Variance Component Estimation by the Method of Least-Squares

P.J.G. Teunissen, A.R. Amiri-Simkooei
Delft institute of Earth Observation and Space systems (DEOS), Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands, e-mail: P.J.G.Teunissen@TUDelft.nl

Abstract. Motivated by the fact that the method of least-squares is one of the leading principles in parameter estimation, we introduce and develop the method of least-squares variance component estimation (LS-VCE). The results are presented both for the model of observation equations and for the model of condition equations. LS-VCE has many attractive features. It provides a unified least-squares framework for estimating the unknown parameters of both the functional and stochastic model. Also, our existing body of knowledge of least-squares theory is directly applicable to LS-VCE. LS-VCE has a similar insightful geometric interpretation as standard least-squares. Properties of the normal equations, estimability, orthogonal projectors, precision of estimators, nonlinearity, and prior information on VCE can be easily established. Also measures of inconsistency, such as the quadratic form of residuals and the w-test statistic can directly be given. This will lead us to apply hypotheses testing to the stochastic model.

Keywords. Least-squares variance component estimation, BIQUE, MINQUE, REML

1 Introduction

Estimation and validation with heterogeneous data requires insight into the random characteristics of the observables. Proper knowledge of the stochastic model of the observables is therefore a prerequisite for parameter estimation and hypothesis testing. In many cases, however, the stochastic model may still contain unknown components. They need to be determined to be able to properly weigh the contribution of the heterogeneous data to the final result. Different methods exist in the geodetic and statistical literature for estimating such unknown (co)variance components. However, the principles on which these methods are based are often unlinked with the principles on which the estimation of the parameters of the functional model is based.

This paper formulates a unified framework for both the estimation and validation problem of the stochastic model. We concentrate on the problem of estimating parts of the stochastic model. The method is based on the least-squares principle which was originally proposed by Teunissen (1988). We will therefore have the possibility of applying one estimation principle, namely our well-known and well understood method of least-squares, to both the problem of estimating the functional model and the stochastic model. We give the results without proof. For proofs we can closely follow Teunissen and Amiri-Simkooei (2007).

We present the weighted least-squares (co)variance component estimation (LS-VCE) formula for which an *arbitrary* symmetric and positive-definite *weight* matrix can be used. Weighted LS-VCE gives *unbiased* estimators. Based on the normal distribution of original observations, we present the covariance matrix of the observables in the stochastic model. We can obtain the minimum variance estimators by taking the weight matrix as the inverse of the covariance matrix. This corresponds to the best linear unbiased estimator (BLUE) of unknown parameters x in the functional model. These estimators are therefore unbiased and of minimum variance. In this paper the property of minimum variance is restricted to normally distributed data. Teunissen and Amiri-Simkooei (2007) derived such estimators for a larger class of elliptical distributions.

We will make use of the *vector* (vec) and *vector-half* (vh) operators, the Kronecker product ($\otimes$), and the commutation (K) and duplication (D) matrices. For a complete reference on the properties and the theorems among these operators and matrices we refer to Magnus (1988).

2 Least-Squares Estimators

Consider the linear model of observation equations

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

with $\underline{y}$ the $m \times 1$ vector of observables (the underline indicates randomness), x the $n \times 1$ vector of unknown parameters, A the $m \times n$ design matrix, Q_y the $m \times m$ covariance matrix of the observables (Q_0 its known part; the $m \times m$ cofactor matrices Q_k are also known but their contributions through σ_k are unknown). The unknowns σ_k are for instance variance or covariance components. The matrices $Q_k,\ k = 1, \ldots, p$ should be linearly *independent*. The second part of (1) can be written as $\mathsf{D}\{\underline{y}\} = \mathsf{E}\{(\underline{y} - Ax)(\underline{y} - Ax)^T\}$. To get rid of the unknown parameters x in $\mathsf{E}\{(\underline{y} - Ax)(\underline{y} - Ax)^T\}$, one can rewrite (1) in terms of the model of condition equations. One can therefore show that (1) can equivalently be reformulated as

$$\mathsf{E}\{\underline{t}\} = 0;\ \mathsf{E}\{\underline{t}\,\underline{t}^T\} - B^T Q_0 B = \sum_{k=1}^{p} \sigma_k B^T Q_k B\,, \quad (2)$$

with the $b \times 1$ vector of misclosures $\underline{t} = B^T \underline{y}$, the $m \times b$ matrix B satisfying $B^T A = 0$. $b = m - n$ is the redundancy of the functional model. The matrices $B^T Q_1 B, \ldots, B^T Q_p B$ should be linearly independent, which is a *necessary and sufficient* condition in order for the VCE model to have a unique solution.

