Variance components in errors-in-variables models: estimability, stability and bias analysis

Peiliang Xu · Jingnan Liu

Abstract Although total least squares has been substantially investigated theoretically and widely applied in practical applications, almost nothing has been done to simultaneously address the estimation of parameters and the errors-in-variables (EIV) stochastic model. We prove that the variance components of the EIV stochastic model are not estimable, if the elements of the random coefficient matrix can be classified into two or more groups of data of the same accuracy. This result of inestimability is surprising as it indicates that we have no way of gaining any knowledge on such an EIV stochastic model. We demonstrate that the linear equations for the estimation of variance components could be ill-conditioned, if the variance components are theoretically estimable. Finally, if the variance components are estimable, we derive the biases of their estimates, which could be significantly amplified due to a large condition number.

Keywords Errors-in-variables model · Estimability · Nonlinear adjustment · Total least squares · Variance components

1 Introduction

Errors-in-variables (EIV) models and total least squares (TLS) have been substantially investigated theoretically for more than a century (see, e.g., Adcock 1877; Kummell 1879;

P. Xu (✉)
Disaster Prevention Research Institute, Kyoto University,
Uji, Kyoto 611-0011, Japan
e-mail: pxu@rcep.dpri.kyoto-u.ac.jp

J. Liu
GNSS Research Center, Wuhan University, Wuhan 430071,
People's Republic of China

Pearson 1901; Deming 1931, 1934; Gerhold 1969; Golub and Loan 1980; Huffel and Vandewalle 1991; Markovsky and Huffel 2007; Schaffrin and Wieser 2008; Xu et al. 2012). Recently, they have been applied to solve a wide variety of science and engineering problems (see, e.g., Huffel and Vandewalle 1991; Markovsky and Huffel 2007; Schaffrin and Wieser 2008; Schaffrin and Felus 2009). The TLS method is designed to optimally estimate the unknown parameters $\boldsymbol{\beta}$ in the following (linear or linearized) EIV model:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

(see, e.g., Seber and Wild 1989; Huffel and Vandewalle 1991), where both $\mathbf{y}$ and $\mathbf{A}$ are measured with random errors, collected in an n-dimensional vector $\boldsymbol{\epsilon}$ and an $(n \times m)$ matrix $\boldsymbol{\epsilon}_A$, respectively, $\boldsymbol{\beta}$ is an m-dimensional deterministic vector of unknown parameters to be estimated. If $\mathbf{A}$ in the EIV model (1) is deterministically given, then (1) returns to a standard linear model (see, e.g., Searle 1971).

When TLS is used to estimate the unknown vector $\boldsymbol{\beta}$ in (1), one simultaneously considers the stochastic nature of both measurements $\mathbf{y}$ and $\mathbf{A}$ instead of treating $\mathbf{A}$ as if it were fixed. However, the sense of TLS optimality has been interpreted differently, either under the framework of approximation theory or from the point of view of statistical estimation theory. The former may be referred to Golub and Loan (1980) and the generalized TLS method by Huffel and Vandewalle (1989) and Markovsky and Huffel (2007), because the so-called positive definite matrix of weighting there is not the inverse of a variance–covariance matrix in the statistical sense, as pointed out by Schaffrin and Wieser (2008) (see also Xu et al. 2012). In this paper, we will affirm to interpret the TLS approach from the statistical point of view. Under the statistical framework, one often assumes zero means for both the random errors of $\mathbf{y}$ and $\mathbf{A}$ in the EIV model (1). The

variance–covariance matrices of the measured data $\mathbf{y}$ and $\mathbf{A}$ are either assumed to be $\mathbf{I}\sigma^2$ and $\mathbf{I}_a\sigma^2$ implicitly (see, e.g., Pearson 1901) or $\mathbf{W}_y^{-1}\sigma^2$ and $\mathbf{W}_a^{-1}\sigma^2$ explicitly (see, e.g., Deming 1931, 1934; Gerhold 1969; Schaffrin and Wieser 2008; Xu et al. 2012), where $\mathbf{I}_a$ is an identity matrix with the dimension corresponding to the vectorized error matrix of $\mathbf{A}$, both $\mathbf{W}_y$ and $\mathbf{W}_a$ are the given weight matrices of $\mathbf{y}$ and $\mathbf{A}$, respectively. Here σ^2 is a positive (unknown) scalar, which is always called the variance of unit weight in the geodetic literature. In general, $\mathbf{y}$ is also assumed to be stochastically independent of $\mathbf{A}$, though such an assumption is not absolutely necessary.

In practice, we can often encounter problems in physical, statistical and engineering sciences in which: (i) $\mathbf{y}$ and $\mathbf{A}$ in (1) are not necessarily of the same types of measurements. In this case, it is inappropriate and/or unreasonable to assume that we know the variance–covariance matrices for $\mathbf{y}$ and $\mathbf{A}$ up to an unknown but identical positive scalar σ^2; and (ii) the elements within each of $\mathbf{y}$ and $\mathbf{A}$ are not necessarily of the same types of measurements either. Even if we assume that $\mathbf{y}$ (and/or $\mathbf{A}$) is of the same type of measurements, they may be measured with instruments of different accuracy. For problems of this kind, the corresponding stochastic model should be more appropriately described as follows:

$$\mathbf{\Sigma}_y = \sum_{i=1}^{m_y} \mathbf{U}_{iy}\sigma_{iy}^2, \tag{2a}$$

$$\mathbf{\Sigma}_a = \sum_{i=1}^{m_a} \mathbf{U}_{ia}\sigma_{ia}^2, \tag{2b}$$

where $\mathbf{\Sigma}_y$ and $\mathbf{\Sigma}_a$ are the variance–covariance matrices of $\mathbf{y}$ and vec($\mathbf{A}$), respectively. Here vec($\mathbf{A}$) stands for the vectorized operation of $\mathbf{A}$ (see, e.g., Magnus and Neudecker 1988). All the $\mathbf{U}_{iy}$ and $\mathbf{U}_{ia}$ are known and positive semidefinite. σ_{iy}^2 ($i = 1, 2, \ldots, m_y$) and σ_{ia}^2 ($i = 1, 2, \ldots, m_a$) are called (unknown) variance components of $\mathbf{y}$ and vec($\mathbf{A}$), respectively. One may also further consider the correlation between $\mathbf{y}$ and $\mathbf{A}$ in the stochastic model (2) (see, e.g., Fang 2011; Snow 2012). Since the inclusion of correlation may not create additional theoretical difficulty, it will not be further pursued in this paper.

Although the general EIV model (1) with the stochastic model (2) should be more appropriate to describe a wide range of real-life EIV problems, little has ever been done to simultaneously address the model parameters and the variance components. Limited related work has been done to estimate both parameters and variance components in spatial and generalized mixed effects models with measurement errors using the maximum likelihood and quasi-likelihood methods (Wang et al. 1998; Li et al. 2009) and the restricted maximum likelihood and pseudo-likelihood methods (Wang and Davidian 1996). Since the EIV model (1) is not the starting model of

these publications, these researchers used likelihood-based methods instead of the TLS principle.

In this paper, we will focus on the weighted TLS estimation of the unknown vector $\boldsymbol{\beta}$ in the functional EIV model (1) and the estimation of both the variance components of $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_A$ in the stochastic model (2). Obviously, the stochastic model (2) is significantly different from any conventional variance component models in the sense that there exist no direct redundant observations of $\mathbf{A}$ for an independent assessment of $\boldsymbol{\epsilon}_A$. From this point of view, we are interested in investigating the following questions: (i) whether the variance components of (2) can be estimated; (ii) does the estimation of variance components exhibit any special numerical and/or statistical behavior; and (iii) Wang and Davidian (1996), Wang et al. (1998) and Li et al. (2009) performed an asymptotic bias analysis of the so-called naive estimator of variance components of the measurements $\mathbf{y}$ for spatial and generalized mixed effects models with measurement errors on the basis of two assumptions: 1) by simply ignoring the measurement errors of $\mathbf{A}$ as if they were known exactly; and 2) by assuming that the number of measurements $\mathbf{y}$ tends to infinity. They did not derive the finite sample biases for the likelihood-based estimated variance components. Unlike Wang and Davidian (1996), Wang et al. (1998) and Li et al. (2009), our emphasis in this paper will be to perform a finite sample bias analysis for the estimated variance components in connection with the EIV models (1) and (2) by fully taking the random errors of $\mathbf{A}$ into account.

To answer the questions posed above, we will start with quadratic forms of the residuals of the measured data $\mathbf{y}$ and $\mathbf{A}$ to estimate the variance components. The paper is thus organized as follows. Section 2 will first reformulate the EIV model as a nonlinear adjustment problem and use the weighted TLS method to estimate the parameters $\boldsymbol{\beta}$. Different solutions methods will be presented. In Sect. 3, we will first discuss adapting variance component estimation to nonlinear models in general and then adapting the minimum norm quadratic unbiased estimation (MINQUE) method to the nonlinear EIV model in particular. Two special structures of variance components will also be discussed. We will show in Sect. 4 that a certain structure of EIV variance components that is commonly assumed and often encountered in practice is not estimable. In other cases, if the variance components are estimable, the estimation of variance components in the EIV models (1) and (2) could become unstable. Instability will then be demonstrated through numerical simulations. To warrant a high quality and stable estimation of variance components, regularization may be needed, which deserves a completely separate paper to discuss and will not be pursued here any further. Unlike the asymptotic bias analysis by Wang and Davidian (1996), Wang et al. (1998) and Li et al. (2009), we will derive the finite sample biases of the estimated variance components in Sect. 5, provided they are estimable.

2 Parameter estimation

2.1 Reformulation of the EIV models (1) and (2) as a nonlinear Gauss–Markoff model with variance components

To investigate the variance component estimation of the functional EIV model (1), we follow Xu et al. (2012) to equivalently rewrite the functional EIV model (1) as the following nonlinear Gauss–Markoff model:

$$\mathbf{y} = \overline{\mathbf{A}}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3a}$$
$$\mathbf{A} = \overline{\mathbf{A}} + \boldsymbol{\epsilon}_A, \tag{3b}$$

which can also be equivalently rewritten in vector form as follows:

$$\mathbf{y} = (\boldsymbol{\beta}^T \otimes \mathbf{I}_n)\overline{\mathbf{a}} + \boldsymbol{\epsilon}, \tag{4a}$$
$$\mathbf{a} = \overline{\mathbf{a}} + \boldsymbol{\epsilon}_a, \tag{4b}$$

or in an even more compact form:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{\beta}^T \otimes \mathbf{I}_n)\overline{\mathbf{a}} \\ \overline{\mathbf{a}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_a \end{bmatrix}, \tag{5}$$

where $\overline{\mathbf{A}}$ is the expectation of $\mathbf{A}$ in the sense that each element of $\overline{\mathbf{A}}$ is the expectation of the corresponding element of $\mathbf{A}$, and the elements of $\boldsymbol{\epsilon}_A$ are the random errors of the observed matrix $\mathbf{A}$, $\otimes$ stands for the Kronecker product, $\mathbf{I}_n$ is an $(n \times n)$ identity matrix, $\mathbf{a} = \mathrm{vec}(\mathbf{A})$, $\overline{\mathbf{a}} = \mathrm{vec}(\overline{\mathbf{A}})$ and $\boldsymbol{\epsilon}_a = \mathrm{vec}(\boldsymbol{\epsilon}_A)$. Schaffrin and Snow (2010) reformulated the functional EIV model (1) as a system of nonlinear Gauss-Helmert condition equations. It is obvious from (4a) that the expectation of $\mathbf{y}$ is a vector of nonlinear functions of both $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$. In case that the elements of $\mathbf{A}$ are not functionally independent or that some of its elements are deterministic, we can follow Xu et al. (2012) to reformulate the functional EIV model (1) as a partial EIV model.

