6.32	2.06	2.21	6.99
	$\sigma_{\hat{\sigma}}$	0.05	0.06	0.14	0.26	0.27	0.57
GRAZ	$\hat{\sigma}$	3.26	4.09	7.75	2.02	2.72	6.07
	$\sigma_{\hat{\sigma}}$	0.06	0.07	0.15	0.29	0.34	0.69
ALGO	$\hat{\sigma}$	2.76	3.25	6.43	2.69	1.91	5.98
	$\sigma_{\hat{\sigma}}$	0.06	0.05	0.13	0.24	0.28	0.57

is in fact a nonlinear stochastic problem that can again be solved by LS-VCE.

The method was applied to 71 globally distributed GPS stations. The average value for the time-scale is $\alpha \approx 0.25$. The mean amplitude of white noise and autoregressive noise were $\sigma_w = 2.3,\ 3.3,\ 6.3$ and $\sigma_a = 1.3,\ 1.8,\ 4.0$ (all in mm) for north, east, and vertical components, respectively. In practice it is more convenient to combine white noise and autoregressive noise into one short-memory process using the average values obtained above. For example, based on the results obtained above, if we assume that the time-scale α and also the relative magnitude of noise components σ_a/σ_w is known, the covariance matrix $Q_y = \sigma_w^2 I + \sigma_a^2 Q_a + \sigma_f^2 Q_f$ can be reformulated as

$$Q_y = \sigma_s^2 Q_s + \sigma_f^2 Q_f \tag{8.20}$$

where $\sigma_s^2 = \sigma_a^2 + \sigma_w^2$ is the variance of the short-memory noise process and Q_s is given as

$$q_{ij}^s = \begin{cases} 1 & \text{if } \tau = 0 \\ \beta e^{-\alpha\tau} & \text{if } \tau \neq 0 \end{cases} \tag{8.21}$$

with $\alpha \approx 0.25$ and $\beta = \sigma_a^2/(\sigma_w^2 + \sigma_a^2) \approx 0.25$. Figure 8.8 shows the weighted mean autocorrelation function of 71 permanent GPS stations, and its approximation based on equation (8.21).

Stochastic model (8.20) is referred to as the short-memory noise and flicker noise model (model IIb), for which two variance components σ_s^2 and σ_f^2 need to be estimated by LS-VCE. We now consider this equation to estimate the magnitude of short-memory (combined WN and AR(1)) and flicker noise process. A correct functional model consisting of annual and semiannual signals, a period of 13.66 days, and periods of 350 days and its fractions 350/n (n=1,...,8) was also used. The results are given in Table 8.7. Compared to the

Table 8.8: Error estimate (formal standard deviation) of slope, intercept and position for different stochastic models (I: white noise only, IIa: white noise plus flicker noise, IIb: short-memory noise plus flicker noise with proper functional model, and III: white noise plus random walk noise).

		KOSG			GRAZ			ALGO		
Error in	Model	N	E	V	N	E	V	N	E	V
Intercept	I	0.12	0.12	0.26	0.13	0.17	0.32	0.13	0.13	0.29
(mm)	IIa	2.87	3.08	7.41	3.04	3.46	8.35	3.25	3.09	8.15
	IIb	1.89	1.37	4.40	1.72	2.16	5.13	2.27	1.62	5.05
	III	1.13	1.23	2.90	1.20	1.42	3.32	1.42	1.49	3.58
Slope	I	0.02	0.02	0.04	0.02	0.03	0.06	0.02	0.02	0.05
(mm/yr)	IIa	0.25	0.27	0.65	0.27	0.31	0.73	0.29	0.27	0.71
	IIb	0.17	0.12	0.39	0.15	0.19	0.46	0.20	0.15	0.45
	III	2.52	2.87	7.39	2.62	2.89	8.12	3.02	2.97	8.11
Position [a]	I	0.06	0.06	0.13	0.06	0.08	0.16	0.06	0.06	0.14
(mm)	IIa	2.58	2.77	6.66	2.73	3.13	7.50	2.92	2.77	7.33
	IIb	1.69	1.22	3.94	1.53	1.93	4.59	2.03	1.45	4.52
	III	12.6	14.3	37.0	13.1	14.5	40.6	15.1	14.9	40.5
Position [b]	I	0.12	0.12	0.25	0.13	0.16	0.31	0.12	0.12	0.28
(mm)	IIa	2.87	3.08	7.41	3.04	3.49	8.35	3.25	3.08	8.15
	IIb	1.88	1.37	4.40	1.72	2.16	5.13	2.27	1.62	5.05
	III	25.2	28.7	73.9	26.1	28.9	81.2	30.1	29.7	81.0

[a] Error in the middle of time series; [b] Error at the end of time series

results given in Table 8.1 for white plus flicker noise model (model IIa) flicker noise shows a reduction of 40% whereas white (in fact short-memory) noise increases about 20%. The difference in the log-likelihood values have also been computed which show an increase of about 10% compared to the values given in Table 8.2 for model IIa. With this strategy not only one can obtain more precise results for parameters of interest but also one will be able to increase the log-likelihood values.

Table 8.9: Minimum and maximum error estimate coefficients of slope, intercept and position for different stochastic models (IIa: white noise plus flicker noise, IIb: short-memory noise plus flicker noise with proper functional model, and III: white noise plus random walk noise) all compared to those of white noise only model (I).

	Error Model			
Error in	I	IIa	IIb	III
Intercept	1-1	16-29	7-21	7-12
Slope	1-1	9-16	5-10	67-185
Position [a]	1-1	31-52	13-38	109-313
Position [b]	1-1	17-30	8 -22	109-324

[a] In middle of series, [b] At end of series

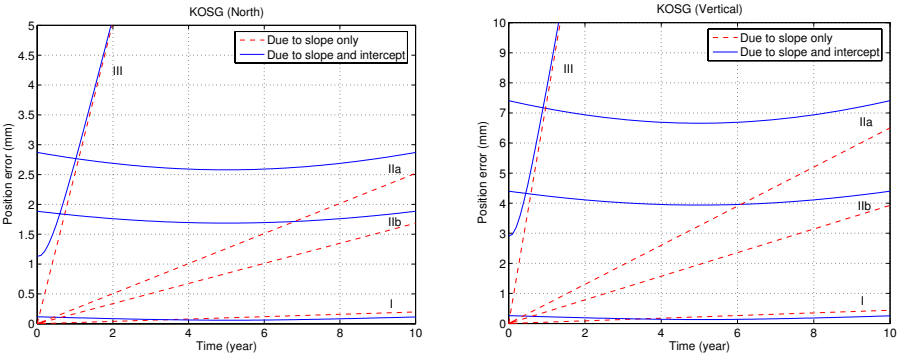


Figure 8.9: Position error (standard deviation) of different stochastic models as a function of time for north and vertical component of KOSG caused by rate uncertainty only (dashed lines) and full covariance matrix of intercept and site velocity (solid lines); WN only (I), WN + FN (IIa), modified WN + FN (IIb), and WN + RW (III).

8.5.6 Error estimate of parameters

The goal here is to estimate and compare the error of four parameters of interest, namely the intercept, slope (rate), position in the middle, and position at the far end of the time series, for different stochastic models. Model I, IIa, IIb, and III are the pure white noise, white noise plus flicker noise, short-memory (combined white and autoregressive) noise plus flicker noise, and white noise plus random walk noise, respectively. Model IIb also includes a few justified harmonic functions (annual and semiannual signals, a period of 13.66 days, and periods of 350/n days, n=1,...,8) in the functional model. We show how an incorrect stochastic model will result in a too optimistic (or too pessimistic) precision description of the parameters of interest. Analytical expressions for the rate uncertainty due to white noise and random walk noise were derived by Zhang et al. (1997). Langbein and Johnson (1995) and Williams (2003a) discussed the effects of sampling rate and time series length on the rate estimates and their error. In this paper however we will rely only on numerical results.

The results given in Table 8.8 are based on $Q'_x = (A^T Q'^{-1}_y A)^{-1}$ (see section 8.3.4), where Q'^{-1}_y is an (in)correct covariance matrix. The pure white noise model (I) gives the most optimistic results and the white plus random walk noise (III) generally gives the least precise results. The error of parameters for different models compared to those for the pure white noise model are larger by the coefficients given in Table 8.9. For example, if a white plus flicker noise model is used instead of a pure white noise model, the velocity error obtained can be larger by factors of 9-16. The corresponding values for model IIb are smaller because the estimated flicker noise is smaller. Among different models the error estimates of site velocity and position for model III are considerably larger (about one order of magnitude) than those for other models.

Compared to the white noise magnitude in the pure white noise model (see Table 8.1), the standard deviation of positions in the middle of the time series are 2%, 80%, 50%, and 400% for models I, IIa, IIb, and III, respectively. These values increase at the end of the time series to 4%, 90%, 55%, and 800%. For all models, except for model III, the minimum error estimate of the position is obtained in the middle and the largest values

are given at both ends of the series. The results of model I are too optimistic. The error contribution of the intercept and the slope on the position seems to be the same (2% in the middle and 4% at both ends). The results obtained from models IIa and IIb appear to be more realistic. In Model IIa the slope error has only a contribution of 10% on the position error estimate (compare 80% in the middle vs 90% at both ends). The error in the intercept plays the main role. This holds also for model IIb. When compared with model IIa, the results with model IIb get improved by a factor of 1.6. The behavior of the white noise plus random walk noise is somehow different. The results are too pessimistic and the error in the slope plays the main role.

In geophysical literature, the site velocity uncertainty is usually of interest and not directly the position error. However, in some geodesy applications (e.g. realization of ITRF) the final goal of the site velocity is in fact the position and its uncertainty. This means that so far we have ignored the uncertainty of the intercept. In models IIa and IIb the intercept plays the main role of error. Figure 8.9 shows the effect of site velocity only and site velocity plus intercept on the position error for different stochastic models. When interested in position error and relying only on the site velocity error, it seems that models IIa and IIb are more appropriate for long-term accuracy but give optimistic results over short periods. On the other hand, model III is likely suitable for short-term accuracy but yields pessimistic results over long periods (see Figure 8.9).

8.6 Summary and conclusions

We assessed the noise characteristics in global time series of daily position estimates by LS-VCE. The method is easily understood, generally applicable and very flexible. The LS-VCE estimators are unbiased and of minimum variance. This method provides the precision of the (co)variance estimators. Based on the results given, the following conclusions can be drawn:

The w-test statistic is a powerful tool to recognize the data noise characteristics in order to construct an appropriate stochastic model. Using the w-test a combination of short-memory (white plus autoregressive) noise and flicker noise was in general found to best describe the noise characteristics of the position components; we hardly observed that the strict white plus random walk was the preferred noise model. These results have also been verified using correlograms of the time series, the frequencies of the harmonic functions, and the signs of the estimated flicker and random walk noise variance components.

The least-squares harmonic estimation method was used to find and consequently remove a set of harmonic functions from the data. These harmonic functions captured unmodelled effects. The results confirm the presence of annual and semiannual signals in the series. We could also observe other periodic effects; for example a period of 13.66, 14.2, and 14.8 days. We observed also significant periodic patterns with periods of 350 days and its fractions 350/n, n=1,2,...,8 which is likely due to the aliased multipath effects in permanent stations. When such variations have underlying physical phenomena (or modeling error), their effects can be considered as systematic periodic signals. It may not be appropriate to capture their effects by a power-law noise process in the stochastic model. They may mistakenly mimic flicker or random walk noise if we neglect them in the functional model. Therefore, neglecting such effects, which may be best described by a deterministic model rather than a power-law noise model, can seriously affect the error

estimate of the site velocity and the position.

There are also some effects in the series that are not of periodic nature. They can most likely be expressed as power-law noise. We however employed the harmonic estimation method to find more frequencies in the series. A significant decrease in the w-test values for power-law noise implies that most parts of this noise are captured by the harmonic functions. This led us to see a duality between the stochastic and functional model; what is not captured in the functional model is captured in the stochastic model. When we include the harmonic functions, almost exclusively white noise remains in the data. Only at very high frequencies a significant time-correlation appeared to be present which can be expressed as an exponential function (e.g. a first-order autoregressive noise process). This noise can be caused by common sources of error like atmospheric effects as well as satellite orbit errors that last over only a few successive days.

The overestimation of the power-law noise was due to the presence of the autoregressive noise and also the justified hidden periodic effects in the series. This means that neither the white noise plus flicker noise model nor the white noise plus random walk noise model is the preferred model. The best model includes in addition to power-law noise also other noise models like autoregressive noise or as Langbein (2004) used the band-pass-filtered noise. Instead of strict white noise model a short-memory noise process was introduced which led to the reduction of the flicker noise magnitude.

9 Conclusions and Recommendations

9.1 Introduction

The least-squares method is one of the leading principles in the estimation of unknown parameters. In this thesis we introduced and used the least-squares variance component estimation (LS-VCE) method. The research objective was focused on the theoretical as well as practical aspects of LS-VCE for which two different applications of LS-VCE to real data were considered.

To see the analogy with the functional model, we started by reviewing the standard least-squares theory with a linear model in chapter 2. This chapter consists of the estimation part which includes the weighted least-squares estimator, the minimum variance estimator (BLUE), and the maximum likelihood estimator (MLE), but also consists of the validation part which includes the w-test, v-test, and overall model test. In chapter 3 we reviewed the many different variance component estimation methods. We can at least mention the MINQUE, BIQUE, and MLE methods. The underlying assumptions of each method were discussed. The methods give identical estimators under the normality assumption. Chapters 4, 5, and 6 form the body of knowledge of our least-squares variance component estimation and validation theory. Chapters 7 and 8 dealt with applications of LS-VCE to real GPS data. In the next section we summarize the key points of our achievements and eventually provide recommendations for further research.

9.2 Summary

Since the LS-VCE method is undeservedly still one of the lesser known VCE-methods, in chapters 4 and 5, we emphasized that LS-VCE is a simple, flexible, and attractive method. The method is simple since it formulates the variance component estimation model in terms of a model of observation equations with the observable vector $\underline{y}_{\mathrm{vh}} = \mathrm{vh}(\underline{t}\,\underline{t}^T - B^TQ_0B)$, the design matrix $A_{\mathrm{vh}} = [\mathrm{vh}(B^TQ_1B), \ldots, \mathrm{vh}(B^TQ_pB)]$, and the positive-definite weight matrix W_{vh}. The unbiased least-squares estimator of the (co)variance component vector follows directly as

$$\hat{\underline{\sigma}} = (A_{\mathrm{vh}}^T W_{\mathrm{vh}} A_{\mathrm{vh}})^{-1} A_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{y}_{\mathrm{vh}} = N^{-1}\underline{l}\,, \tag{9.1}$$

where the $p \times p$ matrix N and the p-vector $\underline{l}$ are obtained from $\underline{t}$, B^TQ_1B,...,B^TQ_pB and W_{vh}. The unbiasedness property follows directly from the weighted least-squares theory and is independent of the (un)specified distribution of the observables either $\underline{y}$ or $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. The method is flexible, since it works with a user-defined weight matrix W_{vh}. One has thus the possibility to use different classes of weight matrices, one of which is given as

$W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D$ with D the duplication matrix and W_t a positive-definite matrix. Another weight matrix class follows from the class of elliptically contoured distributions.

The LS-VCE method is also attractive, since it allows one to directly apply the existing body of knowledge of least-squares theory. We addressed measures of inconsistencies in the stochastic model, the use of a-priori (co)variance component estimation, the estimability of variance components, robust estimation of variance components, non-negative variance components, and nonlinear variance component estimation. The precision description of the (co)variance component estimator easily follows from application of the error propagation law to equation (9.1), as

$$Q_{\hat{\sigma}} = N^{-1}MN^{-1}, \tag{9.2}$$

with $N = A_{\mathrm{vh}}^T W_{\mathrm{vh}} A_{\mathrm{vh}}$, $M = A_{\mathrm{vh}}^T W_{\mathrm{vh}} Q_{\mathrm{vh}} W_{\mathrm{vh}} A_{\mathrm{vh}}$ and Q_{vh} the covariance matrix of $\underline{y}_{\mathrm{vh}}$. Being able to perform a rigorous precision analysis is important for many applications, such as estimability analysis and significance testing.

In an analogous way to chapter 2 we described the weighted least-squares estimator, the minimum variance estimator, and the maximum likelihood estimator to VCE. To implement the minimum variance estimators of (co)variance components, iterative step-by-step procedures were proposed; two equivalent formulation were given. The last part of chapter 4 was focused on inadmissible variance and covariance components. This can be the case for instance when estimated variance components are negative. Some defects of the VCE model in which the (co)variance components are inestimable or poorly estimable were discussed as well. This was achieved by using a few illustrative examples.

We applied hypothesis testing with the stochastic model, which is the subject of discussion in chapter 5. We presented the w-test, v-test, and overall model test for the stochastic model. The w-test statistic is used to see whether or not a certain noise component is likely to be present in the data set at hand, which consequently can be included in the stochastic model. Following the standard least-squares theory, this test statistic is given as

$$\underline{w} = \frac{\langle P_{A_{\mathrm{vh}}}^{\perp} \mathrm{vh}(C_t), P_{A_{\mathrm{vh}}}^{\perp} \mathrm{vh}(\underline{t}\,\underline{t}^T)\rangle}{\|P_{A_{\mathrm{vh}}}^{\perp} \mathrm{vh}(C_t)\|_{Q_{\mathrm{vh}}^{-1}}}, \tag{9.3}$$

where C_t is a cofactor matrix under the alternative hypothesis. Based on the normal distribution of the original observables we determined the mean, variance, and distribution of the w-test statistic. Under the null hypothesis, the mean is zero and the variance is one. The distribution is a linear combination of mutually independent central chi-square distributions each with one degree of freedom.

An equivalent expression was given by introducing the v-test statistic. The goal is to decrease the number of (co)variance components of the stochastic model by testing the significance of the (co)variance components. If one (co)variance component is insignificant, that component can likely be excluded from the stochastic model. The overall model test was also introduced to generally validate a proposed stochastic model.

In chapters 7 and 8 we applied LS-VCE to real data. The LS-VCE method was applied to the GPS geometry-free observation model in chapter 7. This model is linear and can be easily implemented in computer code. We started by formulating the functional and stochastic model of the GPS observables. Mainly, the variance component of different observation types, the satellite elevation dependence of GPS observables precision, and the correlation between different observation types were considered. We showed for instance

that the precision of the GPS observables depends, to a large extent, upon the elevation of satellites. Also, significant correlation between observation types was found. In order to obtain the minimum variance estimators, and for a proper precision description of the estimates, such important issues should be taken into account in the covariance matrix of the GPS observables.

In chapter 8 we proposed a methodology to assess the noise characteristics in time series of daily position estimates for permanent GPS stations in a global reference frame. We also applied the statistical tests to find an appropriate stochastic model of GPS time series. Using the w-test statistic values, a combination of white noise and flicker noise turned out in general to best characterize the noise in all three position components. An interpretation for the colored noise of the series was given. Unmodelled periodic effects in the data were captured by a set of harmonic functions for which we relied on the least-squares harmonic estimation (LS-HE) method and parameter significance testing developed in the same framework as LS-VCE. Having included harmonic functions into the model, practically only white noise was observed to remain in the data. Remaining time-correlation was present only at very high frequencies (spanning a few days only) which was best expressed as a first order autoregressive noise process. This noise can be caused by common and well-known sources of error like unmodelled tropospheric and ionospheric effects as well as satellite orbit errors. The results confirmed the presence of annual and semiannual signals in the series. We observed also other significant periodic patterns in the series.