The first part of (2), i.e., the functional part, consists of all redundant observations as there exists no unknown in this model. The adjustment of this part is trivial because $\hat{\underline{t}} = 0$. We may therefore concentrate on the second part, i.e., the stochastic model. Note also that the condition $\mathsf{E}\{\underline{t}\} = 0$, which implies that there is no misspecification in the functional model, has been used in the second part by default because $Q_t = \mathsf{E}\{\underline{t}\,\underline{t}^T\} - \mathsf{E}\{\underline{t}\}\mathsf{E}\{\underline{t}\}^T$.

Stochastic Model

The matrix equation in the second part of (2) can now be recast into a set of b^2-number of observation equations by stacking the b-number of $b \times 1$ column vectors of $\mathsf{E}\{\underline{t}\,\underline{t}^T\}$ into a $b^2 \times 1$ observation vector. Therefore, just like we interpret the functional model $\mathsf{E}\{\underline{y}\} = Ax$ as a set of m-number of observation equations with the observation vector $\underline{y}$, we are going to interpret the stochastic model $\mathsf{E}\{\underline{t}\,\underline{t}^T - B^T Q_0 B\} = \sum_{k=1}^{p} \sigma_k B^T Q_k B$ as a set of b^2-number of observation equations with the observation matrix $\underline{t}\,\underline{t}^T - B^T Q_0 B$. Since the matrix of observables $\underline{t}\,\underline{t}^T$ is symmetric, its upper triangular elements do not provide new information. There are only $\frac{b(b+1)}{2}$ distinct (functionally independent) elements. We can therefore apply the vh-operator to the second part of (2). This results in the following linear *model of observation equations* (note that both the vh and the E operators are linear):

$$\mathsf{E}\{\underline{y}_{\mathrm{vh}}\} = A_{\mathrm{vh}}\sigma, \quad W_{\mathrm{vh}} \text{ or } Q_{\mathrm{vh}}\,, \quad (3)$$

with $\underline{y}_{\mathrm{vh}} = \mathrm{vh}(\underline{t}\,\underline{t}^T - B^T Q_0 B)$ the observables in the stochastic model, and A_{vh} a $\frac{b(b+1)}{2} \times p$ (design) matrix of the form

$$A_{\mathrm{vh}} = \left[\mathrm{vh}(B^T Q_1 B) \cdots \mathrm{vh}(B^T Q_p B)\right], \quad (4)$$

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

Weighted LS Estimators

Having established these results, we can now apply the method of least-squares to estimate σ. In other words, if the weight matrix W_{vh} is known, we can obtain the weighted least-squares estimators of the (co)variance components. The weighted least-squares estimators of the (co)variance components then read

$$\hat{\underline{\sigma}} = (A_{\mathrm{vh}}^T W_{\mathrm{vh}} A_{\mathrm{vh}})^{-1} A_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{y}_{\mathrm{vh}} = N^{-1} \underline{l}, \quad (5)$$

where $N = A_{\mathrm{vh}}^T W_{\mathrm{vh}} A_{\mathrm{vh}}$, the $p \times p$ normal matrix, and $\underline{l} = A_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{y}_{\mathrm{vh}}$, a p-vector, are of the forms

$$n_{kl} = \mathrm{vh}(B^T Q_k B)^T W_{\mathrm{vh}} \mathrm{vh}(B^T Q_l B)\,, \quad (6)$$

and

$$\underline{l}_k = \mathrm{vh}(B^T Q_k B)^T W_{\mathrm{vh}} \underline{y}_{\mathrm{vh}}\,, \quad (7)$$

respectively, with $k, l = 1, \cdots, p$. Any symmetric and positive-definite matrix W_{vh} can play the role of the weight matrix.