In the remainder of this paper, we will focus on the nonlinear model (3) or (5). The stochastic model for $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_a$ is the same as formulated in (2), which can be equivalently represented in one matrix for both the measurements $\mathbf{y}$ and $\mathbf{A}$ as follows

$$\begin{aligned}
\boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{\Sigma}_y & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_a \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{m_y} \mathbf{U}_{iy}\sigma_{iy}^2 & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^{m_a} \mathbf{U}_{ia}\sigma_{ia}^2 \end{bmatrix} \\
&= \sum_{i=1}^{m_y} \begin{bmatrix} \mathbf{U}_{iy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \sigma_{iy}^2 + \sum_{i=1}^{m_a} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{ia} \end{bmatrix} \sigma_{ia}^2 \\
&= \sum_{i=1}^{m_y} \mathbf{U}_i\sigma_{iy}^2 + \sum_{i=1}^{m_a} \mathbf{U}_i\sigma_{ia}^2
\end{aligned}$$

$$= \sum_{i=1}^{m_y+m_a} \mathbf{U}_i\sigma_i^2, \tag{6}$$

where the matrices $\mathbf{U}_i$ corresponding to σ_{iy}^2 and σ_{ia}^2 are given, respectively, as follows:

$$\mathbf{U}_i = \begin{bmatrix} \mathbf{U}_{iy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

for $i = 1, 2, \ldots, m_y$ and

$$\mathbf{U}_i = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{ia} \end{bmatrix},$$

for $i = (1+m_y), (2+m_y), \ldots, (m_y+m_a)$. In this paper, we will assume that both $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_a$ are invertible. Nevertheless, this assumption is not really necessary. For example, if $\boldsymbol{\Sigma}_a$ is singular, one can follow Xu et al. (2012) to reformulate the original EIV model into a partial EIV model, where the new variance–covariance matrix is only restricted to those independent random elements of $\mathbf{A}$ and becomes invertible. With the reformulation of (5) and the stochastic model (6), it has become clear that the EIV model (1) can be identically treated mathematically as a nonlinear Gauss–Markoff model with unknown variance components; obviously, this is the most natural way to handle the estimation of variance components in EIV models, since conventional nonlinear estimation theory and methods can be adapted naturally. In other words, one can now freely use any numerical methods to solve for the parameter estimate of the nonlinear model and freely adopt any appropriate methods to estimate the variance components.

2.2 The weighted TLS estimate of the parameters without linearization

A procedure of variance component estimation almost always consists of two iteratively looped steps: one to compute the parameters and the other to compute the variance components. Given a set of initial values σ_{0iy}^2 ($i = 1, 2, \ldots, m_y$) and σ_{0ia}^2 ($i = 1, 2, \ldots, m_a$) for the unknown variance components in (2), we can solve the following weighted TLS minimization problem:

$$\begin{aligned}
\min: \quad S(\boldsymbol{\beta}, \overline{\mathbf{a}}) &= [(\boldsymbol{\beta}^T \otimes \mathbf{I}_n)\overline{\mathbf{a}} - \mathbf{y}]^T \boldsymbol{\Sigma}_{0y}^{-1}[(\boldsymbol{\beta}^T \otimes \mathbf{I}_n)\overline{\mathbf{a}} - \mathbf{y}] \\
&\quad + (\overline{\mathbf{a}} - \mathbf{a})^T \boldsymbol{\Sigma}_{0a}^{-1}(\overline{\mathbf{a}} - \mathbf{a}), \tag{7}
\end{aligned}$$

to estimate the parameters $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$, where $\boldsymbol{\Sigma}_{0y}$ and $\boldsymbol{\Sigma}_{0a}$ are the initial variance–covariance matrices of $\mathbf{y}$ and $\mathbf{a}$, with the unknown variance components replaced by the corresponding given initial values σ_{0iy}^2 ($i = 1, 2, \ldots, m_y$) and σ_{0ia}^2 ($i = 1, 2, \ldots, m_a$), respectively.

Differentiating the objective function $S(\boldsymbol{\beta}, \overline{\mathbf{a}})$ of (7) with respect to $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$ yields:

$$\frac{\partial S(\boldsymbol{\beta}, \overline{\mathbf{a}})}{\partial \boldsymbol{\beta}} = 2\overline{\mathbf{A}}^T \boldsymbol{\Sigma}_{0y}^{-1}(\overline{\mathbf{A}}\boldsymbol{\beta} - \mathbf{y}), \tag{8a}$$

$$\frac{\partial S(\boldsymbol{\beta}, \overline{\mathbf{a}})}{\partial \overline{\mathbf{a}}} = 2\boldsymbol{\Sigma}_{0a}^{-1}(\overline{\mathbf{a}} - \mathbf{a})$$
$$+ 2(\boldsymbol{\beta} \otimes \mathbf{I}_n)\boldsymbol{\Sigma}_{0y}^{-1}\{(\boldsymbol{\beta}^T \otimes \mathbf{I}_n)\overline{\mathbf{a}} - \mathbf{y}\}. \tag{8b}$$

By equating the partial derivatives (8) to zero, we can obtain the weighted TLS estimates of $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$, which are, respectively, denoted by $\hat{\boldsymbol{\beta}}$ and $\hat{\overline{\mathbf{a}}}$ (and accordingly $\hat{\overline{\mathbf{A}}}$). They are given as follows:

$$\hat{\boldsymbol{\beta}} = \left[(\hat{\overline{\mathbf{A}}})^T \boldsymbol{\Sigma}_{0y}^{-1}\hat{\overline{\mathbf{A}}}\right]^{-1}(\hat{\overline{\mathbf{A}}})^T \boldsymbol{\Sigma}_{0y}^{-1}\mathbf{y}, \tag{9a}$$

$$\hat{\overline{\mathbf{a}}} = [\boldsymbol{\Sigma}_{0a}^{-1} + \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})^T \otimes \boldsymbol{\Sigma}_{0y}^{-1}]^{-1}[\boldsymbol{\Sigma}_{0a}^{-1}\mathbf{a} + \hat{\boldsymbol{\beta}} \otimes (\boldsymbol{\Sigma}_{0a}^{-1}\mathbf{y})]. \tag{9b}$$

Since the dimension of $\boldsymbol{\Sigma}_{0a}$ is much larger than that of $\boldsymbol{\Sigma}_{0y}$, it is not desirable to directly invert the normal matrix $[\boldsymbol{\Sigma}_{0a}^{-1} + \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})^T \otimes \boldsymbol{\Sigma}_{0y}^{-1}]$ in (9b). By following the inverse of a summed matrix (Magnus and Neudecker 1988), we can rewrite the inverse of this normal matrix:

$$[\boldsymbol{\Sigma}_{0a}^{-1} + \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})^T \otimes \boldsymbol{\Sigma}_{0y}^{-1}]^{-1}$$
$$= \boldsymbol{\Sigma}_{0a} - \boldsymbol{\Sigma}_{0a}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n)\mathbf{E}^{-1}((\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n)\boldsymbol{\Sigma}_{0a},$$

where

$$\mathbf{E} = \boldsymbol{\Sigma}_{0y} + ((\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n)\boldsymbol{\Sigma}_{0a}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n). \tag{9c}$$

As a result, (9b) can finally be simplified after some technical derivations as follows

$$\hat{\overline{\mathbf{a}}} = \mathbf{a} + \boldsymbol{\Sigma}_{0a}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n)\mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}). \tag{9d}$$

Obviously, (9d) should be much more efficient to compute $\hat{\overline{\mathbf{a}}}$ than (9b), in conjunction with (9a).

The weighted TLS estimates of $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$ in (9) are numerically solved iteratively. Such an algorithm is known to be a variant of the Gauss–Newton method (see, e.g., Dennis and Schnabel 1996; Xu et al. 2012) and converges linearly. If both $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_a$ are identity matrices of different orders, then the estimate of $\boldsymbol{\beta}$ can be elegantly solved as the eigenvector problem with the least eigenvalue (see, e.g., Pearson 1901; Golub and Loan 1980); otherwise, alternative algorithms can be found in, e.g., Huffel and Vandewalle (1991), Markovsky and Huffel (2007), Schaffrin and Wieser (2008). We should note that the weighted TLS estimates (9) of $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$ depend on σ_{0iy}^2 $(i = 1, 2, \ldots, m_y)$ and σ_{0ia}^2 $(i = 1, 2, \ldots, m_a)$. Thus they should be iteratively solved using the updated set of the estimated variance components to be described in Sect. 3.

2.3 Alternative iterative solution of the nonlinear Gauss–Markoff model through linearization

An alternative way to solve the EIV model (5) once it has been formulated as a nonlinear Gauss–Markoff model is based on linear approximation to the corresponding nonlinear model; this linear approximation procedure has been well known and elegantly documented in the widely accessible mathematical literature (see, e.g., Marquardt 1963; Bates and Watts 1980; Seber and Wild 1989; Björck 1996; Dennis and Schnabel 1996; Fletcher 2000). Geodesists may know better the publication by Pope (1972), which is based on the procedure of Marquardt (1963). Applying the Taylor expansion to the nonlinear observation equations (5) at the approximate values $\boldsymbol{\beta}_i$ and $\overline{\mathbf{a}}_i$ of the ith iteration and truncating it up to the linear approximation, one can obtain the following linearized observation equations:

$$\begin{bmatrix} \Delta\mathbf{y} \\ \Delta\mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \overline{\mathbf{A}}_i\boldsymbol{\beta}_i \\ \mathbf{a} - \overline{\mathbf{a}}_i \end{bmatrix}$$
$$\approx \begin{bmatrix} \overline{\mathbf{A}}_i & \boldsymbol{\beta}_i^T \otimes \mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{\beta}_i \\ \delta\overline{\mathbf{a}}_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_a \end{bmatrix}, \tag{10}$$

where $\delta\boldsymbol{\beta}_i = \boldsymbol{\beta} - \boldsymbol{\beta}_i$ and $\delta\overline{\mathbf{a}}_i = \overline{\mathbf{a}} - \overline{\mathbf{a}}_i$. As a result, the well-known weighted LS estimate of $\delta\boldsymbol{\beta}_i$ and $\delta\overline{\mathbf{a}}_i$ can be readily obtained, which will be denoted by $\delta\hat{\boldsymbol{\beta}}_i$ and $\delta\hat{\overline{\mathbf{a}}}_i$, respectively; these estimates are then used to compute the approximate values $\boldsymbol{\beta}_{i+1} = \boldsymbol{\beta}_i + \delta\hat{\boldsymbol{\beta}}_i$ and $\overline{\mathbf{a}}_{i+1} = \overline{\mathbf{a}}_i + \delta\hat{\overline{\mathbf{a}}}_i$ for the next iteration, starting from some initial values $\boldsymbol{\beta}_0$ and $\overline{\mathbf{a}}_0$. The iteration will be terminated when $|\delta\hat{\boldsymbol{\beta}}_s| < e$ and $|\delta\hat{\overline{\mathbf{a}}}_s| < e$ at the sth step of iteration, where e is some predefined very small positive quantity. Thus, the final weighted LS estimates are $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{s-1} + \delta\hat{\boldsymbol{\beta}}_s$ and $\hat{\overline{\mathbf{a}}} = \overline{\mathbf{a}}_{s-1} + \delta\hat{\overline{\mathbf{a}}}_s$; they should be essentially equal to the corresponding estimates given in Sect. 2.2.

We may note that during the iteration, say at the ith iteration, one may compute the following quantities:

$$\begin{bmatrix} \hat{\boldsymbol{\epsilon}}_i \\ \hat{\boldsymbol{\epsilon}}_{ai} \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \overline{\mathbf{A}}_i\boldsymbol{\beta}_i \\ \mathbf{a} - \overline{\mathbf{a}}_i \end{bmatrix} - \begin{bmatrix} \overline{\mathbf{A}}_i & \boldsymbol{\beta}_i^T \otimes \mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix} \begin{bmatrix} \delta\hat{\boldsymbol{\beta}}_i \\ \delta\hat{\overline{\mathbf{a}}}_i \end{bmatrix}. \tag{11}$$

We must note that $\hat{\boldsymbol{\epsilon}}_i$ and $\hat{\boldsymbol{\epsilon}}_{ai}$ are not the residuals of measurements $\mathbf{y}$ and $\mathbf{a}$ in the least squares sense. Actually, as is well known (see, e.g., Marquardt 1963; Bates and Watts 1988; Seber and Wild 1989; Björck 1996; Dennis and Schnabel 1996; Fletcher 2000), when the iteration described above converges mathematically, then $\delta\hat{\boldsymbol{\beta}}_s = \mathbf{0}$ and $\delta\hat{\overline{\mathbf{a}}}_s = \mathbf{0}$. In other words, when the solutions converge to the weighted LS estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\overline{\mathbf{a}}}$, the weighted LS estimates of residuals are given by

$$\begin{bmatrix} \hat{\boldsymbol{\epsilon}} \\ \hat{\boldsymbol{\epsilon}}_a \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \hat{\overline{\mathbf{A}}}\hat{\boldsymbol{\beta}} \\ \mathbf{a} - \hat{\overline{\mathbf{a}}} \end{bmatrix}. \tag{12}$$

Since residuals are almost always denoted using the letter **r** in the above widely accessible mathematical and statistical literature, we will follow them and denote the residuals of **y** and **a** by $\mathbf{r}_y$ and $\mathbf{r}_a$, respectively. Thus, we have

$$\mathbf{r}_y = \mathbf{y} - \hat{\bar{\mathbf{A}}}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\epsilon}}, \tag{13a}$$

$$\mathbf{r}_a = \mathbf{a} - \hat{\bar{\mathbf{a}}} = \hat{\boldsymbol{\epsilon}}_a. \tag{13b}$$

3 MINQUE estimation of variance components

For a linear model with a deterministic design matrix **A** and the stochastic model (2a), there exist a number of well-established methods to estimate the variance components of (2a) such as Helmert method (Helmert 1907; Schaffrin 1983; Grafarend 1985), minimum norm quadratic unbiased estimation (MINQUE) (Rao 1971a; Rao and Kleffe 1988), minimum variance quadratic unbiased estimation (see, e.g., Rao 1971b; LaMotte 1973; Koch 1999), maximum likelihood and marginal/restricted maximum likelihood methods (see, e.g., Kubik 1966, 1970; Hartley and Rao 1967; Patterson and Thompson 1975; Koch 1986) as well as the least squares method of variance components (Pukelsheim 1976; Teunissen and Amiri-Simkooei 2008). The first three classes of methods are intuitively based on quadratic forms of measurements to construct an estimator for the variance components, while the maximum likelihood and marginal maximum likelihood methods have to assume that the joint probability distribution of measurements is given. In contrast, the least squares method is purely algebraic in nature. Pukelsheim (1976) used the residuals of measurements to form the derived linear model with variance components as the unknown model parameters. The corresponding estimator of variance components is unbiased. After a further assumption on the fourth moments of measurement errors, Pukelsheim (1976) proved that the least squares estimator of variance components is also of minimum variance (see also Teunissen and Amiri-Simkooei 2008). Pukelsheim and Styan (1978) further extended the least squares method of variance component estimation to a multivariate linear model. For more details on variance component estimation, the reader is referred to Helmert (1907), Rao and Kleffe (1988), Searle et al. (1992) and Koch (1999) in the case of linear models, and Xu et al. (2006) (see also Koch and Kusche 2007; Xu et al. 2007b), Xu (2009) and Eshagh (2010, 2011) in the case of ill-posed linear models.