9.3 Conclusions

On the basis of the presented theory and the numerical results obtained with real data, the following conclusions can be drawn:

Unification of methods The LS-VCE method is capable of unifying many of the existing VCE-methods. These methods can be easily recovered by making appropriate choices for the weight matrix W_{vh}. From equation (9.1) one directly obtains the minimum variance (co)variance component estimator, if the weight matrix is chosen as $W_{\mathrm{vh}} = Q_{\mathrm{vh}}^{-1}$. The covariance matrix Q_{vh} and its inverse were worked out for the normal distributions. We also made some statements for a class of elliptically contoured distributions. Several important distributions are known to belong to this class, the multivariate normal distribution being one such example. In case of the normal distribution, the REML and the BIQUE are recovered. Without any distribution assumption, the MINQUE can also be recovered from the LS-VCE method by making a particular choice for the weight matrix, namely $W_t = Q_t^{-1}$ in $W_{\mathrm{vh}} = D^T(W_t \otimes W_t)D$.

LS-VCE method LS-VCE is a powerful method for the estimation and validation of the stochastic model parameters. Since the method is based on the least-squares principle, we can directly apply the existing body of knowledge of least-squares theory to the VCE problem. For example it allows one to easily obtain the covariance matrix of the estimators, and also measures of discrepancies such as the quadratic form of the residuals and the w-test statistics can readily be obtained. This can play an important role if one wants to apply the standard quality control theory to the stochastic model. The use of a-priori (co)variance component information, robust estimation of variance components, non-negative variance components, and nonlinear covariance functions can also be dealt with in LS-VCE. This

method can thus be introduced as a standard method for estimation and validation of the (co)variance components of the stochastic model.

Estimability In order for the stochastic model to have a unique solution, the columns of the design matrix A_{vh} should be linearly independent. The necessary and sufficient condition, for the stochastic model to have a unique solution, is thus linear independence of the cofactor matrices $B^T Q_1 B, \ldots, B^T Q_p B$. Therefore, linear independence of $Q_1, \ldots, Q_p$ is only a necessary condition, and not a sufficient one. One can apply the idea of the generalized inverses and the S-transformations to the rank-deficient system of equations in the stochastic model. Estimability, admissibility, and ill-posedness of the stochastic model are three key issues that need to be taken care of in practice. Modification of the variance component model, choosing reasonably good a-priori values for (co)variance components and increasing the redundancy of the functional model, and thereby the redundancy of the stochastic model, cure usually such problems.

Distribution In this thesis we introduced hypotheses testing in the stochastic model. The goal is to come up with an appropriate stochastic model which includes the proper noise components into the covariance matrix, for which we rely on the w-test and v-test statistic. In general, the distribution of the w-test and v-test statistic in the stochastic model is not normal if the original observables are normally distributed. However, based on the central limit theorem, for some special cases, if the redundancy of the model is large enough, the distribution can be approximated by the standard normal distribution. Based on the simulation of a few examples (not included in this thesis) we could draw the same conclusion for different forms of the functional and stochastic model. To stay on the safe side, for the rejection of the null hypothesis, one can also use the Chebyschev inequality which is always independent of the distribution. The Chebyschev inequality does not provide us with a sharp confidence interval (usually gives an upper bound for a specified level of significance α). However, there is no danger with using this if our testing results in the rejection of the null hypothesis. For example, lots of the large values of the w-test statistic given in table 8.5 can simply results in the rejection of the null hypothesis using the Chebyschev inequality. We could thereby conclude that the noise in GPS coordinate time series was best described by a combination of white noise, flicker noise, and autoregressive noise.

Geometry-free model It this thesis we considered the GPS geometry-free (double difference) observation model to assess the stochastics of the GPS code and carrier phase observations. As expected, the variance of a GPS observable generally depends on the elevation of the satellite. Also, significant correlation can occur between different observation types, e.g. between the C1 and P2 codes. For high-precision positioning applications, to obtain best (minimum variance) estimators, it is necessary to use an appropriate covariance matrix for the observables, since $W = Q_y^{-1}$. Using the GPS geometry free observation model, we could only assess the noise behavior of the GPS observables partially. Since the geometry-based model is stronger than the geometry-free model, the use of this model is more fruitful for the application of VCE to GPS observables (see next section).

Time series analysis To obtain a realistic precision description of the estimates for the unknown parameters in the functional model, a proper covariance matrix of the data is required. This means that if the covariance matrix is being estimated by VCE-methods, underparameterization in both functional and stochastic model is not allowed. We observed for instance significant periodic patterns with periods of 350 days and its fractions 350/n, n=1,2,...,8 in the GPS coordinate time series. Neglecting these harmonic signals in the

functional model can seriously overestimate the rate uncertainty, with obvious undesired negative consequences for geophysical research. Their effects are to be considered as periodic systematic effects which may not be captured as power-law noise in the stochastic model. We also observed the presence of the high frequency noise in the data series. To avoid the overestimation (upward bias) of power-law noise, autoregressive noise should be taken into account in the stochastic model in this application.

9.4 Recommendations

There are still some issues that we did not address in this thesis. The following topics need more research in the future:

LS-VCE based estimator In case of the linear model $\mathrm{E}\{\underline{y}\} = Ax$, the BLUE of x is given as $\hat{\underline{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} y$. We cannot use this estimator if Q_y is unknown. If Q_y is unknown, one can use $\hat{Q}_y$ obtained from LS-VCE which leads to an estimator $\hat{\underline{x}}'$. The question is then how the randomness of $\hat{Q}_y$ propagates into the mean and covariance matrix of $\hat{x}'$. For example in order for $\hat{\underline{x}}'$ to be an unbiased estimator of x, a symmetric distribution (e.g. elliptical distributions) of the original observations is required (Teunissen and Amiri-Simkooei, 2006). However, to make a proper statement on the higher order moments, and in general its distribution, more research needs to be done in this field.

Nonlinear model Throughout this thesis, we dealt with the linear form of the functional model. In some practical applications we however deal with *nonlinear* functional models. The GPS geometry-based model is an example of the nonlinear functional model. The nonlinearity needs to be looked into for future plans by applying the LS-VCE to nonlinear problems. Our strategy is of course the linearization and is organized as follows. One can start with an initial covariance matrix and apply the least-squares theory to the linearized form of the functional model to estimate the least-squares residuals. Expression $\hat{\sigma} = N^{-1} l$ with equations (4.110) and (4.112) provides us with estimates for the (co)variance components. Using these estimates one obtains a new covariance matrix and thus a new least-squares residual vector $\hat{e}$. The double-iterative procedure is repeated until the estimated residuals and thus the (co)variance components do not change with further iterations.

Simple methods The large number of observations in case of GPS is always an advantage when estimating the stochastic model parameters; the larger the redundancy in the functional model is, the better the precision of the (co)variance estimators in the stochastic model will be. But, on the other hand, when dealing with great amount of GPS data, the computational burden for VCE will be drastically increased since, in general, one needs successive inversion of Q_y when computing N and l. One may then apply some efficient techniques to reduce the computational load. For example, the *repeated model* of observation equations used in section 7.4 could be helpful. The final estimates were simply obtained by taking the average of the individual estimates of different groups. The other alternative is to use the simplified methods introduced in chapter 3. If one for instance uses a block-diagonal covariance matrix with only some variance components, one can significantly decrease the computational load. This model is known as the disjunctive group model, which always provides positive variance components provided the a-priori variance components are chosen properly. At the point of convergence, the results also coincide with

the almost unbiased estimator (AUE). However, for this special structure the estimates are strictly unbiased. The other alternative, to reduce the computational burden, is to use the Monte-Carlo simulation (see Kusche, 2003a).

Distribution The distribution of the w-test and v-test statistic in the stochastic model needs further research. The distribution is a linear combination of mutually independent central chi-square distributions with one degree of freedom. One should be able to obtain critical values which are needed for effectively hypotheses testing. Since no practical closed form expression for the cumulative distribution function of these test statistics is available, one may use the *asymptotic* expressions, or some numerical root finding methods instead. The other alternative would be to use the normal distribution as an approximation. When many observations are involved in the functional model, the distribution of the w-test and v-test statistic of the stochastic model tends to become *normal*. Concerning the overall model test which was introduced to test the appropriateness of the general structure of the selected stochastic model, the distribution in general looks complicated. We proposed an approximation to handle this problem. The other alternative is to rely on some simulation techniques as bootstrapping or Monte-Carlo. More research needs to be done in this field in future.

Geometry-based model In this thesis we used the geometry-free model to assess the stochastics of GPS observables. Because of the special structure of this model we do have some restrictions to estimate (co)variance components (see example 4.10). Apparently, the GPS *geometry-based* model rather than the geometry-free model is more fruitful. For example, it allows one to estimate the noise components of the carrier phase observations, L1 separately from L2. Generally, the redundancy of the geometry-based model is larger than that of the geometry-free model. The geometry-based model can start for instance from a single difference phase observation which is related to three baseline components, one receiver clock error, and a single difference ambiguity. Usually the double difference ambiguities are estimated by the LAMBDA method and kept fixed in the model. If we have m satellites, for one observation type, the redundancy becomes $b = m - 4$. But, the redundancy in the geometry-free model becomes zero in this case. When we have r observation types, the redundancy of the functional model reads: $b = mr - 4$ for the geometry-based model, and $b = m(r - 1)$ for the geometry-free model.

Other applications We finally emphasize that the LS-VCE is generally applicable. One can find many areas in our field that the LS-VCE can be used. We can at least mention 1) combination of InSAR data and leveling data, 2) weighting of different observations in gravity field modeling, 3) estimation of covariance functions in geosciences and geostatistics, 4) combination of different height systems, 5) combination of GPS, SLR, and VLBI observations, and 6) combination of GPS and upcoming Galileo observations.

Mathematical Background

A

A.1 Trace, Kronecker product and vec operator

Assuming all matrices and vectors involved have appropriate dimensions, the following properties hold for the Kronecker product, trace, determinant,rank, and vec-operator:

$$\mathrm{tr}(U) = \mathrm{tr}(U^T) \tag{A.1}$$

$$\mathrm{tr}(UV) = \mathrm{tr}(VU) \tag{A.2}$$

$$(U+V)\otimes S = U\otimes S + V\otimes S \tag{A.3}$$

$$U\otimes(V+S) = U\otimes V + U\otimes S \tag{A.4}$$

$$(UV)\otimes(ST) = (U\otimes S)(V\otimes T) \tag{A.5}$$

$$\mathrm{tr}(U\otimes V) = \mathrm{tr}(U)\mathrm{tr}(V) \tag{A.6}$$

$$\mathrm{rank}(U\otimes V) = \mathrm{rank}(U)\mathrm{rank}(V) \tag{A.7}$$

$$\det(U\otimes V) = \det(U)^n \det(V)^m,\ U: m\times m\ \ V: n\times n \tag{A.8}$$

$$(U\otimes V)^T = U^T\otimes V^T \tag{A.9}$$

$$(U\otimes V)^{-1} = U^{-1}\otimes V^{-1} \tag{A.10}$$

$$\mathrm{vec}(uv^T) = v\otimes u \tag{A.11}$$

$$\mathrm{vec}(U)^T\mathrm{vec}(V) = \mathrm{tr}(U^TV) \tag{A.12}$$

$$\mathrm{vec}(UVS) = (S^T\otimes U)\mathrm{vec}(V) \tag{A.13}$$

$$\mathrm{tr}(UVST) = \mathrm{vec}(T^T)^T(S^T\otimes U)\mathrm{vec}(V) = \mathrm{vec}(T)^T(U\otimes S^T)\mathrm{vec}(V^T) \tag{A.14}$$

A.2 Duplication and commutation matrices

For symmetric matrix S of size n and arbitrary matrices U and V of size $n\times n$, the following formulas holds between the duplication matrix D and the commutation matrix K:

$$\mathrm{vec}(S) = D\mathrm{vh}(S),\ D \text{ has full column-rank} \tag{A.15}$$

$$D^+ = (D^TD)^{-1}D^T \tag{A.16}$$

$$\mathrm{vh}(S) = D^+\mathrm{vec}(S) \tag{A.17}$$

$$D^+D = I, \text{ and } DD^+ = D^{+T}D^T \text{ a projector} \tag{A.18}$$

$$\mathrm{vec}(S) = DD^+\mathrm{vec}(S) \tag{A.19}$$

$$(D^+(S\otimes S)D)^{-1} = D^+(S^{-1}\otimes S^{-1})D \tag{A.20}$$

$$DD^+(U\otimes U)D = (U\otimes U)D \tag{A.21}$$

$$D^+(U \otimes U)DD^+ = D^+(U \otimes U) \tag{A.22}$$

$$(D^T(U \otimes U)D)^{-1} = D^+(U^{-1} \otimes U^{-1})D^{+T} \tag{A.23}$$

$$K(U \otimes V) = (V \otimes U)K \tag{A.24}$$

$$K(U \otimes V)K = V \otimes U \tag{A.25}$$

$$KD = D, \quad D^+K = D^+ \tag{A.26}$$

$$DD^+ = \frac{1}{2}(I + K) \tag{A.27}$$

B Derivation of Equations

B.1 Equation (5.116)

To see how the distribution of the w-test statistic looks like, we need the eigenvalues of $C_y Q_{\hat{e}}^-$ in equation (5.82)

$$C_y Q_{\hat{e}}^- = \frac{1}{\sigma^2}\begin{bmatrix} 0 & 0 \\ 0 & I_m \end{bmatrix}\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} 0 & 0 \\ Q_2 & Q_1 \end{bmatrix} \tag{B.1}$$

with Q_1 and Q_2 as

$$q_{ij}^{(1)} = \begin{cases} \frac{2m-1}{2m} & \text{if} \quad i=j \\ \frac{-1}{2m} & \text{if} \quad i\neq j \end{cases}, \quad \text{and} \quad q_{ij}^{(2)} = \frac{-1}{2m} \;\; \forall\, i,j = 1,...,m \tag{B.2}$$

The matrix $C_y Q_{\hat{e}}^-$ is of rank m, and we just need to compute the eigenvalues of the full rank matrix $\frac{1}{\sigma^2}Q_1$; note that $\frac{1}{\sigma^2}Q_1$ is the block diagonal of $Q_{\hat{e}}^-$. It can be rewritten as

$$\frac{1}{\sigma^2}Q_1 = \frac{1}{2\sigma^2}P_1 + \frac{1}{2\sigma^2}I_m \tag{B.3}$$

with P_1 an orthogonal projector of rank $m-1$ and of the form

$$p_{ij} = \begin{cases} \frac{m-1}{m} & \text{if} \quad i=j \\ \frac{-1}{m} & \text{if} \quad i\neq j \end{cases} \tag{B.4}$$

The eigenvalues of $\frac{1}{\sigma^2}Q_1$ then read

$$\left|\frac{1}{\sigma^2}Q_1 - \gamma I_m\right| = \left|\frac{1}{2\sigma^2}P_1 - (\gamma - \frac{1}{2\sigma^2})I_m\right| = 0 \tag{B.5}$$

or $|P_1 - (2\sigma^2\gamma - 1)I_m| = 0$. Since P_1 is a projector of rank $m-1$, its eigenvalues read $2\sigma^2\gamma_1 - 1 = 0,\;\; 2\sigma^2\gamma_2 - 1 = ... = 2\sigma^2\gamma_m - 1 = 1$, which yields

$$\gamma_1 = \frac{1}{2\sigma^2}, \;\; \gamma_2 = ... = \gamma_m = \frac{1}{\sigma^2} \tag{B.6}$$

This, with equation (5.82), $w^d = \frac{\sqrt{m-1}}{2\sigma^2}$ and $\mathrm{tr}(C_y Q_{\hat{e}}^-) = \frac{2m-1}{2\sigma^2}$ gives the $2m-2$ number of non-zero eigenvalues λ as

$$\lambda_1 = 0; \;\; \lambda_2 = ... = \lambda_m = \frac{1}{2\sqrt{m-1}}; \;\; \lambda_{m+1} = ... = \lambda_{2m-1} = \frac{-1}{2\sqrt{m-1}} \tag{B.7}$$

B.2 Equation (5.123)

To see how the distribution looks like, we need the eigenvalues of $C_y Q_{\hat{e}}^-$ in (5.82)

$$C_y Q_{\hat{e}}^- = \frac{1}{\sigma^2}\begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix}\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix} = \frac{1}{\sigma^2}\begin{bmatrix} Q_2 & Q_1 \\ Q_1 & Q_2 \end{bmatrix} \tag{B.8}$$

This matrix is of rank $2m-1$ and its eigenvalues follow from $|C_y Q_{\hat{e}}^- - \gamma I_{2m}|$, or

$$\left|\begin{bmatrix} Q_2 & Q_1 \\ Q_1 & Q_2 \end{bmatrix} - \sigma^2\gamma\begin{bmatrix} I_m & 0 \\ 0 & I_m \end{bmatrix}\right| = 0 \tag{B.9}$$

The preceding equation, can be rewritten as follows:

$$\left|\begin{bmatrix} Q_2+I_m & Q_1-I_m \\ Q_1-I_m & Q_2+I_m \end{bmatrix} - \begin{bmatrix} I_m & -I_m \\ -I_m & I_m \end{bmatrix} - \sigma^2\gamma\begin{bmatrix} I_m & 0 \\ 0 & I_m \end{bmatrix}\right| = 0 \tag{B.10}$$

or

$$\left|\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix} - \begin{bmatrix} (\sigma^2\gamma+1)I_m & -I_m \\ -I_m & (\sigma^2\gamma+1)I_m \end{bmatrix}\right| = 0 \tag{B.11}$$

or in terms of Kronecker product $\otimes$ as

$$\left|\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix} - \begin{bmatrix} (\sigma^2\gamma+1) & -1 \\ -1 & (\sigma^2\gamma+1) \end{bmatrix} \otimes I_m\right| = 0 \tag{B.12}$$

Singular value decomposition for the matrix $\begin{bmatrix} (\sigma^2\gamma+1) & -1 \\ -1 & (\sigma^2\gamma+1) \end{bmatrix}$ gives

$$\left|\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix} - \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \sigma^2\gamma+2 & 0 \\ 0 & \sigma^2\gamma \end{bmatrix}\begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \otimes I_m\right| = 0 \tag{B.13}$$

or, using the properties of the Kronecker product, as

$$\left|\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -I_m & I_m \\ I_m & I_m \end{bmatrix}\begin{bmatrix} (\sigma^2\gamma+2)I_m & 0 \\ 0 & (\sigma^2\gamma)I_m \end{bmatrix}\begin{bmatrix} -I_m & I_m \\ I_m & I_m \end{bmatrix}\right| = 0 \tag{B.14}$$

or

$$\left|\frac{1}{2}\begin{bmatrix} -I_m & I_m \\ I_m & I_m \end{bmatrix}\begin{bmatrix} Q_1 & Q_2 \\ Q_2 & Q_1 \end{bmatrix}\begin{bmatrix} -I_m & I_m \\ I_m & I_m \end{bmatrix} - \begin{bmatrix} (\sigma^2\gamma+2)I_m & 0 \\ 0 & (\sigma^2\gamma)I_m \end{bmatrix}\right| = 0 \tag{B.15}$$

which simplifies to

$$\left|\begin{bmatrix} Q_1-Q_2 & 0 \\ 0 & Q_1+Q_2 \end{bmatrix} - \begin{bmatrix} (\sigma^2\gamma+2)I_m & 0 \\ 0 & (\sigma^2\gamma)I_m \end{bmatrix}\right| = 0 \tag{B.16}$$