Weight Matrix

From a numerical point of view, an arbitrary weight matrix W_{vh} in (6) and (7) may not be advisable as it is of size $\frac{b(b+1)}{2} \times \frac{b(b+1)}{2}$. For this reason, we now restrict ourselves to those weight matrices which computationally are more efficient. One admissible and, in fact, simple weight matrix W_{vh} has the following form

$$W_{\mathrm{vh}} = D^T (W_t \otimes W_t) D \,, \tag{8}$$

where W_t is an arbitrary positive-definite symmetric matrix of size b and D is the $b^2 \times \frac{b(b+1)}{2}$ *duplication* matrix. Using the properties of the Kronecker product one can show that W_{vh} is in fact positive-definite and therefore can play the role of the weight matrix. Substituting (8) into (6) and (7) gives

$$n_{kl} = \mathrm{tr}(B^T Q_k B W_t B^T Q_l B W_t), \tag{9}$$

and

$$\underline{l}_k = \underline{t}^T W_t B^T Q_k B W_t \underline{t} - \mathrm{tr}(B^T Q_k B W_t B^T Q_0 B W_t). \tag{10}$$

respectively. The weighted least-squares (co)variance component estimation was formulated by rewriting the (co)variance component model into a linear model of observation equations. The above formulation of VCE is based on the weighted least-squares method for which an arbitrary weight matrix W_{vh} (e.g. in form of (8)) can be used. An important feature of the weighted least-squares estimators is the *unbiasedness* property.

Covariance Matrix of $vh(\underline{t}\,\underline{t}^T)$

In order to evaluate the covariance matrix of (co)variance components, i.e. $Q_{\hat{\sigma}}$, we need to know the $\frac{b(b+1)}{2} \times \frac{b(b+1)}{2}$ covariance matrix of $\mathrm{vh}(\underline{t}\,\underline{t}^T)$, namely Q_{vh}. In addition, one can in particular choose the weight matrix W_{vh} as the inverse of Q_{vh} to obtain the minimum variance estimators. Let us first present the covariance matrix of $\mathrm{vec}(\underline{t}\,\underline{t}^T)$ which is based on the following theorem:

Theorem 1. *Let the stochastic vector $\underline{t}$ be normally distributed with mean zero and covariance matrix Q_t, i.e. $\underline{t} \sim \mathrm{N}(0, Q_t)$, then the covariance matrix of the observables* $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ *is given as*

$$Q_{\mathrm{vh}} = 2\, D^+ (Q_t \otimes Q_t) D^{+T} \,, \tag{11}$$

where D is the duplication matrix and D^+ is its pseudo-inverse as $D^+ = (D^T D)^{-1} D^T$.

Proof. Closely follow Teunissen and Amiri-Simkooei (2007).

Using the properties of the duplication matrix and the Kronecker product, the inverse of Q_{vh} is obtained as

$$Q_{\mathrm{vh}}^{-1} = \frac{1}{2} D^T (Q_t^{-1} \otimes Q_t^{-1}) D \,. \tag{12}$$

For normally distributed data, Q_{vh}^{-1} is thus an element of the class of admissible weight matrices defined in (8) with $W_t = \frac{1}{\sqrt{2}} Q_t^{-1}$. This is in fact an interesting result because we can now choose the weight matrix $W_{\mathrm{vh}} = Q_{\mathrm{vh}}^{-1}$ to obtain the minimum variance estimators of the (co)variance components.