In this paper, we will focus on the MINQUE method for the variance component estimation with the EIV models (1) and (2). We will not assume any probability distributions and the fourth moments of the random errors of both **y** and **A**. Although the MINQUE method is reported to possess almost all good properties of variance component estima-

tion (if estimable) in the linear model such as invariance, minimum norm, and further the minimum variance under the extra assumption of normal distributions (see, e.g., Rao 1971b; LaMotte 1973; Pukelsheim and Styan 1978; Rao and Kleffe 1988; Wulff and Birkes 2005), these nice statistical properties are not valid anymore here, since the EIV model (1) is essentially nonlinear. From this point of view, one may choose to use any method of his/her choice for variance component estimation, for two reasons: (i) there exist no superior methods for variance component estimation in linear models that are mathematically proved to perform the best in any situation; and (ii) optimal statistical properties of any variance component estimation methods established for linear models are generally not valid for nonlinear models.

3.1 Adapting variance component estimation methods to nonlinear models

For linear models of the form $\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with a non-random design matrix **A**, $E\{\boldsymbol{\epsilon}\} = \mathbf{0}$ and $E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\} = \boldsymbol{\Sigma} = \sum_{i=1}^{p} \mathbf{U}_i \sigma_i^2$, most variance component estimation methods lead to a set of equations $\mathbf{S}\boldsymbol{\sigma} = \mathbf{q}$ to be solved for $\boldsymbol{\sigma} = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2]$, where **S** and **q** are functions of the positive (semi-)definite matrices $\mathbf{U}_i$, the design matrix **A** and the observations **y** or the residual estimates $\mathbf{r} = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}$. In particular, the design matrix **A** enters through the projection matrix:

$$\mathbf{Z}_A = \mathbf{A}(\mathbf{A}^T \boldsymbol{\Sigma}_0^{-1} \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\Sigma}_0^{-1} = \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T \boldsymbol{\Sigma}_0^{-1}, \tag{14}$$

on the range (column space) $\mathcal{R}(\mathbf{A})$ of **A**, where $\mathbf{N} = \mathbf{A}^T \boldsymbol{\Sigma}_0^{-1} \mathbf{A}$ and $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{p} \mathbf{U}_i \sigma_{i0}^2$ is based on some approximate values σ_{i0}^2 of σ_i^2. For example, in the case of the MINQUE method, we have the elements s_{ij} of **S** and those q_i of **q** as follows:

$$s_{ij} = \text{tr}\{\boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{Z}_A)\mathbf{U}_i \boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{Z}_A)\mathbf{U}_j\}$$
$$= \text{tr}\{\mathbf{P}\mathbf{U}_i\mathbf{P}\mathbf{U}_j\} \tag{15a}$$

for $i, j = 1, 2, \ldots, p$ and,

$$q_i = \mathbf{r}^T \boldsymbol{\Sigma}_0^{-1} \mathbf{U}_i \boldsymbol{\Sigma}_0^{-1} \mathbf{r}, \tag{15b}$$

for $i = 1, 2, \ldots, p$, where

$$\mathbf{P} = \boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{Z}_A)$$
$$= \boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T \boldsymbol{\Sigma}_0^{-1})$$
$$= \boldsymbol{\Sigma}_0^{-1}(\mathbf{I} - \mathbf{H}\boldsymbol{\Sigma}_0^{-1})$$
$$= \boldsymbol{\Sigma}_0^{-1}\mathbf{R}, \tag{15c}$$

with $\mathbf{H} = \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T$ and $\mathbf{R} = \mathbf{I} - \mathbf{H}\boldsymbol{\Sigma}_0^{-1}$. Here $\text{tr}(\cdot)$ stands for the trace of a square matrix.

In the linear case, $\mathbf{A}\hat{\boldsymbol{\beta}} = \mathbf{Z}_A\mathbf{y}$ is the weighted LS estimate of the observables $\mathbf{A}\boldsymbol{\beta}$. It is the closest point to **y** from the

linear subspace $\mathcal{R}(\mathbf{A})$. In the case of a nonlinear model $\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\epsilon}$, the least squares solution seeks the estimate $\hat{\boldsymbol{\beta}}$ such that $\mathbf{f}(\hat{\boldsymbol{\beta}}) = \mathbf{Z}_{\mathcal{M}}(\mathbf{y})$ is the closest point to $\mathbf{y}$ from the nonlinear manifold $\mathcal{M} = \{\mathbf{f}(\boldsymbol{\beta})|\boldsymbol{\beta} \in R^m\} \subset R^n$. In a small neighborhood of $\mathbf{f}(\hat{\boldsymbol{\beta}})$, the manifold $\mathcal{M}$ is close to the linear manifold spanned by the rows of the matrix:

$$\mathbf{A}(\hat{\boldsymbol{\beta}}) = \left.\frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$$

and the nonlinear projector $\mathbf{Z}_{\mathcal{M}}(\mathbf{y})$ gives the same value as $\mathbf{Z}_{A(\hat{\boldsymbol{\beta}})}$, where

$$\mathbf{Z}_{A(\hat{\boldsymbol{\beta}})} = \mathbf{A}(\hat{\boldsymbol{\beta}})\{[\mathbf{A}(\hat{\boldsymbol{\beta}})]^T \boldsymbol{\Sigma}_0^{-1} \mathbf{A}(\hat{\boldsymbol{\beta}})\}^{-1}[\mathbf{A}(\hat{\boldsymbol{\beta}})]^T \boldsymbol{\Sigma}_0^{-1}. \quad (16)$$

This means that a variance component estimation method can be adapted to a nonlinear model within a local type of approximation around $\mathbf{f}(\hat{\boldsymbol{\beta}})$. By simply replacing $\mathbf{Z}_A$ with $\mathbf{Z}_{A(\hat{\boldsymbol{\beta}})}$ in the equations $\mathbf{S}\boldsymbol{\sigma} = \mathbf{q}$ as well as $(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})$ with its nonlinear counterpart $\mathbf{r}(= \mathbf{y} - \mathbf{f}(\hat{\boldsymbol{\beta}}))$. It must be strongly emphasized that the method of linearization used here for obtaining the least squares solution $\hat{\boldsymbol{\beta}}$ and hence the residuals $\mathbf{r}$ and $\mathbf{A}(\hat{\boldsymbol{\beta}})$ is completely irrelevant to the suggested adaptation. We have already given two alternative methods in Sects. 2.2 and 2.3.

We should also note that although the matrix $\mathbf{S}$ is computed at the point of $\hat{\boldsymbol{\beta}}$, it is mathematically admissible to treat $\hat{\boldsymbol{\beta}}$ as some approximate values only to linearize the nonlinear model, as exactly done, for example, by Marquardt (1963) and Pope (1972) . In other words, it is appropriate to treat the matrix $\mathbf{S}$ as non-random. Actually, one must clearly distinguish two basic concepts: numerical solution and statistical evaluation of the solution. From the point of view of numerical solution, the matrix $\mathbf{S}$ is simply a consequence of linearization to iterate for the solution to a nonlinear model. One can completely avoid the matrix $\mathbf{S}$ using other numerical methods such as simulated annealing, genetic algorithms and/or global optimization methods of deterministic type. On the other hand, to statistically evaluate the accuracy of the optimal solution, it is totally incorrect to treat the matrix $\mathbf{S}$ as random and then apply the error propagation law to compute the accuracy of the solution; instead, in this case, one must depend on the differential geometry of the original nonlinear model, as well documented in, for example, Bates and Watts (1980, 1988) and Seber and Wild (1989) for the error estimate of the weighted LS solution in nonlinear models.

3.2 Adapting the MINQUE method to the EIV model

The MINQUE method for estimation of variance components was proposed by Rao (1971a) (see also Rao and Kleffe 1988; Searle et al. 1992) to minimize the Euclidean norm of $\mathbf{B}$ in the quadratic form $\mathbf{y}^T \mathbf{B} \mathbf{y}$ of measurements in the standard linear model with a deterministic design/coefficient matrix $\mathbf{A}$, subject to constraints such as invariance and unbiasedness. Since the functional EIV model (4) is nonlinear, the MINQUE method is not directly applicable to estimate the variance components in (2), and as a result, does not necessarily possess all its optimal properties in the case of linear Gauss–Markoff models. From this point of view, the choice of the MINQUE method in this paper is basically due to its convenience of using the positive (semi-)definite matrices $\mathbf{U}_{iy}$ and $\mathbf{U}_{ja}$ to naturally construct quadratic forms.

For adapting the MINQUE method to the EIV model (5), we need only to note that the design matrix in (10) is accordingly given by

$$\mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{a}}) = \begin{bmatrix} \hat{\bar{\mathbf{A}}} & (\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix}. \quad (17)$$

With the linearized EIV model (10) and the stochastic model (2), we can now adapt the MINQUE method to estimate the variance components. For brevity of notations, we collect the variance components of $\mathbf{y}$ in (2a) and those of $\mathbf{a}$ (or $\mathbf{A}$) in (2b) into two vectors $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$, respectively. The residuals of $\mathbf{y}$ and $\mathbf{a}$ from the nonlinear weighted TLS adjustment by $\mathbf{r}_y$ and $\mathbf{r}_a$ have been defined in (13a) and (13b), respectively.

