Least-squares variance component estimation

P. J. G. Teunissen · A. R. Amiri-Simkooei

Abstract Least-squares variance component estimation (LS-VCE) is a simple, flexible and attractive method for the estimation of unknown variance and covariance components. LS-VCE is simple because it is based on the well-known principle of LS; it is flexible because it works with a user-defined weight matrix; and it is attractive because it allows one to directly apply the existing body of knowledge of LS theory. In this contribution, we present the LS-VCE method for different scenarios and explore its various properties. The method is described for three classes of weight matrices: a general weight matrix, a weight matrix from the unit weight matrix class; and a weight matrix derived from the class of elliptically contoured distributions. We also compare the LS-VCE method with some of the existing VCE methods. Some of them are shown to be special cases of LS-VCE. We also show how the existing body of knowledge of LS theory can be used to one's advantage for studying various aspects of VCE, such as the precision and estimability of VCE, the use of a-priori variance component information, and the problem of nonlinear VCE. Finally, we show how the mean and the variance of the fixed effect estimator of the linear model are affected by the results of LS-VCE. Various examples are given to illustrate the theory.

P. J. G. Teunissen (✉) · A. R. Amiri-Simkooei
Delft Institute of Earth Observation and Space systems (DEOS),
Faculty of Aerospace Engineering, Delft University of Technology,
Kluyverweg 1, 2629 HS Delft, The Netherlands
e-mail: p.j.g.teunissen@tudelft.nl

A. R. Amiri-Simkooei
Department of Surveying Engineering, Faculty of Engineering,
The University of Isfahan, 81744 Isfahan, Iran

1 Introduction

The second-order central moments of the entries of a random vector are described by the entries of its variance matrix. For many applications, it is of importance to have information available on the variance matrix of an observable random vector. Such information allows one to study the different contributing factors of the errors in observations, to describe the precision of random vectors, or functions thereof, by means of an application of the variance propagation law, or to obtain minimum variance estimators of the parameters in a linear model. Often, however, the variance matrix of the observables is only partly known, as a consequence of which the unknown part needs to be estimated from the data. The estimation of these unknown components of a variance matrix is generally referred to as variance component estimation (VCE).

Incomplete knowledge of the variance matrix of the observables occurs in many modern geodetic applications. In the case of Global Navigation Satellite Systems (GNSS), for instance, our knowledge of the stochastic model is still at a rather rudimentary level. This is in contrast to the GNSS functional model (i.e., observation equations), which is sufficiently known and well-documented in the many GNSS textbooks available. Various VCE studies have been conducted to improve our knowledge of the GNSS stochastic model.

Examples of earlier studies are Euler and Goad (1991); Gerdan (1995); Gianniou (1996), and Jin and de Jong (1996), who studied the elevation dependence of the observation variances; Jonkman (1997); Tiberius (1998); Wang et al. (1998), and Tiberius et al. (1999), who also considered time correlation and cross-correlation of the code pseudoranges and carrier phases; and Schaffrin and Bock (1988); Bock (1998); Teunissen (1998), and Brunner et al. (1999), who considered the inclusion of atmospheric uncertainties.

Examples showing that improved stochastic modelling indeed pays off in terms of an increased success rate for GNSS ambiguity resolution, can be found in e.g., Jonkman (1997) and Teunissen et al. (1998). More recent examples of VCE studies dealing with the GNSS stochastic model are Tiberius and Borre (2000); Tiberius and Kenselaar (2000); Tiberius and Kenselaar (2003); Bona (2000); Wang et al. (2002); Satirapod et al. (2002); Radovanovic et al. (2004); Bischoff et al. (2005, 2006), and Amiri-Simkooei and Tiberius (2007). Also in the GNSS position domain, similar studies are ongoing. Examples are Zhang et al. (1997); Mao et al. (1999); Williams et al. (2004), and Amiri-Simkooei et al. (2007), who studied the noise characteristics (e.g., white noise, random walk and flicker noise) in daily GPS coordinate time-series of permanent networks.

VCE is also an important issue in other geodetic fields of application, in particular in applications where heterogeneous data needs to be combined. Examples are Koch and Kusche (2002) and Kusche (2003), who studied the relative weighting of heterogeneous data for gravity field modelling; Fotopoulos (2003, 2005), who considered the combined adjustment of ellipsoidal, orthometric and geoid heights; Chen et al. (1990), who studied the combination of electronic distance measurements and GPS for monitoring tectonic activity; and Sahin et al. (1992) and Lucas and Dillinger (1998), who applied VCE to satellite laser ranging and very long baseline interferometry, respectively.

Many different methods exists for VCE. These 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 variance 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. Of the different 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 approach to VCE.

The MINQUE method (Rao 1971) is probably still one of the more commonly used methods for VCE. This method does not require any distributional assumptions, apart from

the first- and second-order moments of the observables. 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, by e.g., Koch (1978); Caspary (1987); Yu (1992).

The LS-VCE method is based on the least-squares principle and works with a user-defined weight matrix (Teunissen, 1988). The method has been applied to study the stochastics of GPS code and carrier-phase data, and GPS coordinate time-series, respectively, by Tiberius and Kenselaar (2000); Tiberius and Kenselaar (2003); Teunissen and Amiri-Simkooei (2006); Amiri-Simkooei (2007) and Amiri-Simkooei et al. (2006, 2007). The REML method and the Bayesian approach, both require, in contrast to the previous methods, complete information about the probability density function of the observables. Using the normal distribution, maximum likelihood estimators and Bayesian estimators have been derived and studied by, e.g., Koch (1987); Ou (1991); Yu (1996). For a review and comparison of some of the VCE methods, we refer to, e.g., Grafarend (1985); Crocetto et al. (2000); Fotopoulos (2003).

In the present contribution, we further study the LS-VCE method. 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. Being a least-squares estimator, the LS-VCE automatically inherits all the well-known properties of a least-squares estimator. To illustrate the theory, we have included various VCE-examples, some of which are of interest in their own right.

This contribution is organised as follows. In Sect. 2, we introduce the principle of weighted least-squares estimation for the estimation of unknown (co)variance components. We formulate the linear (co)variance component model, define the least-squares variance component estimator and determine its mean and variance. We also show how LS-VCE can be turned into a minimum variance VCE. In Sect. 3, we 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 VCE problem. Topics that are addressed are: measures of inconsistency, the use of a-priori variance component information, estimability of variance components, nonlinear variance component estimation, and robust and non-negative variance component estimation.

In Sects. 4 and 5, we consider the LS-VCE method for two special classes of weight matrices. In Sect. 4, we introduce the unit weight matrix class and show how the LS-VCE simplifies accordingly. In Sect. 5, we use a weight matrix derived from the class of elliptically contoured distributions. The corresponding LS-VCE, which is of minimum variance, is worked out accordingly. The class of elliptically contoured distributions includes the multivariate

normal distribution and the multivariate Student distribution as special cases.

In Sect. 6, we show how the mean and the variance of the fixed effect estimator of the linear model is affected by the results of LS-VCE. This will allow one to judge by how much the precision of the weighted least-squares estimator of the fixed effects is degraded when the estimated variance matrix is used as inverse weight matrix, instead of the true, but unknown, variance matrix. Finally, in Sect. 7, a comparison is made with some of the existing VCE methods, such as MINQUE, BIQUE and REML. It is shown that each of these variance component estimators can be reproduced with the LS-VCE method by using a suitable choice of weight matrix.

2 Weighted least-squares estimation of (co)variance components

2.1 The linear (co)variance component model

First, we will show how one can formulate a linear system of observation equations for the unknown (co)variance components. We start from the linear model

$$\mathsf{E}\{y\} = Ax, \quad \mathsf{D}\{y\} = Q_{yy} \tag{1}$$

with

$$Q_{yy} = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k \tag{2}$$

and where $\mathsf{E}\{.\}$ and $\mathsf{D}\{.\}$ denote the mathematical expectation operator and dispersion operator, respectively. The model in Eq. (1) contains two sets of unknowns: the parameter vector x and the (co)variance components $\sigma_k, k = 1, \ldots, p$.

It is the goal of VCE to construct an estimator for the unknown (co)variance components. The $m \times n$ matrix A is assumed known and to be of full column rank. Also the cofactor matrices $Q_k, k = 0, \ldots, p$, are assumed known and their weighted sum $Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$ is assumed to be positive definite. Matrix Q_0 is the known part of the variance matrix Q_{yy}.

We now introduce a one-to-one transformation of the vector of observables y, such that Eq. (1) takes a simplified form. Let B be a basis matrix of the null space of A^T. Thus, B is an $m \times (m - n)$ matrix of which the $m - n$ linear independent columns span the null space of A^T: $A^T B = 0$ or $B^T A = 0$. Then the following one-to-one correspondence between y and $(\hat{x}^T, t^T)^T$ exists:

$$\begin{bmatrix} \hat{x} \\ t \end{bmatrix} = \begin{bmatrix} Q_{\hat{x}\hat{x}} A^T Q_{yy}^{-1} \\ B^T \end{bmatrix} y \iff$$

$$y = [A, Q_{yy} B Q_{tt}^{-1}] \begin{bmatrix} \hat{x} \\ t \end{bmatrix} \tag{3}$$

with $Q_{\hat{x}\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1}$ and $Q_{tt} = B^T Q_{yy} B$. If we apply this one-to-one transformation to Eq. (1), we obtain the linear model

$$\mathsf{E}\left\{\begin{bmatrix} \hat{x} \\ t \end{bmatrix}\right\} = \begin{bmatrix} x \\ 0 \end{bmatrix},$$

$$\mathsf{D}\left\{\begin{bmatrix} \hat{x} \\ t \end{bmatrix}\right\} = \begin{bmatrix} (A^T Q_{yy}^{-1} A)^{-1} & 0 \\ 0 & B^T Q_{yy} B \end{bmatrix} \tag{4}$$

Note the decoupling between $\hat{x}$ and t. We recognize the n-vector $\hat{x}$ as the best linear unbiased estimator (BLUE) of x. The zero-mean $(m - n)$-vector t is the redundancy vector of misclosures. It consists of the misclosures that follow from using the model of condition equations, $B^T \mathsf{E}\{y\} = 0$. The redundancy of the linear model in Eq. (1) is defined as $b = m - n$. The vector of misclosures t exists on account of redundancy ($b > 0$). The two random vectors $\hat{x}$ and t are uncorrelated. They are independent if y is normally distributed.

From Eqs. (2) and (4), it follows, since $\mathsf{E}\{t\} = 0$ and $\mathsf{D}\{t\} = \mathsf{E}\{tt^T\} = B^T Q_{yy} B$, that $\mathsf{E}\{tt^T - B^T Q_0 B\} = \sum_{k=1}^{p} \sigma_k B^T Q_k B$. This equation can be seen as a *matrix observation equation* for the unknown (co)variance components. The matrix equation consists of b^2 scalar observation equations. We can bring the matrix observation equation into the familiar vector-matrix form, if we stack the b number of $b \times 1$ column vectors of $\mathsf{E}\{tt^T - B^T Q_0 B\}$ into a $b^2 \times 1$ observation vector. This is achieved by the so-called vec-operator.

However, since the observation matrix $tt^T - B^T Q_0 B$ is symmetric, we should only consider the $\frac{1}{2}b(b + 1)$ entries on and below the diagonal of the observation matrix. Otherwise an unwanted duplication of the data occurs. This implies that we should use the vh operator, instead of the vec operator (for a definition and properties of the vec and vh operators, we refer to Appendix A). The use of the vh operator, instead of the vec operator, also avoids the problem of having to work with a singular variance matrix (see Sect. 5.1). The variance matrix of $\text{vec}(tt^T)$ is singular due to the duplication that occurs in the entries of tt^T. With the vh operator, we can bring the matrix observation equation in the familiar vector-matrix form of linear observation equations

$$\mathsf{E}\{y_{\text{vh}}\} = A_{\text{vh}}\sigma \tag{5}$$

where $y_{\text{vh}} = \text{vh}(tt^T - B^T Q_0 B)$, $A_{\text{vh}} = [\text{vh}(B^T Q_1 B), \ldots, \text{vh}(B^T Q_p B)]$, and $\sigma = [\sigma_1, \ldots, \sigma_p]^T$.

The linear model of Eq. (5) will form the basis of our least-squares (co)variance component estimation. The $\frac{1}{2}b(b + 1)$ vector $y_{\text{vh}} = \text{vh}(tt^T - B^T Q_0 B)$ plays the role of the observation vector. Thus, we have $\frac{1}{2}b(b+1)$ observation equations in the p unknown parameters $\sigma_k, k = 1, \ldots, p$. We will assume that the design matrix of Eq. (5) has full column rank p. The redundancy of the above model is then $\frac{1}{2}b(b + 1) - p$.