Minimum Variance Estimators

As with the best linear unbiased estimator (BLUE) in the functional model, the (co)variance components can be estimated according to BLUE with the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. One can obtain such estimators by taking the weight matrix W_{vh} as the inverse of the covariance matrix of the observables, Q_{vh}^{-1}. Then this linear form of the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$ can be rewritten as the best (minimum variance) quadratic unbiased estimator of the misclosures $\underline{t}$. To obtain the minimum variance estimators, one needs to substitute $W_t = \frac{1}{\sqrt{2}} Q_t^{-1}$ in (9) and (10). Such estimators are therefore given as $\underline{\hat{\sigma}} = N^{-1} \underline{l}$ with

$$n_{kl} = \frac{1}{2} \mathrm{tr}(B^T Q_k B Q_t^{-1} B^T Q_l B Q_t^{-1}), \tag{13}$$

and

$$\underline{l}_k = \frac{1}{2} \underline{t}^T Q_t^{-1} B^T Q_k B Q_t^{-1} \underline{t}, \tag{14}$$

in which we assumed $Q_0 = 0$. Since the covariance matrix Q_{vh} in (11) is derived for normally distributed data, the 'best' (minimum variance) property is restricted to the normal distribution.

3 Formulation in Terms of A-Model

Weighted LS Estimators

The least-squares method to (co)variance component estimation can directly be used, if the matrix B is available (model of condition equations). In practice, however, one will usually have the design matrix A available (model of observation equations) instead of B. We now extend the least-squares method for estimation of (co)variance components to the model of observation equations. We consider again the case that the covariance matrix can be split into a known part Q_0 and an unknown (co)variance component model, namely $Q_y = Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$.

To apply the weighted least-squares variance component estimation to the model of observation equations we shall therefore have to rewrite (9) and (10) in terms of the design matrix A. Using the relation between elements of the B and A models and also taking into account the *trace* properties, the matrix N in (9) and the vector $\underline{l}$ in (10) can be reformulated as

$$n_{kl} = \mathrm{tr}(Q_k W P_A^{\perp} Q_l W P_A^{\perp}) \,, \tag{15}$$

and

$$\underline{l}_k = \underline{\hat{e}}^T W Q_k W \underline{\hat{e}} - \text{tr}(Q_k W P_A^\perp Q_0 W P_A^\perp)\,, \qquad (16)$$

respectively, where W is an arbitrary $m \times m$ positive-definite matrix, $\underline{\hat{e}}$ is the least-squares residuals given as $\underline{\hat{e}} = P_A^\perp \underline{y}$, with the orthogonal projector $P_A^\perp = I - A(A^T W A)^{-1} A^T W$. The weighted least-squares estimator is therefore given as $\underline{\hat{\sigma}} = N^{-1}\underline{l}$ with N and $\underline{l}$ given by (15) and (16), respectively.

Minimum Variance Estimators

To obtain the minimum variance estimators, we should choose the weight matrix as the inverse of the covariance matrix. In an analogous way to $W_t = \frac{1}{\sqrt{2}} Q_t^{-1}$, one can use the matrix $W = \frac{1}{\sqrt{2}} Q_y^{-1}$. If we now substitute W into (15) and (16), we will then obtain

$$n_{kl} = \frac{1}{2}\text{tr}(Q_k Q_y^{-1} P_A^\perp Q_l Q_y^{-1} P_A^\perp), \qquad (17)$$

and

$$\underline{l}_k = \frac{1}{2}\underline{\hat{e}}^T Q_y^{-1} Q_k Q_y^{-1} \underline{\hat{e}} - \frac{1}{2}\text{tr}(Q_k Q_y^{-1} P_A^\perp Q_0 Q_y^{-1} P_A^\perp), \quad (18)$$

where $P_A^\perp = I - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$.

Implementation

Equations (17) and (18) with $\hat{\sigma} = N^{-1} l$ show that we need $Q_y = Q_0 + \sum_{k=1}^p \sigma_k Q_k$ in order to compute the estimates $\underline{\hat{\sigma}}_k$. But the (co)variance components σ_k are unknown apriori. The final solution should be sought through an iterative procedure. For this purpose we start with an initial guess for the σ_k. Based on these values, we compute with $\hat{\sigma} = N^{-1} l$ estimates for the σ_k, which in a next iteration are considered the improved values for σ_k. The procedure is repeated until the estimated components do not change by further iteration. Figure 1 gives a straightforward iterative algorithm for implementing LS-VCE in terms of the model of observation equations.