More specifically, we compute all the corresponding matrices such as $\mathbf{N}$, $\mathbf{Z}_A$, $\mathbf{P}$, $\mathbf{H}$ and $\mathbf{R}$ in (15) with the new design matrix $\mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{a}})$ of (17) and denote the corresponding matrix $\mathbf{P}$ of (15c) by:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}.$$

As a result, the corresponding MINQUE estimates of $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$ are derived by solving the following system of linear equations:

$$\begin{bmatrix} \mathbf{S}_y & \mathbf{S}_{ya} \\ \mathbf{S}_{ay} & \mathbf{S}_a \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\sigma}}_y \\ \hat{\boldsymbol{\sigma}}_a \end{bmatrix} = \begin{bmatrix} \mathbf{q}_y \\ \mathbf{q}_a \end{bmatrix}, \quad (18)$$

where the elements of $\mathbf{S}_y$, $\mathbf{S}_a$ and $\mathbf{S}_{ya}$ are computed, respectively, by the following equations:

$$\begin{aligned} s_y^{ij} &= \text{tr}\{\mathbf{P}\mathbf{U}_i\mathbf{P}\mathbf{U}_j\} \\ &= \text{tr}\left\{\begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}\begin{bmatrix} \mathbf{U}_{iy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}\begin{bmatrix} \mathbf{U}_{jy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right\} \\ &= \text{tr}(\mathbf{P}_y\mathbf{U}_{iy}\mathbf{P}_y\mathbf{U}_{jy}), \quad (19a) \end{aligned}$$

for $i, j = 1, 2, \ldots, m_y$,

$$\begin{aligned} s_a^{ij} &= \text{tr}\{\mathbf{P}\mathbf{U}_i\mathbf{P}\mathbf{U}_j\} \\ &= \text{tr}\left\{\begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{ia} \end{bmatrix}\begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{ja} \end{bmatrix}\right\} \\ &= \text{tr}(\mathbf{P}_a\mathbf{U}_{ia}\mathbf{P}_a\mathbf{U}_{ja}), \quad (19b) \end{aligned}$$

for $i, j = 1, 2, \ldots, m_a$, and

$$
\begin{aligned}
s_{ya}^{ij} &= \text{tr}\{\mathbf{P}\mathbf{U}_i\mathbf{P}\mathbf{U}_j\} \\
&= \text{tr}\left\{\begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}\begin{bmatrix} \mathbf{U}_{iy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{ja} \end{bmatrix}\right\} \\
&= \text{tr}(\mathbf{P}_{ay}\mathbf{U}_{iy}\mathbf{P}_{ya}\mathbf{U}_{ja}),
\end{aligned} \tag{19c}
$$

for $i = 1, 2, \ldots, m_y$ and $j = 1, 2, \ldots, m_a$. The matrices $\mathbf{P}_y$, $\mathbf{P}_a$ and $\mathbf{P}_{ya}$ are respectively equal to

$$
\mathbf{P}_y = \boldsymbol{\Sigma}_{0y}^{-1} - \boldsymbol{\Sigma}_{0y}^{-1}\mathbf{H}_y\boldsymbol{\Sigma}_{0y}^{-1},
$$

$$
\mathbf{P}_a = \boldsymbol{\Sigma}_{0a}^{-1} - \boldsymbol{\Sigma}_{0a}^{-1}\mathbf{H}_a\boldsymbol{\Sigma}_{0a}^{-1},
$$

and

$$
\mathbf{P}_{ya} = -\boldsymbol{\Sigma}_{0y}^{-1}\mathbf{H}_{ya}\boldsymbol{\Sigma}_{0a}^{-1}.
$$

The matrices $\mathbf{H}_{ay}$, $\mathbf{H}_a$ and $\mathbf{H}_y$ have been given in the Appendix. The elements of the two (sub-)vectors $\mathbf{q}_y$ and $\mathbf{q}_a$ on the right hand side of (18) are defined as

$$
\begin{aligned}
q_y^i &= \begin{bmatrix} \mathbf{r}_y \\ \mathbf{r}_a \end{bmatrix}^T\begin{bmatrix} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{U}_{iy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{r}_y \\ \mathbf{r}_a \end{bmatrix} \\
&= \mathbf{r}_y^T\boldsymbol{\Sigma}_{0y}^{-1}\mathbf{U}_{iy}\boldsymbol{\Sigma}_{0y}^{-1}\mathbf{r}_y,
\end{aligned} \tag{20a}
$$

for $i = 1, 2, \ldots, m_y$, and

$$
\begin{aligned}
q_a^i &= \begin{bmatrix} \mathbf{r}_y \\ \mathbf{r}_a \end{bmatrix}^T\begin{bmatrix} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{ia} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{r}_y \\ \mathbf{r}_a \end{bmatrix} \\
&= \mathbf{r}_a^T\boldsymbol{\Sigma}_{0a}^{-1}\mathbf{U}_{ia}\boldsymbol{\Sigma}_{0a}^{-1}\mathbf{r}_a,
\end{aligned} \tag{20b}
$$

for $i = 1, 2, \ldots, m_a$.

Substituting (19) and (20) into (18), we can readily obtain the MINQUE estimates of the variance components $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$:

$$
\begin{bmatrix} \hat{\boldsymbol{\sigma}}_y \\ \hat{\boldsymbol{\sigma}}_a \end{bmatrix} = \begin{bmatrix} \mathbf{S}_y & \mathbf{S}_{ya} \\ \mathbf{S}_{ay} & \mathbf{S}_a \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{q}_y \\ \mathbf{q}_a \end{bmatrix}, \tag{21}
$$

if the coefficient matrix of (18) is regular or invertible.

Before closing this subsection, we make four remarks on the MINQUE estimates of the variance components $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$. (i) Although we have derived the estimates of the variance components through the linearization of the nonlinear EIV model, an alternative approach is to directly construct MINQUE-like quadratic forms with the nonlinear residuals, expand the nonlinear residuals to the linear approximation, apply the expectation operators to the quadratic forms with the linearized residuals and finally derive the equations (18) by removing the expectation operators, as can be clearly seen in Sect. 5; (ii) following the method of estimability analysis by Xu et al. (2007a), the total number of variance components $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$ cannot be larger than $(n-m)$; otherwise, the coefficient matrix in (18) would not be regular anymore and, as a result, $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$ are not estimable; (iii) the MINQUE estimates of $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$ are exact only under the assumption

of linear models. Since the original EIV model (1), or equivalently (4), is essentially nonlinear, we can only adapt the MINQUE method to estimate $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$, as given by (21). Thus, from the theoretical point of view, all the good statistical properties of the MINQUE estimate of variance components in the linear model such as invariance, unbiasedness and minimum variance (in the case of normal distributions) do not hold anymore in the nonlinear EIV model (4); and (iv) If the variance components are estimable, one can start with an initial set of variance components σ_{0iy}^2 ($i = 1, 2, \ldots, m_y$) and σ_{0ia}^2 ($i = 1, 2, \ldots, m_a$), and then use (18) to compute the iterative MINQUE estimates of $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$ numerically, with the elements of $\mathbf{q}_y$ and $\mathbf{q}_a$ given by (20).

3.3 Two special structures of variance components

The first special structure of variance components assumes only two variance components in (2), one for the measurements $\mathbf{y}$ and the other for the measured matrix data $\mathbf{a}$. As a result, the variance component model (2) can now be rewritten as follows:

$$
\boldsymbol{\Sigma}_y = \mathbf{W}_y^{-1}\sigma_y^2, \tag{22a}
$$

$$
\boldsymbol{\Sigma}_a = \mathbf{W}_a^{-1}\sigma_a^2, \tag{22b}
$$

where $\mathbf{W}_y$ and $\mathbf{W}_a$ are the weight matrices of $\mathbf{y}$ and $\mathbf{a}$, respectively.

With the stochastic structure (22), according to Horn et al. (1975) and Rao and Kleffe (1988), it is easy to prove that when the MINQUE estimates of σ_y^2 and σ_a^2 using (18) converge in association with the linearized observation model (10), and after some (light) derivations, we must equivalently have the following simplified equations:

$$
\widehat{\sigma_y^2} = \mathbf{r}_y^T\mathbf{W}_y\mathbf{r}_y/\text{tr}(\mathbf{R}_y), \tag{23a}
$$

to estimate the variance component σ_y^2, and

$$
\widehat{\sigma_a^2} = \mathbf{r}_a^T\mathbf{W}_a\mathbf{r}_a/\text{tr}(\mathbf{R}_a), \tag{23b}
$$

to estimate the variance component σ_a^2. $\mathbf{R}_y$ and $\mathbf{R}_a$ have been defined in (49) of the Appendix, except that now the initial values of σ_y^2 and σ_a^2 are replaced by $\widehat{\sigma_y^2}$ and $\widehat{\sigma_a^2}$, respectively. Actually, the two quantities $\text{tr}(\mathbf{R}_y)$ and $\text{tr}(\mathbf{R}_a)$ serve as the degrees of freedom for $\widehat{\sigma_y^2}$ and $\widehat{\sigma_a^2}$, respectively, and are often called the redundant observational numbers in the geodetic literature. Since both $\mathbf{R}_y$ and $\mathbf{R}_a$ contain $\widehat{\sigma_y^2}$ and $\widehat{\sigma_a^2}$, the estimates of variance components as given by (23) have to be solved iteratively.

The second special structure of variance components assumes one variance component for the measured matrix data $\mathbf{a}$ only and further assume that the measurements $\mathbf{y}$ consist of a number of independent subgroups or different types of measurements. Then the corresponding variance component model (2) can now be rewritten as follows:

$$\mathbf{\Sigma}_y = \begin{bmatrix} \mathbf{W}_1^{-1}\sigma_{1y}^2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2^{-1}\sigma_{2y}^2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{m_y}^{-1}\sigma_{m_y y}^2 \end{bmatrix}, \quad (24\text{a})$$

$$\mathbf{\Sigma}_a = \mathbf{W}_a^{-1}\sigma_a^2, \quad (24\text{b})$$

where $\mathbf{W}_i$ are the weight matrices of the corresponding subgroups or types of measurements $\mathbf{y}$, respectively, and σ_{iy}^2 $(i = 1, 2, \ldots, m_y)$ are the corresponding variance components of $\mathbf{y}$.

Following the same rationale as in the derivation of (23), we can readily obtain the estimates of the variance components σ_{iy}^2 $(i = 1, 2, \ldots, m_y)$ and σ_a^2 below:

$$\widehat{\sigma_{iy}^2} = \mathbf{r}_{iy}^T \mathbf{W}_i \mathbf{r}_{iy} / \text{tr}(\mathbf{R}_{iy}), \quad (25\text{a})$$

for $i = 1, 2, \ldots, m_y$, and

$$\widehat{\sigma_a^2} = \mathbf{r}_a^T \mathbf{W}_a \mathbf{r}_a / \text{tr}(\mathbf{R}_a), \quad (25\text{b})$$

where $\mathbf{r}_{iy}$ and $\mathbf{R}_{iy}$ are the residual sub-vectors of $\mathbf{r}_y$ and the diagonal submatrices of $\mathbf{R}_y$, corresponding to the ith subgroup or type of measurements $\mathbf{y}$, respectively.

4 Estimability and stability analysis

From the point of view of parameter estimation, the EIV model (1) can be equivalently reformulated as a nonlinear Gauss–Markoff model. However, as far as the variance components are concerned, the combination of the functional model (1) with the stochastic model (2) is very special. Looking at the second part (3b) or (4b) of the EIV models, we clearly see that the measurements $\mathbf{A}$ contribute no direct redundant measurements to the estimation of the variance components of $\mathbf{A}$. In this section, we will investigate how this peculiarity would affect the estimation of variance components described in (2a) and (2b). We emphasize that the following proof of inestimability of variance components, in association with the EIV model (4) with the stochastic model (27), depends solely on quadratic forms of the residuals of the measurements $\mathbf{y}$ and $\mathbf{A}$; the proof is independent of methods used for variance component estimation, though MINQUE and/or Helmert methods may be formally applied to derive an estimator of variance components, as is the case in Sect. 3.

4.1 Estimability analysis

For a standard linear model, namely, the linear model (1) with a deterministic (or non-random) coefficient matrix $\mathbf{A}$, together with the stochastic model (2a), a linear function $\mathbf{f}^T\boldsymbol{\sigma}$ of the variance components is said to be unbiasedly

estimable, if there exists a quadratic estimator $\mathbf{y}^T\mathbf{By}$ such that

$$E(\mathbf{y}^T\mathbf{By}) = \mathbf{f}^T\boldsymbol{\sigma}, \quad (26)$$

where $\mathbf{f}$ is a given vector, $\boldsymbol{\sigma}$ is the vector consisting of all the variance components of $\mathbf{y}$ to be estimated, and $\mathbf{B}$ is symmetric and should be independent of both $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ (or $\mathbf{\Sigma}_y$). Simply speaking, variance components are estimable, if they can be uniquely determined from the derived normal equations for the variance components. In other words, the normal matrix of the linear equations for the variance components can be regularly inverted. For more details on the concept of estimability, the reader is referred to Pincus (1974), Schaffrin (1983), Rao and Kleffe (1988) and Xu et al. (2007a).

In this section, the question of our concern is whether the variance components of the EIV model (1) can be uniquely determined from the given measurements $\mathbf{y}$ and $\mathbf{A}$. To start with an estimability analysis, let us assume a commonly encountered stochastic structure (2a) for the measurements $\mathbf{y}$ and further assume that the elements at each column of $\mathbf{A}$ are of the same type of measurements with the same accuracy. In other words, the variance structure for the elements at the jth column of $\mathbf{A}$ can be uniquely represented using one unknown variance component, i.e., $\mathbf{\Sigma}_{aj} = \mathbf{I}\sigma_{ja}^2$. Accordingly, the stochastic model (2b) for the random matrix $\mathbf{A}$ can be rewritten as follows:

$$\mathbf{\Sigma}_a = \sum_{i=1}^m \mathbf{U}_{ia}\sigma_{ia}^2 = \text{diag}(\sigma_{ia}^2) \otimes \mathbf{I}_n, \quad (27)$$

where $\mathbf{U}_{ia}$ is a diagonal block matrix with its ith block being equal to the identity matrix $\mathbf{I}_n$, and $\text{diag}(\sigma_{ia}^2)$ is a diagonal matrix with its ith diagonal element being equal to σ_{ia}^2.