2.2 The weighted least-squares (co)variance component estimator

Now that we have the above linear model available, we can immediately determine the weighted least-squares solution for the unknown (co)variance components.

Theorem 1 (weighted LS-VCE) *Let*

$$\mathsf{E}\{y_{\mathrm{vh}}\} = A_{\mathrm{vh}}\sigma \tag{6}$$

be the linear (co)variance component model and define the weighted LS-VCE of σ as $\hat{\sigma} = \arg\min_\sigma (y_{\mathrm{vh}} - A_{\mathrm{vh}}\sigma)^T W_{\mathrm{vh}}(y_{\mathrm{vh}} - A_{\mathrm{vh}}\sigma)$, where W_{vh} is the weight matrix. Then

$$\hat{\sigma} = (A_{\mathrm{vh}}^T W_{\mathrm{vh}} A_{\mathrm{vh}})^{-1} A_{\mathrm{vh}}^T W_{\mathrm{vh}} y_{\mathrm{vh}} \tag{7}$$

Proof Follows from standard least-squares theory.

The weighted LS-VCE $\hat{\sigma}$ has many attractive features. Since $\hat{\sigma}$ is a least-squares estimator, we can make a direct use of the existing body of knowledge of least-squares theory. This holds true for the numerical aspects (e.g., use can be made of standard least-squares software packages), as well as for many of the statistical aspects.

First we give, for later use, two different but equivalent expressions for the LS-VCE system of normal equations.

Corollary 1 (LS-VCE normal equations)
Let the system of normal equations of the LS-VCE of Eq. (7) be given as

$$N\hat{\sigma} = r \tag{8}$$

with normal matrix $N = (A_{\mathrm{vh}}^T W_{\mathrm{vh}} A_{\mathrm{vh}})$ and right-hand side $r = A_{\mathrm{vh}}^T W_{\mathrm{vh}} y_{\mathrm{vh}}$. Then the entries n_{kl} of the normal matrix N and the entries r_k of the right-hand side vector r, are given as

$$n_{kl} = \mathrm{vec}(B^T Q_k B)^T [D^{+T} W_{\mathrm{vh}} D^+] \mathrm{vec}(B^T Q_l B)$$
$$= \sum_{i=1}^p \sum_{j=1}^p c_i^T B^T Q_k B [D^{+T} W_{\mathrm{vh}} D^+]_{ij} B^T Q_l B c_j \tag{9}$$

and

$$r_k = \mathrm{vec}(B^T Q_k B)^T [D^{+T} W_{\mathrm{vh}} D^+] \mathrm{vec}(tt^T - B^T Q_0 B)$$
$$= \sum_{i=1}^p \sum_{j=1}^p c_i^T B^T Q_k$$
$$\times B[D^{+T} W_{\mathrm{vh}} D^+]_{ij}[t_j t - B^T Q_0 B c_j] \tag{10}$$

with D the duplication matrix, D^+ its pseudo-inverse, c_i the canonical unit vector having a one as its ith entry, and $D^{+T} W_{\mathrm{vh}} D^+ = ([D^{+T} W_{\mathrm{vh}} D^+]_{ij})$, $i, j = 1, \dots, b$ (for a definition of the Kronecker product and duplication matrix and their relations with the vec and vh operator, we refer to Appendix A).

Proof From $n_{kl} = \mathrm{vh}(B^T Q_k B)^T W_{\mathrm{vh}} \mathrm{vh}(B^T Q_l B)$ and $\mathrm{vh}(.) = D^+ \mathrm{vec}(.)$ the first part is trivial. One can now rewrite the $b^2 \times b^2$ matrix $D^{+T} W_{\mathrm{vh}} D^+$ as the sum $\sum_{i=1}^p \sum_{j=1}^p c_i c_j^T \otimes [D^{+T} W_{\mathrm{vh}} D^+]_{ij}$ with $[D^{+T} W_{\mathrm{vh}} D^+]_{ij}$ the appropriate $b \times b$ submatrices. Substitution of this expression gives $n_{kl} = \sum_{i=1}^p \sum_{j=1}^p \mathrm{tr}([D^{+T} W_{\mathrm{vh}} D^+]_{ij} B^T Q_l B c_j c_i^T B^T Q_k B) = \sum_{i=1}^p \sum_{j=1}^p c_i^T B^T Q_k B[D^{+T} W_{\mathrm{vh}} D^+]_{ij} B^T Q_l B c_j$. In a similar way, the right-hand side r_k is obtained. ∎

To see the weighted LS-VCE at work, consider the following two simple examples.

Example 1 (An unbiased estimator of the variance factor) Let $Q_{yy} = \sigma^2 Q$, with matrix Q known and scalar σ^2 unknown. Thus we have one unknown variance component (the variance factor of unit weight): $Q_0 = 0$ and $p = 1$, with $\sigma_1 = \sigma^2$. As a weight matrix, we take $W_{\mathrm{vh}} = D^T D$, with D the duplication matrix. Then, since $[D^{+T} W_{\mathrm{vh}} D^+] = D^{+T} D^T D D^+$ and $DD^+ \mathrm{vec}(.) = \mathrm{vec}(.)$, the normal scalar reads $n = \mathrm{vec}(B^T Q B)^T [D^{+T} W_{\mathrm{vh}} D^+] \mathrm{vec}(B^T Q B) = \mathrm{tr}(B^T Q B B^T Q B)$ and the right-hand side as $r = \mathrm{vec}(B^T Q B)^T [D^{+T} W_{\mathrm{vh}} D^+] \mathrm{vec}(tt^T) = \mathrm{tr}(B^T Q B tt^T) = t^T B^T Q B t$. Hence, the weighted LS-VCE follows as

$$\hat{\sigma}^2 = \frac{r}{n} = \frac{t^T B^T Q B t}{\mathrm{tr}(B^T Q B B^T Q B)} \tag{11}$$

Note that this is an unbiased estimator of σ^2. ∎

Example 2 (A natural variance matrix estimator) Let y be a zero mean random m-vector, with variance matrix Q_{yy} of which all the entries are unknown. Hence, matrix A of Eq. (1) is zero, and therefore $B = I$ and $t = y$. There are $p = \frac{1}{2}m(m + 1)$ unknown (co)variance components and as many observation equations in $\mathsf{E}\{y_{\mathrm{vh}}\} = A_{\mathrm{vh}}\sigma$. Thus $\hat{\sigma} = A_{\mathrm{vh}}^{-1} y_{\mathrm{vh}} = A_{\mathrm{vh}}^{-1} \mathrm{vh}(yy^T)$, provided the square matrix A_{vh} is invertible.

To determine A_{vh}, we write the variance matrix of y in the unknown (co)variance components as $Q_{yy} = \sum_{k=1}^p \sigma_k Q_k$, where the cofactor matrices Q_k are of form $c_i c_i^T$, for the variances, or of the form $c_i c_j^T + c_j c_i^T$, for the covariances, with c_i being the canonical unit vector having a one as its ith entry. From this, it follows that $A_{\mathrm{vh}} = I_{m(m+1)/2}$. We therefore have $\hat{\sigma} = A_{\mathrm{vh}}^{-1} \mathrm{vh}(yy^T) = \mathrm{vh}(yy^T)$, and thus

$$\hat{Q}_{yy} = yy^T \tag{12}$$

This shows that, in the absence of redundancy in the stochastic model ($p = \frac{1}{2}m(m + 1)$), the LS-VCE reduces to the natural estimator of Q_{yy}. ∎

Since $\hat{\sigma}$ is a least-squares estimator, its mean and variance matrix follow from Eq. (7).

Corollary 2 (LS-VCE mean and variance) *The mean of the LS-VCE $\hat{\sigma}$ is given as $\mathsf{E}\{\hat{\sigma}\} = \sigma$ and its variance matrix is*

$$Q_{\hat{\sigma}\hat{\sigma}} = N^{-1} M N^{-1} \tag{13}$$

with $M = A_{vh}^T W_{vh} Q_{vh} W_{vh} A_{vh}$ and Q_{vh} the variance matrix of y_{vh}.

Proof $E\{\hat{\sigma}\} = \sigma$ follows from taking the expectation of Eq. (7), knowing that $E\{y_{vh}\} = A_{vh}\sigma$. The variance matrix follows from applying the variance propagation law to Eq. (7).

This result shows that $\hat{\sigma}$ is an unbiased estimator for any choice of the weight matrix W_{vh}. This property is a direct consequence of the fact that $\hat{\sigma}$ is a linear least-squares estimator of σ with respect to the linear model $E\{y_{vh}\} = A_{vh}\sigma$. The LS-VCE $\hat{\sigma}$ is linear in y_{vh} and therefore a quadratic function of y. It is now easily verified that $\hat{\sigma}$ is a quadratic unbiased estimator of σ with respect to the model of Eqs. (1) and (2). For the conditions that a quadratic unbiased estimator needs to fulfill, we refer to Rao and Kleffe (1988).

The above corollary also gives us readily a precision description of the (co)variance component estimator $\hat{\sigma}$. In order to be able to evaluate this expression, we need to know the variance matrix of y_{vh} and thus of $vh(tt^T)$. Note, however, that both the property of unbiasedness as well as the variance expression of Eq. (13), hold true for any probability distribution that y_{vh} or t might have.

Equation (13) shows how the precision of $\hat{\sigma}$ changes for different choices of the weight matrix W_{vh}. From least-squares theory, we know that the least-squares estimator becomes identical to the BLUE if the weight matrix is chosen as the inverse of the variance matrix of the observables. In that case, the estimator will have the smallest possible variance. This property can be directly applied to our weighted LS-VCE.

Corollary 3 (LS-VCE as minimum variance estimator) *Let*

$$E\{y_{vh}\} = A_{vh}\sigma, \quad D\{y_{vh}\} = Q_{vh} \tag{14}$$

Then

$$\hat{\sigma} = (A_{vh}^T Q_{vh}^{-1} A_{vh})^{-1} A_{vh}^T Q_{vh}^{-1} y_{vh} \tag{15}$$

with variance matrix

$$Q_{\hat{\sigma}\hat{\sigma}} = (A_{vh}^T Q_{vh}^{-1} A_{vh})^{-1} \tag{16}$$

is the BLUE of σ.

Proof Follows from standard least-squares theory.

The minimum variance property is attained in the class of linear unbiased functions of y_{vh}. Thus, since y_{vh} is quadratic in t, the minimum variance property is attained in the class of quadratic unbiased functions of t. The minimum variance property is independent of the particular distribution that y_{vh} or t might have. In Sect. 5, we will consider the class of elliptically contoured distributions, of which the normal distribution is a special case.

Example 3 (A minimum variance estimator of the variance factor) Let $E\{t\} = E\{B^T y\} = E\{y\} = 0$. Thus, $B = I$ and $A = 0$. Further let the m entries of y be independent and identically distributed. The distribution, however, is assumed to be unknown. Then $Q_{yy} = \sigma^2 I$, with σ^2 being the unknown variance component. We will now apply Corollary 3 and determine the minimum variance estimator of σ^2.

First, we need to determine the variance matrix Q_{vh}. The only non-zero central moments (up to degree four) are given by

$$E\{y_i^2\} = \mu_2, \quad E\{y_i^4\} = \mu_4, \quad E\{y_i^2 y_j^2\} = \begin{cases} \mu_4 & \text{if } i = j \\ \mu_2^2 & \text{if } i \neq j \end{cases} \tag{17}$$

From this, it follows that Q_{vh} is given by the diagonal matrix

$$Q_{vh} = \text{diag}(\mu_4 - \mu_2^2, D_{m-1}, \mu_4 - \mu_2^2, D_{m-2}, \ldots, \mu_4 - \mu_2^2) \tag{18}$$

with the $i \times i$ diagonal matrix $D_i = \text{diag}(\mu_2^2, \ldots, \mu_2^2)$.

Furthermore we have, since $B = I$ and $Q_{yy} = \sigma^2 I$, that $A_{vh} = vh(I)$. Hence, the elements of the normal equation are

$$N = n_{11} = \frac{m}{\mu_4 - \mu_2^2}, \quad r = r_1 = \frac{y^T y}{\mu_4 - \mu_2^2} \tag{19}$$

from which the minimum variance LS-VCE of σ^2 follows as

$$\hat{\sigma}^2 = \frac{y^T y}{m} \tag{20}$$

Note that we did not need to specify the distribution of the y_i to determine this estimator. The variance of the estimator is given as $\sigma_{\hat{\sigma}^2}^2 = N^{-1} = n_{11}^{-1}$, for which we need the second and fourth central moments. □

3 Application of least-squares theory to LS-VCE

In this section, we show some of the attractive features of the LS-VCE method. We illustrate 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 VCE problem.

3.1 Measures of inconsistency

Existing methods of VCE provide means of estimating σ, but do not generally provide the user with measures to infer how well the data supports the chosen (co)variance component model. Since our approach is based on the least-squares principle, such measures of inconsistency, or lack of model fit, are readily available for the LS-VCE.