Since the matrices are in a block-diagonal form, the preceding equation can be split into two parts, i.e.

$$\left|Q_1-Q_2-(\sigma^2\gamma+2)I_m\right| = 0\,, \quad \left|Q_1+Q_2-(\sigma^2\gamma)I_m\right| = 0 \tag{B.17}$$

in which $Q_1-Q_2 = \frac{m-1}{m}I_m$, and hence its eigenvalues read

$$\sigma^2\gamma+2 = 1 \;\rightarrow\; \gamma_1 = ... = \gamma_m = -\frac{1}{\sigma^2} \tag{B.18}$$

and Q_1+Q_2 is a projector of rank $m-1$, thus

$$\sigma^2\gamma = 1 \;\rightarrow\; \gamma_{m+1} = ... = \gamma_{2m-1} = \frac{1}{\sigma^2},\;\; \gamma_{2m} = 0 \tag{B.19}$$

This, with equation (5.82), $w^d = \frac{1}{\sigma^2}\sqrt{\frac{2m(m-1)}{2m-1}}$ and $\mathrm{tr}(C_y Q_{\hat{e}}^-) = \frac{-1}{\sigma^2}$, gives the $2m-1$ number of non-zero eigenvalues λ as

$$\lambda_1 = ... = \lambda_m = -\frac{m-1}{\sqrt{2m(2m-1)(m-1)}};\; \lambda_{m+1} = ... = \lambda_{2m-1} = \frac{m}{\sqrt{2m(2m-1)(m-1)}} \tag{B.20}$$

B.3 Equation (5.158)

Second term: The expectation of the second term in equation (5.156) reads:

$$\begin{aligned} \mathsf{E}\{(\underline{t}^T Q_t^{-1}\underline{t})^2(\underline{l}^T N^{-1}\underline{l})\} &= \mathsf{E}\{(\underline{t}^T Q_t^{-1}\underline{t})(\underline{t}^T Q_t^{-1}\underline{t})(\underline{l}_i n_{ij}^{-1}\underline{l}_j)\} \\ &= n_{ij}^{-1}\mathsf{E}\{(\underline{t}^T Q_t^{-1}\underline{t})(\underline{t}^T Q_t^{-1}\underline{t})(\underline{l}_i)(\underline{l}_j)\}\,, \end{aligned} \tag{B.21}$$

with $\underline{l}_i = \frac{1}{2}\underline{t}^T Q_t^{-1} Q_{t_i} Q_t^{-1}\underline{t}$ and $\underline{l}_j = \frac{1}{2}\underline{t}^T Q_t^{-1} Q_{t_j} Q_t^{-1}\underline{t}$. This equation consists of multiplication of *four quadratic forms* in $\underline{t}$. The expectation of this multiplication, consisting of 24 terms, can be obtained from equation (C.50). Note that here we have $\underline{t} \sim \mathsf{N}(0, Q_t)$ instead of $\underline{e} \sim \mathsf{N}(0, Q_y)$ used in appendix C. Also, we have $A = B = Q_t^{-1}$, $C = \frac{1}{2}Q_t^{-1}Q_{t_i}Q_t^{-1}$ and $D = \frac{1}{2}Q_t^{-1}Q_{t_j}Q_t^{-1}$. The terms 1 through 7 in equation (C.50) then read

$$\begin{aligned} F1 &= b\,b\,l_i l_j n_{ij}^{-1} = b^2(l^T N^{-1} l) = b^3/2\,; \\ F2 &= 2b\,l_i l_j n_{ij}^{-1} = 2b(l^T N^{-1} l) = b^2\,; \\ F3 &= 2l_i\,b\,l_j n_{ij}^{-1} = 2b(l^T N^{-1} l) = b^2\,; \\ F4 &= 2l_j\,b\,l_i n_{ij}^{-1} = 2b(l^T N^{-1} l) = b^2\,; \\ F5 &= 2l_i\,b\,l_j n_{ij}^{-1} = 2b(l^T N^{-1} l) = b^2\,; \\ F6 &= 2l_j\,b\,l_i n_{ij}^{-1} = 2b(l^T N^{-1} l) = b^2\,; \\ F7 &= n_{ij}\,b\,b\,n_{ij}^{-1} = b^2\delta_{ii} = b^2 p\,, \end{aligned} \tag{B.22}$$

in which equation (5.152), i.e. $l^T N^{-1} l = \frac{b}{2}$, has been used. The terms 8 through 18 in equation (C.50) then read

$$\begin{aligned} F8 &= 4l_j l_i n_{ij}^{-1} = 4l^T N^{-1} l = 2b\,; \\ F9 &= 4l_j l_i n_{ij}^{-1} = 4l^T N^{-1} l = 2b\,; \\ F10 &= 4l_i l_j n_{ij}^{-1} = 4l^T N^{-1} l = 2b\,; \\ F11 &= 4l_i l_j n_{ij}^{-1} = 4l^T N^{-1} l = 2b\,; \\ F12 &= 2b\,n_{ij} n_{ij}^{-1} = 2b\delta_{ii} = 2b\,p\,; \\ F13 &= 2b\,n_{ij} n_{ij}^{-1} = 2b\delta_{ii} = 2b\,p\,; \\ F14 &= 2b\,n_{ij} n_{ij}^{-1} = 2b\delta_{ii} = 2b\,p\,; \\ F15 &= 2b\,n_{ij} n_{ij}^{-1} = 2b\delta_{ii} = 2b\,p\,; \\ F16 &= 2b\,n_{ij} n_{ij}^{-1} = 2b\delta_{ii} = 2b\,p\,; \\ F17 &= 4l_i l_j n_{ij}^{-1} = 4l^T N^{-1} l = 2b\,; \\ F18 &= 4l_j l_i n_{ij}^{-1} = 4l^T N^{-1} l = 2b\,. \end{aligned} \tag{B.23}$$

Similarly, the terms 19 through 24 in equation (C.50) can then be evaluated as

$$\begin{aligned} F19 &= 4n_{ij} n_{ij}^{-1} = 4p\,; \\ F20 &= 4n_{ij} n_{ij}^{-1} = 4p\,; \\ F21 &= 4n_{ij} n_{ij}^{-1} = 4p\,; \\ F22 &= 4n_{ij} n_{ij}^{-1} = 4p\,; \\ F23 &= 4n_{ij} n_{ij}^{-1} = 4p\,; \\ F24 &= 4n_{ij} n_{ij}^{-1} = 4p\,. \end{aligned} \tag{B.24}$$

Adding up the terms in equations (B.22), (B.23) and (B.24), i.e. $F1$ through $F24$, the expectation of the second term in equation (5.156) can be simplified to

$$\mathsf{E}\{(\underline{t}^T Q_t^{-1}\underline{t})^2(\underline{l}^T N^{-1}\underline{l})\} = \frac{b^3}{2} + 5b^2 + b^2 p + 12b + 10bp + 24p\,. \tag{B.25}$$

B.4 Equation (5.159)

Third term: The expectation of the third term in equation (5.156) reads

$$\mathsf{E}\{(\underline{l}^T N^{-1}\underline{l})^2\} = \mathsf{E}\{\underline{l}_i n_{ij}^{-1}\underline{l}_j\, \underline{l}_k n_{kl}^{-1}\underline{l}_l\} = n_{ij}^{-1}n_{kl}^{-1}\mathsf{E}\{\underline{l}_i\underline{l}_j\underline{l}_k\underline{l}_l\}\,, \tag{B.26}$$

where $\underline{l}_i = \frac{1}{2}\underline{t}^T Q_t^{-1}Q_{t_i}Q_t^{-1}\underline{t}$ and so are the other terms. The expectation of the term $\underline{l}_i\underline{l}_j\underline{l}_k\underline{l}_l$ in equation (B.26), as before, consists of the multiplication of *four quadratic forms* in $\underline{t}$. The expectation of this multiplication can be obtained from equation (C.50) but this time with $Q_y = Q_t$, $A = \frac{1}{2}Q_t^{-1}Q_{t_i}Q_t^{-1}$, $B = \frac{1}{2}Q_t^{-1}Q_{t_j}Q_t^{-1}$, $C = \frac{1}{2}Q_t^{-1}Q_{t_k}Q_t^{-1}$ and $D = \frac{1}{2}Q_t^{-1}Q_{t_l}Q_t^{-1}$. The terms 1 through 7 in equation (C.50) then read

$$\begin{array}{lllllllll}
F1 &=& n_{ij}^{-1}n_{kl}^{-1}l_il_jl_kl_l &=& l_in_{ij}^{-1}l_j\,l_kn_{kl}^{-1}l_l &=& (l^TN^{-1}l)^2 &=& b^2/4\,;\\
F2 &=& n_{ij}^{-1}n_{kl}^{-1}n_{ij}l_kl_l &=& n_{ij}^{-1}n_{ji}l_kn_{kl}^{-1}l_l &=& p\,l_kn_{kl}^{-1}l_l &=& p\,b/2\,;\\
F3 &=& n_{ij}^{-1}n_{kl}^{-1}n_{ik}l_jl_l &=& \delta_{il}n_{ij}^{-1}l_jl_l &=& l_in_{ij}^{-1}l_j &=& b/2\,;\\
F4 &=& n_{ij}^{-1}n_{kl}^{-1}n_{il}l_jl_k &=& \delta_{ik}n_{ij}^{-1}l_jl_k &=& l_in_{ij}^{-1}l_j &=& b/2\,;\\
F5 &=& n_{ij}^{-1}n_{kl}^{-1}n_{jk}l_il_l &=& \delta_{ik}n_{kl}^{-1}l_il_l &=& l_kn_{kl}^{-1}l_l &=& b/2\,;\\
F6 &=& n_{ij}^{-1}n_{kl}^{-1}n_{jl}l_il_k &=& \delta_{il}n_{kl}^{-1}l_il_k &=& l_kn_{kl}^{-1}l_l &=& b/2\,;\\
F7 &=& n_{ij}^{-1}n_{kl}^{-1}n_{kl}l_il_j &=& n_{kl}^{-1}n_{lk}l_in_{ij}^{-1}l_j &=& p\,l_in_{ij}^{-1}l_j &=& p\,b/2\,.
\end{array} \tag{B.27}$$

If we now introduce the *third-order tensor* V as

$$v_{ijk} = \frac{1}{2}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j}Q_t^{-1}Q_{t_k}),\ i,j,k = 1,2,...,p\,, \tag{B.28}$$

it is simple to verify that the following equality holds true between the entries of V

$$v_{ijk} = v_{jki} = v_{kij},\ i,j,k = 1,2,...,p\,. \tag{B.29}$$

We will need the term $N^{-1} : VN^{-1}l$ in the subsequent derivations, in which : denotes the *double inner (dot) product* of two tensors. The double dot product between the second order tensor (matrix) N^{-1} and the third order tensor V produces a vector $s = N^{-1} : V$, with components $s_k = n_{ij}^{-1}v_{jik},\ k = 1,2,...,p$ (only one free index k). The term $N^{-1} : VN^{-1}l$ simplifies to

$$\begin{array}{lll}
N^{-1} : VN^{-1}l &=& N^{-1} : V\sigma = n_{ij}^{-1}v_{ijk}\sigma_k\\
&=& \frac{1}{2}n_{ij}^{-1}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j}Q_t^{-1}Q_{t_k})\sigma_k\\
&=& \frac{1}{2}n_{ij}^{-1}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j}Q_t^{-1}[\sigma_kQ_{t_k}])\\
&=& \frac{1}{2}n_{ij}^{-1}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j})\\
&=& \frac{1}{2}n_{ij}^{-1}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j})\\
&=& n_{ij}^{-1}n_{ij} = \delta_{ii} = p\,.
\end{array} \tag{B.30}$$

With these in mind, the terms 8 through 18 in equation (C.50) can now be simplified as follows (see equation (B.30)):

$$\begin{array}{lllllllll}
F8 &=& n_{ij}^{-1}n_{kl}^{-1}l_lv_{ijk} &=& n_{ji}^{-1}v_{ijk}n_{kl}^{-1}l_l &=& N^{-1} : VN^{-1}l &=& p\,;\\
F9 &=& n_{ij}^{-1}n_{kl}^{-1}l_lv_{ikj} &=& n_{ij}^{-1}v_{jik}n_{kl}^{-1}l_l &=& N^{-1} : VN^{-1}l &=& p\,;\\
F10 &=& n_{ij}^{-1}n_{kl}^{-1}l_kv_{ijl} &=& n_{ji}^{-1}v_{ijl}n_{lk}^{-1}l_k &=& N^{-1} : VN^{-1}l &=& p\,;\\
F11 &=& n_{ij}^{-1}n_{kl}^{-1}l_kv_{ilj} &=& n_{ij}^{-1}v_{jil}n_{lk}^{-1}l_k &=& N^{-1} : VN^{-1}l &=& p\,;\\
F12 &=& n_{ij}^{-1}n_{kl}^{-1}l_jv_{ikl} &=& n_{lk}^{-1}v_{kli}n_{ij}^{-1}l_j &=& N^{-1} : VN^{-1}l &=& p\,;\\
F13 &=& n_{ij}^{-1}n_{kl}^{-1}l_jv_{ilk} &=& n_{kl}^{-1}v_{lki}n_{ij}^{-1}l_j &=& N^{-1} : VN^{-1}l &=& p\,;\\
F14 &=& n_{ij}^{-1}n_{kl}^{-1}l_iv_{jkl} &=& n_{lk}^{-1}v_{klj}n_{ji}^{-1}l_i &=& N^{-1} : VN^{-1}l &=& p\,;\\
F15 &=& n_{ij}^{-1}n_{kl}^{-1}l_iv_{jlk} &=& n_{kl}^{-1}v_{lkj}n_{ji}^{-1}l_i &=& N^{-1} : VN^{-1}l &=& p\,;\\
F16 &=& n_{ij}^{-1}n_{kl}^{-1}n_{ij}n_{kl} &=& n_{ij}^{-1}n_{ji}n_{kl}^{-1}n_{lk} &=& \delta_{ii}\delta_{kk} &=& p^2\,;\\
F17 &=& n_{ij}^{-1}n_{kl}^{-1}n_{ik}n_{jl} &=& n_{ij}^{-1}n_{jl}n_{lk}^{-1}n_{ki} &=& \delta_{il}\delta_{li} &=& p\,;\\
F18 &=& n_{ij}^{-1}n_{kl}^{-1}n_{il}n_{jk} &=& n_{ij}^{-1}n_{jk}n_{kl}^{-1}n_{li} &=& \delta_{ik}\delta_{ki} &=& p\,.
\end{array} \tag{B.31}$$

Introducing the *fourth-order tensor* Z as

$$z_{ijkl} = \frac{1}{2}\mathrm{tr}(Q_t^{-1}Q_{t_i}Q_t^{-1}Q_{t_j}Q_t^{-1}Q_{t_k}Q_t^{-1}Q_{t_l}),\ i,j,k,l = 1,2,...,p\,, \tag{B.32}$$

gives $z_{ijkl} = z_{jkli} = z_{klij} = z_{lijk},\ i,j,k,l = 1,2,...,p$. Note that here $N^{-1} : Z$ is a second order tensor (a matrix), and $N^{-1} : Z : N^{-1}$ is a scalar. Taking these circumstances into account, the terms 19 through 24 in equation (C.50) read then

$$\begin{array}{rclcll} F19 & = & n_{ij}^{-1}n_{kl}^{-1}z_{ijkl} & = & n_{ji}^{-1}z_{ijkl}n_{lk}^{-1} & = & N^{-1} : Z : N^{-1}; \\ F20 & = & n_{ij}^{-1}n_{kl}^{-1}z_{ijlk} & = & n_{ji}^{-1}z_{ijlk}n_{kl}^{-1} & = & N^{-1} : Z : N^{-1}; \\ F21 & = & n_{ij}^{-1}n_{kl}^{-1}z_{ikjl} & = & n_{ji}^{-1}z_{ijkl}^{23}n_{lk}^{-1} & = & N^{-1} : Z^{23} : N^{-1}; \\ F22 & = & n_{ij}^{-1}n_{kl}^{-1}z_{iklj} & = & n_{ij}^{-1}z_{jikl}n_{lk}^{-1} & = & N^{-1} : Z : N^{-1}; \\ F23 & = & n_{ij}^{-1}n_{kl}^{-1}z_{iljk} & = & n_{ji}^{-1}z_{ijlk}^{23}n_{kl}^{-1} & = & N^{-1} : Z^{23} : N^{-1}; \\ F24 & = & n_{ij}^{-1}n_{kl}^{-1}z_{ilkj} & = & n_{ij}^{-1}z_{jilk}n_{kl}^{-1} & = & N^{-1} : Z : N^{-1}, \end{array} \tag{B.33}$$

where Z^{23} denotes the Z matrix in which its *columns*, notated by 2, and *heights*, notated by 3, have been transposed. Unfortunately, the terms $N^{-1} : Z : N^{-1}$ and $N^{-1} : Z^{23} : N^{-1}$ can not be simplified any more. Adding up the terms in equations (B.27), (B.31) and (B.33), i.e. $F1$ through $F24$, the expectation of the third term in equation (5.156) reads

$$\mathsf{E}\{(\underline{l}N^{-1}\underline{l})^2\} = \frac{b^2}{4} + p\,b + 2b + 10p + p^2 + 4N^{-1} : Z : N^{-1} + 2N^{-1} : Z^{23} : N^{-1}\,. \tag{B.34}$$

B.5 Equation (5.174)

Equation (5.172) can be reformulated as

$$2\underline{T}_{\mathrm{df}} + 2\underline{\epsilon} = \left[\underline{t}^T Q_t^{-1}\underline{t} - \frac{1}{2}\right]^2 - \frac{1}{4}\,, \tag{B.35}$$

or after a few simple operations as

$$\frac{1}{2}\left(\sqrt{8(\underline{T}_{\mathrm{df}} + \underline{\epsilon}) + 1} + 1\right) = \underline{t}^T Q_t^{-1}\underline{t}\,. \tag{B.36}$$

The term $\sqrt{8(\underline{T}_{\mathrm{df}} + \underline{\epsilon}) + 1}$ can be rewritten as

$$\sqrt{8(\underline{T}_{\mathrm{df}} + \underline{\epsilon}) + 1} = \sqrt{8\underline{T}_{\mathrm{df}} + 1}\left(1 + \frac{8\underline{\epsilon}}{8\underline{T}_{\mathrm{df}} + 1}\right)^{1/2}. \tag{B.37}$$

Using the Taylor series expansion of $(1+x)^{1/2} \approx 1 + \frac{x}{2}$, the preceding formula can be approximated as

$$\sqrt{8(\underline{T}_{\mathrm{df}} + \underline{\epsilon}) + 1} \approx \sqrt{8\underline{T}_{\mathrm{df}} + 1} + \frac{4\underline{\epsilon}}{\sqrt{8\underline{T}_{\mathrm{df}} + 1}}\,. \tag{B.38}$$

Substitution of equation (B.38) into equation (B.36) yields

$$\frac{1}{2}\left(\sqrt{8\underline{T}_{\mathrm{df}} + 1} + 1\right) + \frac{2\underline{\epsilon}}{\sqrt{8\underline{T}_{\mathrm{df}} + 1}} \approx \underline{t}^T Q_t^{-1}\underline{t} \tag{B.39}$$