Under the assumption of (27) and given a set of initial values σ_{0iy}^2 $(i = 1, 2, \ldots, m_y)$ and σ_{0ia}^2 $(i = 1, 2, \ldots, m)$, we can obtain

$$\mathbf{\Sigma}_{0a}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n) = (\text{diag}(\sigma_{0ia}^2) \otimes \mathbf{I}_n)(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n) = \mathbf{d}_{0\beta} \otimes \mathbf{I}_n, \quad (28\text{a})$$

where

$$\mathbf{d}_{0\beta} = \text{diag}(\sigma_{0ia}^2)\hat{\boldsymbol{\beta}} = \begin{bmatrix} \sigma_{01a}^2\hat{\beta}_1 \\ \sigma_{02a}^2\hat{\beta}_2 \\ \vdots \\ \sigma_{0ma}^2\hat{\beta}_m \end{bmatrix}.$$

Inserting (28a) into (9c) yields

$$\mathbf{E} = \mathbf{\Sigma}_{0y} + c_\beta \mathbf{I}_n, \quad (28\text{b})$$

where

$$c_\beta = \mathbf{d}_{0\beta}^T \hat{\boldsymbol{\beta}} = \sum_{i=1}^m \sigma_{0ia}^2\hat{\beta}_i^2.$$

Substituting (28a) and (28b) into (9d), we can readily obtain the weighted TLS estimate for each column of $\overline{\mathbf{A}}$, which is denoted by $\hat{\bar{\mathbf{a}}}_i$ and given as follows:

$$\hat{\bar{\mathbf{a}}}_i = \mathbf{a}_i + \sigma_{0ia}^2 \hat{\beta}_i \mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}) \tag{29}$$

for $i = 1, 2, \ldots, m$, where $\hat{\bar{\mathbf{a}}}_i$ and $\mathbf{a}_i$ are the ith columns of $\hat{\overline{\mathbf{A}}}$ and $\mathbf{A}$, respectively. As a result, the residual vector for each column of $\mathbf{A}$ is equal to

$$\mathbf{r}_{ia} = -\sigma_{0ia}^2 \hat{\beta}_i \mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}) \tag{30}$$

for $i = 1, 2, \ldots, m$. It is surprising to see from (30) that all the residual vectors $\mathbf{r}_{ia}$ of columns of $\mathbf{A}$ are proportional to the transformed vector $\mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})$. In other words, under the stochastic models (2a) and (27), the correction vector to each column of $\mathbf{A}$ is proportional to $\mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})$ and its varying coefficient depends on the initial value of the variance component and the estimated parameter of the corresponding column elements of $\mathbf{A}$, i.e., σ_{0ia}^2 and $\hat{\beta}_i$. More precisely speaking, if we denote the ratios of the corresponding elements among the residual vectors $\mathbf{r}_{ia}$ by $\mathbf{r}_{1a} : \mathbf{r}_{2a} : \ldots : \mathbf{r}_{ma}$, then we have

$$\mathbf{r}_{1a} : \mathbf{r}_{2a} : \ldots : \mathbf{r}_{ma} = (\sigma_{01a}^2 \hat{\beta}_1) : (\sigma_{02a}^2 \hat{\beta}_2) : \ldots : (\sigma_{0ma}^2 \hat{\beta}_m). \tag{31}$$

Since the residual vectors $\mathbf{r}_{ia}$ of (30) are the starting point of any variance component estimation, and after taking into account the following equality:

$$E(\mathbf{r}_a^T \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{r}_a) = E(\mathbf{r}_{ia}^T \mathbf{r}_a)/\sigma_{0ia}^4,$$

we have

$$E(\mathbf{r}_{1a}^T \mathbf{r}_{1a}/\sigma_{01a}^4) : E(\mathbf{r}_{2a}^T \mathbf{r}_{2a}/\sigma_{02a}^4) : \ldots : E(\mathbf{r}_{ma}^T \mathbf{r}_{ma}/\sigma_{0ma}^4)$$
$$= \beta_1^2 : \beta_2^2 : \ldots : \beta_m^2, \tag{32}$$

up to the second-order approximation of $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_a$. The ratio (32) is mathematically equivalent to saying that the submatrix $\mathbf{S}_a$ in the system of equations for the variance components is singular. Thus, we can immediately conclude from (30), (31) and (32) that the variance components under the combination of the stochastic models (2a) and (27) are not estimable. This result of inestimability indicates that we have no way of gaining any knowledge about such an EIV stochastic model from the measurements $\mathbf{y}$ and $\mathbf{A}$. Our result of inestimability remains valid even if there are only two unknown variance components for the elements of $\mathbf{A}$ so far as each group of elements is of the same accuracy. The proof is trivial, since one can simply add constraints on σ_{ia}^2 ($i = 1, 2, \ldots, m$) in (27).

If we further assume $\boldsymbol{\Sigma}_y = \sigma_y^2 \mathbf{I}_n$, then (29) and (30), respectively, become

$$\hat{\bar{\mathbf{a}}}_i = \mathbf{a}_i + \frac{\sigma_{0ia}^2 \hat{\beta}_i}{\sigma_{0y}^2 + c_\beta}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}) \tag{33}$$

and

$$\mathbf{r}_{ia} = -\frac{\sigma_{0ia}^2 \hat{\beta}_i}{\sigma_{0y}^2 + c_\beta}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}) \tag{34}$$

for $i = 1, 2, \ldots, m$. Obviously, the residual vectors (34) also satisfy the relation of ratios (31).

For the EIV model of regression type, namely,

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{A}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \tag{35}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ have been defined as in (1), $\mathbf{1}$ is a column vector with each element being equal to unity, β_0 and $\boldsymbol{\beta}_1$ are the scalar and ($m \times 1$) unknown parameters to be estimated, and the elements of $\mathbf{A}_1$ are measured with random errors, our result of inestimability remains valid, if $\mathbf{A}_1$ has the same stochastic structure as defined by (27) (with a varying number of variance components σ_{ia}^2 between two and m). The proof can be completed by following the partial EIV formulation of Xu et al. (2012) and then using the same rationales as used to prove the inestimability in this section. Since all the derivations are technical, we will not repeat the proof here.

4.2 Stability analysis

For arbitrary positive semi-definite matrices $\mathbf{U}_{iy}$ and $\mathbf{U}_{ia}$ in the stochastic models (2a) and (2b), we cannot obtain elegant inestimability results, as given in the above for the stochastic model (27). Nevertheless, after taking (9d) into account, we can rewrite the residuals (13) of the measurements $\mathbf{y}$ and $\mathbf{a}$ as follows:

$$\mathbf{r}_y = \mathbf{y} - \hat{\overline{\mathbf{A}}}\hat{\boldsymbol{\beta}}, \tag{36a}$$

$$\mathbf{r}_a = -\boldsymbol{\Sigma}_{0a}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n)\mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}). \tag{36b}$$

Since $\hat{\overline{\mathbf{A}}}$ is essentially the addition of $\mathbf{A}$ and its correction matrix $\delta\mathbf{A}$ which can be directly reconstructed from $-\mathbf{r}_a$, i.e., $\text{vec}(\delta\mathbf{A}) = -\mathbf{r}_a$, we can rewrite (36a) as

$$\mathbf{r}_y = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}} - \delta\mathbf{A}\hat{\boldsymbol{\beta}}. \tag{36c}$$

Obviously, if $\delta\mathbf{A}$ is small, then both the residual vectors $\mathbf{r}_y$ and $\mathbf{r}_a$ will be dominated by $(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})$. It is also clear from (36b) that the ($n \times m$) residual elements of $\mathbf{A}$ are determined completely through the combination of only the n elements of $(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})$, which are further related to the residuals $\mathbf{r}_y$ of $\mathbf{y}$. From this point of view, the condition of the coefficient matrix in (18) for the estimation of variance components may not be very good.

In this case, even if the variance components of the stochastic models (2a) and (2b) are theoretically estimable, we would like to know whether they can be accurately estimated. A proper measure to describe this aspect of estimation prop-

erty is the stability of the estimated variance components or, more precisely, the condition number of the coefficient matrix of the linear equations (18). Actually, if there exist zero eigenvalues for the coefficient matrix of (18), then the variance components are not estimable. Obviously, the condition number can be computed as soon as the functional EIV model (1) and its corresponding stochastic models (2a) and (2b) are given. To give the reader an impression of stability, we will simulate some examples and show their condition numbers in this section.

In fact, before we start the research on this topic, we have in mind an application of the methods developed in this paper to the Atlantic sea surface temperature data reported first by Mann and Emanuel (2006) and used by Schaffrin and Wieser (2008). With the above estimability analysis, it becomes clear now that we cannot gain the knowledge on the stochastic nature of this data set, since a reasonable assumption for data of this type would be the stochastic model (27). Examples of this type include dam deformation analysis, for example. Thus, instead, we will use problems of this type with only two model parameters β_1 and β_2 to demonstrate the stability of the linear equations (18).

More specifically, simulated examples can be represented by the following equations

$$y_i = a_{i1}\beta_1 + a_{i2}\beta_2 + \epsilon_i, \quad (i = 1, 2, \ldots, n). \quad (37)$$

We assume three variance components, namely, one for y_i, one for a_{i1} and one for a_{i2}. The number of measurements $\mathbf{y}$ is set to 120. For each experiment, we randomly generate two true values for the parameters β_1 and β_2 from the interval $[-5.0, 5.0]$ and those for a_{i1} and a_{i2} are from $[0.1, 100.0]$. Although the stochastic model (27) has been known to be not estimable, we will use this simulated example to numerically demonstrate the inestimability. For this purpose, we use the uniform distribution over $[0.1, 1.0]$ to generate the weights for y_i and assume the stochastic model $\mathbf{\Sigma}_{ai} = \mathbf{I}_n\sigma_{ia}^2(i = 1, 2)$ for the random elements of each column of $\mathbf{A}$. Applying the MINQUE method to the simulated example, we obtain the following system of equations for estimation of the variance components σ_y^2, σ_{1a}^2 and σ_{2a}^2, namely,

$$\begin{bmatrix} 33.900162 & 0.725844 & 3.113569 \\ 0.725844 & 3.935471 & 16.881530 \\ 3.113569 & 16.881530 & 72.414734 \end{bmatrix} \begin{bmatrix} \widehat{\sigma_y^2} \\ \widehat{\sigma_{1a}^2} \\ \widehat{\sigma_{2a}^2} \end{bmatrix}$$

$$= \begin{bmatrix} 184.037631 \\ 926.564864 \\ 3974.577300 \end{bmatrix}. \quad (38)$$

It is easy to show that the coefficient matrix of (38), denoted by $\mathbf{S}$, is numerically singular, with the three eigenvalues being equal to 0.000000, 33.660731 and 76.589635, respectively. The elements of its last two rows of $\mathbf{S}$, together

with the last two elements of the constant vector on the right hand side of (38), satisfy the following relationship:

$$\frac{\mathbf{S}(3, 1)}{\mathbf{S}(2, 1)} = \frac{\mathbf{S}(3, 2)}{\mathbf{S}(2, 2)} = \frac{\mathbf{S}(3, 3)}{\mathbf{S}(2, 3)} = \frac{3974.577300}{926.564864} = 4.28958.$$

Now we will use non-identity weight matrices for a_{i1} and a_{i2} to show the instability aspect of variance component estimations in the EIV model (1). The weight matrices for y_i, a_{i1} and a_{i2} are assumed to be diagonal and generated using uniform distributions. More specifically, the weights for y_i are drawn from $[0.1, 1.0]$ and those for a_{i1} and a_{i2} are from $[0.1, 100.0]$. The experiment is randomly repeated 200 times. Since we may have no prior knowledge on the values of variance components, we get the first set of 200 condition numbers by setting initial values of variance components to unity. For the purpose of comparison, we also use the first iteration result of the componentwise positive variance estimators (25) to get a set of better approximate values for the variance components and then compute the 200 condition numbers for the coefficient matrix of the linear equations (18). These two sets of condition numbers in logarithm with different approximate values of variance components are shown in Fig. 1, with the blue solid line corresponding to the first set of approximate values and the dotted red line to the second set of approximate values. It is clear from the blue solid line of this figure that the coefficient matrix of (18) can often be ill-conditioned, but if very good approximate values of the variance components are available, the condition of the coefficient matrix can be significantly improved. Nevertheless, the coefficient matrix can still become more ill-conditioned from time to time. We should like to warn, however, that a small condition number may not automati-
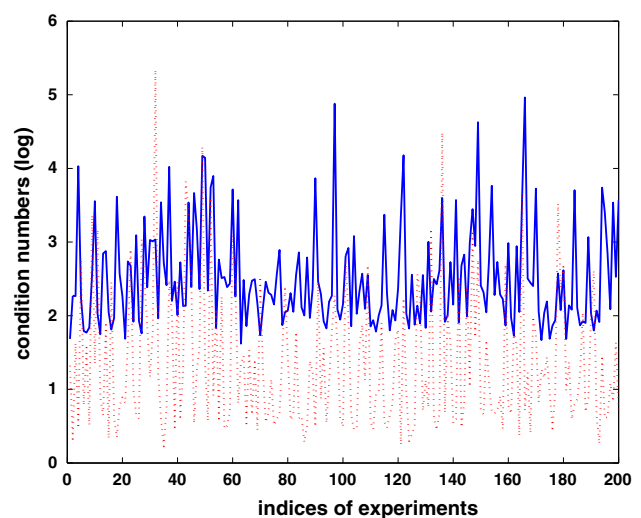

Fig. 1 Condition numbers of the coefficient matrix from the MINQUE estimation of variance components with two different sets of approximate variance components

cally mean a good estimation of variance components, since all the residuals of $\mathbf{y}$ and $\mathbf{A}$ are determined either by $(\mathbf{y} - \hat{\bar{\mathbf{A}}}\hat{\boldsymbol{\beta}})$ or $(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}})$.