One such measure is given by the weighted sum of squared least-squares residuals, $\hat{e}_{vh}^T W_{vh} \hat{e}_{vh}$, in which

$\hat{e}_{vh} = y_{vh} - A_{vh}\hat{\sigma} = P_{A_{vh}}^{\perp} y_{vh}$, with the orthogonal projector $P_{A_{vh}}^{\perp} = I - A_{vh}(A_{vh}^T W_{vh} A_{vh})^{-1} A_{vh}^T W_{vh}$. For an arbitrary weight matrix W_{vh}, the quadratic form of residuals works out as

$$\hat{e}_{vh}^T W_{vh} \hat{e}_{vh} = \text{vh}(tt^T - B^T Q_0 B)^T$$
$$\times [W_{vh} - W_{vh} A_{vh}(A_{vh}^T W_{vh} A_{vh})^{-1} A_{vh}^T W_{vh}]$$
$$\times \text{vh}(tt^T - B^T Q_0 B) \qquad (21)$$

By using the analogy with hypothesis testing in linear models, other statistics can be derived. In order to be able to judge the significance of the inconsistency, we need to have some information on the distributional properties of the statistics. This is a nontrivial problem, because of the complicated nature in which they depend on t. In some special cases, depending on the distribution of t and the structure of B and Q_{yy}, it is possible to determine the exact distribution of Eq. (21), see e.g., Kuhri et al. (1998). In general, however, one will have to rely on alternative computer-based techniques, such as Monte Carlo simulation or bootstrapping, e.g., Efron and Tibshirani (1993).

3.2 The use of a-priori variance component information

In many applications, one already has some fair idea about the variance matrix Q_{yy} or parts thereof. In this case, one often would like the a-priori information to be included and weighted accordingly. The LS-VCE method is easily adapted to accommodate such a-priori information. Here, one can work with the extended linear (co)variance component model,

$$\mathsf{E}\left\{\begin{bmatrix} y_{vh} \\ \sigma_0 \end{bmatrix}\right\} = \begin{bmatrix} A_{vh} \\ C_0^T \end{bmatrix}\sigma \qquad (22)$$

Thus, the original model is then simply extended with additional (co)variance component observation equations. Let $\hat{\sigma}$ and $\hat{\hat{\sigma}}$ be the LS-VCE of σ based on the original and extended observation equations, respectively, and let $N = A_{vh} W_{vh} A_{vh}$. The weighted LS-VCE $\hat{\hat{\sigma}}$ of σ follows then as

$$\hat{\hat{\sigma}} = \hat{\sigma} + N^{-1} C_0 (C_0^T N^{-1} C_0 + W_0^{-1})^{-1}(\sigma_0 - C_0^T \hat{\sigma}) \quad (23)$$

in which W_0 is the weight matrix for σ_0.

Note that the contribution of the a-priori information is described by the second term on the right-hand side of Eq. (23). Also note that if the weight matrix is set such that $W_0^{-1} = 0$, then the solution is obtained that corresponds to using the hard constraints $\sigma_0 = C_0^T \sigma$. Finally note that if one has a-priori information on all the (co)variance components, one can set the design matrix such that $C_0 = I_p$, in which case Eq. (23) simplifies to

$$\hat{\hat{\sigma}} = \hat{\sigma} + N^{-1}(N^{-1} + W_0^{-1})^{-1}(\sigma_0 - \hat{\sigma})$$
$$= \sigma_0 + (N + W_0)^{-1} N(\hat{\sigma} - \sigma_0) \qquad (24)$$

3.3 Estimability of the variance components

So far, the design matrix $A_{vh} = [\text{vh}(B^T Q_1 B), \dots, \text{vh}(B^T Q_p B)]$ of the variance component model was assumed to be of full column rank, see Eq. (6). The full column rank of A_{vh} or the possible lack thereof, touches the estimability problem of variance components.

Let us first recall the definition of estimability, phrased with respect to the linear model of Eq. (6): a linear function of the variance components, say $f^T \sigma$, is said to be unbiased estimable if a linear function of the data vector exists, say $l^T y_{vh}$, such that $f^T \sigma = \mathsf{E}\{l^T y_{vh}\}$ for all σ. From this definition, it follows that $f^T \sigma$ is unbiased estimable if and only if f lies in the row space of A_{vh}, i.e., if an l exists such that $f = A_{vh}^T l$. It will be clear that this condition is satisfied for any f if A_{vh} has full column rank, $\text{rank}(A_{vh}) = p$. In this case, every linear function of σ is unbiased estimable. For this to be true, the column vectors $\text{vh}(B^T Q_k B)$ of A_{vh} need to be linear independent.

This shows that the linear independence of the cofactor matrices Q_k, $k = 1, \dots, p$, in the variance matrix $Q_{yy} = Q_0 + \sum_{k=1}^{p} \sigma Q_k$, is a necessary condition for estimability. It is, however, not a sufficient condition. To ensure estimability, one also has to take the interaction between matrix B and the cofactor matrices into account. Since B is a basis matrix of the null space of A^T ($A^T B = 0$), it is here where the interaction between the functional model $\mathsf{E}\{y\} = Ax$ and the stochastic model $\mathsf{D}\{y\} = Q_{yy}$ takes place, see Eq. (1).

Let us now assume that A_{vh} fails to be of full column rank and, thus, that not all variance components are unbiased estimable. In this case, one can follow the same approach as is used when solving rank-deficient least-squares problems. That is, for solving the rank-deficient system of normal equations $(A_{vh}^T W_{vh} A_{vh})\hat{\sigma} = A_{vh}^T W_{vh} y_{vh}$, one can either apply the theory of generalized inverses (e.g., Rao and Mitra, 1971), or one can apply the theory of S-transformations (Baarda, 1973; Teunissen, 1985). As a result one will obtain a solution that can be interpreted again as being an estimable linear function of σ.

The simplest route using the theory of S-transformations would be to eliminate as many columns of A_{vh} as is necessary to eliminate the rank deficiency. Thus, if $\text{rank}(A_{vh}) = p - q$, then q columns are eliminated, such that the resulting $(p - q)$ number of normal equations have a unique solution again. If the entries of σ are ordered such that they are the last q columns of A_{vh} that are eliminated, then the solution of the normal equations is the same as the solution one would obtain when solving the rank-deficient least-squares problem with the minimal constraints $(0, I_q)\sigma = 0$. Hence, the LS-VCE of Q_{yy} is then given as

$$\hat{Q}_{yy} = Q_0 + \sum_{k=1}^{p-q} \hat{\sigma}_k Q_k \qquad (25)$$

A different, but equivalent, LS-VCE of Q_{yy} is obtained if another set of q minimal constraints is chosen. These different, but equivalent, solutions can be transformed to one another by means of an S-transformation.

It is not always clear-cut whether A_{vh} is rank-deficient or not. Matrix A_{vh} could theoretically be of full rank, but instead show a numerical near rank-deficiency. Such near rank-deficiencies are indicators for poor estimability. They can be traced by considering the singular value decomposition of A_{vh} or the eigenvalue decomposition of the normal matrix $N = A_{vh}^T W_{vh} A_{vh}$. The eigenvectors corresponding to the small singular values or the small eigenvalues are then the linear functions of σ, which are prone to be poorly estimable.

The presence of small singular values or small eigenvalues will inflate the entries of N^{-1} and therefore also the entries of the LS-VCE variance matrix $Q_{\hat{\sigma}\hat{\sigma}}$. To avoid such an inflation, one can now follow the same route as one is used to when solving least-squares problems with near rank-deficiencies. For a discussion of such least-squares regularization procedures, we refer to, e.g., Hansen (1998).

3.4 Nonlinear VCE

The LS-VCE method also lends itself to solving nonlinear VCE problems. Such problems occur, for instance, in time-series analysis when determining the parameters of nonlinear covariance functions.

Let the variance matrix of y be given as $Q_{yy} = Q(\sigma)$, in which the entries of Q are nonlinear functions of the unknown components of σ. The LS-VCE of σ is then defined as

$$\hat{\sigma} = \arg \min_{\sigma} [y_{vh} - A_{vh}(\sigma)]^T W_{vh} [y_{vh} - A_{vh}(\sigma)] \quad (26)$$

with $A_{vh}(\sigma) = vh(B^T Q(\sigma) B)$. This nonlinear least-squares problem can be solved by means of the Gauss–Newton iteration method. It amounts to solving a linear least-squares problem in each iteration step. Let σ^i be an approximation of σ and let $A_{vh}(\sigma)$ be approximated as

$$A_{vh}(\sigma) \approx A_{vh}(\sigma^i) + \partial_\sigma A_{vh}(\sigma^i) \Delta \sigma^i \quad (27)$$

with $A_{vh}(\sigma^i) = vh(B^T Q(\sigma^i) B)$, $\partial_\sigma A_{vh}(\sigma^i) = vh(B^T \partial_\sigma Q(\sigma^i) B)$ and $\Delta \sigma^i = \sigma - \sigma^i$. Then the Gauss–Newton iteration, which starts with an initial value σ^0, is given as

$$\sigma^{i+1} = \sigma^i + \arg \min_{\Delta \sigma^i} [\Delta y_{vh}^i - \partial_\sigma A_{vh}(\sigma^i) \Delta \sigma^i]^T W_{vh} \quad (28)$$

$$\times [\Delta y_{vh}^i - \partial_\sigma A_{vh}(\sigma^i) \Delta \sigma^i] \quad (29)$$

for $i = 0, \ldots$, where $\Delta y_{vh}^i = y_{vh} - A_{vh}(\sigma^i)$. The Gauss–Newton iteration is known to have a linear rate of convergence. Its speed of convergence is dictated by the normal curvature of the nonlinear manifold $A_{vh}(\sigma)$, see Teunissen (1990).

3.5 Robust estimation

The application of robust estimation techniques to the linear model of Eq. (6) can be important if one needs to be guarded against outlying data. Various approaches exist for implementing robust estimation. One is again based on the least-squares principle, but now with an adaptable weight matrix, which downplays the influence of outlying data, see e.g. Myers (1986). One may also think of an L_1 norm minimization problem for which we refer to e.g. Amiri-Simkooei (2003).

Example 4 (A robust estimation of variance factor) In Example 3, one can apply the idea of L_1 norm minimization problem. The stochastic model expressed in terms of a model of observation equations reads

$$\mathsf{E}\{y_{vh}\} = vh(I)\sigma^2, \quad W_{vh} = I \quad (30)$$

The L_1 norm principle states that the sum of the absolute values of the 'residual' vector $vh(yy^T) - vh(I)\sigma^2$ should be minimized. Because $\frac{m(m-1)}{2}$ elements of $vh(I)$ are zeros, they do not affect the minimization problem. Therefore, the L_1 norm minimization problem is reduced to: $\sum_{i=1}^m |y_i^2 - \sigma^2| \to \min$.

This is very similar to the problem of measuring one unknown quantity, where the observables are y_i^2, $i = 1, \ldots, m$. It is 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 $y_1^2, y_2^2, \ldots, y_m^2$, namely

$$\breve{\sigma}^2 = y_{(m+1)/2}^2 \quad (31)$$

whereas the least-squares gives the *mean* of $y_1^2, y_2^2, \ldots, y_m^2$ (see Example 3). □

3.6 Non-negative VCE

To ensure that the estimated variances are non-negative, one can incorporate this information by adding non-negativity constraints $\sigma \geq 0$ to the linear model of Eq. (6). Such constraints will, of course, affect the unbiasedness of the variance component estimator. On the other hand, without these constraints, one runs the risk, in particular when the estimators lack good precision, to obtain estimates that are negative.

When the constraints are included, the LS-VCE becomes a least-squares problem with inequality constraints. Such least-squares problems are discussed in, e.g., Lawson and Hanson (1974); Rao and Toutenburg (1995). An alternative approach, suggested by Teunissen (1988), is to reparametrize the model such that non-negativity of the variance components is automatically ensured. For instance, if σ_k is such an unknown variance, one can reparametrize it as

$$\sigma_k = \exp(s_k) \quad (32)$$

and solve for s_k, instead of for σ_k. The solution $\hat{\sigma}_k = \exp(\hat{s}_k)$ is then non-negative by definition. This approach turns a linear LS-VCE problem into a nonlinear LS-VCE problem.

4 LS-VCE for a special class of weight matrices

In this and the next section, we consider the LS-VCE method for some important special cases. In the present section, we consider a special class of weight matrices and show how the weighted LS-VCE specializes accordingly. As an example, we show that the unbiased empirical or sample covariance function is a LS-VCE for a particular choice of weight matrix.