Substitution of the second term $\sqrt{8\underline{T}_{\mathrm{df}} + 1}$ in equation (B.39) with the approximation $2(\underline{t}^T Q_t^{-1}\underline{t} - \frac{1}{2})$ gives

$$\underline{T}_b' + \underline{\epsilon}' \approx \underline{t}^T Q_t^{-1}\underline{t}\,, \tag{B.40}$$

with

$$\underline{T}_b' = \frac{1}{2}\left(\sqrt{8\underline{T}_{\mathrm{df}} + 1} + 1\right); \quad \underline{\epsilon}' = \frac{\underline{\epsilon}}{\underline{t}^T Q_t^{-1}\underline{t} - \frac{1}{2}}\,. \tag{B.41}$$

Substituting for $\underline{\epsilon}$ in equation (B.41) from equation (5.173) yields the (bias) statistic $\underline{\epsilon}'$ as

$$\underline{\epsilon}' = \frac{1}{2}\left[\frac{\underline{t}^T Q_t^{-1}\hat{Q}_t Q_t^{-1}\underline{t} - \frac{1}{2}}{\underline{t}^T Q_t^{-1}\underline{t} - \frac{1}{2}} - 1\right], \tag{B.42}$$

or approximately

$$\underline{\epsilon}' \approx \frac{1}{2}\left[\frac{\underline{t}^T Q_t^{-1}\hat{Q}_t Q_t^{-1}\underline{t}}{\underline{t}^T Q_t^{-1}\underline{t}} - 1\right]. \tag{B.43}$$

C Moments of Normally Distributed Data

C.1 Moment generating function

Let $\underline{y}$ be an $m \times 1$ random vector with the multivariate probability density function $f(y)$. Then, the *moment generating function* $M_y(s)$, with s an arbitrary m-vector, of $\underline{y}$ reads

$$M_y(s) = \mathsf{E}\{e^{s^T \underline{y}}\} = \mathsf{E}\{\exp(s_1 \underline{y}_1 + ... + s_m \underline{y}_m)\}\,, \tag{C.1}$$

or

$$M_y(s) = \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} e^{s_1 y_1 + ... + s_m y_m} f(y_1, ..., y_m) dy_1 ... dy_m\,, \tag{C.2}$$

from which the k^{th} *moment* of y, namely $\mathsf{E}\{\underline{y}_1^{k_1} ... \underline{y}_m^{k_m}\}$ with $k = \sum_{i=1}^m k_i$, can be obtained by differentiations of $M_y(s)$ k times with respect to s and setting s equal to zero, since

$$\begin{aligned} \left.\frac{\partial^k M_y(s)}{\partial s_1^{k_1} ... \partial s_m^{k_m}}\right|_{s=0} &= \left. \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} y_1^{k_1} ... y_m^{k_m} e^{s^T y} f(y) dy_1 ... dy_m \right|_{s=0} \\ &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} y_1^{k_1} ... y_m^{k_m} f(y) dy_1 ... dy_m \\ &= \mathsf{E}\{\underline{y}_1^{k_1} ... \underline{y}_m^{k_m}\}\,. \end{aligned} \tag{C.3}$$

C.2 Moment generating function for normal distribution

If the $m \times 1$ stochastic vector $\underline{y}$ is normally distributed, i.e. $\underline{y} \sim \mathsf{N}_m(\mu_y, Q_y)$, its probability density function reads

$$f(y) = \frac{1}{(2\pi)^{\frac{m}{2}} (\det Q_y)^{\frac{1}{2}}} \exp[-\frac{1}{2}(y - \mu_y)^T Q_y^{-1} (y - \mu_y)] \tag{C.4}$$

The moment generating function of $\underline{y}$ is then

$$M_y(s) = \frac{1}{(2\pi)^{\frac{m}{2}} (\det Q_y)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \exp[s^T y - \frac{1}{2}(y - \mu_y)^T Q_y^{-1} (y - \mu_y)] dy_1 ... dy_m \tag{C.5}$$

By rewriting

$$s^T y - \frac{1}{2}(y - \mu_y)^T Q_y^{-1} (y - \mu_y) = s^T \mu_y + \frac{1}{2} s^T Q_y s - \frac{1}{2}(y - \mu_y - Q_y s)^T Q_y^{-1} (y - \mu_y - Q_y s) \tag{C.6}$$

and considering

$$\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \exp[\frac{1}{2}(y - \mu_y - Q_y s)^T Q_y^{-1} (y - \mu_y - Q_y s)] dy_1 ... dy_m = (2\pi)^{\frac{m}{2}} (\det Q_y)^{\frac{1}{2}} \tag{C.7}$$

we find that

$$\boxed{M_y(s) = \exp[s^T \mu_y + \frac{1}{2} s^T Q_y s]} \tag{C.8}$$

Note that $M_y(s)$ only depends on μ_y and Q_y. So all moments of a normally distributed vector $\underline{y}$ are fully determined by its expectation (first moment) and dispersion (second central moment).

C.3 First four moments of normally distributed data

Using the results of the previous subsection, we can now derive the first four moments of a normally distributed vector. If we write the moment generating function of the m-dimensional normal distribution as

$$M_y(s) = \exp\left[\sum_{\alpha=1}^{m} s_\alpha \mu_{y_\alpha} + \frac{1}{2}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{m} s_\alpha s_\beta Q_{y_{\alpha\beta}}\right] \tag{C.9}$$

and consider that $M_y(s=0) = 1$, we may derive straightforwardly:

$$\mathsf{E}\{\underline{y}_i\} = \left.\frac{\partial M_y(s)}{\partial s_i}\right|_{s=0} = \left(\mu_{y_i} + \sum_{\alpha}^{m} s_\alpha Q_{y_{\alpha i}}\right) M_y(s)\Big|_{s=0} = \mu_{y_i} \tag{C.10}$$

In the sequel, for the sake of convenience, we denote $\mu_{y_i} = \mu_{y_i}$ and $Q_{y_{ij}} = q_{ij}$ and in addition we disregard the summation Σ (see NOTE 5.1). The second, third, and fourth moments are obtaied as

$$\begin{array}{rcl} \mathsf{E}\{\underline{y}_i\underline{y}_j\} & = & \left.\frac{\partial^2 M_y(s)}{\partial s_i \partial s_j}\right|_{s=0} \\ & = & q_{ij}M_y(s) + (\mu_i + s_\alpha q_{\alpha i})\,(\mu_j + s_\alpha q_{\alpha j})\,M_y(s)|_{s=0} = q_{ij} + \mu_i\mu_j \end{array} \tag{C.11}$$

and

$$\mathsf{E}\{\underline{y}_i\underline{y}_j\underline{y}_k\} = \left.\frac{\partial^3 M_y(s)}{\partial s_i \partial s_j \partial s_k}\right|_{s=0} = q_{ij}\mu_k + q_{ik}\mu_j + q_{jk}\mu_i + \mu_i\mu_j\mu_k \tag{C.12}$$

and

$$\begin{array}{rcl} \mathsf{E}\{\underline{y}_i\underline{y}_j\underline{y}_k\underline{y}_k\} & = & \left.\frac{\partial^4 M_y(s)}{\partial s_i \partial s_j \partial s_k \partial s_l}\right|_{s=0} \\ & = & q_{ij}q_{kl} + q_{ij}\mu_k\mu_l + q_{ik}q_{jl} + q_{ik}\mu_j\mu_l + q_{jk}q_{il} \\ & + & q_{jk}\mu_i\mu_l + q_{il}\mu_j\mu_k + q_{jl}\mu_i\mu_k + q_{kl}\mu_i\mu_j + \mu_i\mu_j\mu_k\mu_l \end{array} \tag{C.13}$$

respectively. If we now define the residual vector e as (not to be confused with the vector of *least-squares residuals* $\underline{\hat{e}}$)

$$\underline{e} = \underline{y} - \mu_y \tag{C.14}$$

it can be concluded that $\hat{e}$ is also normally distributed. Its shape, expressed by covariance matrix Q_y, is the same as the shape of the distribution of $\underline{y}$ but its mean has moved to the origin

$$\underline{e} \sim \mathsf{N}_m(0, Q_y) \tag{C.15}$$

The preceding first four moments (central moments) become then (set $\mu_i = \mu_j = \mu_k = \mu_l = 0$)

$$\begin{array}{rcl} \mathsf{E}\{\underline{e}_i\} & = & 0 \\ \mathsf{E}\{\underline{e}_i\underline{e}_j\} & = & q_{ij} \\ \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\} & = & 0 \\ \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\} & = & q_{ij}q_{kl} + q_{ik}q_{jl} + q_{il}q_{jk} \end{array} \tag{C.16}$$

The aforementioned formulation for the moment generating functions and the moments are usually given in the literatures. However, in the remainder of this appendix, as an extension, we will derive the moments of higher orders (six and eight) and consequently to evaluate the expectation of the multiplication of the quadratic forms, more than two terms, namely three and four terms.

C.4 Sixth and eighth central moments

As before, let $\underline{e}$ be normally distributed as $\underline{e} \sim \mathrm{N}_m(0, Q_y)$. The moment generating function of $\underline{e}$ is then

$$M_e(s) = \exp[\frac{1}{2} s^T Q_y s] \tag{C.17}$$

This shows that all central moments of a normally distributed vector $\underline{e}$ are fully determined by its dispersion (second central moment)

$$M_e(s) = \exp\left[\frac{1}{2}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{m} s_\alpha s_\beta q_{\alpha\beta}\right] = \exp[\phi] \tag{C.18}$$

where

$$\phi = \frac{1}{2}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{m} s_\alpha s_\beta q_{\alpha\beta} = \frac{1}{2} s_\alpha s_\beta q_{\alpha\beta} \tag{C.19}$$

One can simply show that the partial derivatives of ϕ read

$$\frac{\partial\phi}{\partial s_i} = \dot{\phi}_i = s_\alpha q_{\alpha i}, \quad \frac{\partial^2\phi}{\partial s_i \partial s_j} = q_{ij}, \quad \frac{\partial^3\phi}{\partial s_i \partial s_j \partial s_j} = ... = 0 \tag{C.20}$$

The partial derivatives of the moment generating function read then

First derivative:

$$\frac{\partial M_e(s)}{\partial s_i} = \frac{\partial\phi}{\partial s_i} M_e(s) = \dot{\phi}_i M_e(s) \tag{C.21}$$

Second derivative:

$$\frac{\partial^2 M_e(s)}{\partial s_i \partial s_j} = \frac{\partial^2\phi}{\partial s_i \partial s_j} M_e(s) + \frac{\partial\phi}{\partial s_i}\frac{\partial\phi}{\partial s_j} M_e(s) = q_{ij} M_e(s) + \dot{\phi}_i \dot{\phi}_j M_e(s) \tag{C.22}$$

Third derivative:

$$\frac{\partial^3 M_e(s)}{\partial s_i \partial s_j \partial s_k} = (q_{ij}\dot{\phi}_k + q_{ik}\dot{\phi}_j + q_{jk}\dot{\phi}_i) M_e(s) + \dot{\phi}_i \dot{\phi}_j \dot{\phi}_k M_e(s) \tag{C.23}$$

Fourth derivative:

$$\begin{aligned}\frac{\partial^4 M_e(s)}{\partial s_i \partial s_j \partial s_k \partial s_l} = {} & (q_{ij}q_{kl} + q_{ik}q_{jl} + q_{jk}q_{il}) M_e(s) + (q_{ij}\dot{\phi}_k\dot{\phi}_l + q_{ik}\dot{\phi}_j\dot{\phi}_l + q_{jk}\dot{\phi}_i\dot{\phi}_l + \\ & q_{il}\dot{\phi}_j\dot{\phi}_k + q_{jl}\dot{\phi}_i\dot{\phi}_k + q_{kl}\dot{\phi}_i\dot{\phi}_j) M_e(s) + \dot{\phi}_i\dot{\phi}_j\dot{\phi}_k\dot{\phi}_l M_e(s)\end{aligned} \tag{C.24}$$

Fifth derivative: here and later on do not confuse m, the free index, with the size of vector $\underline{y}$.

$$\begin{aligned}\frac{\partial^5 M_e(s)}{\partial s_i \partial s_j \partial s_k \partial s_l \partial s_m} = {} & \Big([q_{ij}q_{kl} + q_{ik}q_{jl} + q_{jk}q_{il}]\dot{\phi}_m + \\ & [q_{ij}q_{km} + q_{ik}q_{jm} + q_{jk}q_{im}]\dot{\phi}_l + [q_{ij}q_{lm} + q_{il}q_{jm} + q_{jl}q_{im}]\dot{\phi}_k + \\ & [q_{ik}q_{lm} + q_{il}q_{km} + q_{kl}q_{im}]\dot{\phi}_j + [q_{jk}q_{lm} + q_{jl}q_{km} + q_{kl}q_{jm}]\dot{\phi}_i\Big) M_e(s) \quad + \\ & \Big(q_{ij}\dot{\phi}_k\dot{\phi}_l\dot{\phi}_m + q_{ik}\dot{\phi}_j\dot{\phi}_l\dot{\phi}_m + q_{il}\dot{\phi}_j\dot{\phi}_k\dot{\phi}_m + q_{im}\dot{\phi}_j\dot{\phi}_k\dot{\phi}_l + \\ & q_{jk}\dot{\phi}_i\dot{\phi}_l\dot{\phi}_m + q_{jl}\dot{\phi}_i\dot{\phi}_k\dot{\phi}_m + q_{jm}\dot{\phi}_i\dot{\phi}_k\dot{\phi}_l + q_{kl}\dot{\phi}_i\dot{\phi}_j\dot{\phi}_m + \\ & q_{km}\dot{\phi}_i\dot{\phi}_j\dot{\phi}_l + q_{lm}\dot{\phi}_i\dot{\phi}_j\dot{\phi}_k\Big) M_e(s) \quad + \quad \dot{\phi}_i\dot{\phi}_j\dot{\phi}_k\dot{\phi}_l\dot{\phi}_m M_e(s)\end{aligned} \tag{C.25}$$

Sixth derivative:

$$
\begin{aligned}
&\tfrac{\partial^6 M_e(s)}{\partial s_i \partial s_j \partial s_k \partial s_l \partial s_m \partial s_n} = ([q_{ij}q_{kl} + q_{ik}q_{jl} + q_{jk}q_{il}]q_{mn} + \\
&[q_{ij}q_{km} + q_{ik}q_{jm} + q_{jk}q_{im}]q_{ln} + [q_{ij}q_{lm} + q_{il}q_{jm} + q_{jl}q_{im}]q_{kn} + \\
&[q_{ik}q_{lm} + q_{il}q_{km} + q_{kl}q_{im}]q_{jn} + [q_{jk}q_{lm} + q_{jl}q_{km} + q_{kl}q_{jm}]q_{in})\, M_e(s) + \\
&\Big([q_{ij}q_{kl} + q_{ik}q_{jl} + q_{jk}q_{il}]\dot\phi_m\dot\phi_n + [q_{ij}q_{km} + q_{ik}q_{jm} + q_{jk}q_{im}]\dot\phi_l\dot\phi_n + \\
&[q_{ij}q_{lm} + q_{il}q_{jm} + q_{jl}q_{im}]\dot\phi_k\dot\phi_n + [q_{ik}q_{lm} + q_{il}q_{km} + q_{kl}q_{im}]\dot\phi_j\dot\phi_n + \\
&[q_{jk}q_{lm} + q_{jl}q_{km} + q_{kl}q_{jm}]\dot\phi_i\dot\phi_n + [q_{ij}q_{kn} + q_{ik}q_{jn} + q_{jk}q_{in}]\dot\phi_l\dot\phi_m + \\
&[q_{ij}q_{ln} + q_{il}q_{jn} + q_{jl}q_{in}]\dot\phi_k\dot\phi_m + [q_{ik}q_{ln} + q_{kl}q_{in} + q_{il}q_{kn}]\dot\phi_j\dot\phi_m + \\
&[q_{jk}q_{ln} + q_{jl}q_{kn} + q_{kl}q_{jn}]\dot\phi_i\dot\phi_m + [q_{ij}q_{mn} + q_{im}q_{jn} + q_{jm}q_{in}]\dot\phi_k\dot\phi_l + \\
&[q_{ik}q_{mn} + q_{im}q_{kn} + q_{km}q_{in}]\dot\phi_j\dot\phi_l + [q_{jk}q_{mn} + q_{jm}q_{kn} + q_{km}q_{jn}]\dot\phi_i\dot\phi_l + \\
&[q_{il}q_{mn} + q_{im}q_{ln} + q_{lm}q_{in}]\dot\phi_j\dot\phi_k + [q_{jl}q_{mn} + q_{jm}q_{ln} + q_{lm}q_{jn}]\dot\phi_i\dot\phi_k + \\
&[q_{kl}q_{mn} + q_{km}q_{ln} + q_{lm}q_{kn}]\dot\phi_i\dot\phi_j\Big)\, M_e(s) + \\
&\Big(q_{ij}\dot\phi_k\dot\phi_l\dot\phi_m\dot\phi_n + q_{ik}\dot\phi_j\dot\phi_l\dot\phi_m\dot\phi_n + q_{il}\dot\phi_j\dot\phi_k\dot\phi_m\dot\phi_n + q_{im}\dot\phi_j\dot\phi_k\dot\phi_l\dot\phi_n + \\
&q_{jk}\dot\phi_i\dot\phi_l\dot\phi_m\dot\phi_n + q_{jl}\dot\phi_i\dot\phi_k\dot\phi_m\dot\phi_n + q_{jm}\dot\phi_i\dot\phi_k\dot\phi_l\dot\phi_n + q_{kl}\dot\phi_i\dot\phi_j\dot\phi_m\dot\phi_n + \\
&q_{km}\dot\phi_i\dot\phi_j\dot\phi_l\dot\phi_n + q_{lm}\dot\phi_i\dot\phi_j\dot\phi_k\dot\phi_n\Big)\, M_e(s) \;+\; \dot\phi_i\dot\phi_j\dot\phi_k\dot\phi_l\dot\phi_m\dot\phi_n M_e(s)
\end{aligned}
\tag{C.26}
$$

The seventh and eighth derivatives can be obtained in a similar way. For the sake of convenience, we will not give these formulas here. To obtain the central moments, we need to set s equal to zero in the preceding partial derivatives, i.e.

$$
\begin{aligned}
\mathsf{E}\{\underline{e}_i\} &= 0 \\
\mathsf{E}\{\underline{e}_i\underline{e}_j\} &= q_{ij} \\
\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\} &= 0 \\
\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\} &= q_{ij}q_{kl} + q_{ik}q_{jl} + q_{il}q_{jk} \\
\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_m\} &= 0 \\
\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_m\underline{e}_n\} &= [q_{ij}q_{kl}q_{mn} + q_{ik}q_{jl}q_{mn} + q_{jk}q_{il}q_{mn}] \\
&+ [q_{ij}q_{km}q_{ln} + q_{ik}q_{jm}q_{ln} + q_{jk}q_{im}q_{ln}] \\
&+ [q_{ij}q_{lm}q_{kn} + q_{il}q_{jm}q_{kn} + q_{jl}q_{im}q_{kn}] \\
&+ [q_{ik}q_{lm}q_{jn} + q_{il}q_{km}q_{jn} + q_{kl}q_{im}q_{jn}] \\
&+ [q_{jk}q_{lm}q_{in} + q_{jl}q_{km}q_{in} + q_{kl}q_{jm}q_{in}]
\end{aligned}
\tag{C.27}
$$

From the above results, the eighth central moment can be predicted. The number of terms for this central moment is $7\times5\times3\times1 = 105$ (24 independent terms, see later on). In terms of the sixth central moment, it reads

$$
\begin{aligned}
\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_m\underline{e}_n\underline{e}_o\underline{e}_p\} &= \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_m\underline{e}_n\}\, q_{op} + \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_m\underline{e}_o\}\, q_{np} \\
&+ \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_o\underline{e}_n\}\, q_{mp} + \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_o\underline{e}_m\underline{e}_n\}\, q_{lp} \\
&+ \mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_o\underline{e}_l\underline{e}_m\underline{e}_n\}\, q_{kp} + \mathsf{E}\{\underline{e}_i\underline{e}_o\underline{e}_k\underline{e}_l\underline{e}_m\underline{e}_n\}\, q_{jp} \\
&+ \mathsf{E}\{\underline{e}_o\underline{e}_j\underline{e}_k\underline{e}_l\underline{e}_m\underline{e}_n\}\, q_{ip}
\end{aligned}
\tag{C.28}
$$

C.5 Quadratic forms in normal variables

A *quadratic form* of a stochastic $m\times1$ vector $\underline{y}$ is defined as $\underline{y}^T A\underline{y}$ with A a *symmetric* $m\times m$ matrix. In the following, the expectation and the (co)variance of quadratic forms of a normally distributed $m\times1$ vector $\underline{y}\sim\mathsf{N}_m(\mu_y, Q_y)$ are derived.