5 Bias analysis

The weighted TLS estimate of $\boldsymbol{\beta}$ in the EIV model (1) is a solution to a nonlinear system of equations, as can be seen in Sect. 2.2 and from various different algorithms to solve TLS problems (see, e.g., Pearson 1901; Gerhold 1969; Golub and Loan 1980; Huffel and Vandewalle 1991; Markovsky and Huffel 2007; Schaffrin and Wieser 2008; Fang 2011; Xu et al. 2012), even if the stochastic model for both $\mathbf{y}$ and $\mathbf{A}$ is exactly given without any unknown variance and/or covariance parameters. In fact, the reformulated EIV model (4) is a nonlinear adjustment/regression model with peculiar features. Bates and Watts (1980) developed geometrical measures to diagnose the nonlinearity of a nonlinear regression model; applications in geodesy and geophysics can be found, e.g., in Teunissen (1989b). For a general nonlinear regression model, statistical consequences of model nonlinearity have been substantially investigated. Approximate confidence regions and the bias of the nonlinear LS estimate of $\boldsymbol{\beta}$ were worked out for a general nonlinear regression model by Beale (1960) and Box (1971), respectively. Clarke (1980) derived the first and second moments of the nonlinear LS estimate of $\boldsymbol{\beta}$. For some more details on nonlinear regression, the reader can refer to, e.g., Ratkowsky (1983) and Seber and Wild (1989). The first and second moments of the nonlinear LS estimate can also be found in the geodetic literature (see, e.g., Teunissen 1989a). Xu et al. (2012) applied the theory and methods by Beale (1960) and Box (1971) to investigate the statistical aspects of parameter estimation in the nonlinear EIV model (4).

However, nothing has been done to analyze finite sample statistical aspects of variance component estimation in a nonlinear regression/adjustment model, to our best knowledge. Wang and Davidian (1996), Wang et al. (1998) and Li et al. (2009) derived an asymptotic bias analysis of the naive estimator of variance components of $\mathbf{y}$. Nevertheless, these authors ignored the effect of measurement errors of $\mathbf{A}$ and assumed that the number of measurements $\mathbf{y}$ was infinite. In this paper, we will directly follow the weighted TLS method to handle the EIV model (1) and carry out the finite sample bias analysis. Following the same rationale as in Xu et al. (2012), we know that the estimates of variance components, as derived in Sect. 3, cannot be unbiased. We will analyze the local biases of the approximate MINQUE estimates (21) of variance components $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$, if they are estimable. In the remainder of this section, we will assume: (i) that both $\mathbf{y}$ and $\mathbf{A}$ are normally distributed, and (ii) that $\mathbf{y}$ and $\mathbf{A}$ are statistically independent.

By saying that an estimator of variance components is *locally unbiased*, we follow the standard definition in the literature of variance component estimation (see, e.g., Rao and Kleffe 1988; Koch 1999) to mean that the estimator is unbiased under a given set of initial values of the variance components. To derive the biases of the approximate MINQUE variance components $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_a$, we apply the expectation operator to (21):

$$E\left\{\begin{bmatrix} \hat{\sigma}_y \\ \hat{\sigma}_a \end{bmatrix}\right\} = \begin{bmatrix} \mathbf{S}_y & \mathbf{S}_{ya} \\ \mathbf{S}_{ay} & \mathbf{S}_a \end{bmatrix}^{-1} E\left\{\begin{bmatrix} \mathbf{q}_y \\ \mathbf{q}_a \end{bmatrix}\right\} \qquad (39)$$

if the coefficient matrix is invertible, where the elements of $\mathbf{q}_y$ and $\mathbf{q}_a$ are given by (20a) and (20b), respectively. To proceed to analyze the biases of the estimated variance components under the given initial values σ_{0iy}^2 ($i = 1, 2, \ldots, m_y$) and σ_{0ia}^2 ($i = 1, 2, \ldots, m_a$), we will have to focus on the two quadratic forms q_y^i of (20a) and q_a^i of (20b), which further depend on the residuals $\mathbf{r}_y$ of (13a) and $\mathbf{r}_a$ of (13b), respectively. Obviously, in the case of a linear model, the MINQUE estimates of variance components are also unbiased (Rao and Kleffe 1988).

5.1 Expanding the residuals $\mathbf{r}_y$ and $\mathbf{r}_a$ in terms of ϵ and ϵ_a

We now expand the residuals $\mathbf{r}_y$ and $\mathbf{r}_a$ of (13) at the true values of $\boldsymbol{\beta}$ and $\bar{\mathbf{a}}$ and truncate the Taylor's expansions up to the second-order approximation of ϵ and ϵ_a. In other words, all the terms of expansion with the third- and higher-order partial derivatives are neglected. As a result, we have

$$\begin{bmatrix} \mathbf{r}_y \\ \mathbf{r}_a \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \hat{\bar{\mathbf{A}}}\hat{\boldsymbol{\beta}} \\ \mathbf{a} - \hat{\bar{\mathbf{a}}} \end{bmatrix}$$
$$= \epsilon_{ya} - \mathbf{F}(\boldsymbol{\beta}, \bar{\mathbf{a}})\boldsymbol{\phi} - \frac{1}{2}\mathbf{J}\mathbf{K}\epsilon_{ya}, \qquad (40a)$$

where

$$\epsilon_{ya} = \begin{bmatrix} \epsilon \\ \epsilon_a \end{bmatrix}, \qquad (40b)$$

$$\mathbf{F}(\boldsymbol{\beta}, \bar{\mathbf{a}}) = \mathbf{F} = \begin{bmatrix} \bar{\mathbf{A}} & (\boldsymbol{\beta}^T \otimes \mathbf{I}_n) \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix}, \qquad (40c)$$

$$\boldsymbol{\phi} = \mathbf{K}\epsilon_{ya} + \mathbf{q}_{\beta\bar{a}}, \qquad (40d)$$
$$\mathbf{K} = \mathbf{N}^{-1}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}, \qquad (40e)$$
$$\mathbf{J} = \{\mathbf{G}_1\mathbf{K}\epsilon_{ya}, \mathbf{G}_2\mathbf{K}\epsilon_{ya}, \ldots, \mathbf{G}_n\mathbf{K}\epsilon_{ya}, \mathbf{0}, \ldots, \mathbf{0}\}^T, \qquad (40f)$$

and the matrices $\mathbf{G}_i$ of the second partial derivatives from the first nonlinear observation equations (4a) can be readily written as follows:

$$\mathbf{G}_i = \frac{\partial^2 y_i}{\partial \begin{bmatrix} \boldsymbol{\beta} \\ \bar{\mathbf{a}} \end{bmatrix} \partial (\boldsymbol{\beta}^T, \bar{\mathbf{a}}^T)}$$
$$= \begin{bmatrix} \mathbf{0} & (\mathbf{I}_m \otimes \mathbf{e}_i) \\ (\mathbf{I}_m \otimes \mathbf{e}_i^T) & \mathbf{0} \end{bmatrix} \qquad (40g)$$

for $i = 1, 2, \ldots n$. Here $\mathbf{e}_i = (0, \ldots 0, 1, 0, \ldots, 0)$ is an n-dimensional natural row vector, i.e., all the elements of $\mathbf{e}_i$ are equal to zero, except for the ith element being equal to unity. Except for the first n terms with non-zero values, all the remaining terms of $\mathbf{J}$ are equal to zero. The reason is this: because the expectation functions of $\mathbf{a}$ in (4b) are linear with respect to $\bar{\mathbf{a}}$ and have nothing to do with $\boldsymbol{\beta}$, the corresponding second partial derivatives with respect to $\bar{\mathbf{a}}$ and $\boldsymbol{\beta}$ are all equal to zero. We should note that when expanding the residuals $\mathbf{r}_y$ and $\mathbf{r}_a$ of (13), one should not represent their Taylor's expansions first up to the second-order approximation in terms of the estimated errors of $\boldsymbol{\beta}$ and $\bar{\mathbf{a}}$ and then use the second-order approximations of these errors $\boldsymbol{\phi}$ in (40d) to derive the final truncated expansion (40a); otherwise, one will end up with an expansion of the residuals $\mathbf{r}_y$ and $\mathbf{r}_a$, which correctly uses the second-order partial derivatives with respect to $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_a$ but incorrectly contains the third- and fourth-order terms of $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_a$, as is the case of Box (1971).

5.2 Expectations of the quadratic forms $\mathbf{q}_y$ and $\mathbf{q}_a$

To compute the expectations of $\mathbf{q}_y$ and $\mathbf{q}_a$ in (20), we insert the second-order truncated residuals (40a) into the quadratic forms (20a) and (20b) and apply the expectation operator to both of them, respectively. In the case of (20a), under the assumption of normal distribution, we have

$$
\begin{aligned}
E(q_y^i) &= E(\mathbf{r}_y^T \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{r}_y) \\
&= E\Big\{ \big(\boldsymbol{\epsilon} - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon} - \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a - \mathbf{d}_y\big)^T \\
&\quad \times \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \big(\boldsymbol{\epsilon} - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon} - \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a - \mathbf{d}_y\big) \Big\} \\
&= E\Big\{ \big(\boldsymbol{\epsilon} - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon} - \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a\big)^T \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \\
&\quad \times \big(\boldsymbol{\epsilon} - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon} - \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a\big) \Big\} + u_{y\sigma}^i \\
&= E\Big\{ \big(\boldsymbol{\epsilon} - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon}\big)^T \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \big(\boldsymbol{\epsilon} - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon}\big) \Big\} \\
&\quad + E\Big\{ \big(\mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a\big)^T \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a \Big\} \\
&\quad + u_{y\sigma}^i \\
&= \mathrm{tr}\{\mathbf{P}_y \mathbf{U}_{iy} \mathbf{P}_y \boldsymbol{\Sigma}_y\} \\
&\quad + \mathrm{tr}\Big\{ \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\Sigma}_a \Big\} + u_{y\sigma}^i \\
&= \mathbf{s}_y^i \boldsymbol{\sigma}_y + \mathbf{s}_{ya}^i \boldsymbol{\sigma}_a + u_{y\sigma}^i \quad (41a)
\end{aligned}
$$

for $i = 1, 2, \ldots, m_y$, where

$$
\begin{aligned}
u_{y\sigma}^i &= E\Big\{ \mathbf{d}_y^T \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{d}_y \Big\} \\
&= \mathrm{tr}\Big\{ \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{U}_{iy} \boldsymbol{\Sigma}_{0y}^{-1} E(\mathbf{d}_y \mathbf{d}_y^T) \Big\}, \quad (41b)
\end{aligned}
$$

$$
\mathbf{d}_y = \mathbf{F}_y \mathbf{q}_{\beta\bar{a}} + \frac{1}{2} \mathbf{J}_y \mathbf{K} \boldsymbol{\epsilon}_{ya}, \quad (41c)
$$

$\mathbf{F}_y = [\bar{\mathbf{A}}, (\boldsymbol{\beta}^T \otimes \mathbf{I}_n)]$, and $\mathbf{J}_y$ is the column vector of dimension n, corresponding to the first n none zero elements of $\mathbf{J}$ in (40f), and $\mathbf{s}_y^i$ and $\mathbf{s}_{ya}^i$ are the ith row vector of $\mathbf{S}_y$ and $\mathbf{S}_{ya}$ in Eq. (21), respectively.