4.1 A special class of weight matrices

Consider the entries of the LS-VCE normal matrix, $n_{kl} = \mathrm{vec}(B^T Q_k B)^T [D^{+T} W_{\mathrm{vh}} D^+] \mathrm{vec}(B^T Q_l B)$, see Eq. (9). Note that if the weight matrix W_{vh} is chosen as $W_{\mathrm{vh}} = D^T D$, then $n_{kl} = \mathrm{vec}(B^T Q_k B)^T [I_b \otimes I_b] \mathrm{vec}(B^T Q_l B)$, since $DD^+ \mathrm{vec}(.) = \mathrm{vec}(.)$. Hence, this choice corresponds to using the unit matrix as the weight matrix for $\mathrm{vec}(tt^T)$. We now define a class of weight matrices, which, by means of a linear transformation of t, can always be transformed back to the unit matrix case. This class of weight matrices will be called the *unit weight matrix class*.

Definition 1 (Unit weight matrix class) Let W_{vh}^t be the LS-VCE weight matrix of $\mathrm{vh}(tt^T)$. Then W_{vh}^t is said to be a member of the unit weight matrix class, if it can be expressed as

$$W_{\mathrm{vh}}^t = D^T (W_t \otimes W_t) D \qquad (33)$$

for some positive definite matrix W_t and with D the duplication matrix.

This definition is motivated as follows. Let W_{vh} be the weight matrix used for the data vector y_{vh}. Then, in order to obtain the same LS-VCE solution, matrix $T^{-T} W_{\mathrm{vh}} T^{-1}$ should be used as weight matrix, if $T y_{\mathrm{vh}}$ (with matrix T invertible) is used as data vector. Let us now assume that t is transformed into s using the invertible $b \times b$ matrix R, i.e., $s = Rt$ or $t = R^{-1}s$. Then $\mathrm{vec}(tt^T) = \mathrm{vec}(R^{-1}ss^T R^{-T}) = (R^{-1} \otimes R^{-1})\mathrm{vec}(ss^T)$, or, since $\mathrm{vec}(.) = D\mathrm{vh}(.)$ and $\mathrm{vh}(.) = D^+ \mathrm{vec}(.)$, with D the duplication matrix (see Appendix A), $\mathrm{vh}(tt^T) = D^+(R^{-1} \otimes R^{-1})D\mathrm{vh}(ss^T) = T\mathrm{vh}(ss^T)$.

With $W_{\mathrm{vh}}^t = T^{-T} W_{\mathrm{vh}}^s T^{-1}$, this gives $W_{\mathrm{vh}}^t = D^T (R^T \otimes R^T)D^{+T} W_{\mathrm{vh}}^s D^+(R \otimes R)D$. Hence, if $W_{\mathrm{vh}}^s = D^T D$, it follows, since $DD^+(R \otimes R)D = (R \otimes R)D$, that

$$W_{\mathrm{vh}}^t = D^T (R^T \otimes R^T)(R \otimes R)D = D^T (R^T R \otimes R^T R)D \qquad (34)$$

Thus, since R in $s = Rt$ can be any invertible matrix, the matrix $W_t = R^T R$ can be any positive definite matrix.

4.2 The normal equations

We are now in the position to show how the LS-VCE works out for this special class of weight matrices.

Theorem 2 (LS-VCE for the unit weight matrix class) *Let the weight matrix of the LS-VCE $\hat{\sigma}$ be given by Eq. (33). The entries of the system of normal equations of Eq. (8) are then given as*

$$n_{kl} = \mathrm{tr}(B^T Q_k B W_t B^T Q_l B W_t), \quad k, l = 1, \ldots, p \qquad (35)$$

and

$$r_k = t^T W_t B^T Q_k B W_t t - \mathrm{tr}(B^T Q_k B W_t B^T Q_0 B W_t),$$
$$\times k = 1, \ldots, p \qquad (36)$$

respectively.

Proof Substitution of $D^{+T} W_{\mathrm{vh}} D^+ = D^{+T} D^T (W_t \otimes W_t) D D^+$ into Eq. (9), gives (since $DD^+ \mathrm{vec}(.) = \mathrm{vec}(.)$) $n_{kl} = \mathrm{vec}(B^T Q_k B)^T [W_t \otimes W_t] \mathrm{vec}(B^T Q_l B) = \mathrm{tr}(B^T Q_k B W_t B^T Q_l B W_t)$. Equation (36) can be derived in a similar manner. $\square$

The following example shows that the empirical or sample autocovariance function is a weighted LS-VCE, if the weight matrix is chosen from the above given unit weight matrix class.

Example 5 (The unbiased sample autocovariance 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 data points are available, we have, with $y = (y_1, \ldots, y_m)^T$, that

$$\mathsf{E}\{y\} = 0 \quad \text{and} \quad \mathsf{D}\{y\} = Q_{yy} = \sum_{\tau=0}^{m-1} \sigma_\tau Q_\tau \qquad (37)$$

with the $m \times m$ cofactor matrices

$$Q_\tau = a_\tau \sum_{i=1}^{m-\tau} (c_i c_{i+\tau}^T + c_{i+\tau} c_i^T) \qquad (38)$$

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 may now apply Theorem 2, in order to estimate $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_{m-1})^T$. In our case, we have $t = B^T y$ with $B = I$. We will use the diagonal matrix $W_t = W_y = \mathrm{diag}(w_1, \ldots, w_m)$ as the weight matrix. Substitution of $B = I$ and $W_t = W_y = \mathrm{diag}(w_1, \ldots, w_m)$ into Eq. (35), gives, with Eq. (38),

$$n_{\tau,\kappa} = \mathrm{tr}(Q_\tau W_y Q_\kappa W_y)$$
$$= \begin{cases} 2a_\tau \sum_{i=1}^{m-\tau} w_i w_{i+\tau} & \text{if } \tau = \kappa \\ 0 & \text{if } \tau \neq \kappa \end{cases} \qquad (39)$$

Hence, 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, $r_\tau = 2a_\tau \sum_{i=1}^{m-\tau} w_i w_{i+\tau} y_i y_{i+\tau}$. The LS-VCE solution for the unknown (co)variance components therefore follows as

$$\hat{\sigma}_\tau = \frac{\sum_{i=1}^{m-\tau} w_i w_{i+\tau} y_i y_{i+\tau}}{\sum_{i=1}^{m-\tau} w_i w_{i+\tau}}, \quad \tau = 0, 1, \ldots, m-1 \quad (40)$$

If $w_1 = \cdots = w_m = 1$, then

$$\hat{\sigma}_\tau = \frac{\sum_{i=1}^{m-\tau} y_i y_{i+\tau}}{m-\tau}, \quad \tau = 0, 1, \ldots, m-1 \quad (41)$$

which is known as the unbiased sample autocovariance function of a zero-mean stationary time-series y_i, see e.g. Wei (1990).

We have thus shown that the sample autocovariance function is a LS-VCE, if the weight matrix is chosen as the identity matrix. In the next section (see Example 6), we will determine the precision of this LS-VCE. □

The normal equations of Theorem 2 are expressed in the basis matrix B of the null space of A^T. They can, however, also be expressed in matrix A, instead of in matrix B. In order to show this, we first recall some elements from least-squares theory.

Let a partitioned linear model $\mathsf{E}\{y\} = A_1 x_1 + A_2 x_2$ be solved in a least-squares sense with weight matrix W. Now let B_2 be a basis matrix of the null space of A_2^T. Thus, $B_2^T A_2 = 0$. The solution for x_1 is then identical to the solution one would get when solving $\mathsf{E}\{B_2^T y\} = B_2^T A_1 x_1$ in a least-squares sense, but now with the weight matrix $W_2 = (B_2^T W^{-1} B_2)^{-1}$ (e.g., Teunissen, 2000). With the use of this analogy from least-squares theory, we can determine the following alternative expression for the normal equations of Eqs. (35) and (36).

Corollary 4 (an alternative expression) *If $W_{\mathrm{vh}}^y = D^T(W_y \otimes W_y)D$ is used as weight matrix for $\mathrm{vh}(yy^T)$, then the normal equations of Eqs. (35) and (36) can be expressed in y, W_y and A as*

$$n_{kl} = \mathrm{tr}(Q_k W_y P_A^\perp Q_l W_y P_A^\perp), \quad k, l = 1, \ldots, p \quad (42)$$

and

$$r_k = \hat{e}^T W_y Q_k W_y \hat{e} - \mathrm{tr}(Q_k W_y P_A^\perp Q_0 W_y P_A^\perp),$$
$$k = 1, \ldots, p \quad (43)$$

with $\hat{e} = P_A^\perp y$ and $P_A^\perp = I - A(A^T W_y A)^{-1} A^T W_y$.

Proof Since $\mathrm{vh}(tt^T) = D^+(B^T \otimes B^T)D\mathrm{vh}(yy^T)$, we let B_2^T in $W_2 = (B_2^T W^{-1} B_2)^{-1}$ be given as $B_2^T = D^+(B^T \otimes B^T)D$. Furthermore, we set $W_2 = W_{\mathrm{vh}}^t = D^T(W_t \otimes W_t)D$ and $W = W_{\mathrm{vh}}^y = D^T(W_y \otimes W_y)D$. It then follows upon substitution in $W_2 = (B_2^T W^{-1} B_2)^{-1}$, that $W_t = (B^T W_y^{-1} B)^{-1}$

and therefore that $BW_t B^T = B(B^T W_y^{-1} B)^{-1} B^T = W_y P_A^\perp$. Substitution of $BW_t B^T = W_y P_A^\perp$ into $n_{kl} = \mathrm{tr}(B^T Q_k BW_t B^T Q_l BW_t) = \mathrm{tr}(Q_k BW_t B^T Q_l BW_t B^T)$, gives the stated result of Eq. (42). Recognizing that $t = B^T y = B^T \hat{e}$, a similar derivation can be given for the right-hand side r_k. □

5 LS-VCE in case of elliptically contoured distributions

In this section, we consider the LS-VCE method for the case the probability distribution of the vector of observables y in the linear model of Eq. (1) is elliptically contoured. The class of elliptically contoured distributions includes the multivariate normal distribution as a special case. Since linear functions of elliptically distributed random vectors are again elliptically distributed (e.g., Gupta and Varga, 1993), we will work directly with $t = B^T y$. First, we will determine the variance matrix of $\mathrm{vh}(tt^T)$, then we will consider the different versions of the LS-VCE method.

5.1 The variance matrix of $\mathrm{vh}(tt^T)$

Definition 2 (elliptically contoured distribution) A random b vector t is said to have an elliptically contoured distribution if its PDF is of the form

$$f_t(t) = |\Lambda_t|^{-1/2} g((t - \mu_t)^T \Lambda_t^{-1}(t - \mu_t)), \quad t \in R^b \quad (44)$$

where $g : R \mapsto [0, \infty)$ is nonincreasing. This distribution is denoted as $E_b(\mu_t, \Lambda_t)$.

Several important distributions are known to belong to this family of distributions. The multivariate normal distribution can be shown to be a member of this family by choosing $g(x) = (2\pi)^{-n/2} \exp\{-(1/2)x\}$. Another member of this family is the multivariate Student distribution. Note that the mean of t is given by μ_t. The variance matrix of t can be obtained from the characteristic function, which is of the form $\phi(\theta) = \exp(i\theta^T \mu_t)\psi(\theta^T \Lambda_t \theta)$, for some function ψ, where $i = \sqrt{-1}$. The variance matrix is then given as $\mathsf{D}\{t\} = Q_{tt} = -2\psi'(0)\Lambda_t$.

Since $\mathrm{vh}(tt^T) = D^+\mathrm{vec}(tt^T)$, we can determine the variance matrix of $\mathrm{vh}(tt^T)$ from the variance matrix of $\mathrm{vec}(tt^T)$, the entries of which are given by

$$Q_{\mathrm{vec}}^{ijkl} = \mathsf{E}\{(t_i t_j - \mathsf{E}\{t_i t_j\})(t_k t_l - \mathsf{E}\{t_k t_l\})\}$$
$$= \mathsf{E}\{t_i t_j t_k t_l\} - \mathsf{E}\{t_i t_j\}\mathsf{E}\{t_k t_l\} \quad (45)$$
$$i, j, k, l = 1, 2, \ldots, b$$

Hence, we need the second- and fourth-order moments of the elliptically contoured distribution. They are given by the following Lemma.

Lemma 1 *Let* $t \sim E_b(0, \Lambda_t)$. *Then the odd-order moments of* t *are zero, and the second- and fourth-order moments are of the form*

$$
\begin{aligned}
\mathsf{E}\{t_i t_j\} &= q_{ij} \\
\mathsf{E}\{t_i t_j t_k t_l\} &= (\kappa + 1)(q_{ij} q_{kl} + q_{ik} q_{jl} + q_{jk} q_{il}) \\
&\quad i, j, k, l = 1, 2, \ldots, b
\end{aligned}
\tag{46}
$$

where q_{ij} *represents* Q_{tt} *in index notation, i.e.,* $q_{ij} = (Q_{tt})_{ij}$, *and* κ *is the kurtosis parameter.*

Proof For the proof, we refer to Berkane and Bentler (1986); Maruyama and Seo (2003).