Expectation If we use the trace operator and the moments of $\underline{y}$ as derived in the previous section, the expectation of a quadratic form can be derived as

$$
\boxed{\mathsf{E}\{\underline{y}^T A\underline{y}\} = \mathrm{tr}(A\mathsf{E}\{\underline{y}^T\underline{y}\}) = \mathrm{tr}(AQ_y) + \mu_y^T A\mu_y}
\tag{C.29}
$$

(Co)variance The covariance $\mathsf{C}\{.,.\}$ between two quadratic forms in $\underline{y}$ is defined as

$$\mathsf{C}\{\underline{y}^T A\underline{y}, \underline{y}^T B\underline{y}\} = \mathsf{E}\{(\underline{y}^T A\underline{y} - \mathsf{E}\{\underline{y}^T A\underline{y}\})(\underline{y}^T B\underline{y} - \mathsf{E}\{\underline{y}^T B\underline{y}\})\} \tag{C.30}$$

with A and B symmetric $m \times m$ matrices (C and D as well in the sequel). With $\underline{y} = \mu_y + \underline{e}$, one obtains

$$\begin{array}{rcl} \mathsf{C}\{\underline{y}^T A\underline{y}, \underline{y}^T B\underline{y}\} & = & \mathsf{E}\{\underline{e}^T A\underline{e}\underline{e}^T B\underline{e}\} + 2\mu_y^T B\mathsf{E}\{\underline{e}\underline{e}^T A\underline{e}\} + 2\mu_y^T A\mathsf{E}\{\underline{e}\underline{e}^T B\underline{e}\} \\ & + & 4\mu_y^T AQ_y B\mu_y - \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y) \end{array} \tag{C.31}$$

The m-vector $\mathsf{E}\{\underline{e}\underline{e}^T A\underline{e}\}$ reads

$$\mathsf{E}\{\underline{e}\underline{e}^T A\underline{e}\} = \mathsf{E}\{\underline{e}_k\underline{e}_i a_{ij}\underline{e}_j\} = a_{ij}\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\} = 0 \tag{C.32}$$

and so does $\mathsf{E}\{\underline{e}\underline{e}^T B\underline{e}\} = 0$. Furthermore

$$\begin{array}{rcl} \mathsf{E}\{\underline{e}^T A\underline{e}\underline{e}^T B\underline{e}\} & = & \mathsf{E}\{\underline{e}_i a_{ij}\underline{e}_j\underline{e}_k b_{kl}\underline{e}_l\} \;=\; a_{ij}b_{kl}\mathsf{E}\{\underline{e}_i\underline{e}_j\underline{e}_k\underline{e}_l\} \\ & = & a_{ij}b_{kl}q_{ij}q_{kl} + a_{ij}b_{kl}q_{ik}q_{jl} + a_{ij}b_{kl}q_{il}q_{jk} \\ & = & a_{ij}q_{ji}b_{kl}q_{lk} + a_{ij}q_{jl}b_{lk}q_{ki} + a_{ij}q_{jk}b_{kl}q_{li} \\ & = & \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y) + \mathrm{tr}(AQ_yBQ_y) + \mathrm{tr}(AQ_yBQ_y) \\ & = & \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y) + 2\mathrm{tr}(AQ_yBQ_y) \end{array} \tag{C.33}$$

So the covariance between two quadratic forms in $\underline{y}$ is given as:

$$\boxed{\mathsf{C}\{\underline{y}^T A\underline{y}, \underline{y}^T B\underline{y}\} = 2\mathrm{tr}(AQ_yBQ_y) + 4\mu_y^T AQ_yB\mu_y} \tag{C.34}$$

For special case $A = B$, i.e. for the variance of a quadratic form, one obtains

$$\boxed{\mathsf{D}\{\underline{y}^T A\underline{y}\} = 2\mathrm{tr}(AQ_yAQ_y) + 4\mu_y^T AQ_yA\mu_y} \tag{C.35}$$

In case $A = B = Q_y^{-1}$, it follows that

$$\mathrm{tr}(AQ_y) = m \text{ and } \mathrm{tr}(AQ_yBQ_y) = m \tag{C.36}$$

which with equation (C.33) gives

$$\boxed{\mathsf{E}\{(\underline{e}^T Q_y^{-1}\underline{e})^2\} = m(m+2)} \tag{C.37}$$

and thus from equation (C.35), it follows

$$\boxed{\mathsf{D}\{\underline{e}^T Q_y^{-1}\underline{e}\} = 2m} \tag{C.38}$$

It is repeated that $\underline{e} \sim N_m(0, Q_y)$ and is not to be confused with $\hat{e}$, the least-squares residuals.

C.6 Multiplication of four quadratic forms

Sometimes we need to know the expectation of the multiplication of four quadratic forms. This is the case if one wants to derive the variance of the quadratic form of the residuals in the stochastic model. For this purpose, the eighth moment ($2 \times 2 \times 2 = 8^{th}$) of the residual vector is needed (one 2 for stochastic model; one 2 for quadratic form; one 2 for variance). For the sake of convenience, let us firstly derive the expectation of the multiplication of three quadratic forms, and then generalize it to four.

$$\mathsf{E}\{\underline{e}^T A\underline{e}\,\underline{e}^T B\underline{e}\,\underline{e}^T C\underline{e}\} \;=\; \mathsf{E}\{\underline{e}_i a_{ij}\underline{e}_j\underline{e}_k b_{kl}\underline{e}_l\underline{e}_m c_{mn}\underline{e}_n\} \tag{C.39}$$

It can be rewritten as

$$\mathsf{E}\{\underline{e}^T A \underline{e}\, \underline{e}^T B \underline{e}\, \underline{e}^T C \underline{e}\} = a_{ij} b_{kl} c_{mn} \mathsf{E}\{\underline{e}_i \underline{e}_j \underline{e}_k \underline{e}_l \underline{e}_m \underline{e}_n\} \tag{C.40}$$

Substitution from equation (C.27) yields

$$\begin{aligned}\mathsf{E}\{\underline{e}^T A \underline{e}\, \underline{e}^T B \underline{e}\, \underline{e}^T C \underline{e}\} &= a_{ij} b_{kl} c_{mn} [q_{ij} q_{kl} q_{mn} + q_{ik} q_{jl} q_{mn} + q_{jk} q_{il} q_{mn}] \\ &+ a_{ij} b_{kl} c_{mn} [q_{ij} q_{km} q_{ln} + q_{ik} q_{jm} q_{ln} + q_{jk} q_{im} q_{ln}] \\ &+ a_{ij} b_{kl} c_{mn} [q_{ij} q_{lm} q_{kn} + q_{il} q_{jm} q_{kn} + q_{jl} q_{im} q_{kn}] \\ &+ a_{ij} b_{kl} c_{mn} [q_{ik} q_{lm} q_{jn} + q_{il} q_{km} q_{jn} + q_{kl} q_{im} q_{jn}] \\ &+ a_{ij} b_{kl} c_{mn} [q_{jk} q_{lm} q_{in} + q_{jl} q_{km} q_{in} + q_{kl} q_{jm} q_{in}]\end{aligned} \tag{C.41}$$

or, if we denote $\mathsf{E} = \mathsf{E}\{\underline{e}^T A \underline{e}\, \underline{e}^T B \underline{e}\, \underline{e}^T C \underline{e}\}$

$$\begin{aligned}\mathsf{E} &= a_{ij} b_{kl} c_{mn} q_{ij} q_{kl} q_{mn} + a_{ij} b_{kl} c_{mn} q_{ik} q_{jl} q_{mn} + a_{ij} b_{kl} c_{mn} q_{jk} q_{il} q_{mn} \\ &+ a_{ij} b_{kl} c_{mn} q_{ij} q_{km} q_{ln} + a_{ij} b_{kl} c_{mn} q_{ik} q_{jm} q_{ln} + a_{ij} b_{kl} c_{mn} q_{jk} q_{im} q_{ln} \\ &+ a_{ij} b_{kl} c_{mn} q_{ij} q_{lm} q_{kn} + a_{ij} b_{kl} c_{mn} q_{il} q_{jm} q_{kn} + a_{ij} b_{kl} c_{mn} q_{jl} q_{im} q_{kn} \\ &+ a_{ij} b_{kl} c_{mn} q_{ik} q_{lm} q_{jn} + a_{ij} b_{kl} c_{mn} q_{il} q_{km} q_{jn} + a_{ij} b_{kl} c_{mn} q_{kl} q_{im} q_{jn} \\ &+ a_{ij} b_{kl} c_{mn} q_{jk} q_{lm} q_{in} + a_{ij} b_{kl} c_{mn} q_{jl} q_{km} q_{in} + a_{ij} b_{kl} c_{mn} q_{kl} q_{jm} q_{in}\end{aligned} \tag{C.42}$$

Rearranging the preceding equation yields

$$\begin{aligned}\mathsf{E} &= a_{ij} q_{ji} b_{kl} q_{lk} c_{mn} q_{nm} + a_{ij} q_{jl} b_{lk} q_{ki} c_{mn} q_{nm} + a_{ij} q_{jk} b_{kl} q_{li} c_{mn} q_{nm} \\ &+ a_{ij} q_{ji} b_{kl} q_{ln} c_{nm} q_{mk} + a_{ij} q_{jm} c_{mn} q_{nl} b_{lk} q_{ki} + a_{ij} q_{jk} b_{kl} q_{ln} c_{nm} q_{mi} \\ &+ a_{ij} q_{ji} b_{kl} q_{lm} c_{mn} q_{nk} + a_{ij} q_{jm} c_{mn} q_{nk} b_{kl} q_{li} + a_{ij} q_{jl} b_{lk} q_{kn} c_{nm} q_{mi} \\ &+ a_{ij} q_{jn} c_{nm} q_{ml} b_{lk} q_{ki} + a_{ij} q_{jn} c_{nm} q_{mk} b_{kl} q_{li} + a_{ij} q_{jn} c_{nm} q_{mi} b_{kl} q_{lk} \\ &+ a_{ij} q_{jk} b_{kl} q_{lm} c_{mn} q_{ni} + a_{ij} q_{jl} b_{lk} q_{km} c_{mn} q_{ni} + a_{ij} q_{jm} c_{mn} q_{ni} b_{kl} q_{lk}\end{aligned} \tag{C.43}$$

or equivalently

$$\begin{aligned}\mathsf{E} &= \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y)\mathrm{tr}(CQ_y) + \mathrm{tr}(AQ_yBQ_y)\mathrm{tr}(CQ_y) + \mathrm{tr}(AQ_yBQ_y)\mathrm{tr}(CQ_y) \\ &+ \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_yCQ_y) + \mathrm{tr}(AQ_yCQ_yBQ_y) + \mathrm{tr}(AQ_yBQ_yCQ_y) \\ &+ \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_yCQ_y) + \mathrm{tr}(AQ_yCQ_yBQ_y) + \mathrm{tr}(AQ_yBQ_yCQ_y) \\ &+ \mathrm{tr}(AQ_yCQ_yBQ_y) + \mathrm{tr}(AQ_yCQ_yBQ_y) + \mathrm{tr}(AQ_yCQ_y)\mathrm{tr}(BQ_y) \\ &+ \mathrm{tr}(AQ_yBQ_yCQ_y) + \mathrm{tr}(AQ_yBQ_yCQ_y) + \mathrm{tr}(AQ_yCQ_y)\mathrm{tr}(BQ_y)\end{aligned} \tag{C.44}$$

or finally

$$\begin{aligned}\mathsf{E}\{\underline{e}^T A \underline{e}\, \underline{e}^T B \underline{e}\, \underline{e}^T C \underline{e}\} &= \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y)\mathrm{tr}(CQ_y) + 2\mathrm{tr}(AQ_y)\mathrm{tr}(BQ_yCQ_y) \\ &+ 2\mathrm{tr}(BQ_y)\mathrm{tr}(AQ_yCQ_y) + 2\mathrm{tr}(CQ_y)\mathrm{tr}(AQ_yBQ_y) \\ &+ 4\mathrm{tr}(AQ_yBQ_yCQ_y) + 4\mathrm{tr}(AQ_yCQ_yBQ_y)\end{aligned} \tag{C.45}$$

In case Q_y is of full rank, and $A = B = C = Q_y^{-1}$ one obtains

$$\mathrm{tr}(AQ_y) = m, \quad \mathrm{tr}(AQ_yBQ_y) = m \quad \text{and} \quad \mathrm{tr}(AQ_yBQ_yCQ_y) = m \tag{C.46}$$

This with equation (C.45) gives

$$\boxed{\mathsf{E}\{(\underline{e}^T Q_y^{-1} \underline{e})^3\} = m(m+2)(m+4)} \tag{C.47}$$

To derive the expectation of the multiplication of four quadratic forms, one needs to compute the following:

$$\mathsf{E}\{\underline{e}^T A \underline{e}\, \underline{e}^T B \underline{e}\, \underline{e}^T C \underline{e}\, \underline{e}^T D \underline{e}\} = \mathsf{E}\{\underline{e}_i a_{ij} \underline{e}_j \underline{e}_k b_{kl} \underline{e}_l \underline{e}_m c_{mn} \underline{e}_n \underline{e}_o d_{op} \underline{e}_p\} \tag{C.48}$$

It can be rewritten as

$$\mathsf{E}\{\underline{e}^T A \underline{e}\, \underline{e}^T B \underline{e}\, \underline{e}^T C \underline{e}\, \underline{e}^T D \underline{e}\} = a_{ij} b_{kl} c_{mn} d_{op} \mathsf{E}\{\underline{e}_i \underline{e}_j \underline{e}_k \underline{e}_l \underline{e}_m \underline{e}_n \underline{e}_o \underline{e}_p\} \tag{C.49}$$

In a similar, but very very long, way to what we did for the multiplication of three quadratic forms, one obtains

$$
\begin{aligned}
\mathsf{E}\{\underline{e}^T A\underline{e}\,\underline{e}^T B\underline{e}\,\underline{e}^T C\underline{e}\,\underline{e}^T D\underline{e}\} &= \mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y)\mathrm{tr}(CQ_y)\mathrm{tr}(DQ_y) && (F1)\\
&+ 2\mathrm{tr}(AQ_yBQ_y)\mathrm{tr}(CQ_y)\mathrm{tr}(DQ_y) && (F2)\\
&+ 2\mathrm{tr}(AQ_yCQ_y)\mathrm{tr}(BQ_y)\mathrm{tr}(DQ_y) && (F3)\\
&+ 2\mathrm{tr}(AQ_yDQ_y)\mathrm{tr}(BQ_y)\mathrm{tr}(CQ_y) && (F4)\\
&+ 2\mathrm{tr}(BQ_yCQ_y)\mathrm{tr}(AQ_y)\mathrm{tr}(DQ_y) && (F5)\\
&+ 2\mathrm{tr}(BQ_yDQ_y)\mathrm{tr}(AQ_y)\mathrm{tr}(CQ_y) && (F6)\\
&+ 2\mathrm{tr}(CQ_yDQ_y)\mathrm{tr}(AQ_y)\mathrm{tr}(BQ_y) && (F7)\\
&+ 4\mathrm{tr}(DQ_y)\mathrm{tr}(AQ_yBQ_yCQ_y) && (F8)\\
&+ 4\mathrm{tr}(DQ_y)\mathrm{tr}(AQ_yCQ_yBQ_y) && (F9)\\
&+ 4\mathrm{tr}(CQ_y)\mathrm{tr}(AQ_yBQ_yDQ_y) && (F10)\\
&+ 4\mathrm{tr}(CQ_y)\mathrm{tr}(AQ_yDQ_yBQ_y) && (F11)\\
&+ 4\mathrm{tr}(BQ_y)\mathrm{tr}(AQ_yCQ_yDQ_y) && (F12)\\
&+ 4\mathrm{tr}(BQ_y)\mathrm{tr}(AQ_yDQ_yCQ_y) && (F13)\\
&+ 4\mathrm{tr}(AQ_y)\mathrm{tr}(BQ_yCQ_yDQ_y) && (F14)\\
&+ 4\mathrm{tr}(AQ_y)\mathrm{tr}(BQ_yDQ_yCQ_y) && (F15)\\
&+ 4\mathrm{tr}(AQ_yBQ_y)\mathrm{tr}(CQ_yDQ_y) && (F16)\\
&+ 4\mathrm{tr}(AQ_yCQ_y)\mathrm{tr}(BQ_yDQ_y) && (F17)\\
&+ 4\mathrm{tr}(AQ_yDQ_y)\mathrm{tr}(BQ_yCQ_y) && (F18)\\
&+ 8\mathrm{tr}(AQ_yBQ_yCQ_yDQ_y) && (F19)\\
&+ 8\mathrm{tr}(AQ_yBQ_yDQ_yCQ_y) && (F20)\\
&+ 8\mathrm{tr}(AQ_yCQ_yBQ_yDQ_y) && (F21)\\
&+ 8\mathrm{tr}(AQ_yCQ_yDQ_yBQ_y) && (F22)\\
&+ 8\mathrm{tr}(AQ_yDQ_yBQ_yCQ_y) && (F23)\\
&+ 8\mathrm{tr}(AQ_yDQ_yCQ_yBQ_y) && (F24)
\end{aligned}
\tag{C.50}
$$

consisting of 24 independent terms. In case $A = B = C = D = Q_y^{-1}$, the preceding equation simplifies to

$$
\boxed{\mathsf{E}\{(\underline{e}^T Q_y^{-1}\underline{e})^4\} = m(m+2)(m+4)(m+6)} \tag{C.51}
$$

Mixed model with hard constraints

D

D.1 Model representation $\mathsf{E}\{\underline{y}\} = Ax$ with $B^T x = c$

Following chapter 5 of *Adjustment Theory*, our starting point is the representation (Teunissen, 2000a)

$$\mathsf{E}\{\underline{y}\} = Ax;\ B^T x = c;\quad D\{\underline{y}\} = Q_y \tag{D.1}$$

with c a constant q-vector and the $n \times q$ (constraints) matrix B. This representation is in the form of observation equations with hard constraints on the parameter vector x. In order to derive the least squares estimators $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$, we will first transform equation (D.1) into a form with observation equations only. This is done by finding the parametric representation for:

$$B^T x = c \tag{D.2}$$

It is well known that its solution is given as the sum of a particular solution and the homogeneous solution. A particular solution of the above is given as

$$x_p = B(B^T B)^{-1} c \tag{D.3}$$

In order to find the solution of the homogeneous equation

$$B^T x = 0, \tag{D.4}$$

we denote the $n \times (n-q)$ matrix of which the column vectors are orthogonal to B by $B^\perp$. Then:

$$B^T B^\perp = 0 \tag{D.5}$$

With this, the parametric representation of the homogeneous equation $B^T x = 0$ becomes:

$$x_h = B^\perp \lambda \tag{D.6}$$

with λ an $(n-q)$-vector. The general solution of the inhomogeneous equation, is therefore given by the sum of the homogeneous and particular solutions, i.e.