There are two ways of estimating (co)variance components. The first way is to consider the cofactor matrices as a whole and try to estimate unknown unit *factors* (scale factors). That is, in each iteration we modify the cofactor matrices by multiplying them with the estimated factors. After a few iterations we expect the factors to converge to ones. In the *second* way, we consider the cofactor matrices to be fixed. In each iteration, the (co)variance *components* rather than the cofactor matrices are modified. After some iterations, the modified (co)variance components converge so that their values do not change by further iterations. For example, consider the covariance matrix as $Q_y = \sigma_1 Q_1 + \sigma_2 Q_2$. At the point of convergence, the above strategies look as follows: In the first way, we obtain the factors f_1 and f_2, therefore $Q_y = \hat{f}_1 \times \sigma_1 Q_1 + \hat{f}_2 \times \sigma_2 Q_2$ where $\hat{f}_1 = \hat{f}_2 = 1$ and in the second way we estimate the components σ_1 and σ_2, therefore $Q_y = \hat{\sigma}_1 \times Q_1 + \hat{\sigma}_2 \times Q_2$.

Implementation of LS-VCE (A-model)

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

begin
  check for presence of gross errors in observations;
  set iteration counter $i = 0$;
  begin
    evaluate matrix $Q_y = Q_0 + \sum_{k=1}^p \sigma_k Q_k$;
    calculate N and l from (17) and (18);
    solve for a new $\hat{\sigma}$ from equations $N\hat{\sigma} = l$;
    $i \leftarrow i + 1$;
    update vector $\sigma^i \leftarrow \hat{\sigma}$;
    while $\|\sigma^i - \sigma^{i-1}\|_{Q_{\hat{\sigma}}^{-1}} > \epsilon$ **repeat**;
  end
  obtain $\hat{\sigma}$ and its covariance matrix $Q_{\hat{\sigma}} = N^{-1}$.
end

Fig. 1. Symbolic algorithm for implementation of least-squares variance component estimation in terms of linear model of observation equations (A model); σ^i is the vector of (co)variance components estimated in iteration i.

4 Properties of Proposed Method

Since we have obtained the least-squares (co)variance estimators based on a model of observation equations, see (3), the following features can easily be established:

4.1 Unification of Methods

To obtain the weighted least-squares solutions, no assumption on the distribution of $\text{vh}(\underline{t}\,\underline{t}^T)$ is required. Also, we know without any additional derivation that the estimators are *unbiased*. This property is independent of the distribution of the observable vector $\text{vh}(\underline{t}\,\underline{t}^T)$. This makes the LS-VCE method more flexible as we can now use a class of weight matrices as $W_{\text{vh}} = D^T (W_t \otimes W_t) D$ where W_t is an arbitrary positive definite matrix and D the duplication matrix.

In a special case where one takes the weight matrix as the inverse of the covariance matrix, i.e. $W_{\text{vh}} = Q_{\text{vh}}^{-1}$, one can simply obtain the minimum variance estimators. Therefore, LS-VCE is capable of unifying many of the existing VCE methods such as minimum norm quadratic unbiased estimator (MINQUE) (see

Rao, 1971, Rao and Kleffe, 1988, Sjöberg, 1983), best invariant quadratic unbiased estimator (BIQUE) (see Caspary, 1987, Koch, 1978, 1999, Schaffrin, 1983), and restricted maximum likelihood (REML) estimator (see Koch, 1986).

4.2 Similarity with Standard LS

LS-VCE has a similar insightful geometric interpretation as the standard least-squares. Properties of the normal matrix, estimability of (co)variance components, and the orthogonal projectors can easily be established. Also, in an analogous way to the functional model in which one deals with redundancy $b = m - n$, one can define the redundancy (or here the degrees of freedom df) in the stochastic model. From (4) it follows that $df = \frac{b(b+1)}{2} - p$, when the design matrix A_{vh} of the stochastic model is assumed to be of full rank, and with p, as before, being the number of unknown (co)variances components. This implies that the maximum number of estimable (co)variance components is $p = \frac{b(b+1)}{2}$, which leads to $df = 0$ (see also Xu *et al.*, 2007).