Elliptical distributions are characterized by the kurtosis parameter κ, which can be used to describe the relation of central fourth-order to second-order moments of any random variable as $\mu_4 = 3(1 + \kappa)\mu_2^2$. For the multivariate normal distribution, we have $\kappa = 0$, and for the multivariate Student distribution with b degrees of freedom, the kurtosis parameter is given as $\kappa = \frac{2}{b-4}$.

With the use of Lemma 1, Eq. (45) can be written in terms of the kurtosis parameter κ and the entries of the variance matrix Q_{tt}, as

$$
Q_{\text{vec}}^{ijkl} = (\kappa + 1)(q_{ik} q_{jl} + q_{jk} q_{il}) + \kappa q_{ij} q_{kl}
\tag{47}
$$

With this result, we are able to determine the variance matrix of $\text{vh}(tt^T)$ and its inverse.

Theorem 3 (variance matrix of $\text{vh}(tt^T)$) *Let the* b *vector* t *have an elliptically contoured distribution with zero mean and variance matrix* Q_{tt}. *The variance matrix of* $\text{vh}(tt^T)$, *and its inverse, are then given as*

$$
\begin{aligned}
Q_{\text{vh}} = {}&2(\kappa + 1) D^+ (Q_{tt} \otimes Q_{tt}) D^{+T} \\
&+ \kappa D^+ \text{vec}(Q_{tt}) \text{vec}(Q_{tt})^T D^{+T}
\end{aligned}
\tag{48}
$$

and

$$
\begin{aligned}
Q_{\text{vh}}^{-1} = \frac{1}{2(\kappa + 1)} \Big[&D^T (Q_{tt}^{-1} \otimes Q_{tt}^{-1}) D \\
&- \kappa' \, D^T \text{vec}(Q_{tt}^{-1}) \text{vec}(Q_{tt}^{-1})^T D \Big]
\end{aligned}
\tag{49}
$$

with

$$
\kappa' = \frac{\kappa}{2(\kappa + 1) + \kappa b}
\tag{50}
$$

Proof see Appendix B. □

Note, in case t is normally distributed ($\kappa = 0$), that the inverse of the variance matrix is a member of the special class of weight matrices considered earlier, with $W_t = \frac{1}{\sqrt{2}} Q_{tt}^{-1}$. This is not true, however, when $\kappa \neq 0$.

5.2 The weighted LS-VCE

With the variance matrix of $\text{vh}(tt^T)$ available, we are now in the position to evaluate the precision of the weighted LS-VCE

in case of an elliptically contoured distribution. The variance matrix $Q_{\hat{\sigma}\hat{\sigma}}$ follows from substituting Q_{vh} of Eq. (48) into Eq. (13). This variance matrix can be worked out further for the case the weight matrix is a member of the unit weight matrix class.

Corollary 5 (precision of (co)variance components) *Let the zero mean random* b *vector* $t = B^T y$ *have an elliptically contoured distribution with variance matrix* $Q_{tt} = B^T Q_{yy} B$, *and let the weight matrix of the LS-VCE* $\hat{\sigma}$ *be given by* $W_{\text{vh}} = D^T (W_t \otimes W_t) D$, *with* $W_t = (B^T W_y^{-1} B)^{-1}$. *Then the variance matrix of* $\hat{\sigma}$ *is*

$$
Q_{\hat{\sigma}\hat{\sigma}} = N^{-1} M N^{-1}
\tag{51}
$$

where the entries of N *and* M *are*

$$
n_{kl} = \text{tr}(B^T Q_k B W_t B^T Q_l B W_t) = \text{tr}(Q_k W_y P_A^\perp Q_l W_y P_A^\perp)
\tag{52}
$$

and

$$
\begin{aligned}
m_{kl} = {}&2(\kappa + 1)\text{tr}(B^T Q_k B W_t Q_{tt} W_t B^T Q_l B W_t Q_{tt} W_t) \\
&+ \kappa \text{tr}(B^T Q_k B W_t Q_{tt} W_t)\text{tr}(B^T Q_l B W_t Q_{tt} W_t) \\
= {}&2(\kappa+1)\text{tr}(Q_k W_y P_A^\perp Q_{yy} W_y P_A^\perp Q_l W_y P_A^\perp Q_{yy} W_y P_A^\perp) \\
&+ \kappa \text{tr}(Q_k W_y P_A^\perp Q_{yy} W_y P_A^\perp)\text{tr}(Q_l W_y P_A^\perp Q_{yy} W_y P_A^\perp)
\end{aligned}
\tag{53}
$$

with $P_A^\perp = I - A(A^T W_y A)^{-1} A^T W_y = W_y^{-1} B(B^T W_y^{-1} B)^{-1} B^T$.

Proof Expressions $Q_{\hat{\sigma}\hat{\sigma}} = N^{-1} M N^{-1}$ and $n_{kl} = \text{tr}(B^T Q_k B W_t B^T Q_l B W_t) = \text{tr}(Q_k W_y P_A^\perp Q_l W_y P_A^\perp)$ are trivial. Matrix M is given as $M = A_{\text{vh}}^T W_{\text{vh}} Q_{\text{vh}} W_{\text{vh}} A_{\text{vh}} = m_{kl} = \text{vh}(B^T Q_k B)^T W_{\text{vh}} Q_{\text{vh}} W_{\text{vh}} \text{vh}(B^T Q_l B)$, where $W_{\text{vh}} = D^T (W_t \otimes W_t) D$ and Q_{vh} for an elliptical distribution is given by Eq. (48). Substitution of these terms, with $\text{vh}(.) = D^+ \text{vec}(.)$ and $DD^+ \text{vec}(.) = \text{vec}(.)$, gives the first part of the expression in terms of B. In order to obtain the alternative expression in terms of W_y, Q_{yy} and A, we substitute $Q_{tt} = B^T Q_{yy} B$ and then $B W_t B^T = W_y P_A^\perp$. □

To evaluate the precision in case y, and thus t, are normally distributed, we have to set $\kappa = 0$. For a Student distribution with b degrees of freedom, we have $\kappa = \frac{2}{b-4}$. Hence, if the degrees of freedom is sufficiently large, one may use $\kappa = 0$ as an approximation.

In the following example, we determine the precision of the sample autocovariance function.

Example 6 (Example 5 continued) To determine the precision of the sample autocovariance function, we will assume that y is normally distributed ($\kappa = 0$). The entries of matrix N in the expression of the variance matrix $Q_{\hat{\sigma}\hat{\sigma}} = N^{-1} M N^{-1}$ of $\hat{\sigma} = (\hat{\sigma}_0, \ldots, \hat{\sigma}_{m-1})^T$, are given

in Eq. (39). To determine the entries of matrix M, we substitute $\kappa = 0$, $W_t = W_y = \text{diag}(w_1, \ldots, w_m)$ and $B = I$ (or $P_A^\perp = I$) into Eq. (53). This gives

$$m_{\tau\kappa} = 2\text{tr}(Q_\tau W_y Q_{yy} W_y Q_\kappa W_y Q_{yy} W_y) \qquad (54)$$

Hence, the entries of the variance matrix are given as

$$(Q_{\hat\sigma\hat\sigma})_{\tau\kappa} = \frac{\text{tr}(Q_\tau W_y Q_{yy} W_y Q_\kappa W_y Q_{yy} W_y)}{2a_\tau^2(\sum_{i=1}^{m-\tau} w_i w_{i+\tau})^2} \qquad (55)$$

This expression is easily evaluated numerically.

A simple, but approximate, analytical expression can be obtained, if we approximate the variance matrix of y by $Q_{yy} \approx \sigma_0 I$. This is a good approximation, the closer the randomness of the time-series resembles that of white noise. With this approximation, the variance matrix of Eq. (55) becomes a diagonal matrix, with entries

$$\sigma_{\hat\sigma_\tau}^2 = \frac{\sigma_0^2 \sum_{i=1}^{m-\tau} w_i^2 w_{i+\tau}^2}{a_\tau(\sum_{i=1}^{m-\tau} w_i w_{i+\tau})^2}, \quad \tau = 0, 1, \ldots, m-1 \qquad (56)$$

If $w_1 = \cdots = w_m = 1$, then

$$\sigma_{\hat\sigma_\tau}^2 = \frac{\sigma^4}{a_\tau(m-\tau)}, \quad \tau = 0, 1, \ldots, m-1 \qquad (57)$$

with $\sigma^4 = \sigma_0^2$. This shows that the precision of the autocovariance function gets poorer with increasing time-lag τ. This also makes sense, since less data are used when τ increases. □

5.3 The minimum variance VCE

We have seen that the LS-VCE becomes a minimum variance estimator if the weight matrix is chosen as the inverse of the variance matrix, $W_{\text{vh}} = Q_{\text{vh}}^{-1}$. With Eq. (49), we are now in the position to determine the minimum variance estimator of σ in case of elliptically contoured distributions.

Theorem 4 (minimum variance estimator) *Let the zero mean random vector $t = B^T y$ have an elliptically contoured distribution, with variance matrix $Q_{tt} = B^T Q_{yy} B$. The system of normal equations of the minimum variance LS-VCE $\hat\sigma$ is then given by*

$$N\hat\sigma = r \qquad (58)$$

where the entries of N and r are

$$n_{kl} = \frac{\text{tr}(Q_k Q_{yy}^{-1} P_A^\perp Q_l Q_{yy}^{-1} P_A^\perp)}{2(\kappa+1)}$$
$$\times \left[1 - \frac{\kappa'\text{tr}(Q_k Q_{yy}^{-1} P_A^\perp)\text{tr}(Q_l Q_{yy}^{-1} P_A^\perp)}{\text{tr}(Q_k Q_{yy}^{-1} P_A^\perp Q_l Q_{yy}^{-1} P_A^\perp)}\right] \qquad (59)$$

and

$$r_k = \frac{\hat e^T Q_{yy}^{-1} Q_k Q_{yy}^{-1} \hat e - \text{tr}(Q_k Q_{yy}^{-1} P_A^\perp Q_0 Q_{yy}^{-1} P_A^\perp)}{2(\kappa+1)}$$
$$\times \left[1 - \frac{\kappa'\text{tr}(Q_k Q_{yy}^{-1} P_A^\perp)[\hat e^T Q_{yy}^{-1} \hat e - \text{tr}(Q_0 Q_{yy}^{-1} P_A^\perp)]}{\hat e^T Q_{yy}^{-1} Q_k Q_{yy}^{-1} \hat e - \text{tr}(Q_k Q_{yy}^{-1} P_A^\perp Q_0 Q_{yy}^{-1} P_A^\perp)}\right] \qquad (60)$$

with $\hat e = P_A^\perp y = Q_{yy} B Q_{tt}^{-1} t$ and $P_A^\perp = I - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} = Q_{yy} B(B^T Q_{yy} B)^{-1} B^T$.

Proof To obtain the minimum variance estimator in case of elliptical distributions one can rely on Corollary 3, with Q_{vh}^{-1} given in Eq. (49). Substitution of the related terms, with $DD^+\text{vec}(.) = \text{vec}(.)$ gives

$$n_{kl} = \frac{1}{2(\kappa+1)}\text{vec}(B^T Q_k B)^T \left[Q_{tt}^{-1} \otimes Q_{tt}^{-1}\right.$$
$$\left. - \kappa'\text{vec}(Q_{tt}^{-1})\text{vec}(Q_{tt}^{-1})^T\right]\text{vec}(B^T Q_l B)$$
$$= \frac{\text{tr}(B^T Q_k B Q_{tt}^{-1} B^T Q_l B Q_{tt}^{-1})}{2(\kappa+1)}$$
$$\times \left[1 - \frac{\kappa'\text{tr}(B^T Q_k B Q_{tt}^{-1})\text{tr}(B^T Q_l B Q_{tt}^{-1})}{\text{tr}(B^T Q_k B Q_{tt}^{-1} B^T Q_l B Q_{tt}^{-1})}\right] \qquad (61)$$

In order to obtain the alternative expression in terms of Q_{yy} and A we use again $B Q_{tt}^{-1} B^T = Q_{yy}^{-1} P_A^\perp$ in Eq. (61). For the right-hand side $r = A_{\text{vh}}^T Q_{\text{vh}}^{-1} \text{vh}(tt^T - B^T Q_0 B)$, follow a similar procedure and use $t = B^T y = B^T \hat e$. □

Note that in this case, the variance matrix of $\hat\sigma$ is directly given by the inverse of the normal matrix, $Q_{\hat\sigma\hat\sigma} = N^{-1}$. Also note that if $\kappa = \kappa' = 0$, i.e. t is normally distributed, the structure of the above system of normal equations becomes identical to the structure of the system of normal equations of the weighted LS-VCE, with weight matrix $W_{\text{vh}} = D^T(W_t \otimes W_t)D$ (see Theorem 2 and Corollary 3). If the weight matrix is chosen such that $W_t = \frac{1}{\sqrt{2}}Q_{tt}^{-1}$ and $W_y = \frac{1}{\sqrt{2}}Q_{yy}^{-1}$, then Eqs. (35) and (36) become identical to Eqs. (59) and (60), respectively.