Following the approach of bias analysis by Box (1971) (see also Xu and Shimada 2000), we can further write the quadratic term $\mathbf{q}_{\beta\bar{a}}$ in (40d) as follows

$$
\mathbf{q}_{\beta\bar{a}} = -\frac{1}{2} \mathbf{K} \mathbf{J} \mathbf{K} \boldsymbol{\epsilon}_{ya} + \mathbf{Q} \mathbf{C}^T (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{H} \boldsymbol{\Sigma}^{-1}) \boldsymbol{\epsilon}_{ya}, \quad (42)
$$

where the matrix $\mathbf{C}$ is the first-order correction to the matrix $\mathbf{F}(\hat{\boldsymbol{\beta}}, \hat{\bar{\mathbf{a}}})$ expanded at the point of $(\boldsymbol{\beta}, \bar{\mathbf{a}})$ and is given by

$$
\mathbf{C} = \begin{bmatrix} \mathbf{C}_y \\ \mathbf{C}_a \end{bmatrix}.
$$

$\mathbf{C}_a = \mathbf{0}$, since the measurements $\mathbf{a}$ are not functions of $\boldsymbol{\beta}$, and the submatrix $\mathbf{C}_y$ is equal to

$$
\mathbf{C}_y = [\mathbf{C}_y^1 \mathbf{K} \boldsymbol{\epsilon}_{ya}, \mathbf{C}_y^2 \mathbf{K} \boldsymbol{\epsilon}_{ya}, \ldots, \mathbf{C}_y^{(n+1)m} \mathbf{K} \boldsymbol{\epsilon}_{ya}].
$$

Here, the first m submatrices $\mathbf{C}_y^i$ ($i = 1, 2, \ldots, m$) and the remaining $(n \times m)$ submatrices $\mathbf{C}_y^j$ $\{i = (m+1), (m+2), \ldots, (n+1)m\}$ correspond to $\boldsymbol{\beta}$ and $\bar{\mathbf{a}}$, respectively, and are computed as follows:

$$
\mathbf{C}_y^i = \frac{\partial^2 \mathbf{y}}{\partial \beta_i \partial (\boldsymbol{\beta}, \bar{\mathbf{a}})^T} = [\mathbf{0}, (\mathbf{e}_i^T \otimes \mathbf{I}_n)]
$$

for $i = 1, 2, \ldots, m$, and

$$
\mathbf{C}_y^{m+j} = \frac{\partial^2 \mathbf{y}}{\partial \bar{a}_j \partial (\boldsymbol{\beta}, \bar{\mathbf{a}})^T} = [(\mathbf{I}_m \otimes \mathbf{I}_n) \mathbf{e}_j, \mathbf{0}]
$$

for $j = 1, 2, \ldots, (n \times m)$.

To compute u_σ^i of (41b), we first need to compute the expectation of $\mathbf{d}_y \mathbf{d}_y^T$, namely,

$$
\begin{aligned}
E(\mathbf{d}_y \mathbf{d}_y^T) &= E(\mathbf{F}_y \mathbf{q}_{\beta\bar{a}} \mathbf{q}_{\beta\bar{a}}^T \mathbf{F}_y^T) + \frac{1}{2} E(\mathbf{F}_y \mathbf{q}_{\beta\bar{a}} \boldsymbol{\epsilon}_{ya}^T \mathbf{K}^T \mathbf{J}_y^T) \\
&\quad + \frac{1}{2} E(\mathbf{J}_y \mathbf{K} \boldsymbol{\epsilon}_{ya} \mathbf{q}_{\beta\bar{a}}^T \mathbf{F}_y^T) \\
&\quad + \frac{1}{4} E(\mathbf{J}_y \mathbf{K} \boldsymbol{\epsilon}_{ya} \boldsymbol{\epsilon}_{ya}^T \mathbf{K}^T \mathbf{J}_y^T) \\
&= \mathbf{D}_{y1} + \mathbf{D}_{y2} + \mathbf{D}_{y3} + \mathbf{D}_{y4}, \quad (43)
\end{aligned}
$$

where

$$
\mathbf{D}_{y1} = E(\mathbf{F}_y \mathbf{q}_{\beta\bar{a}} \mathbf{q}_{\beta\bar{a}}^T \mathbf{F}_y^T),
$$

$$
\mathbf{D}_{y2} = \frac{1}{2} E(\mathbf{F}_y \mathbf{q}_{\beta\bar{a}} \boldsymbol{\epsilon}_{ya}^T \mathbf{K}^T \mathbf{J}_y^T),
$$

$$
\mathbf{D}_{y3} = \mathbf{D}_{y2}^T,
$$

$$
\mathbf{D}_{y4} = \frac{1}{4} E(\mathbf{J}_y \mathbf{K} \boldsymbol{\epsilon}_{ya} \boldsymbol{\epsilon}_{ya}^T \mathbf{K}^T \mathbf{J}_y^T).
$$

Because both $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_a$ have been assumed to be normally distributed with zero mean and to be statistically independent, we can use statistical results of quadratic forms (Searle

[1971](#)) and obtain, after some technical derivations and keeping in mind that $(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\mathbf{H})\boldsymbol{\Sigma}^{-1}\mathbf{F} = \mathbf{0}$, the expectations of the four matrices $\mathbf{D}_{y1}$, $\mathbf{D}_{y2}$, $\mathbf{D}_{y3}$ and $\mathbf{D}_{y4}$ of ([43](#)) as follows:

$$
\begin{aligned}
\mathbf{D}_{y1} &= E(\mathbf{F}_y \mathbf{q}_{\beta\bar{a}} \mathbf{q}_{\beta\bar{a}}^T \mathbf{F}_y^T) \\
&= \mathbf{F}_y E(\mathbf{q}_{\beta\bar{a}} \mathbf{q}_{\beta\bar{a}}^T) \mathbf{F}_y^T \\
&= \frac{1}{4} \mathbf{F}_y \mathbf{K} \mathbf{M}_q \mathbf{K}^T \mathbf{F}_y^T,
\end{aligned}
\tag{44a}
$$

$$
\begin{aligned}
\mathbf{D}_{y2} &= -\frac{1}{4} \overline{\mathbf{A}} \mathbf{Q}_\beta \overline{\mathbf{A}}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{M}_{jy} \\
&\quad -\frac{1}{4} (\boldsymbol{\beta}^T \otimes \mathbf{I}_n) \mathbf{Q}_{\bar{\mathbf{a}}\beta} \overline{\mathbf{A}}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{M}_{jy},
\end{aligned}
\tag{44b}
$$

$$
\mathbf{D}_{y4} = \frac{1}{4} \mathbf{M}_{jy},
\tag{44c}
$$

$\mathbf{M}_q$ is given by

$$
\mathbf{M}_q = \begin{bmatrix} \mathbf{M}_{jy} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},
\tag{44d}
$$

which is a square matrix with the dimension equal to the total number of elements of $\boldsymbol{\beta}$ and $\mathbf{A}$, and $\mathbf{M}_{jy}$ is an $(n \times n)$ matrix with its element given by

$$
\mathbf{M}_{jy}(i, j) = \mathrm{tr}(\mathbf{G}_i \mathbf{Q} \mathbf{G}_j \mathbf{Q}), \quad \text{for} \quad i, j = 1, 2, \dots, n.
\tag{44e}
$$

Finally, by inserting ([43](#)) and ([44](#)) into ([41b](#)), we can compute each element of u_σ^i ($i = 1, 2, \dots, n$).

In the similar manner to ([41a](#)), we compute the expectation $E(q_a^i)$ of the ith element of $\mathbf{q}_a$ as follows:

$$
\begin{aligned}
E(q_a^i) &= E(\mathbf{r}_a^T \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{r}_a) \\
&= E\Big\{ \big(\boldsymbol{\epsilon}_a - \mathbf{H}_a \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a - \mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon}\big)^T \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \big(\boldsymbol{\epsilon}_a \\
&\quad - \mathbf{H}_a \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a - \mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon}\big)\Big\} + u_{a\sigma}^i \\
&= E\Big\{ \big(\boldsymbol{\epsilon}_a - \mathbf{H}_a \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a\big)^T \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \big(\boldsymbol{\epsilon}_a \\
&\quad - \mathbf{H}_a \boldsymbol{\Sigma}_{0a}^{-1} \boldsymbol{\epsilon}_a\big)\Big\} \\
&\quad + E\Big\{ \big(\mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon}\big)^T \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\epsilon}\Big\} + u_{a\sigma}^i \\
&= \mathrm{tr}\{\mathbf{P}_a \mathbf{U}_{ia} \mathbf{P}_a \boldsymbol{\Sigma}_a\} \\
&\quad + \mathrm{tr}\Big\{ \boldsymbol{\Sigma}_{0y}^{-1} \mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} \boldsymbol{\Sigma}_y \Big\} \\
&\quad + u_{a\sigma}^i \\
&= \mathbf{s}_{ay}^i \boldsymbol{\sigma}_y + \mathbf{s}_a^i \boldsymbol{\sigma}_a + u_{a\sigma}^i
\end{aligned}
\tag{45a}
$$

for $i = 1, 2, \dots, m_a$, where $\mathbf{F}_a = [\mathbf{0}, \mathbf{I}_a]$ and $u_{a\sigma}^i$ is given by

$$
\begin{aligned}
u_{a\sigma}^i &= E\{(\mathbf{F}_a \mathbf{q}_{\beta\bar{a}})^T \mathbf{F}_a \mathbf{q}_{\beta\bar{a}}\} \\
&= \frac{1}{4} \mathrm{tr}\left\{ \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \mathbf{U}_{ia} \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix} \mathbf{K} \mathbf{M}_q \mathbf{K}^T \right\}.
\end{aligned}
\tag{45b}
$$

5.3 Biases of the estimated variance components

By collecting the expectation of $\mathbf{q}_y$ in ([41a](#)) and that of $\mathbf{q}_a$ in ([45a](#)) together and writing them in matrix form, we have

$$
E\left\{ \begin{bmatrix} \mathbf{q}_y \\ \mathbf{q}_a \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{S}_y & \mathbf{S}_{ya} \\ \mathbf{S}_{ay} & \mathbf{S}_a \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma}_y \\ \boldsymbol{\sigma}_a \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{y\sigma} \\ \mathbf{u}_{a\sigma} \end{bmatrix},
\tag{46}
$$

where the elements of both $\mathbf{u}_{y\sigma}$ and $\mathbf{u}_{a\sigma}$ have been derived and given by ([41b](#)) and ([45b](#)), respectively. It is now immediately clear that the approximate MINQUE estimate ([21](#)) of the variance components is biased. The biases of $\hat{\boldsymbol{\sigma}}_y$ and $\hat{\boldsymbol{\sigma}}_a$, denoted, respectively, by $\mathbf{b}_{y\sigma}$ and $\mathbf{b}_{a\sigma}$, are given as follows:

$$
\begin{aligned}
\begin{bmatrix} \mathbf{b}_{y\sigma} \\ \mathbf{b}_{a\sigma} \end{bmatrix} &= E\left\{ \begin{bmatrix} \hat{\boldsymbol{\sigma}}_y \\ \hat{\boldsymbol{\sigma}}_a \end{bmatrix} \right\} - \begin{bmatrix} \boldsymbol{\sigma}_y \\ \boldsymbol{\sigma}_a \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{S}_y & \mathbf{S}_{ya} \\ \mathbf{S}_{ay} & \mathbf{S}_a \end{bmatrix}^{-1} E\left\{ \begin{bmatrix} \mathbf{q}_y \\ \mathbf{q}_a \end{bmatrix} \right\} - \begin{bmatrix} \boldsymbol{\sigma}_y \\ \boldsymbol{\sigma}_a \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{S}_y & \mathbf{S}_{ya} \\ \mathbf{S}_{ay} & \mathbf{S}_a \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u}_{y\sigma} \\ \mathbf{u}_{a\sigma} \end{bmatrix},
\end{aligned}
\tag{47}
$$

if the coefficient matrix is invertible. Obviously, if this coefficient matrix is ill-conditioned, the condition number will be very large, as shown in Fig. [1](#), and as a result, the biases of the estimated variance components can become significantly large.

We should like to remark that the values $u_{y\sigma}^i$ in ([41b](#)) and $u_{a\sigma}^i$ in ([45b](#)) are always positive. As a result, all the biases of the componentwise positive estimators of variance components, as discussed in Sect. [3.3](#), are obviously positive as well. In other words, the approximate MINQUE (componentwise positive) estimators of variance components always statistically produce larger values for all the variance components. While the asymptotic biases of the naive estimator for the variance components of $\mathbf{y}$ are due to the fact that the random errors of $\mathbf{A}$ are ignored in the construction of such an estimator ([Wang and Davidian 1996](#); [Wang et al. 1998](#); [Li et al. 2009](#)), the finite sample bias analysis of variance components in this paper is valid for both the variance components of $\mathbf{y}$ and $\mathbf{A}$ and shows that the source of biases comes from the induced-model nonlinearity of the random errors of $\mathbf{A}$. Finally, we would like to note that the biases of the estimated variance components will result in incorrect determination of weights of the measurements. As a consequence, the variance–covariance matrices of the weighted TLS estimates of $\boldsymbol{\beta}$ and $\overline{\mathbf{a}}$ will be affected, the extent of which will depend on the sizes of the biases of the estimated variance components. If the biases of the estimated variance components are significantly amplified due to instability, the corresponding weight matrices of the measurements will simply be erroneous and the formal variance–covariance matrices of the weighted TLS estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\overline{\mathbf{a}}}$ cannot properly reflect their corresponding accuracy any more.