$$x = B^\perp \lambda + B(B^T B)^{-1} c \tag{D.7}$$

Substituting the preceding equation into equation (D.1), yields

$$\mathsf{E}\{\underline{y}\} = AB^\perp \lambda + AB(B^T B)^{-1} c;\quad D\{\underline{y}\} = Q_y \tag{D.8}$$

or equivalently

$$\mathsf{E}\{\underline{y} - AB(B^T B)^{-1} c\} = AB^\perp \lambda;\quad D\{\underline{y}\} = Q_y \tag{D.9}$$

This representation is completely equivalent to the original equation (D.1), and it in the form of *observation equation only*. The least squares estimator $\underline{\hat{\lambda}}$ of λ reads

$$\underline{\hat{\lambda}} = (B^{\perp T} A^T Q_y^{-1} A B^\perp)^{-1} B^{\perp T} A^T Q_y^{-1} \left(\underline{y} - AB(B^T B)^{-1} c\right) \tag{D.10}$$

Now let us consider the solution of (D.1) *without* the constraints on the parameter vector assuming that the design matrix A is of full rank. This is our standard model of observation equations. As we know, the solution of equation (D.1) without the constraints on the parameter vector satisfies the normal equations:

$$A^T Q_y^{-1} A \underline{\hat{x}}_A = A^T Q_y^{-1} \underline{y} \tag{D.11}$$

We have given $\underline{\hat{x}}_A$ of equation (D.11) the subscript A to emphasize that $\underline{\hat{x}}_A$ is not the solution of equation (D.1) but of

$$\mathsf{E}\{\underline{y}\} = Ax; \quad D\{\underline{y}\} = Q_y \tag{D.12}$$

The covariance matrix of $\underline{\hat{x}}_A$ reads

$$Q_{\hat{x}_A} = (A^T Q_y^{-1} A)^{-1} \tag{D.13}$$

With equation (D.11) and equation (D.13), equation (D.10) reads

$$\underline{\hat{\lambda}} = (B^{\perp T} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp T} Q_{\hat{x}_A}^{-1} \hat{x}_A - (B^{\perp T} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp T} Q_{\hat{x}_A}^{-1} B (B^T B)^{-1} c \tag{D.14}$$

This result together with equation (D.7) gives the estimator of x for model in equation (D.1) as

$$\underline{\hat{x}} = B^{\perp} (B^{\perp T} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp T} Q_{\hat{x}_A}^{-1} \hat{x}_A + \left[I - B^{\perp} (B^{\perp T} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp T} Q_{\hat{x}_A}^{-1} \right] B (B^T B)^{-1} c \tag{D.15}$$

The projector $B^{\perp} (B^{\perp T} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp T} Q_{\hat{x}_A}^{-1}$ can be rewritten in terms of the matrix B as

$$B^{\perp} (B^{\perp T} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp T} Q_{\hat{x}_A}^{-1} = I - Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T \tag{D.16}$$

Substitution of (D.16) into (D.15) gives

$$\boxed{\underline{\hat{x}} = \left[I - Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T \right] \underline{\hat{x}}_A + Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} c} \tag{D.17}$$

with the covariance matrix

$$\boxed{Q_{\hat{x}} = \left[I - Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T \right] Q_{\hat{x}_A}} \tag{D.18}$$

Recalling from *adjustment theory*, one gets $\underline{\hat{y}} = A \underline{\hat{x}}$, or

$$\underline{\hat{y}} = A \left[I - Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T \right] \underline{\hat{x}}_A + A Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} c \tag{D.19}$$

with the covariance matrix

$$Q_{\hat{y}} = A Q_{\hat{x}_A} A^T - A Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T Q_{\hat{x}_A} A^T \tag{D.20}$$

and finally the residuals $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$, or as

$$\underline{\hat{e}} = \underline{y} - A \left[I - Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T \right] \hat{x}_A - A Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} c \tag{D.21}$$

with the covariance matrix

$$Q_{\hat{e}} = Q_y - A Q_{\hat{x}_A} A^T + A Q_{\hat{x}_A} B (B^T Q_{\hat{x}_A} B)^{-1} B^T Q_{\hat{x}_A} A^T \tag{D.22}$$

D.2 Parameter significance test (v-test)

In this section we will consider a testing problem that, though mathematically equivalent to the well-known w-test statistic, occurs when we want to test the *significance of the parameters*. The idea is to test whether or not it is possible to reduce the number of unknowns (one at a time), e.g. by introducing one hard constraint on the parameters. We will derive the appropriate simple likelihood ratio test of size α and the corresponding test statistic. The following two hypotheses are considered (cf. Teunissen (2000b)):

$$\mathrm{H}_o : \mathsf{E}\{\underline{y}\} = Ax;\ d^T x = c_0 \quad \text{versus} \quad \mathrm{H}_a : \mathsf{E}\{\underline{y}\} = Ax;\ d^T x = c_A \neq c_0 \tag{D.23}$$

with d in the bold face to indicate that it is different from scalar b, the redundancy of the functional model. The preceding two hypotheses differ in the sense that under H_o it is assumed that the linear function of x, $d^T x$, is identical to c_0, whereas under H_a, this function is identical to the unknown parameter $c_A \neq c_0$. Thus what we want to find out is whether $d^T x = c_0$ or not. Note that the model $\mathsf{E}\{\underline{y}\} = Ax$ with the constraint $d^T x = c$ is of the form of the mixed model discussed in the previous section. In order to be able to apply the theory of the w-test statistic, we will write the general solution of the inhomogeneous equation $d^T x = c$ as (see (D.7) in previous section)

$$x = B^\perp \lambda + d(d^T d)^{-1} c \tag{D.24}$$

with $d^T B^\perp = 0$. Applying the above parametric representation to the null and the alternative hypotheses gives

$$\mathrm{H}_o : \mathsf{E}\{\underline{y}\} = AB^\perp \lambda + Ad(d^T d)^{-1} c_0 \quad \text{versus} \quad \mathrm{H}_a : \mathsf{E}\{\underline{y}\} = AB^\perp \lambda + Ad(d^T d)^{-1} c_A,\ c_A \neq c_0 \tag{D.25}$$

or

$$\mathrm{H}_o : \mathsf{E}\{\underline{y} - Ad(d^T d)^{-1} c_0\} = \overbrace{AB^\perp}^{A} \lambda \tag{D.26}$$

versus

$$\mathrm{H}_a : \mathsf{E}\{\underline{y} - Ad(d^T d)^{-1} c_0\} = AB^\perp \lambda + \overbrace{Ad(d^T d)^{-1}}^{c_y} \nabla,\ \nabla \neq 0 \tag{D.27}$$

with $\nabla = c_A - c_0$. Comparison of the preceding hypotheses with those of the w-test statistic shows the equivalent structure. That is, the matrix $AB^\perp$ plays the role of A in the hypotheses for the w-test statistic, and the vector $Ad(d^T d)^{-1}$ plays the role of the vector c_y. Because of this equivalence in structure of the hypotheses, the simple likelihood ratio test for the present testing problem have the same structure as the w-test statistic. The corresponding test statistic, denoted as $\underline{v}$, follows then if we replace c_y with $Ad(d^T d)^{-1}$:

$$\underline{\hat{v}} = \frac{(d^T d)^{-1} d^T A^T Q_y^{-1} \underline{\hat{e}}}{\sqrt{(d^T d)^{-1} d^T A^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} Ad(d^T d)^{-1}}} = \frac{d^T A^T Q_y^{-1} \underline{\hat{e}}}{\sqrt{d^T A^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} Ad}} \tag{D.28}$$

with $\underline{\hat{e}}$ and $Q_{\hat{e}}$ as those given in the previous section. It is not difficult to show that

$$A^T Q_y^{-1} \underline{\hat{e}} = d(d^T Q_{\hat{x}_A} d)^{-1} [d^T \underline{\hat{x}}_A - c_0] \tag{D.29}$$

and

$$A^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} A = d(d^T Q_{\hat{x}_A} d)^{-1} d^T \tag{D.30}$$

which with (D.28) give

$$\boxed{\underline{v} = \frac{d^T \underline{\hat{x}}_A - c_0}{\sqrt{d^T Q_{\hat{x}_A} d}}} \tag{D.31}$$

The corresponding simple likelihood ratio test of size α for testing problem (D.23) reads therefore:

$$\text{reject } \mathrm{H}_o \text{ if } |v| > k_{\alpha/2} \tag{D.32}$$

with $k_{\alpha/2}$ the critical value of the test obtained from the standard normal distribution.

D.3 Derivation of w-test from v-test statistic

Let us consider the following two hypotheses

$$\mathrm{H}_o : \mathsf{E}\{\underline{y}\} = Ax, \quad \text{versus} \quad \mathrm{H}_a : \mathsf{E}\{\underline{y}\} = Ax + c_y\nabla, \quad \text{and} \quad \nabla \neq 0 \tag{D.33}$$

In order to obtain the well-known w-test statistic from the v-test presented in the previous section, we may reformulate equation (D.33) as

$$\mathrm{H}_o : \mathsf{E}\{\underline{y}\} = [A\, c_y] \begin{bmatrix} x \\ \nabla \end{bmatrix}, \ \nabla = 0 \ \text{ versus } \ \mathrm{H}_a : \mathsf{E}\{\underline{y}\} = [A\, c_y] \begin{bmatrix} x \\ \nabla \end{bmatrix}, \ \nabla \neq 0 \tag{D.34}$$

This formulation is identical to equation (D.23) with

$$b = [0, ..., 0, 1]^T, \quad c_0 = 0 \tag{D.35}$$

The least squares solution of the model $\mathsf{E}\{\underline{y}\} = [A\, c_y] \begin{bmatrix} x \\ \nabla \end{bmatrix}$ read

$$\begin{bmatrix} \hat{\underline{x}} \\ \hat{\underline{\nabla}} \end{bmatrix} = \begin{bmatrix} A^T Q_y^{-1} A & A^T Q_y^{-1} c_y \\ c_y^T Q_y^{-1} A & c_y^T Q_y^{-1} c_y \end{bmatrix}^{-1} \begin{bmatrix} A^T Q_y^{-1} \underline{y} \\ c_y^T Q_y^{-1} \underline{y} \end{bmatrix} \tag{D.36}$$

The inverse of the normal matrix reads then

$$\begin{bmatrix} A^T Q_y^{-1} A & A^T Q_y^{-1} c_y \\ c_y^T Q_y^{-1} A & c_y^T Q_y^{-1} c_y \end{bmatrix}^{-1} = \begin{bmatrix} N_{11}^{-1} & N_{12}^{-1} \\ N_{21}^{-1} & N_{22}^{-1} \end{bmatrix} \tag{D.37}$$

with

$$\begin{array}{rcl} N_{11}^{-1} & = & (A^T Q_y^{-1} A)^{-1} - \sigma^2_{\hat{\nabla}} (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} c_y c_y^T Q_y^{-1} A (A^T Q_y^{-1} A)^{-1} \\ N_{12}^{-1} & = & -\sigma^2_{\hat{\nabla}} . (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} c_y \\ N_{21}^{-1} & = & -\sigma^2_{\hat{\nabla}} . c_y^T Q_y^{-1} A (A^T Q_y^{-1} A)^{-1} \\ N_{22}^{-1} & = & \sigma^2_{\hat{\nabla}} \end{array} \tag{D.38}$$

with

$$\begin{array}{rcl} \sigma^2_{\hat{\nabla}} & = & (c_y Q_y^{-1} c_y - c_y Q_y^{-1} A (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} c_y)^{-1} \\ & = & (c_y Q_y^{-1} (I - A (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}) c_y)^{-1} \\ & = & (c_y Q_y^{-1} P_A^{\perp} c_y)^{-1} \\ & = & \left\| P_A^{\perp} c_y \right\|^{-2}_{Q_y^{-1}} \end{array} \tag{D.39}$$

The estimator $\hat{\underline{\nabla}}$ can be obtained as

$$\begin{array}{rcl} \hat{\underline{\nabla}} & = & N_{21}^{-1} A^T Q_y^{-1} \underline{y} + N_{22}^{-1} c_y^T Q_y^{-1} \underline{y} \\ & = & \sigma^2_{\hat{\nabla}} c_y^T Q_y^{-1} \underline{y} - \sigma^2_{\hat{\nabla}} . c_y^T Q_y^{-1} A (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \\ & = & \sigma^2_{\hat{\nabla}} c_y^T Q_y^{-1} \left(I - A (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \right) \underline{y} \\ & = & \sigma^2_{\hat{\nabla}} c_y^T Q_y^{-1} P_A^{\perp} \underline{y} \end{array} \tag{D.40}$$

The v-test statistic (D.31), denoted here as w-test statistic, with equation (D.35), can be written as

$$\underline{w} = \frac{\hat{\nabla}}{\sigma_{\hat{\nabla}}} \tag{D.41}$$

which with equations (D.39) and (D.40) proves the claim

$$\boxed{\underline{w} = \frac{c_y^T Q_y^{-1} P_A^{\perp} \underline{y}}{\left\| P_A^{\perp} c_y \right\|_{Q_y^{-1}}}} \tag{D.42}$$

Bibliography

Abbasi M (1999). Comparison of Fourier, least-squares and wavelet spectral analysis methods, tested on Persian Gulf tidal data. MSc thesis, Surveying Engineering Department, K.N. Toosi University of Technology, Tehran, Iran.

Agnew D (1992). The time domain behavior of power law noises. *Geophysical Research Letters*, 19: 333–336.

Agnew DC, Larson KM (2007). Finding the repeat times of the GPS constellation. *GPS Solutions*, 11(1): 71–76.

Altamimi Z, Sillard P, Boucher C (2002). ITRF2000: A new release of the International Terrestrial Reference Frame for earth science applications. *Journal of Geophysical Research*, 107(B10, 2214).

Amiri-Simkooei AR (2003). Formulation of L_1 norm minimization in Gauss-Markov models. *Journal of Surveying Engineering*, 129(1): 37–43.

Amiri-Simkooei AR (2005). Separating receiver noise and multipath effects in time series of GPS baselines using harmonic functions. In *ION GNSS 18th International Technical Meeting of the Satellite Division*, September 13-16 2005, Long Beach, CA, pp. 1034–1045. (sponsored student paper).

Amiri-Simkooei AR, Tiberius CCJM (2007). Assessing receiver noise using GPS short baseline time series. *GPS Solutions*, 11(1): 21–35.

Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2006). Assessment of noise in GPS coordinate time series: methodology and results. *Journal of Geophysical Research*, submitted.

Asgari J, Harmel A (2005). Least squares spectral analysis of GPS derived ionospheric data. In *The ION National Technical Meeting*, Jan. 24-26 San Diego, California.

Baarda W (1968). A testing procedure for use in geodetic networks. Technical report, Netherlands Geodetic Commission, Publ. on Geodesy, New Series, Vol. 2(5), Delft.

Baarda W (1973). S-transformations and criterion matrices. Technical report, Netherlands Geodetic Commission, Publ. on Geodesy, New Series, Vol. 5(1), Delft.

Barbarella M, Pieri L (1983). I pesi nella compensazione di reti topografiche. *Boll Geod Sci Affini*, XLII(3): 317–355.

Barnes JB (2002). *Real time kinematic GPS and multipath : characterisation and improved least squares modelling*. Ph. D. thesis, Department of Geomatics, University of Newcastle upon Tyne.

Barnes JB, Ackroyd N, Cross PA (1998). Stochastic modelling for very high precision real-time kinematic GPS in an engineering environment. In *Proceedings of XXI International Congress of the International Federation of Surveyors (FIG)*, Brighton, pp. 61–76.

Barnes JB, Cross PA (1998). Processing models for very high accuracy GPS positioning. *Journal of Navigation*, 51(2): 180–193.

Beavan J (2005). Noise properties of continuous GPS data from concrete pillar geodetic monuments in New Zealand and comparison with data from U.S. deep drilled braced monuments. *Journal of Geophysical Research*, 110(B08410).

Berkane M, Bentler PM (1986). Moments of elliptically distributed random variates. *Statistics & Probability Letters*, 4: 333–335.

Beutler G (2006). Geodetic monitoring of the Earth and the role of geodetic theory in the age of space geodesy. In *VI Hotine-Marussi Symposium of Theoretical and Computational Geodesy: Challenge and Role of Modern Geodesy*, 29 May - 2 June, Wuhan University, China.

Beutler G, Rothacher M, Schaer S, Springer TA, Kouba J, Neilan RE (1999). The International GPS Service (IGS): an interdisciplinary service in support of Earth sciences. *Advances in Space Research*, 23(4): 631–635.

Bischoff W, Heck B, Howind J, Teusch A (2005). A procedure for testing the assumption of homoscedasticity in least squares residuals: a case study of GPS carrier-phase observations. *Journal of Geodesy*, 78: 397–404.

Bischoff W, Heck B, Howind J, Teusch A (2006). A procedure for estimating the variance function of linear models and for checking the appropriateness of estimated variances: a case study of GPS carrier-phase observations. *Journal of Geodesy*, 79: 694–704.

Blewitt G, Lavallée D (2002). Effect of annual signals on geodetic velocity. *Journal of Geophysical Research*, 107(B7).

Bock Y, Nikolaidis RM, de Jonge PJ, Bevis M (2000). Instantaneous geodetic positioning at medium distances with the Global Positioning System. *Journal of Geophysical Research*, 105(B12): 28223–28253.

Bock Y, S.Wdowinski, Fang P, Zhang J, Williams S, Johnson H, Behr J, Genrich J, Dean J, van Domselaar M, Agnew D, Wyatt F, Stark K, Oral B, Hudnut K, King R, Herring T, Dinardo S, Young W, Jackson D, Gurtner W (1997). Southern California Permanent GPS Geodetic Array: Continuous measurements of regional crustal deformation between the 1992 Landers and 1994 Northridge earthquakes. *Journal of Geophysical Research*, 102(B8): 18013–18033.

Bona P (2000). Precision, cross correlation, and time correlation of GPS phase and code observations. *GPS Solutions*, 4(2): 3–13.

Bona P, Tiberius CCJM (2000). An experimental comparison of noise characteristics of seven high-end dual frequency GPS receiver-sets. In *Proc., IEEE PLANS2000*, pp. 237–244.

Calais E (1999). Continuous GPS measurements across the Western Alps, 1996–1998. *Geophysical Journal International*, 138: 221–230.

Caspary WF (1987). Concepts of network and deformation analysis. Technical report, School of Surveying, The University of New South Wales, Kensington.

Chen YQ, Chrzanowski A, Kavouras M (1990). Assessment of observations using minimum norm quadratic unbiased estimation (MINQUE). *CISM Journal of ACSGS*, 44: 39–46.

Craymer MR (1998). *The least-squares spectrum, its inverse transform and autocorrelation function: Theory and some applications in geodesy*. Ph. D. thesis, University of Toronto, Department of Civil Engineering, Toronto, Canada.

Crocetto N, Gatti M, Russo P (2000). Simplified formulae for the BIQUE estimation of variance components in disjunctive observation groups. *Journal of Geodesy*, 74: 447–457.

Davies P, Blewitt G (2000). Methodology for global geodetic time series estimation: A new tool for geodynamics. *Journal of Geophysical Research*, 105(B5): 11083–11100.