In the following example, a minimum variance estimator is given of the variance matrix of repeated measurements.

Example 7 (Variance matrix estimation from repeated experiments) Consider the following linear model

$$\mathsf{E}\{y_i\} = Ax_i, \quad \mathsf{D}\{y_i, y_j\} = \sigma_{ij}Q \quad i, j = 1, 2, \ldots, r \qquad (62)$$

It consists of r groups of observation equations. The y_i, $i = 1, \ldots, r$, are normally distributed m-vectors of observables and the x_i, $i = 1, \ldots, r$, are the n-vectors of unknowns. The $m \times n$ design matrix A is assumed known and of full column rank. In addition to the x_i, also the $r(r+1)/2$ (co)variance components σ_{ij} are assumed unknown. The $m \times m$ cofactor matrix Q is assumed known and positive definite. The goal is to determine the minimum variance LS-VCE solution for the σ_{ij}. Note, if $r = 1$, that the above model reduces to the standard linear model, with the variance factor of unit weight as the single unknown variance component.

If we define $Y = (y_1, \ldots, y_r)$, $x = (x_1^T, \ldots, x_r^T)^T$ and $(\Sigma)_{ij} = \sigma_{ij}$, we may write the above model, with the use of the Kronecker product, as

$$\mathsf{E}\{\text{vec}(Y)\} = (I \otimes A)x, \quad \mathsf{D}\{\text{vec}(Y)\} = \Sigma \otimes Q \qquad (63)$$

with

$$\Sigma = \sum_{k=1}^{r(r+1)/2} \sigma_k C_k \tag{64}$$

where σ_k is respectively σ_1^2, σ_{12}, σ_{13}, ..., σ_r^2, and $C_k = c_i c_i^T$, for $\sigma_k := \sigma_i^2$, and $C_k = c_i c_j^T + c_j c_i^T$, for $\sigma_k := \sigma_{ij}$ ($i \neq j$). Thus, Eq. (63) may also be written as

$$\mathsf{E}\{\mathrm{vec}(Y)\} = (I \otimes A)x,$$
$$\mathsf{D}\{\mathrm{vec}(Y)\} = \sum_{k=1}^{r(r+1)/2} \sigma_k (C_k \otimes Q) \tag{65}$$

We may now apply Theorem 4 (with $\kappa = \kappa' = 0$) to find the minimum variance estimator of $\sigma = (\sigma_1, \ldots, \sigma_{r(r+1)/2}) = \mathrm{vh}(\Sigma)$. We have $Q_{\mathrm{vec}(Y)\mathrm{vec}(Y)}^{-1} = \Sigma^{-1} \otimes Q^{-1}$, $Q_k = C_k \otimes Q$ and $\hat{e} = P_{I_r \otimes A}^{\perp} \mathrm{vec}(Y) = (I_r \otimes P_A^{\perp})\mathrm{vec}(Y) = \mathrm{vec}(P_A^{\perp}Y)$, with $P_A^{\perp} = I_m - A(A^T Q^{-1} A)^{-1} A^T Q^{-1}$. Substitution into $r_k = \frac{1}{2}\hat{e}^T Q_{\mathrm{vec}(Y)\mathrm{vec}(Y)}^{-1} Q_k Q_{\mathrm{vec}(Y)\mathrm{vec}(Y)}^{-1} \hat{e}$, gives, with $\hat{E} = P_A^{\perp}Y$, for the entries of the right-hand side vector

$$\begin{aligned} r_k &= \frac{1}{2}\mathrm{vec}(\hat{E})^T [\Sigma^{-1} C_k \Sigma^{-1} \otimes Q^{-1}]\mathrm{vec}(\hat{E}) \\ &= \frac{1}{2}\mathrm{tr}(\Sigma^{-1} C_k \Sigma^{-1} \hat{E}^T Q^{-1} \hat{E}) \\ &= \frac{1}{2}\mathrm{vec}(C_k)^T [\Sigma^{-1} \otimes \Sigma^{-1}]\mathrm{vec}(\hat{E}^T Q^{-1} \hat{E}) \\ &= \frac{1}{2}\mathrm{vh}(C_k)^T [D^T (\Sigma^{-1} \otimes \Sigma^{-1})D]\mathrm{vh}(\hat{E}^T Q^{-1} \hat{E}) \end{aligned} \tag{66}$$

and thus

$$r = \frac{1}{2}[D^T (\Sigma^{-1} \otimes \Sigma^{-1})D]\mathrm{vh}(\hat{E}^T Q^{-1} \hat{E}) \tag{67}$$

since $\mathrm{vh}(C_k)$ is the canonical unit vector having the one as its kth entry.

In a similar way we find for the entries of the normal matrix,

$$\begin{aligned} n_{kl} &= \frac{1}{2}\mathrm{tr}(C_k \Sigma^{-1} C_l \Sigma^{-1} \otimes P_A^{\perp}) \\ &= \frac{1}{2}\mathrm{tr}(C_k \Sigma^{-1} C_l \Sigma^{-1})\mathrm{tr}(P_A^{\perp}) \\ &= \frac{1}{2}\mathrm{vh}(C_k)^T [D^T (\Sigma^{-1} \otimes \Sigma^{-1})D]\mathrm{vh}(C_l)\,\mathrm{tr}(P_A^{\perp}) \end{aligned} \tag{68}$$

and thus

$$N = \frac{m-n}{2}D^T (\Sigma^{-1} \otimes \Sigma^{-1})D \tag{69}$$

since $\mathrm{tr}(P_A^{\perp}) = \mathrm{rank}(P_A^{\perp}) = m-n$. With Eqs. (67) and (69), we obtain the minimum variance LS-VCE $\hat{\Sigma}$ from $\hat{\sigma} = N^{-1}r$ and $\hat{\sigma} = \mathrm{vh}(\hat{\Sigma})$, as

$$\hat{\Sigma} = \frac{\hat{E}^T Q^{-1} \hat{E}}{m-n} \tag{70}$$

Note that this result is a generalized form of the familiar estimator for the variance factor of unit weight.

We can also determine the precision of the entries of $\hat{\Sigma}$. Since the variance matrix of $\hat{\sigma} = \mathrm{vh}(\hat{\Sigma})$ is given by the inverse of the normal matrix, it follows from Eq. (69) that

$$Q_{\hat{\sigma}\hat{\sigma}} = N^{-1} = \frac{2}{m-n}D^+ (\Sigma \otimes \Sigma)D^{+T} \tag{71}$$

From this result, the variances and covariances of the estimated (co)variance components follow as

$$\sigma_{\hat{\sigma}_{ij}\hat{\sigma}_{kl}} = \frac{\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}}{m-n}; \quad \sigma_{\hat{\sigma}_{ij}}^2 = \frac{\sigma_i^2\sigma_j^2 + \sigma_{ij}^2}{m-n} \tag{72}$$

Thus, the variances of the variance components are given as

$$\sigma_{\hat{\sigma}_i^2}^2 = \frac{2\sigma_i^4}{m-n}, \quad i = 1, ..., r. \tag{73}$$

$\square$

6 The mean and the variance of the LS-VCE based estimator of x

In case of the linear model of Eq. (1), the BLUE of x is $\hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} y$. This estimator can not be used if Q_{yy} is unknown. In this section, we investigate what can be said about the mean and the variance of the parameter estimator when Q_{yy} is replaced by its estimator.

6.1 The mean

Once the (co)variance component estimator $\hat{\sigma}$ has been obtained, one can estimate Q_{yy} as

$$\hat{Q}_{yy} = Q_0 + \sum_{k=1}^{p} \hat{\sigma}_k Q_k \tag{74}$$

When $\hat{Q}_{yy}$ is used to replace Q_{yy} in the expression for the BLUE of x, we obtain the estimator

$$\hat{x}' = (A^T \hat{Q}_{yy}^{-1} A)^{-1} A^T \hat{Q}_{yy}^{-1} y \tag{75}$$

where the matrix $\hat{Q}_{yy}$ and the vector y are both random.

The question is then how this randomness propagates into the mean and variance of $\hat{x}'$. For this purpose, we first rewrite Eq. (75) into a more convenient form. With the use of the decomposition $y = A\hat{x} + Q_{yy}B(B^T Q_{yy}B)^{-1}t$, see Eq. (3), we may express $\hat{x}'$ as

$$\hat{x}' = \hat{x} + M(t)t \tag{76}$$

with

$$M(t) = (A^T \hat{Q}_{yy}^{-1} A)^{-1} A^T \hat{Q}_{yy}^{-1} Q_{yy}B(B^T Q_{yy}B)^{-1} \tag{77}$$

Note that since the entries of $\hat{Q}_{yy}$ are functions of the misclosure vector t, this also holds true for the entries of the matrix $M(t)$. Also note that since $A^T B = 0$, $M(t)$ will be small if $\hat{Q}_{yy}$ is a good approximation of Q_{yy}.

With the above decomposition, it is now possible to see when $\hat{x}'$ is unbiased or not. It follows from Eq. (76) that $\hat{x}'$ is an unbiased estimator of x, i.e., $\mathsf{E}\{\hat{x}'\} = \mathsf{E}\{\hat{x}\} = x$, if and only if $\mathsf{E}\{M(t)t\} = 0$. Thus for $\hat{x}'$ to be unbiased, we need that

$$\mathsf{E}\{M(t)t\} = \int M(t)t f_t(t)\mathrm{d}t = 0 \tag{78}$$

holds true, where $f_t(t)$ denotes the probability density function (PDF) of t. This condition is satisfied if $M(t)f_t(t)$ is symmetric with respect to the origin, thus if $M(-t)f_t(-t) = M(t)f_t(t)$. Since $\hat{Q}_{yy}$, and therefore also $M(t)$, is a function of $\mathrm{vh}(tt)^T$, it follows that $M(t)$ is always symmetric with respect to the origin, $M(-t) = M(t)$.

The conclusion is therefore, that $\hat{x}'$ is an unbiased estimator of x, whenever the PDF $f_t(t)$ is symmetric with respect to its mean $\mathsf{E}\{t\} = 0$. This is a gratifying result, since this holds true for many of the distributions used in practice (such as the class of elliptically contoured distributions).

6.2 The variance

Let us now consider the variance of $\hat{x}'$. The evaluation of the variance is more problematic. We will assume that Eq. (78) holds true and that $\hat{x}$ is uncorrelated with $M(t)t$. The latter is true, for instance, when y is normally distributed, in which case $\hat{x}$ and t, and thus also $\hat{x}$ and $M(t)t$, are independent. With these assumptions, an application of the variance propagation law to Eq. (76) gives

$$Q_{\hat{x}'\hat{x}'} = Q_{\hat{x}\hat{x}} + \int M(t)tt^T M(t)^T f_t(t)\mathrm{d}t \tag{79}$$

Equation (79) shows by how much the variance of $\hat{x}'$ differs from the smallest possible variance, i.e. the variance of the BLUE of x. Although it is difficult to evaluate the second term on the right-hand side of Eq. (79) exactly, a first-order approximation can be obtained if one makes use of the linearization $M(t) = M(0) + \partial_t M(0)t$. Substitution into Eq. (79), gives

$$Q_{\hat{x}'\hat{x}'} = Q_{\hat{x}\hat{x}} + M(0)Q_{tt}M(0)^T$$
$$+ \sum_{i=1}^{b}\sum_{j=1}^{b} \partial_t M_i(0)\mathsf{E}\{t_i t_j tt^T\}\partial_t M_j(0)^T \tag{80}$$

This shows that the difference between $Q_{\hat{x}'\hat{x}'}$ and $Q_{\hat{x}\hat{x}}$ depends on the second- and fourth-order moments of t. They are given in Lemma 1 for elliptically contoured distributions.

7 LS-VCE compared to MINQUE, BIQUE and MLE

In this section, we compare the LS-VCE method with some of the existing and well-known methods of VCE. A comparison is made with the following estimators: (1) the minimum norm quadratic unbiased estimator (MINQUE); (2) the best invariant quadratic unbiased estimator (BIQUE), also known as the minimum variance quadratic unbiased estimator (MIVQUE); and (3) the maximum likelihood estimator (MLE) of the (co)variance components.

In the following, we will describe the underlying assumptions of the different methods and show under which circumstances the different estimators coincide.