4.3 Covariance Matrix of Estimators

Since the weighted least-squares estimators are in a linear form of the observables $\underline{y}_{\mathrm{vh}}$, applying the error propagation law to $\underline{\hat{\sigma}} = N^{-1}A_{\mathrm{vh}}^T W_{\mathrm{vh}} \underline{y}_{\mathrm{vh}}$ automatically gives us the covariance matrix of the estimated (co)variance components, namely $Q_{\hat{\sigma}} = N^{-1}MN^{-1}$ where the $p \times p$ matrix M is given as $m_{kl} = 2\mathrm{tr}(B^T Q_k B W_t Q_t W_t B^T Q_l B W_t Q_t W_t) = 2\mathrm{tr}(Q_k W P_A^{\perp} Q_y W P_A^{\perp} Q_l W P_A^{\perp} Q_y W P_A^{\perp})$. This equation can therefore provide us with the precision of the estimators. This is in fact an important feature of the least-squares variance component estimation. In case of minimum variance estimators ($W = \frac{1}{\sqrt{2}} Q_y^{-1}$), one can simply show that $M = N$, and therefore $Q_{\hat{\sigma}} = N^{-1}$.

4.4 Measures of Inconsistency

Since the approach is based on the least-squares principle, parts of the standard quality-control theory can be applied to the model in (3). One can in particular apply the idea of hypotheses testing to the stochastic model. For example, one can deal with the w-test statistic and the quadratic form of the residuals in the stochastic model. As an important measure of any least-squares solution, one can compute the quadratic form of the residuals. This also holds true for the LS-VCE. The quadratic form of the residuals is then given as

$$\underline{\hat{e}}_{\mathrm{vh}}^T Q_{\mathrm{vh}}^{-1} \underline{\hat{e}}_{\mathrm{vh}} = \frac{1}{2}(\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}})^2 - \underline{l}^T N^{-1} \underline{l}\,, \qquad (19)$$

in which we assumed $Q_0 = 0$. One can also obtain the w-test statistic to identify the proper noise components of the stochastic model. Further discussion on this topic is beyond the scope of the present contribution. For more information we refer to Amiri-Simkooei (2007).

4.5 Nonlinear Stochastic Model

LS-VCE has the capability of applying to a nonlinear (co)variance component model, namely $Q_y = Q(\sigma)$. To overcome the nonlinearity, one can expand the stochastic model into a Taylor series, for which one needs the initial values of the unknown vector σ, namely σ^0. When expanded into Taylor series, the covariance matrix can be written as $Q_y = Q(\sigma) \approx Q_0 + \sum_{k=1}^{p} \sigma_k Q_k$. We can now apply the LS-VCE to estimate σ. The estimated $\underline{\hat{\sigma}}$ can then be considered as a new update for σ^0 and the same procedure can be repeated. We can iterate until the estimated (co)variance components do not change by further iterations. The applied iteration is the Gauss-Newton iteration which has a linear rate of convergence (see Teunissen, 1990).

4.6 Prior Information

In some cases, we may have prior information about the (co)variance components. Such information can be provided by equipment manufacturers or from a previous process. Let us assume that this information can be expressed as $\mathsf{E}\{\underline{\sigma}_0\} = \sigma\,;\ \mathsf{D}\{\underline{\sigma}_0\} = Q_{\sigma_0}$, which means that the (co)variance components $\underline{\sigma}_0$ are earlier estimators available with the covariance matrix Q_{σ_0}. One important feature of the LS-VCE is the possibility of incorporating such prior information with the observables $\mathrm{vh}(\underline{t}\,\underline{t}^T)$. Without additional derivations, one can obtain the least-squares (co)variance estimators as $\underline{\hat{\sigma}} = (N + Q_{\sigma_0}^{-1})^{-1}(\underline{l} + Q_{\sigma_0}^{-1}\underline{\sigma}_0)$. Note that the covariance matrix of these estimators is simply given as $Q_{\hat{\sigma}} = (N + Q_{\sigma_0}^{-1})^{-1}$.