Ding XL, Zheng DW, Dong DN, Ma C, Chen YQ, Wang GL (2005). Seasonal and secular positional variations at eight co-located GPS and VLBI stations. *Journal of Geodesy*, 79(1–3): 71–81.

Dong D, Fang P, Bock Y, Cheng MK, Miyazaki S (2002). Anatomy of apparent seasonal variations from GPS-derived site position time series. *Journal of Geophysical Research*, 107.

Efron B, Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Euler HJ, Goad C (1991). On optimal filtering of GPS dual frequency observations without using orbit information. *Bull Géod*, 65: 130–143.

Förstner W (1979). Ein verfahren zur schätzung von varianz und kovarianzkomponenten. *Allgemeine Vermessungs Nachrichten*: 446–453.

Fotopoulos G (2003). *An analysis on the optimal combination of geoid, orthometric and ellipsoidal height data*. Ph. D. thesis, University of Calgary, Department of Geomatics Engineering, Calgary, Alberta.

Fotopoulos G (2005). Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetric geoid height data. *Journal of Geodesy*, 79: 111–123.

Gerdan GP (1995). A comparison of four methods of weighting double difference pseudo range measurements. *Trans Tasman Surv*, 1(1): 60–66.

Gianniou M (1996). *Genauigkeitssteigerung bei kurzzeit-statischen und kinematischen Satellitenmessungen bis hin zu Echtzeitanwendung*. Ph. D. thesis, Deutsche Geodaetische Kommission, Reihe C, No. 458, Muenchen.

Grafarend EW (1984). Variance-covariance component estimation of Helmert type in the Gauss-helmert model. *ZFV*, 109: 34–44.

Grafarend EW (1985). Variance-covariance component estimation: theoretical results and geodetic applications. *Statistics & Decisions*, Supplement Issue 2: 407–441.

Grodecki J (1999). Generalized maximum-likelihood estimation of variance components with inverted gamma prior. *Journal of Geodesy*, 73: 367–374.

Grodecki J (2001). Generalized maximum-likelihood estimation of variance-covariance components with non-informative prior. *Journal of Geodesy*, 75: 157–163.

Harville DA (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72: 320–340.

Haskell KH, Hanson RJ (1981). An algorithm for linear least squares problems with equality and nonnegativity constraints. *Mathematical Programming*, 21: 98–118.

Helmert FR (1907). Die ausgleichungsrechnung nach der methode der kleinsten quadrate. 2. Auflage, Verlag von BG Teubner, Leipzig/Berlin.

Helmert FR (1924). Die ausgleichungsrechnung nach der methode der kleinsten quadrate. 3. AUFL., Leipzig/Berlin.

Horn SD, Horn RA, Duncan DB (1975). Estimating heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 70: 380–385.

Hsu R (1999). An alternative expression for the variance factors in using iterated almost unbiased estimation. *Journal of Geodesy*, 73: 173–179.

Hsu R (2001). Helmert method as equivalent of iterated almost unbiased estimation. *Journal of Surveying Engineering*, 127(3): 79–89.

Imhof JP (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3 and 4): 419–426.

Jin XX, de Jong CD (1996). Relationship between satellite elevation and precision of GPS code observations. *J Navig*, 49: 253–265.

Johnson HO, Agnew DC (2000). Correlated noise in geodetic time series. U.S. Geol. Surv. Final Tech. Rep., FTR-1434-HQ-97-GR-03155.

Johnson HO, Wyatt FK (1994). Geodetic network design for fault-mechanics studies. *Manuscripta Geodaetica*, 19: 309–323.

Jonkman NF (1998). Integer GPS ambiguity estimation without the receiver-satellite geometry. LGR-Series 18, Geodetic Computing Centre, Delft.

Kenyeres A, Bruyninx C (2004). EPN coordinate time series monitoring for reference frame maintenance. *GPS Solutions*, 8: 200–209.

Kim D, Langley RB (2001a). Estimation of the stochastic model for long-baseline kinematic GPS applications. In *Proceedings of The Institute of Navigation 2001 National Technical Meeting*, Long Beach, CA, USA, 22-24 January 2001, pp. 586–595.

Kim D, Langley RB (2001b). Quality control techniques and issues in GPS applications: stochastic modelling and reliability testing. In *Proceedings of the 8th GNSS Workshop - 2001 International Symposium on GPS/GNSS*, Jeju Island, Korea, 7-9 November, Tutorial/Domestic Session, pp. 76–85.

Kleijer F, Elósegui P, Davis JL (2004). Characterizing atmospheric turbulence with GPS. In *16th Symposium on Boundary Layers and Turbulence*, pp. 297–303.

Koch KR (1978). Schätzung von varianzkomponenten. *Allgemeine Vermessungs Nachrichten*, 85: 264–269.

Koch KR (1986). Maximum likelihood estimate of variance components. *Bull. Géod.*, 60: 329–338. Ideas by A.J. Pope.

Koch KR (1987). Bayesian inference for variance components. *Manuscripta Geodaetica*, 12: 309–313.

Koch KR (1990). *Bayesian inference with geodetic applications*. Springer, Berlin, Heidelberg, New York.

Koch KR (1999). *Parameter estimation and hypothesis testing in linear models*. Springer Verlag, Berlin.

Koch KR, Kusche J (2002). Regularization of geopotential determination from satellite data by variance components. *Journal of Geodesy*, 76: 259–268.

Kubik K (1970). The estimation of the weights of measured quantities within the method of least squares. *Bull. Géod.*, 95: 21–40.

Kusche J (2003a). A Monte-Carlo technique for weight estimation in satellite geodesy. *Journal of Geodesy*, 76: 641–652.

Kusche J (2003b). Noise variance estimation and optimal weight determination for GOCE gravity recovery. *Advances in Geosciences*, 1: 81–85.

Kusche J, Schrama JO (2005). Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data. *Journal of Geophysical Research*, 110.

Kusche J, van Loon JP (2005). Statistical assessment of CHAMP data and models using the energy balance approach. In C. Reigber, H. Lühr, P. Schwintzer, and J. Wickert (Eds.), *Earth observation with CHAMP: Results from three years in orbit*, Springer, Berlin, pp. 133–138.

Langbein J (2004). Noise in two-color electronic distance meter measurements revisited. *Journal of Geophysical Research*, 109.

Langbein J, Bock Y (2004). High-rate real-time GPS network at Parkfield: Utility for detecting fault slip and seismic displacements. *Geophysical Research Letters*, 31.

Langbein J, Johnson H (1995). Noise level of geodetic monuments. *Eos Trans. AGU*, 76(46). Fall Meet. Suppl., F142.

Langbein J, Johnson H (1997). Correlated errors in geodetic time series: Implications for time-dependent deformation. *Journal of Geophysical Research*, 102(B1): 591–603.

Langley RB (1997). GPS observation noise. *GPS World*, 8(6): 40–45.

Langville AN, Stewart WJ (2004). The Kronecker product and stochastic automata networks. *Journal of Computational and Applied Mathematics*, 167: 429–447.

Lawson CL, Hanson RJ (1974). *Solving least-squares problems*. Englewood Cliffs, N. J.: Prentice-Hall, Inc. series in automatic computation.

Lucas JR, Dillinger WH (1998). MINQUE for block diagonal bordered systems such as those encountered in VLBI data analysis. *Journal of Geodesy*, 72: 343–349.

Magnus JR (1988). *Linear Structures*. London School of Economics and Political Science, Charles Griffin & Company LTD, London: Oxford University Press.

Mandelbrot B (1983). *The Fractal Geometry of Nature*. New York: W. H. Freeman.

Mandelbrot B, van Ness J (1968). Fractional Brownian motions, fractional noises, and applications. *SIAM Review*, 10: 422–439.

Mao A, Harrison CGA, Dixon TH (1999). Noise in GPS coordinate time series. *Journal of Geophysical Research*, 104(B2): 2797–2816.

Maruyama Y, Seo T (2003). Estimation of moment parameters in elliptical distributions. *Journal of the Japan Statistical Society*, 33(2): 215–229.

Odijk D (2002). *Fast precise GPS positioning in the presence of ionospheric delays*. Ph. D. thesis, Delft University of Technology, Publication on Geodesy, 52, Netherlands Geodetic Commission, Delft.

Ou Z (1989). Estimation of variance and covariance components. *Bull. Géod.*, 63: 139–148.

Ou Z (1991). Approximative bayes estimation for variance components. *Manuscripta Geodaetica*, 16: 168–172.

Ou Z, Koch KR (1994). Analytical expressions for Bayes estimates of variance components. *Manuscripta Geodaetica*, 19: 284–293.

Patterson HD, Thompson R (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58.

Penna NT, Stewart MP (2003). Aliased tidal signatures in continuous GPS height time series. *Geophysical Research Letters*, 30(23).

Perfetti N (2006). Detection of station coordinate discontinuities within the Italian GPS Fiducial Network. *Journal of Geodesy*, 80(7): 381–396.

Petrov L, Ma C (2003). Study of harmonic site position variations determined by very long baseline interferometry. *Journal of Geophysical Research*, 108(B4, 2190).

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1992). *Numerical Recipes*. New York: Cambridge University Press.

Priestley MB (1981). *Spectral analysis and time series*. Academic Press, San Diego. Probability and Mathematical Statistics, A Series of Monographs and Textbooks.

Rao CR (1971a). Estimation of variance and covariance components - MINQUE theory. *Journal of multivariate analysis*, 1: 257–275.

Rao CR (1971b). Minimum variance quadratic unbiased estimation of variance components. *Journal of multivariate analysis*, 1: 445–456.

Rao CR (1973). *Linear statistical inference and its applications*. John Wiley & Sons, Inc. New York, London, Sydney, Toronto.

Rao CR, Kleffe J (1988). *Estimation of variance components and applications*, Volume 3. North-Holland. Series in Statistics and Probability.

Rao CR, Mitra SK (1971). *Generalized Inverses of Matrices and Its Applications*. John Wiley and Sons.

Ray J (2006). Systematic errors in GPS position estimates. In *IGS Workshop 2006*, 8-12 May, ESOC Darmstadt, Germany. Session 11.

Sahin M, Cross PA, Sellers PC (1992). Variance component estimation applied to satellite laser ranging. *Bull. Géod.*, 66: 284–295.

Satirapod C, Wang J, Rizos C (2002). A simplified MINQUE procedure for the estimation of variance-covariance components of GPS observables. *Survey Review*, 36(286): 582–590.

Schaffrin B (1981). Best invariant covariance component estimators and its application to the generalized multivariate adjustment of heterogeneous deformation observations. *Bull. Géod.*, 55: 73–85.

Schaffrin B (1983). Varianz-kovarianz-komponenten-schätzung bei der ausgleichung heterogener wiederholungsmessungen. C282, Deutsche Geodätische Kommission, München.

Sjöberg LE (1983). Unbiased estimation of variance-covariance components in condition adjustment with unknowns – a MINQUE approach. *Zeitschrift für Vermessungswesen*, 108(9): 382–387.

Sjöberg LE (1984). Non-negative variance component estimation in the Gauss-Helmert adjustment model. *Manuscripta Geodaetica*, 9: 247–280.

Sjöberg LE (1995). The best quadratic minimum bias non-negative definite estimator for an additive two variance component model. *Manuscripta Geodaetica*, 20: 139–144.

Stewart MP, Penna NT, Lichti DD (2005). Investigating the propagation mechanism of unmodelled systematic errors on coordinate time series estimated using least squares. *Journal of Geodesy*, 79: 479–489.

Strang G (1988). *Linear algebra and its application* (Second ed.). San Diego, USA: Harcourt Brace Jovanovich Publishers.

Teunissen PJG (1985). Generalized inverses, adjustment, the datum problem and S-transformations. Technical report, In: Optimization and Design of Geodetic Networks, EW Grafarend and F Sanso (Eds).

Teunissen PJG (1988). Towards a least-squares framework for adjusting and testing of both functional and stochastic model. Internal research memo, Geodetic Computing Centre, Delft. A reprint of original 1988 report is also available in 2004, No. 26, http://www.lr.tudelft.nl/mgp.

Teunissen PJG (1990). Nonlinear least-squares. *Manuscripta Geodetica*, 15(3): 137–150.

Teunissen PJG (1993). Least squares estimation of the integer GPS ambiguities. In *Invited lecture*, Section IV, Theory and Methodology, IAG General Meeting, Beijing, China.

Teunissen PJG (1995). The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. *Journal of Geodesy*, 70(1–2): 65–83.

Teunissen PJG (1998). The ionosphere-weighted GPS baseline precision in canonical form. *Journal of Geodesy*, 72: 107–117.

Teunissen PJG (2000a). *Adjustment theory: an introduction*. Website http://www.vssd.nl: Delft University Press. Series on Mathematical Geodesy and Positioning.

Teunissen PJG (2000b). *Testing theory: an introduction*. Website: http://www.vssd.nl: Delft University Press. Series on Mathematical Geodesy and Positioning.

Teunissen PJG, Amiri-Simkooei AR (2006). Least-squares variance component estimation. *Journal of Geodesy*. Submitted.

Teunissen PJG, Simons DG, Tiberius CCJM (2005). *Probability and observation theory*. Faculty of Aerospace Engineering, Delft University of Technology: Delft University. Lecture notes AE2-E01.

Tiberius CCJM, Jonkman N, Kenselaar F (1999). The stochastic of GPS observables. *GPS World*, 10(2): 49–54.

Tiberius CCJM, Kenselaar F (2000a). Estimation of the stochastic model for GPS code and phase observables. *Survey Review*, 35(277): 441–454.

Tiberius CCJM, Kenselaar F (2000b). Variance component estimation for precise GPS positioning. *Geodezija Kartografija*, 26(4): 152–159.

Tiberius CCJM, Kenselaar F (2003). Variance component estimation and precise GPS positioning: Case study. *Journal of Surveying Engineering*, 129(1): 11–18.

Urschl C, Gurtner GBW, Hugentobler U, Schaer S (2006). Validation of GNSS orbits using SLR observations. In *IGS Workshop 2006*, 8-12 May, ESOC Darmstadt, Germany. Session 11.

Urschl C, Gurtner W, Hugentobler U, Schaer S, Beutler G (2005). Validation of GNSS orbits using SLR observations. *Advances in Space Research*, 36(3): 412–417.

van Dam T, Wahr J, Milly PCD, Shmakin AB, Blewitt G, Lavallée D, Larson KM (2001). Crustal displacements due to continental water loading. *Geophysical Research Letters*, 28(4): 651–654.

van Loon JP, Kusche J (2005). Statistical model validation of satellite gravity data: A test with CHAMP pseudo-observations. In C. Jekeli, L. Bastos, and J. Fernandes (Eds.), *Gravity, Geoid and Space Missions*, Springer, Berlin, pp. 24–29.

Vanícek P (1969). Approximate spectral analysis by least-squares fit. *Astrophysics and Space Science*, 4: 387–391.

Verhagen AA (2005). *The GNSS integer ambiguities: estimation and validations*. Ph. D. thesis, Delft University of Technology, Publication on Geodesy, 58, Netherlands Geodetic Commission, Delft.

Wang J, Stewart MP, Tsakiri M (1998). Stochastic modeling for static GPS baseline data processing. *Journal of Surveying Engineering*, 124(4): 171–181.

Ware R, Lad F (2003). Approximating the distribution for sums of products of normal variables. Technical Report UCDMS 2003/15, Mathematics and Statistics Department, University of Canterbury.

Williams SDP (2003a). The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. *Journal of Geodesy*, 76: 483–494.

Williams SDP (2003b). Offsets in Global Positioning System time series. *Journal of Geophysical Research*, 108.

Williams SDP (2006). Personal communication. Proudman Oceanographic Laboratory, Bidston, UK.

Williams SDP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Prawirodirdjo L, Miller M, Johnson DJ (2004). Error analysis of continuous GPS position time series. *Journal of Geophysical Research*, 109.

Wyatt FK (1982). Displacement of surface monuments: horizontal motion. *Journal of Geophysical Research*, 87: 979–989.

Wyatt FK (1989). Displacement of surface monuments: vertical motion. *Journal of Geophysical Research*, 94: 1655–1664.

Wyatt FK, Bolton H, Bralla S, Agnew DC (1989). New designs of geodetic monuments for use with GPS. *Eos Trans. AGU*, 70: 1054–1055.

Xu PL, Liu YM, Shen YZ (2007). Estimability analysis of variance and covariance components. *Journal of Geodesy*.

Xu PL, Shen YZ, Fukuda Y, Liu YM (2006). Variance component estimation in linear inverse ill-posed models. *Journal of Geodesy*, 80: 69–81.

Yu Z (1992). A generalization theory of estimation of variance-covariance components. *Manuscripta Geodaetica*, 17: 295–301.

Yu Z (1996). A universal formula of maximum likelihood estimation of variance-covariance components. *Journal of Geodesy*, 70: 233–240.

Zhang J, Bock Y, Johnson H, Fang P, Williams S, Genrich J, Wdowinski S, Behr J (1997). Southern California Permanent GPS Geodetic Array: Error analysis of daily position estimates and site velocitties. *Journal of Geophysical Research*, 102: 18035–18055.

Zumberge JF, Heflin MB, Jefferson DC, Watkins MM, Webb FH (1997). Precise point positioning for the efficient and robust analysis of GPS data from large networks. *Journal of Geophysical Research*, 102: 5005–5017.