7.1 MINQUE

The MINQUE method (Rao 1971) is a popular method for estimating unknown (co)variance components. The underlying model of MINQUE is given by Eqs. (1) and (2), with $Q_0 = 0$ and $Q_k = Z_k Z_k^T$. As with the LS-VCE method, MINQUE has the advantage that no distributional assumptions need to be made about the higher order moments of y or t.

Here we show that the MINQUE method is a special case of the LS-VCE method. That is, if the special class of weight matrices $W_{\mathrm{vh}} = D^T(W_y \otimes W_y)D$ is used and W_y is chosen as

$$W_y = \left(\sum_{k=1}^{p} w_k Q_k\right)^{-1} = \left(\sum_{k=1}^{p} w_k Z_k Z_k^T\right)^{-1} \tag{81}$$

in which the w_k, $k = 1, \ldots, p$, are user-defined scalar weights, then the LS-VCE degenerates to the MINQUE.

MINQUE is defined as a quadratic estimator of σ, say $y^T S y$ (with S symmetric), which is unbiased (i.e., $\mathrm{tr}(Z_k^T S Z_k) = 1$, $k = 1, \ldots, p$) and invariant for changes in x (i.e., $SA = 0$). Furthermore, matrix S is chosen such that it minimizes, in the sense of a Euclidean norm, the difference between the MINQUE and a so-called 'natural' estimator of σ. These conditions of Rao (1971) lead to the following minimization problem:

$$\min_{S} \sum_{k=1}^{p}\sum_{l=1}^{p} \mathrm{tr}(w_k w_l [Z_k^T S Z_l][Z_l^T S Z_k]) \tag{82}$$

subject to the conditions of unbiasedness and invariance. Rao (1971) has shown that the minimizing symmetric matrix $\hat{S}$, and therefore the MINQUE $\hat{\sigma} = y^T \hat{S} y$, is given as

$$\hat{S} = \sum_{l=1}^{p} \lambda_l P_A^{\perp T} W_y Z_l Z_l^T W_y P_A^{\perp} \tag{83}$$

where the λ_k are given as the solutions to the set of equations $\sum_{k=1}^{p} \lambda_k \mathrm{tr}(Z_k Z_k^T W_y P_A^{\perp} Z_l Z_l^T W_y P_A^{\perp}) = 1$, for $l = 1, \ldots, p$, and where W_y is given by Eq. (81).

A comparison of the above solution with that of Corollary 4 (setting $Q_0 = 0$ and $Q_k = Z_k Z_k^T$), shows that the LS-VCE reduces to the MINQUE, in case the special

class of weight matrices is used with the W_y of Eq. (81). Sjöberg (1983) has shown how the MINQUE works out in case the model of condition equations is used, instead of the model of observation equations. In that case, MINQUE becomes identical to the LS-VCE solution of Theorem 2, with $W_t = (B^T W_y^{-1} B)^{-1}$.

7.2 BIQUE

As with MINQUE, the underlying model of BIQUE is given by Eqs. (1) and (2), with $Q_0 = 0$ and $Q_k = Z_k Z_k^T$. Also, like MINQUE, the BIQUE is a quadratic estimator $y^T S y$ that satisfies the conditions of unbiasedness and invariance, $\operatorname{tr}(Z_k^T S Z_k) = 1$, $k = 1, \ldots, p$, and $SA = 0$, respectively.

The two methods differ, however, with respect to the objective function that needs to be minimized. In the case of BIQUE, the objective function is given by the variance of the quadratic estimator. This implies, as for the minimum variance LS-VCE (see Corollary 3 and Sect. 5.3), that distributional information needs to be available of the observables, in particular the central moments up to and including order four.

BIQUE has been derived and studied by various authors under the assumption of normally distributed observables, see e.g. Koch (1978); Sjöberg (1984); Caspary (1987); Yu (1992); Koch (1998). Based on the normal distribution, the minimization of the variance of the invariant quadratic unbiased estimator leads to the following minimization problem:

$$\min_S \ \operatorname{tr}(S Q_{yy} S Q_{yy}) \tag{84}$$

subject to the conditions of unbiasedness and invariance.

Solving for S gives a solution that is identical to the minimum variance LS-VCE solution of Theorem 4 for $\kappa = \kappa' = 0$. This solution is also identical to MINQUE, in case the scalar weights are chosen as $w_k = \sigma_k$. However, note that, in contrast with the LS-VCE, MINQUE cannot accommodate minimum variance estimators for which the inverse variance matrix Q_{vh}^{-1} has a structure other than that of Eq. (33).

7.3 MLE

In contrast with LS-VCE, MINQUE and BIQUE, MLE requires complete knowledge of the PDF of the observables. For (co)variance component estimation based on the linear model of Eqs. (1) and (2), one usually assumes normally distributed observables.

The MLE of the (co)variance components is known to be a biased estimator, see e.g. (Rao and Kleffe, 1988; Rao, 1997). This bias is due to the fact that the MLE does not take into account the loss of degrees of freedom resulting from the estimation of x. To remedy this situation, the restricted

maximum likelihood (REML) estimator has been suggested Patterson and Thompson (1971); Harville (1977).

The REML considers the likelihood function of the distribution of $t = B^T y$, instead of the likelihood function of the distribution of y. Thus, with $t \sim N(0, Q_{tt}(\sigma))$, where $Q_{tt}(\sigma) = \sum_{k=1}^p \sigma_k B^T Q_k B$, the likelihood function of the REML is given as

$$\ln L(t; \sigma) = -\frac{b}{2} \ln 2\pi - \frac{1}{2} \ln \det(Q_{tt}(\sigma)) - \frac{1}{2} t^T Q_{tt}(\sigma)^{-1} t \tag{85}$$

The REML of σ is then defined as the maximizer of $L(t; \sigma)$,

$$\hat{\sigma} = \arg \max_\sigma L(t; \sigma) \tag{86}$$

It can be shown that the REML is identical to the BIQUE (e.g., Koch 1986), and therefore also identical to the minimum variance LS-VCE. Alternative derivations and studies of maximum likelihood variance component estimation, can be found in e.g. Kubik (1970); Ou (1989); Yu (1996).

8 Concluding remarks

In this contribution, we studied the weighted least-squares approach to variance component estimation. Since it is believed that the LS-VCE method is undeservedly still one of the lesser known VCE methods, we have emphasized the simplicity, flexibility and attractiveness of the method, together with the fact that the method is capable of unifying many of the existing VCE methods. The method is very simple indeed, since with the data vector and the design matrix given as $y_{vh} = \operatorname{vh}(t t^T - B^T Q_0 B)$ and $A_{vh} = [\operatorname{vh}(B^T Q_1 B), \ldots, \operatorname{vh}(B^T Q_p B)]$, the unbiased least-squares estimator of the (co)variance component vector follows directly as

$$\hat{\sigma} = (A_{vh}^T W_{vh} A_{vh})^{-1} A_{vh}^T W_{vh} y_{vh} \tag{87}$$

in which W_{vh} is a user-defined weight matrix. The method is flexible, since it gives the user, through the user-defined weight matrix W_{vh}, considerable leeway to weight the data according to the needs of their application. This has been shown for different classes of weight matrices, one of which was derived 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 have discussed various such examples, ranging from the formulation of test statistics for hypothesis testing to the solution of nonlinear (co)variance component estimation problems. Another example is the precision description of the (co)variance component estimator. It readily follows from an application of the variance propagation law to Eq. (87), as

$$Q_{\hat{\sigma}\hat{\sigma}} = N^{-1} M N^{-1} \tag{88}$$

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 variance matrix of y_{vh}.

Being able to perform a rigorous precision analysis is important for many applications, such as estimability studies or significance testing. Clearly, the computations, as needed for Eqs. (87) and (88), are straightforward and easily executed with standard least-squares software packages. This is also true for other least-squares VCE diagnostics.

It was also emphasized that the LS-VCE method is capable of unifying many of the existing VCE-methods. It was shown that these methods could be easily recovered by making the appropriate choice for the weight matrix W_{vh}. From Eq. (87) 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 variance matrix Q_{vh} and its inverse were worked out for the 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, REML and BIQUE are recovered. It was also shown that in this respect MINQUE is a rather restrictive estimator. It can easily be recovered from the LS-VCE method, by making the appropriate choice for the weight matrix. MINQUE itself, however, is not capable of handling elliptically contoured distributions, other then the normal distribution. Since the LS-VCE method can handle elliptically contoured distribution, it can be applied to data, for instance, of which the distribution has tails that are heavier than those of the normal distribution.

Appendix A

Definition of matrix operators

Definition 3 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 $s_{ij} T$ which is seen to be a matrix of order $uq \times vr$. It thus has uv blocks, the ij^{th} block is the matrix $s_{ij} T$ of the order $q \times r$.

Definition 4 Let $S = [s_1 \ s_2 ... \ s_v]$ be an arbitrary matrix of size $u \times v$, with s_i its ith column (vector) of size u. Then the *vec*-operator on this matrix is defined as $\mathrm{vec}(S) = \mathrm{vec}([s_1 \ s_2 ... \ s_v]) = \left[s_1^T \ s_2^T \ ... \ s_v^T \right]^T$. 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.

Definition 5 The *commutation matrix* K_{uv} is the $uv \times uv$ matrix with the property that $K_{uv}\mathrm{vec}(S) = \mathrm{vec}(S^T)$ 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 form $K = \sum_{i=1}^u \sum_{k=1}^u c_i c_k^T \otimes c_k c_i^T$ with $c_i = (0 \cdots 0 \ 1 \ 0 \cdots 0)^T$, the canonical unit vector which contains zeros except a *one* at ith position.

Definition 6 Let $S = (s_{ij})$ be any arbitrary square matrix of size u. The *vh-operator* of S is obtained in the similar way that $\mathrm{vec}(S)$ is defined, but it starts each column at its diagonal element. That is, $\mathrm{vh}(S)$ contains the $\frac{1}{2}u(u+1)$ elements s_{ij}, $i \geq j$, only the elements on and below the diagonal of S.

Definition 7 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 $D\mathrm{vh}(S) = \mathrm{vec}(S)$

Properties of matrix operators

Assuming all matrices and vectors involved have appropriate dimensions, the following properties hold for the Kronecker product, trace, and vec-operator:

$$\mathrm{tr}(UV) = \mathrm{tr}(VU) \tag{89}$$
$$(U + V) \otimes S = U \otimes S + V \otimes S \tag{90}$$
$$(UV) \otimes (ST) = (U \otimes S)(V \otimes T) \tag{91}$$
$$\mathrm{tr}(U \otimes V) = \mathrm{tr}(U)\mathrm{tr}(V) \tag{92}$$
$$(U \otimes V)^T = U^T \otimes V^T \tag{93}$$
$$(U \otimes V)^{-1} = U^{-1} \otimes V^{-1} \tag{94}$$
$$\mathrm{vec}(uv^T) = v \otimes u \tag{95}$$
$$\mathrm{vec}(U)^T \mathrm{vec}(V) = \mathrm{tr}(U^T V) \tag{96}$$
$$\mathrm{vec}(UVS) = (S^T \otimes U)\mathrm{vec}(V) \tag{97}$$
$$\mathrm{tr}(UVST) = \mathrm{vec}(T^T)^T (S^T \otimes U)\mathrm{vec}(V)$$
$$= \mathrm{vec}(T)^T (U \otimes S^T)\mathrm{vec}(V^T) \tag{98}$$

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:

$$\mathrm{vec}(S) = D\mathrm{vh}(S), \ \mathrm{vh}(S) = D^+\mathrm{vec}(S) \text{ with}$$
$$D^+ = (D^T D)^{-1} D^T \tag{99}$$
$$D^+ D = I, \quad \text{and} \quad D D^+ = D^{+T} D^T \text{ a projector} \tag{100}$$
$$\mathrm{vec}(S) = D D^+ \mathrm{vec}(S) \tag{101}$$
$$(D^+(S \otimes S)D)^{-1} = D^+(S^{-1} \otimes S^{-1})D \tag{102}$$
$$D D^+(U \otimes U)D = (U \otimes U)D \tag{103}$$
$$D^+(U \otimes U)D D^+ = D^+(U \otimes U) \tag{104}$$
$$(D^T(U \otimes U)D)^{-1} = D^+(U^{-1} \otimes U^{-1})D^{+T} \tag{105}$$
$$K(U \otimes V) = (V \otimes U)K \tag{106}$$
$$K(U \otimes V)K = V \otimes U \tag{107}$$
$$KD = D, \quad D^+ K = D^+ \tag{108}$$
$$D D^+ = \frac{1}{2}(I + K) \tag{109}$$

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, we refer to Magnus (1988).