4.7 Robust Estimation

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

5 Simple Examples

Example 1 (Minimum variance estimator)**.** As a simple application of LS-VCE, assume that there is only one variance component in the stochastic model, namely $Q_y = \sigma^2 Q$. If our original observables $\underline{y}$ are normally distributed, it follows with (17) and (18) from $\underline{\hat{\sigma}} = N^{-1}\underline{l}$ that

$$\underline{\hat{\sigma}}^2 = \frac{l}{n} = \frac{\frac{1}{2}\underline{\hat{e}}^T Q_y^{-1} Q Q_y^{-1} \underline{\hat{e}}}{\frac{1}{2}\mathrm{tr}(Q Q_y^{-1} P_A^{\perp} Q Q_y^{-1} P_A^{\perp})}, \tag{20}$$

Using $Q_y = \sigma^2 Q$, $P_A^{\perp} P_A^{\perp} = P_A^{\perp}$, and $\mathrm{tr}(P_A^{\perp}) = \mathrm{rank}(P_A^{\perp}) = m - n = b$, the preceding equation, its mean, and its variance simplify to

$$\underline{\hat{\sigma}}^2 = \frac{\underline{\hat{e}}^T Q^{-1} \underline{\hat{e}}}{m-n}; \quad \mathsf{E}\{\underline{\hat{\sigma}}^2\} = \sigma^2; \quad \mathsf{D}\{\underline{\hat{\sigma}}^2\} = \frac{2\sigma^4}{m-n}, \tag{21}$$

respectively. These are the well-known results for the estimator of the variance of unit weight. This estimator can thus be obtained from the least-squares residuals without iteration. This estimator is *unbiased* and of minimum variance. The variance of the estimator was simply obtained by $\mathsf{D}\{\underline{\hat{\sigma}}^2\} = N^{-1} = \frac{2\sigma^4}{\underline{m-n}}$. ■

Example 2 (Weighted LS estimator). To see an important application of the weighted LS-VCE, we derive the empirical autocovariance function in a time series (e.g. to estimate the time correlation of a time series). For simplicity we assume that (1) we measure a functionally known quantity (e.g. a zero baseline measured by GPS receivers), and (2) the cofactor matrices are side-diagonal with equal values which implies that the covariance between observations i and j is only a function of time-lag $\tau = |j - i|$, i.e. $\sigma_{ij} = \sigma_\tau$.

The covariance matrix can thus be written as a linear combination of m cofactor matrices as

$$Q_y = \sigma^2 I + \sum_{\tau=1}^{m-1} \sigma_\tau Q_\tau, \tag{22}$$

where $Q_\tau = \sum_{i=1}^{m-\tau} c_i c_{i+\tau}^T + c_{i+\tau} c_i^T$, $\tau = 1, \ldots, m-1$, with c_i the canonical unit vector, are some cofactor matrices and σ^2 is the unknown variance of the noise process.

We can now apply the weighted least-squares approach to estimate the (co)variance components. One particular choice of the weight matrix W is the unit matrix, $W = I$. Since the design matrix A is empty, it follows that $P_A^{\perp} = I$. To estimate the (co)variance components $\underline{\hat{\sigma}}$ one needs to obtain N and $\underline{l}$ from (15) and (16), respectively. One can show that the (co)variance components σ_τ are estimated as

$$\underline{\hat{\sigma}}_\tau = \frac{\underline{l}_\tau}{n_{\tau,\tau}} = \frac{\sum_{i=1}^{m-\tau} \underline{\hat{e}}_i \underline{\hat{e}}_{i+\tau}}{m-\tau}, \quad \tau = 0, 1, \ldots, m-1, \tag{23}$$

where $\underline{\hat{e}}_i$ is the ith least-squares residual, $\underline{\hat{\sigma}}_0 = \underline{\hat{\sigma}}^2$ is the variance, and $\underline{\hat{\sigma}}_\tau$, $\tau = 1, \ldots, m-1$ are the covariances. One can also derive the covariance matrix of these estimators using $Q_{\hat{\sigma}} = N^{-1} M N^{-1}$. ■