Index

Notation and Symbols

Mathematical Notation and Operators

$\mathbb{R}^m$	real Euclidean space of dimension m
$I \equiv I_m$	identity matrix of order m
c_i	canonical unit vector which contains zeros except a one at position i
$c_{y_i} = c_i$	canonical unit vector $c_i = [0, ..., 0, 1, 0, ..., 1]^T$
C_k, or C_α	matrix of form $C_k = c_i c_i^T$ if $i = j$, and $C_k = c_i c_j^T + c_j c_i^T$ if $i \neq j$
W_t	a $b \times b$ symmetric and positive definite matrix
W	an $m \times m$ symmetric and positive definite matrix
Q	an $m \times m$ symmetric and positive definite matrix
$u \equiv u_m$	m-vector with ones (summation vector)
$\delta_{ij} = c_i^T c_j$	Kronecker delta which is one if $i = j$, and zero elsewhere
D	duplication matrix
K	commutation matrix $K = \sum_{i=1}^{u} \sum_{k=1}^{u} c_i c_k^T \otimes c_k c_i^T$
$\mathrm{tr}(.)$	trace of a matrix (sum of the diagonal elements)
$\mathrm{rank}(.)$	rank of a matrix (independent columns or rows of a matrix)
$\det(.)$	determinant of a matrix
$(.)^T$	transpose of a matrix
$(.)^{-1}$	inverse of a matrix
$(.)^-$	reflexive inverse of a matrix
$(.)^+$	pseudo (Moore-Penrose) inverse of a matrix
:	double inner (dot) product, e.g. $S : T = \mathrm{tr}(ST)$ for matrices S and T
$\|.\|_W$	squared norm of a vector as $(.)^T W (.)$
$\mathrm{diag}(.)$	diagonal elements of a matrix
$\mathrm{blkdiag}(.)$	block-diagonal operator
$\mathcal{R}(.)$	range space of a matrix
$\mathcal{N}(.)$	null space of a matrix
$\oplus$	direct sum of two subspaces
$\perp$	orthogonal complement (is orthogonal to)
$\otimes$	Kronecker product
vec	vector operator (vec-operator)
vh	vector-half operator (vh-operator)
c	condition number of a matrix $c = \lambda_{\max}/\lambda_{\min}$
$\lambda_{\max}$	maximum eigenvalue of a matrix
$\lambda_{\min}$	minimum eigenvalue of a matrix
λ_i or λ_k	eigenvalues of a matrix

Statistical Notation and Operators (Functional Model)

$\mathsf{E}\{.\}$	expectation operator
$\mathsf{D}\{.\}$	dispersion operator
$\mathsf{C}\{.\}$	covariance operator
$\mathsf{P}\{x = x_o\}$	probability that x will be equal to x_o
$\mathsf{N}(x, Q_x)$	normal distribution with mean x and covariance matrix Q_x
$\chi^2(q, \lambda)$	chi-squared distribution with q degrees of freedom and non-centrality parameter λ
H_o	null hypothesis
H_a	alternative hypothesis
α	type I error probability
$\beta = 1 - \gamma$	type II error probability
$f_y(y\|x)$	probability density function of observables $\underline{y}$
$f_y(y\|x)$	likelihood function of y for a given observed y
$L(y; x, \sigma)$	likelihood function associated with $\underline{y}$
$x \in \Phi_0$	x belongs to subset Φ_0
$x \in \Phi \backslash \Phi_o$	x belongs to subset Φ that is complementary to Φ_0
$\underline{y}$	m-vector of observables
x	n-vector of unknown parameters
$b = m - n$	redundancy of functional model
A	$m \times n$ design matrix of functional model $\mathsf{E}\{\underline{y}\} = Ax$
W	$m \times m$ weight matrix
Q_y	$m \times m$ covariance matrix of observables
C_x	$n \times d$ constraint matrix on unknown parameters: $C_x^T x = c_o$
$\underline{\hat{x}}$	n-vector (estimator of x)
$\hat{x}$	n-vector (estimate of x)
$\underline{\hat{\epsilon}}$	n-vector of estimation error $\underline{\hat{\epsilon}} = \underline{\hat{x}} - x$
$Q_{\hat{x}}$	covariance matrix of $\underline{\hat{x}}$
$\underline{t} = B^T \underline{y}$	b-vector of misclosures
B	$m \times b$ coefficient matrix in condition model $B^T \mathsf{E}\{\underline{y}\} = \mathsf{E}\{\underline{t}\} = 0$
Q_t	$b \times b$ covariance matrix of misclosures
$P_A = P_{Q_y B}^{\perp}$	orthogonal projector–projects to range space of A ($\mathcal{R}(A)$)
$P_A^{\perp} = P_{Q_y B}$	orthogonal projector–projects to $\mathcal{R}(A)^{\perp}$
e	m-vector of measurement error
$\underline{\hat{e}}$	least-squares estimator of residuals, $\underline{\hat{e}} = P_A^{\perp} \underline{y}$
$\hat{e}_o$	least-squares estimate of residuals under H_o
$\hat{e}_a$	least-squares estimate of residuals under H_a
$\underline{\hat{y}}$	least-squares estimator of observables, $\underline{\hat{y}} = P_A \underline{y}$
$Q_{\hat{e}}$	covariance matrix of $\underline{\hat{e}}$, $Q_{\hat{e}} = P_A^{\perp} Q_y P_A^{\perp T}$
$Q_{\hat{y}}$	covariance matrix of $\underline{\hat{y}}$, $Q_{\hat{y}} = P_A Q_y P_A^T$
$\underline{T}_q$	T-test statistic with q degrees of freedom
$\underline{w}$	w-test statistic
$\underline{v}$	v-test statistic
C_y	$m \times q$ matrix for extension of A to $[A\, C_y]$ under H_a
c_y	$m \times 1$ vector $c_y = C_y$, where $q = 1$
$\sigma_{\hat{e}_i}$	standard deviation of least-squares residual i: $\sigma_{\hat{e}_i} = (Q_{\hat{e}})_{ii}^{1/2}$

Stochastic Model

σ^2	variance of unit weight in $Q_y = \sigma^2 Q$, Q is cofactor matrix
$\underline{\hat{\sigma}}^2$	estimator of variance of unit weight: $\underline{\hat{\sigma}}^2 = \frac{\underline{\hat{e}}^T Q^{-1} \underline{\hat{e}}}{m-n}$
σ	p-vector of (co)variance components $\sigma = [\sigma_1, ..., \sigma_p]^T$
σ^i	(co)variance components $\sigma^i = [\sigma_1^i, ..., \sigma_p^i]^T$ at iteration i
H	Helmert matrix to VCE
E_k	submatrices for multinomial inverse of Q_y
$\underline{\hat{\sigma}}_k^B$, $\underline{\hat{\sigma}}_k^M$	Bayes and MAP estimators of variance components
Q_k	k^{th} cofactor matrix in: $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$
Q_0	known part of covariance matrix Q_y
$\underline{y}^T M \underline{y}$	quadratic form of observable vector $\underline{y}$
$f^T \sigma$	linear function of (co)variance components
$\Lambda = \lambda_{ij}$	matrix of Lagrange multipliers
$\lambda = \lambda_i$	vector of Lagrange multipliers
$\underline{y}_{\text{vh}}$	vector of observables in stochastic model: $\underline{y}_{\text{vh}} = \text{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)$
$\underline{\hat{e}}_{\text{vh}}$	least-squares residuals in stochastic model
A_{vh}	design matrix in stochastic model
W_{vh}	weight matrix in stochastic model
Q_{vh}	covariance matrix of observables $\underline{y}_{\text{vh}}$ in stochastic model
N	$p \times p$ normal matrix in stochastic model: e.g. $N = A_{\text{vh}}^T W_{\text{vh}} A_{\text{vh}}$
$\underline{l}$	right-hand side p-vector in $N\underline{\hat{\sigma}} = \underline{l}$
μ_2, μ_4	second and fourth central moments, respectively
κ	kurtosis parameter
$\underline{\hat{\sigma}} = N^{-1}\underline{l}$	least-squares estimator of σ
$Q_{\hat{\sigma}}$	covariance matrix of $\underline{\hat{\sigma}}$: $Q_{\hat{\sigma}} = N^{-1} M N^{-1}$ or $Q_{\hat{\sigma}} = N^{-1}$
df	redundancy of stochastic model $df = \frac{b(b+1)}{2} - p$
A_J	contains the first $j-1$ columns of A_{vh}
a_j	column j of A_{vh}
N_{JJ}	normal matrix corresponding to A_J
$P_{A_{\text{vh}}}$	orthogonal projector in stochastic model $P_{A_{\text{vh}}} = I - P_{A_{\text{vh}}}^{\perp}$
$P_{A_{\text{vh}}}^{\perp}$	orthogonal projector in stochastic model
c_{vh}	a $\frac{b(b+1)}{2}$ vector given as $c_{\text{vh}} = \text{vh}(C_t)$
C_t	a $b \times b$ symmetric matrix
C_y	an $m \times m$ symmetric matrix
Q_{t_k}	transformed cofactor matrices $Q_{t_k} = B^T Q_k B,\ k = 0, ..., p$
g	a b-vector of form $g_k = \frac{1}{2}\text{tr}(C_t Q_t^{-1} Q_{t_k} Q_t^{-1})$
$Q_{\hat{e}}^-$	reflexive inverse of $Q_{\hat{e}} = P_A^{\perp} Q_y$, $Q_{\hat{e}}^- = Q_y^{-1} P_A^{\perp}$
$\underline{w}$	w-test statistic in stochastic model
$\underline{v}$	v-test statistic in stochastic model
$\underline{T}_{\text{df}}$	overall model test in stochastic model
w^n, w^d	numerator and denominator of w-test statistic, respectively
$d^T \sigma = c_o$	a linear (hard) constraints on (co)variance components
$\Sigma = \sigma_{ij}$	an $r \times r$ matrix describing covariances between repeated models
$\hat{\rho}_{ij}$	correlation coefficient between models, or between time instants
$\sigma_{\hat{\rho}}$	standard deviation of correlation coefficient

GPS and Time Series

$p_{rj,L}^{sk}$	double difference pseudo ranges on L1 or L2
$\phi_{rj,L}^{sk}$	double difference carrier phase on L1 or L2
$a_{rj,L}^{sk}$	double difference integer carrier phase ambiguities on L1 or L2
σ_p	standard deviation of pseudo range
σ_ϕ	standard deviation of carrier phase
$\sigma_{p\phi}$	covariance between pseudo range and carrier phase
$P_y(f)$	power spectrum of a noise process
f	temporal frequency
κ	spectral index
y_0	intercept in the linear regression model $\mathsf{E}\{\underline{y}(t)\} = y_0 + r\,t$
r	rate (slope) in the linear regression model $\mathsf{E}\{\underline{y}(t)\} = y_0 + r\,t$
s_w^2	variance of a white noise process
s_f^2	variance of a flicker noise process
s_{rw}^2	variance of a random walk noise process
s_a^2	variance of an autoregressive noise process
s_s^2	variance of a short-memory noise process
Q_f	covariance matrix of a flicker noise process
Q_{rw}	covariance matrix of a random walk noise process
Q_a	covariance matrix of an autoregressive noise process
Q_s	covariance matrix of a short-memory noise process

Abstract (in Dutch)

Samenvatting

Gegevensverwerking in geodetische toepassingen vindt doorgaans plaats op basis van de kleinste kwadraten methode. Hiervoor is een goed stochastisch model van de waarnemingsgrootheden nodig. Met een dergelijke realistische covariantie-matrix wordt ten eerste de beste (minimum variantie) lineaire zuivere schatter voor de onbekende parameters verkregen, en kan ten tweede een realistische precisiebeschrijving van de parameters gegeven worden, en kan ten derde, op basis van de verdeling van de waarnemingsgrootheden, hypothese toetsing correct uitgevoerd worden, en kunnen maten voor kwaliteitscontrole bepaald worden, zoals voor betrouwbaarheid. In veel praktische toepassingen is de covariantie-matrix slechts gedeeltelijk bekend. De covariantie-matrix wordt gewoonlijk uitgedrukt als een onbekende lineaire combinatie van een aantal bekende cofactor matrices. Het schatten van de onbekende (co)variantie componenten wordt in het algemeen variantiecomponentenschatting (VCS) genoemd, en ook wel kansmodelschatting.

In dit proefschrift bestuderen we de methode van kleinste kwadraten variantiecomponenten-schatting (KK-VCS) en werken we theoretische en praktische aspecten uit. We laten zien dat de KK-VCS methode een eenvoudige, flexibele en aantrekkelijke methode is voor VCS. De KK-VCS is eenvoudig, daar ze gebaseerd is op het bekende kleinste kwadraten principe. Met deze methode is de schatting van de (co)variantiecomponenten gebaseerd op een lineair model van waarnemingsvergelijkingen. De methode is flexibel omdat ze werkt met een door de gebruiker gedefinieerde gewichtsmatrix. Verschillende klassen van gewichtsmatrices kunnen gedefinieerd worden, die allemaal automatisch tot zuivere schatters voor de (co)variantie componenten leiden. KK-VCS is aantrekkelijk omdat men de bestaande kleinste kwadraten theorie kan toepassen op het probleem van variantie-componenten-schatting. Met deze methode kan men 1) maten voor discrepantie in het stochastisch model verkrijgen, 2) de covariantie-matrix bepalen van de (co)variantie componenten, 3) de minimum variantie schatter verkrijgen voor de (co)variantie componenten door de inverse van de covariantie-matrix als gewichtsmatrix te nemen, 4) a-priori informatie over de (co)variantie componenten in rekening brengen, 5) een niet-lineair (co)variantie componenten model oplossen, 6) robuuste schatting toepassen op variantie componenten, 7) de schatbaarheid van (co)variantie componenten evalueren, en 8) het probleem van negatieve variantie componenten voorkomen.

Met KK-VCS kunnen vele bestaande VCS methoden, zoals MINQUE, BIQUE, en REML in één raamwerk geplaatst worden. Deze methoden worden verkregen door speciale keuzes voor de gewichtsmatrix te maken. Een belangrijke eigenschap van de KK-VCS methode is de mogelijkheid om hypothese toetsing toe te passen op het stochastisch model. We gebruiken hiervoor de w-toets, de v-toets, en de globale toets. Ons doel is om een geschikte

structuur te vinden voor het stochastisch model, dat alle relevante ruiscomponenten in de covariantie-matrix herbergt. De w-toets wordt geïntroduceerd om te kunnen vaststellen of een bepaalde ruiscomponent met waarschijnlijkheid aanwezig is in de waarnemingen, en dientengevolge opgenomen moet worden in het stochastisch model. Gebaseerd op de normale verdeling voor oorspronkelijke waarnemingsgrootheden leiden we de verwachting en de variantie van de w-toetsgrootheid af, welke nul en één zijn, respectievelijk. De verdeling is een lineaire combinatie van onderling onafhankelijke centrale chi-kwadraat verdelingen, elk met één vrijheidsgraad. In een aantal speciale gevallen kan deze verdeling benaderd worden door een standaard normale verdeling. Als een equivalente uitdrukking voor de w-toets wordt de v-toetsgrootheid gegeven. Het doel is om het aantal (co)variantie componenten in het stochastisch model te reduceren, door de significantie van de componenten te toetsen. De globale toets functioneert als een algemene toets op de geschiktheid van het aangenomen stochastisch model.

KK-VCS is toegepast op meetgegevens uit twee GPS toepassingen. Als eerste is de methode toegepast op het GPS geometrievrije model. Daartoe worden het functiemodel en het stochastisch model opgesteld. De variantiecomponenten van verschillende waarnemingstypen, de satelliet-elevatie afhankelijkheid van de precisie van GPS waarnemingsgrootheden, en de correlatie tussen verschillende waarnemingstypen worden geschat met KK-VCS. We laten zien dat de precisie van GPS waarnemingsgrootheden duidelijk afhangt van de elevatie van de satelliet. Ook is er een significante correlatie tussen de waarnemingstypen. Als tweede toepassing worden de ruiskarakteristieken in tijdreeksen van dagelijkse coördinaten van permanente GPS stations bepaald. De KK-VCS is toegepast om de amplitudes van witte ruis en power-law ruis (flicker ruis en random walk ruis) in deze tijdreeksen te schatten. De resultaten bevestigen dat de tijdreeksen behoorlijk gecorreleerd in de tijd zijn. We hebben ook de w-toetsgrootheid gebruikt om een geschikt stochastisch model voor de GPS tijdreeksen te vinden. Een combinatie van witte ruis, autoregressieve ruis, en flicker ruis karakteriseert in het algemeen de ruis in alle drie positie componenten het beste. Ongemodelleerde periodieke effecten in de metingen worden beschreven door een stel harmonische functies. Deze worden geschat met behulp van de kleinste kwadraten methode, in hetzelfde raamwerk als KK-VCS. De resultaten bevestigen de aanwezigheid van jaarlijkse en halfjaarlijkse signalen in de reeksen, als ook andere significante periodieke patronen. Om onzuivere schatting van de variantiecomponenten te voorkomen, dienen dergelijke sinusvormige signalen in het functiemodel opgenomen te worden, alvorens KK-VCS toe te passen.

Curriculum Vitae

AliReza Amiri-Simkooei was born on 24 March 1971 in Bafgh, Yazd, Iran. He graduated from high school in 1989 and, at the same time, he was admitted to the University of Isfahan, Isfahan, Iran. In Spring 1994 he graduated with high honor–first position–with a Bachelor of Science degree in Surveying Engineering. A few months later, he started his master study in Geodesy at K.N. Toosi University of Technology, Tehran, Iran. He worked on 'analytical methods in optimization and design of geodetic networks' and graduated as an outstanding student in 1998. At the same time he was offered a permanent position at the University of Isfahan, Isfahan, Iran. He worked there for 4 years mainly lecturing for bachelor students.

In 2002, he was entitled to an overseas PhD scholarship from the Iranian Ministry of Science, Research and Technology. He worked under supervision of Prof. Peter Teunissen and Dr. Christian Tiberius at the Delft institute of Earth Observation and Space systems (DEOS) of Delft University of Technology, the Netherlands. This thesis covers his PhD study. While finalizing his PhD thesis, in late 2006, he started working as a research associate at the same institute.

when	what	where
9.1985 - 6.1989	High school	Shariati high school, Kerman
9.1989 - 3.1994	BSc in Surveying	The University of Isfahan
9.1994 - 9.1998	MSc in Geodesy	K.N. Toosi University of Technology
9.1998 - 9.2002	Lecturer	The University of Isfahan
10.2002-4.2007	PhD in Geodesy	Delft University of Technology
10.2006- ...	Research associate	Delft University of Technology

Acknowledgements

This thesis could not be realized without the sincere help and support from many people. It is difficult to thank all the individuals who really need thanking for making this work possible.

In the first place, I would like to thank my promotor, Prof. Peter Teunissen, for his complete scientific support. I was impressed with his theory on least-squares variance component estimation which was originally developed in 1988. I really enjoyed studying and elaborating on theoretical and practical aspects of this theory. His valuable suggestions to improve the early versions of this thesis are gratefully acknowledged.

I would also like to express my deepest thanks to my supervisor, Dr. Christian Tiberius, for his excellent guidances during the last four years working and studying at the Delft institute of Earth Observation and Space systems (DEOS) at TU Delft. His open-door policy, accessibility, continuous supervision, quick feedback, friendly behavior and meticulous reviewing of the earlier versions of my works along with his valuable comments are kindly appreciated.

I am also very much thankful to all members of the examination committee for their reviews and useful comments on my thesis.

In this place I would like to thank all of my colleagues at DEOS and specially in MGP for their support and contribution to a good research environment. A special thanks goes to the following people: Prof. Dick Simons for his help while Prof. Peter Teunissen was on sabbatical in Australia; Dr. Hans van der Marel for his satellite positioning course and his general support on GPS-related topics; Ria Scholtes for the administrative support; Jasper van Loon for proofreading this thesis; Peter Joosten for the computer support and for his Matlab scripts; Dr. Sandra Verhagen for lending me her LaTeXstyle file adapted for this thesis and for her general support; Frans Schröder for helping me make the cover of the thesis; and my former colleague Frank Kenselaar at the start of my PhD research.

I would also like to mention the name of other friends who helped me indirectly to complete this work. I can mention at least the names of Dr. Ramon Hanssen, Dr. Mirjam Snellen, Dr. Maxim Keshin, Dr. Dennis Odijk, Dr. Frank Kleijer, Dr. Roderik Lindenbergh, Dr. Rossen Grebenitcharsky, Dr. Zbigniew Perski, Anh Quan Le, Yongjin Moon, Xianglin Liu, Petar Marinkovic, Benoît Muth, Junlin Yan, Freek van Leijen, Gini Ketelaar, and Mahmut Arikan. The best atmosphere creates the best work. I am also indebted to my Iranian friends Hossein Moosavi Alkazemi, Mohammad Ali Sharifi, Masoud Shirazian, Yaghoub Hatam Chavari, Saeed Torkzaban, Sirous Ebrahimi, Mahdi Motagh, Hossein Mansouri, Asadollah Shahbahrami, Shohre Shahnoori, Ata Abdollahi Kakroodi, Gholamreza Shams, Ali Heidary, Yahya Memarzadeh, Fatemeh Karimi Nejadasl, and Amir Abbas Zadpoor.

The Iranian Ministry of Science, Research, and Technology as well as Delft University of Technology (Department of DEOS) are gratefully acknowledged for their financial support to pursue my PhD research.

Finally, I would like to express my sincerest appreciation to my whole family. A sincere appreciation goes to my brothers, sisters and specially to my mother, Fatemeh, whom I really missed during this period. I also express my deepest gratitude to my wife and my daughter, Shahin and Shamim, for their understanding when I was really busy with my work and for being so patient all those evenings that I was away.