Appendix B

Proof of Theorem 3

The $b^2 \times b^2$ variance matrix Q_{vec} of Eq. (47), as a fourth-order tensor, can be rewritten as $Q_{\text{vec}} = \sum_{i=1}^{b} \sum_{k=1}^{b} c_i c_k^T \otimes Q^{i.k.}$, where the $b \times b$ submatrix $Q^{i.k.}$ is of the form

$$Q^{i.k.} = (\kappa + 1)(c_i^T Q_{tt} c_k Q_{tt} + Q_{tt} c_k c_i^T Q_{tt}) + \kappa Q_{tt} c_i c_k^T Q_{tt} \tag{110}$$

with c_i the canonical unit vector. Q_{vec} gives, with Eq. (110)

$$Q_{\text{vec}} = (\kappa+1) \sum_{i=1}^{b} \sum_{k=1}^{b} c_i c_k^T \otimes (c_i^T Q_{tt} c_k Q_{tt} + Q_{tt} c_k c_i^T Q_{tt})$$

$$+ \kappa \sum_{i=1}^{b} \sum_{k=1}^{b} c_i c_k^T \otimes Q_{tt} c_i c_k^T Q_{tt} \tag{111}$$

or

$$Q_{\text{vec}} = (\kappa + 1)\left[Q_{tt} \otimes Q_{tt} + \sum_{i=1}^{b} \sum_{k=1}^{b} (I \otimes Q_{tt}) \right.$$

$$\times (c_i c_k \otimes c_k c_i^T)(I \otimes Q_{tt}) \Big]$$

$$+ \kappa \sum_{i=1}^{b} \sum_{k=1}^{b} (I \otimes Q_{tt})(c_i c_k \otimes c_i c_k^T)(I \otimes Q_{tt}) \tag{112}$$

or

$$Q_{\text{vec}} = (\kappa+1)[Q_{tt} \otimes Q_{tt} + (I \otimes Q_{tt})K(I \otimes Q_{tt})]$$

$$+ \kappa (I \otimes Q_{tt}) \sum_{i=1}^{b} \sum_{k=1}^{b} \left(c_i c_k \otimes c_i c_k^T \right)(I \otimes Q_{tt}) \tag{113}$$

with K the commutation matrix of size $b^2 \times b^2$. Equation (113) is reformulated as

$$Q_{\text{vec}} = (\kappa + 1)[I + K]Q_{tt} \otimes Q_{tt}$$

$$+ \kappa I \otimes Q_{tt} \sum_{i=1}^{b} \sum_{k=1}^{b} (c_i c_k \otimes c_i c_k^T) I \otimes Q_{tt} \tag{114}$$

The term $\sum_{i=1}^{b} \sum_{k=1}^{b} c_i c_k \otimes c_i c_k^T$ can be written as

$$\sum_{i=1}^{b} \sum_{k=1}^{b} c_i c_k^T \otimes c_i c_k^T = \sum_{i=1}^{b} c_i \otimes c_i \sum_{k=1}^{b} c_k^T \otimes c_k^T \tag{115}$$

or

$$\sum_{i=1}^{b} \sum_{k=1}^{b} c_i c_k^T \otimes c_i c_k^T = \text{vec}\left(\sum_{i=1}^{b} c_i c_i^T \right) \text{vec}\left(\sum_{k=1}^{b} c_k c_k^T \right)^T$$

$$= \text{vec}(I)\text{vec}(I)^T \tag{116}$$

Equation (114) with Eq. (116) follows that

$$Q_{\text{vec}} = 2(\kappa + 1)DD^+(Q_{tt} \otimes Q_{tt}) + \kappa \text{vec}(Q_{tt})\text{vec}(Q_{tt})^T \tag{117}$$

Applying the variance propagation law to $\text{vh}(t\,t^T) = D^+\text{vec}(t\,t^T)$ results in the variance matrix of the observable vector $\text{vh}(t\,t^T)$

$$Q_{\text{vh}} = 2(\kappa + 1)D^+(Q_{tt} \otimes Q_{tt})D^{+T} + \kappa D^+\text{vec}(Q_{tt})\text{vec}(Q_{tt})^T D^{+T} \tag{118}$$

which completes the first part.

Using the identity

$$(U + uv^T)^{-1} = U^{-1} - \frac{U^{-1}uv^T U^{-1}}{1 + v^T U^{-1}u} \tag{119}$$

with $U = 2(\kappa+1)D^+(Q_{tt} \otimes Q_{tt})D^{+T}$, $u = \kappa D^+\text{vec}(Q_{tt})$ and $v = D^+\text{vec}(Q_{tt})$ the inverse of Q_{vh} is simply obtained.

References

Amiri-Simkooei AR (2003) Formulation of L1 norm minimization in Gauss–Markov models. J Surv Eng 129(1):37–43
Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2006) Noise characteristics in high precision GPS positioning. In: VI Hotine-Marussi symposium of theoretical and computational geodesy: challenge and role of modern geodesy, 29 May–2 June, Wuhan University, China
Amiri-Simkooei AR (2007) Least-squares variance component estimation: theory and GPS applications, Ph.D. thesis, Delft University of Technology, Publication on Geodesy, 64, Netherlands Geodetic Commission, Delft
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 (2007) Assessment of noise in GPS coordinate time series: methodology and results. J Geophys Res (in press). doi:10.1029/2006JB004913
Baarda W (1973) S-Transformations and criterion matrices, Netherlands Geodetic Commission, Publ. on Geodesy, New Series, Vol. 5(1), Delft
Berkane M, Bentler PM (1986) Moments of elliptically distributed random variates. Statist Probab Lett 4:333–335
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. J Geod 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. J Geod 79(12):694–704
Bock Y (1998) Medium distance GPS measurements. In: Teunissen PJG, Kleusberg A (eds) GPS for geodesy, 2nd edn. Springer, Heidelberg

Bona P (2000) Precision, cross correlation, and time correlation of GPS phase and code observations. GPS Solutions 4(2):3–13

Brunner FK, Hartinger H, Troyer L (1999) GPS signal diffraction modelling: the stochastic SIGMA-Δ model. J Geod 73:259–267

Caspary WF (1987) Concepts of network and deformation analysis. school of surveying, The University of New South Wales, Kensington, pp 97–111

Chen YQ, Chrzanowski A, Kavouras M (1990) Assessment of observations using minimum norm quadratic unbiased estimation (MINQUE), CISM J ACSGS 44:39–46

Crocetto N, Gatti M, Russo P (2000) Simplified formulas for the BIQUE estimation of variance components in disjunctive observation groups. J Geod 73:173–179

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall/CRC

Euler HJ, Goad C (1991) On optimal filtering of GPS dual frequency observations without using orbit information. Bull Geod 65:130–143

Fotopoulos G (2003) An analysis on the optimal combination of geoid, orthometric and ellipsoidal height data, PhD thesis, Department of Geomatics Engineering, University of Calgary, Alberta, Calgary

Fotopoulos G (2005) Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetric geoid height data. J Geod 79(1–3):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. PhD-thesis, Deutsche Geodaetische Kommission, Reihe C, No. 458, Muenchen

Grafarend EW (1985) Variance–covariance component estimation: theoretical results and geodetic applications. Statist Decis Suppl 2:407–441

Gupta AK, Varga T (1993) Elliptically contoured models in statistics. Kluwer, Dordrecht

Hansen PC (1998) Rank-deficient and discrete Ill-posed problems: numerical aspects of linear inversion (Siam Monographs on Mathematical Modeling and Computation), Philadelphia

Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related 2097 problems. J Am Statist Assoc 72:320–340

Jin XX, de Jong CD (1996) Relationship between satellite elevation and precision of GPS code observations. J Navig 49: 253–265

Jonkman NF (1997) Integer GPS-ambiguity estimation without the receiver-satellite geometry. MGP-Series, No. 18, Delft University of Technology

Kuhri AI, Mathew T, Sinha BK (1998) Statistical tests for mixed linear models. Wiley, New York

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

Koch KR (1987) Bayesian inference for variance components. Manuscr Geod 12:309–313

Koch KR (1998) Parameter estimation and hypothesis testing in linear models. Springer, Berlin

Koch KR, Kusche J (2002) Regularization of geopotential determination from satellite data by variance components. J Geod 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 (2003) Noise variance estimation and optimal weight determination for GOCE gravity recovery. Adv Geosci 1:81–85

Lawson CL, Hanson RJ (1974) Solving least-squares problems. Prentice-Hall, Series in Automatic Computation

Lucas JR, Dillinger WH (1998) MINQUE for block diagonal bordered systems such as those encountered in VLBI data analysis. J Geod 72:343–349

Magnus JR (1988) Linear structures. London School of Economics and Political Science, Charles Griffin & Company LTD, London, Oxford University Press, New York

Mao A, Harrison CGA, Dixon TH (1999) Noise in GPS coordinates time series. J Geophys Res 104:2797–2816

Maruyama Y, Seo T (2003) Estimation of moment parameters in elliptical distributions. J Japan Statist Soc 33(2):215–229

Myers RH (1986) Classical and modern regression with applications. Duxburry Press

Ou Z (1989) Estimation of variance and covariance components. Bull Géod 63:139–148

Ou Z (1991) Approximative Bayes estimation for variance components. Manuscr Geod 16:168–172

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58

Radovanovic RS, El-Sheimy N, Teskey WF (2004) Variance-covariance modeling of atmospheric errors for satellite-based network positioning. J Inst Navig 51(2):161–170

Rao CR (1971) Estimation of variance and covariance components—MINQUE theory. J Multivar Anal 1:257–275

Rao CR, Mitra SK (1971) Generalized inverses of matrices and its applications. Wiley, New York

Rao CR, Kleffe J (1988) Estimation of variance components and applications, Vol. 3, North Holland, Series in Statistics and Probability

Rao CR, Toutenburg H (1995) Linear models: least squares and alternatives, Springer Series in Statistics

Rao PSRS (1997) Variance component estimation: mixed models, methodologies and applications. Chapman & Hall

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 estimation of variance–covariance components of GPS observables. Surv Rev 36(286):582–590

Schaffrin B, Bock Y (1988) A unified scheme for processing GPS dual-band observations. Bull Geod 62:142–160

Sjöberg LE (1983) Unbiased estimation of variance–covariance components in condition adjustment with unknowns—a MINQUE approach. ZfV 108(9):382–387

Sjöberg LE (1984) Non negative variance component estimation in the Gauss-Helmert adjustment model. Manuscr Geod 9:247–280

Teunissen PJG (1985) Generalized inverses, adjustment, the datum problem and S-transformations. In: Grafarend EW, Sanso F. (eds) Optimization and design of geodetic networks

Teunissen PJG (1988) Towards a least-squares framework for adjusting and testing of both functional and stochastic model. Geodetic Computing Centre, Delft, MGP Series No. 26, 2004 (Reprint 1988)

Teunissen PJG (1990) Nonlinear least-squares. Manus Geod 15(3):137–150

Teunissen PJG (1998) The ionosphere-weighted GPS baseline precision in canonical form. J Geod 72:107–117

Teunissen PJG, Jonkman NF, Tiberius CCJM (1998) Weighting GPS dual frequency observations: bearing the cross of cross-correlation. GPS Solutions 2(2):28–37

Teunissen PJG (2000) Adjustment theory: an introduction. Delft University Press, Delft University of Technology, Series on Mathematical Geodesy and Positioning

Teunissen PJG, Amiri-Simkooei AR (2006) Variance component estimation by the method of least-squares. In: VI Hotine-Marussi symposium of theoretical and computational geodesy: challenge and Role of Modern Geodesy 29 May–2 June, Wuhan University, China

Tiberius CCJM (1998) Recursive data processing for kinematic GPS surveying. Publications on Geodesy, No. 45 Netherlands Geodetic Commission, Delft, The Netherlands

Tiberius CCJM, Jonkman N, Kenselaar F (1999) The stochastics of GPS observables. GPS World Innon 10(2):49–54

Tiberius CCJM, Borre K (2000) Time series analysis of GPS observables. Proc. ION-GPS2000, 1885–1894

Tiberius CCJM, Kenselaar F (2000) Estimation of the stochastic model for GPS code and phase observables. Surv Rev 35(277): 441–454

Tiberius CCJM, Kenselaar F (2003) Variance component estimation and precise GPS positioning: case study. J Surv Eng 129(1):11–18

Wang J, Stewart MP, Sakiri M (1998) Stochastic modeling for static GPS baseline data processing. J Surv Eng 124(4):171–181

Wang J, Satirapod C, Rizos C (2002) Stochastic assessment of GPS carrier phase measurements for precise static relative positioning. J Geod 76(2):95–104

Wei WWS (1990) Time series analysis, Addison-Wesley, Redwood City

Williams SDP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Miller M, Johnson DJ (2004) Error analysis of continuous GPS position time series. J Geophys Res 109:B03412

Yu Z (1992) A generalization theory of estimation of variance–covariance components. Manuscr Geod 17:295–301

Yu Z (1996) A universal formula of maximum likelihood estimation of variance-covariance components. J Geod 70:233–240

Zhang J, Bock Y, Johnson, 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. J Geophys Res 102:18035–18055