6 Concluding Remarks

There are various VCE formulas based on optimality properties as unbiasedness, minimum variance, minimum norm, and maximum likelihood. In this paper we introduced the method of least-squares for estimating the stochastic model for which *any* symmetric and positive-definite weight matrix can be used. The method is easily understood and very flexible. It can be used for estimation of both variance and covariance components in the A-model and the B-model, both for linear and nonlinear stochastic models. Since the method is based on the least-squares principle, we know without any additional derivation that the estimators are unbiased. One advantage of this technique over other methods of VCE is that the weighted least-squares solution can be obtained without any supposition regarding the distribution of the data. This holds true also for the property of unbiasedness of the estimators. We then simply presented the minimum variance estimators by taking the weight matrix as the inverse of the covariance matrix of observables.

Since we formulated the LS-VCE based on a linear model of observation equations, the proposed method has special and unique features. LS-VCE allows one to apply the existing body of knowledge of least-squares theory to the problem of (co)variance component estimation. With this method, one can (1) obtain measures of discrepancies in the stochastic model, (2) determine the covariance matrix of the (co)variance components, (3) obtain the minimum variance estimator of (co)variance components by choosing the weight matrix as the inverse of the covariance matrix, (4) take the a-priori information on the (co)variance component into account, (5) solve for a nonlinear (co)variance component model, (6) apply the idea of robust estimation to (co)variance components, (7) evaluate the estimability of the (co)variance components, and (8) avoid the problem of obtaining negative variance components.

Acknowledgments

The second author would like to kindly acknowledge Christian Tiberius for his valued contribution and comments on earlier versions of this work. We also thank Frank Kleijer for reading a draft version of this paper.

References

Amiri-Simkooei, A. R. (2003). Formulation of L_1 norm minimization in Gauss-Markov models. *Journal of Surveying Engineering*, **129**(1), 37–43.

Amiri-Simkooei, A. R. (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.

Caspary, W. F. (1987). Concepts of network and deformation analysis. Technical report, School of Surveying, The University of New South Wales, Kensington.

Koch, K. R. (1978). Schätzung von varianzkomponenten. *Allgemeine Vermessungs Nachrichten*, **85**, 264–269.

Koch, K. R. (1986). Maximum likelihood estimate of variance components. *Bulletin Geodesique*, **60**, 329–338. Ideas by A.J. Pope.

Koch, K. R. (1999). *Parameter estimation and hypothesis testing in linear models*. Springer Verlag, Berlin.

Magnus, J. R. (1988). *Linear Structures*. Oxford University Press, London School of Economics and Political Science, Charles Griffin & Company LTD, London.

Rao, C. R. (1971). Estimation of variance and covariance components - MINQUE theory. *Journal of multivariate analysis*, **1**, 257–275.

Rao, C. R. and Kleffe, J. (1988). *Estimation of variance components and applications*, volume 3. North-Holland. Series in Statistics and Probability.

Schaffrin, B. (1983). Varianz-kovarianz-komponentenschätzung bei der ausgleichung heterogener wiederholungsmessungen. C282, Deutsche Geodätische Kommission, München.

Sjöberg, L. E. (1983). Unbiased estimation of variance-covariance components in condition adjustment with unknowns – a MINQUE approach. *Zeitschrift für Vermessungswesen*, **108**(9), 382–387.

Teunissen, P. J. G. (1988). Towards a least-squares framework for adjusting and testing of both functional and stochastic model. Internal research memo, Geodetic Computing Centre, Delft. A reprint of original 1988 report is also available in 2004, No. 26, http://www.lr.tudelft.nl/mgp.

Teunissen, P. J. G. (1990). Nonlinear least-squares. *Manuscripta Geodetica*, **15**(3), 137–150.

Teunissen, P. J. G. and Amiri-Simkooei, A. R. (2008). Least-squares variance component estimation. *Journal of Geodesy* (in press), doi 10.1007/s00190-007-0157-x.

Xu, P. L., Liu, Y. M., and Shen, Y. Z. (2007). Estimability analysis of variance and covariance components. *Journal of Geodesy*, **81**(9), 593–602, doi 10.1007/s00190-006-0122-0.