6 Concluding remarks

Total least squares has been substantially developed for more than one century and found a wide variety of applications, either from the point of view of approximation (see, e.g., Golub and Loan 1980; Huffel and Vandewalle 1991) or from the statistical point of view (see, e.g., Pearson 1901; Deming 1931, 1934; Schaffrin and Wieser 2008; Xu et al. 2012). Very often, one assumes that the variance–covariance matrices of $\mathbf{y}$ and $\mathbf{A}$ are given, or at least, are given up to an unknown variance of unit weight. In practice, however, we have to solve problems in physical, statistical and engineering sciences in which both $\mathbf{y}$ and $\mathbf{A}$ in (1) are not necessarily of the same types of measurements. As a result, we have to simultaneously estimate the EIV model parameters $\boldsymbol{\beta}$ and the variance components in the EIV stochastic models. We have proved that the variance components in the EIV stochastic models are not estimable, if all the elements of $\mathbf{A}$ are random without any functional constraints and can be classified into, at least, two groups of data of the same accuracy. The same is true for the EIV model of regression type (35). This result of inestimability statistically implies that we cannot do anything to gain any knowledge on such EIV stochastic models. From this point of view, to gain knowledge on the EIV stochastic models, one will have to directly collect repeated measurements and/or alike of the elements of $\mathbf{A}$ for such a stochastic model. Otherwise, if the variance components are estimable, we have derived the MINQUE estimates of the variance components in EIV models. For block-wise $\mathbf{U}_{iy}$ and $\mathbf{U}_{ia}$, we have also derived the componentwise positive estimates of the variance components. The estimation of variance components in the EIV model may be unstable, however, as confirmed by the simulated numerical examples. Finally, we have worked out the finite sample biases of the variance components, if they are estimable. As a result of equation instability, the biases of the estimated variance components could be significantly amplified due to a large condition number.

7 Appendix: the matrices required for the MINQUE estimation of variance components

To apply the MINQUE method to estimate the variance components of (2) in association with the linearized EIV model (10), with the design matrix $\mathbf{A}(\hat{\boldsymbol{\beta}}, \hat{\bar{\mathbf{a}}})$ in (17), we have to preparatorily compute the normal, the projection and other necessary matrices. We should note that all these matrices, including any quantity computed with these matrices, depend on the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\bar{\mathbf{a}}}$. For conciseness of notations, we will not write down such dependence without confusion. By definition, the normal matrix, denoted by $\mathbf{N}$, is given as follows:

$$
\begin{aligned}
\mathbf{N} &= \begin{bmatrix} (\hat{\bar{\mathbf{A}}})^T & \mathbf{0} \\ (\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n) & \mathbf{I}_a \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\bar{\mathbf{A}}} & (\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix} \\
&= \begin{bmatrix} (\hat{\bar{\mathbf{A}}})^T \boldsymbol{\Sigma}_{0y}^{-1} \hat{\bar{\mathbf{A}}} & (\hat{\boldsymbol{\beta}})^T \otimes (\hat{\bar{\mathbf{A}}})^T \boldsymbol{\Sigma}_{0y}^{-1} \\ \hat{\boldsymbol{\beta}} \otimes (\boldsymbol{\Sigma}_{0y}^{-1} \hat{\bar{\mathbf{A}}}) & \boldsymbol{\Sigma}_{0a}^{-1} + \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})^T \otimes \boldsymbol{\Sigma}_{0y}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{N}_\beta & \mathbf{N}_{\beta\bar{a}} \\ \mathbf{N}_{\bar{a}\beta} & \mathbf{N}_{\bar{a}} \end{bmatrix}.
\end{aligned}
\tag{48a}
$$

The inverse of the normal matrix $\mathbf{N}$ with the four blocks is symbolically represented by the matrix $\mathbf{Q}$, namely,

$$
\mathbf{Q} = \mathbf{N}^{-1} = \begin{bmatrix} \mathbf{Q}_\beta & \mathbf{Q}_{\beta\bar{a}} \\ \mathbf{Q}_{\bar{a}\beta} & \mathbf{Q}_{\bar{a}} \end{bmatrix},
\tag{48b}
$$

with the size of each of the blocks $\mathbf{Q}_\beta$, $\mathbf{Q}_{\beta\bar{a}}$, $\mathbf{Q}_{\bar{a}\beta}$ and $\mathbf{Q}_{\bar{a}}$ corresponding to that of $\mathbf{N}_\beta$, $\mathbf{N}_{\beta\bar{a}}$, $\mathbf{N}_{\bar{a}\beta}$ and $\mathbf{N}_{\bar{a}}$, respectively.

To compute the projection matrix, we first compute the following $\mathbf{H}$ matrix:

$$
\begin{aligned}
\mathbf{H} &= \begin{bmatrix} \hat{\bar{\mathbf{A}}} & (\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix} \begin{bmatrix} \mathbf{Q}_\beta & \mathbf{Q}_{\beta\bar{a}} \\ \mathbf{Q}_{\bar{a}\beta} & \mathbf{Q}_{\bar{a}} \end{bmatrix} \begin{bmatrix} (\hat{\bar{\mathbf{A}}})^T & \mathbf{0} \\ (\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n) & \mathbf{I}_a \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{H}_y & \mathbf{H}_{ya} \\ \mathbf{H}_{ay} & \mathbf{H}_a \end{bmatrix},
\end{aligned}
\tag{48c}
$$

or explicitly by

$$
\begin{aligned}
\mathbf{H}_y &= \hat{\bar{\mathbf{A}}} \mathbf{Q}_\beta (\hat{\bar{\mathbf{A}}})^T + \hat{\bar{\mathbf{A}}} \mathbf{Q}_{\beta\bar{a}}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n) + (\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n \mathbf{Q}_{\bar{a}\beta}(\hat{\bar{\mathbf{A}}})^T \\
&\quad + (\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n \mathbf{Q}_{\bar{a}}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n), \\
\mathbf{H}_{ya} &= \hat{\bar{\mathbf{A}}} \mathbf{Q}_{\beta\bar{a}} + (\hat{\boldsymbol{\beta}})^T \otimes \mathbf{I}_n \mathbf{Q}_{\bar{a}}, \\
\mathbf{H}_{ay} &= \mathbf{H}_{ya}^T = \mathbf{Q}_{\bar{a}\beta}(\hat{\bar{\mathbf{A}}})^T + \mathbf{Q}_{\bar{a}}(\hat{\boldsymbol{\beta}} \otimes \mathbf{I}_n), \\
\mathbf{H}_a &= \mathbf{Q}_{\bar{a}}.
\end{aligned}
$$

With (14) in mind, the projection matrix in a linear model is defined by $\mathbf{R} = \mathbf{Z}_A^\perp = \mathbf{I}_{ya} - \mathbf{Z}_A$, where $\mathbf{I}_{ya}$ is an identity matrix with the dimension of sizes $\mathbf{y}$ and $\mathbf{A}$. As a result, we can finally obtain the projection matrix as follows:

$$
\begin{aligned}
\mathbf{R} &= \mathbf{I}_{ya} - \mathbf{H}\boldsymbol{\Sigma}_0^{-1} \\
&= \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_a \end{bmatrix} - \mathbf{H} \begin{bmatrix} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_n - \mathbf{H}_y \boldsymbol{\Sigma}_{0y}^{-1} & -\mathbf{H}_{ya} \boldsymbol{\Sigma}_{0a}^{-1} \\ -\mathbf{H}_{ay} \boldsymbol{\Sigma}_{0y}^{-1} & \mathbf{I}_a - \mathbf{H}_a \boldsymbol{\Sigma}_{0a}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}_y & \mathbf{R}_{ya} \\ \mathbf{R}_{ay} & \mathbf{R}_a \end{bmatrix}.
\end{aligned}
\tag{49}
$$

Finally, the matrix $\mathbf{P}$ in Sect. 3.2 is given as follows:

$$
\begin{aligned}
\mathbf{P} &= \mathbf{\Sigma}_0^{-1} \mathbf{R} \\
&= \begin{bmatrix} \mathbf{\Sigma}_{0y}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{0a}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n - \mathbf{H}_y \mathbf{\Sigma}_{0y}^{-1} & -\mathbf{H}_{ya} \mathbf{\Sigma}_{0a}^{-1} \\ -\mathbf{H}_{ay} \mathbf{\Sigma}_{0y}^{-1} & \mathbf{I}_a - \mathbf{H}_a \mathbf{\Sigma}_{0a}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{\Sigma}_{0y}^{-1} - \mathbf{\Sigma}_{0y}^{-1} \mathbf{H}_y \mathbf{\Sigma}_{0y}^{-1} & -\mathbf{\Sigma}_{0y}^{-1} \mathbf{H}_{ya} \mathbf{\Sigma}_{0a}^{-1} \\ -\mathbf{\Sigma}_{0a}^{-1} \mathbf{H}_{ay} \mathbf{\Sigma}_{0y}^{-1} & \mathbf{\Sigma}_{0a}^{-1} - \mathbf{\Sigma}_{0a}^{-1} \mathbf{H}_a \mathbf{\Sigma}_{0a}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_y & \mathbf{P}_{ya} \\ \mathbf{P}_{ay} & \mathbf{P}_a \end{bmatrix} .
\end{aligned}
\tag{50}
$$

References

Adcock RJ (1877) Note on the method of least squares. Analyst 4:183–184

Bates D, Watts DG (1980) Relative curvature measures of nonlinearity (with discussions). J R Stat Soc B42:1–25

Bates D, Watts DG (1988) Nonlinear regression analysis and its applications. Wiley, New York

Beale EML (1960) Confidence regions in non-linear estimation (with discussions). J R Stat Soc B22:41–88

Björck Â (1996) Numerical methods for least squares problems. SIAM, Philadelphia

Box MJ (1971) Bias in nonlinear estimation (with discussions). J Roy Stat Soc B33:171–201

Clarke GPY (1980) Moments of the least squares estimators in a nonlinear regression model. J R Stat Soc B42:227–237

Deming WE (1931) The application of least squares. Philos Mag 11:146–158

Deming WE (1934) On the application of least squares–II. Philos Mag 17:804–829

Deming WE (1964) Statistical adjustment of data. Dover, New York

Dennis JE Jr, Schnabel RB (1996) Numerical methods for unconstrained optimization and nonlinear equations. SIAM classics in applied mathematics, SIAM, Philadelphia

Eshagh M (2010) Variance component estimation in linear ill-posed problems: TSVD issue. Acta Geod Geophys Hung 45:184–194

Eshagh M (2011) On the estimation of variance in unstable condition adjustment models. Acta Geod Geophys Hung 46:71–83

Fang X (2011) Weighted total least squares solutions for applications in geodesy. Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik, Ph.D. Dissertation Nr.294, Leibniz University Hannover

Fletcher R (2000) Practical methods of optimization, 2nd edn. Wiley, New York

Gerhold GA (1969) Least-squares adjustment of weighted data to a general linear equation. Am J Phys 37:156–161

Golub GH, van Loan CF (1980) An analysis of the total least squares problem. SIAM J Numer Anal 17:883–893

Grafarend EW (1985) Variance-covariance component estimation: theoretical results and geodetic applications. Stat Decis 4:407–441

Hartley HO, Rao JNK (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika 54:93–108

Helmert FR (1907) Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate, Zweite edn. Teubner, Leipzig

Horn SD, Horn RA, Duncan DB (1975) Estimating heteroscedastic variances in linear models. J Am Stat Assoc 70:380–385

Koch KR (1986) Maximum likelihood estimate of variance components. Bull Géod 60:329–338

Koch KR (1999) Parameter estimation and hypothesis testing in linear models, 2nd edn. Springer, Berlin

Koch KR, Kusche J (2007) Comments on Xu et al. (2006) Variance component estimation in linear inverse ill-posed models, J Geod 80(1):69–81, J Geod 81:629–631

Kubik K (1966) Schätzfunktionen für Varianzen, Kovarianzen und für andere Parameter in Ausgleichsaufgaben, Ph.D. thesis, ITC, Delft

Kubik K (1970) The estimation of the weights of measured quantities within the method of least squares. Bull Géod 44:21–40

Kummell CH (1879) Reduction of observation equations which contain more than one observed quantity. Analyst 6:97–105

LaMotte LR (1973) Quadratic estimation of variance components. Biometrics 29:311–330

Li Y, Tang H, Lin X (2009) Spatial linear mixed models with covariate measurement errors. Stat Sin 19:1077–1093

Magnus JR, Neudecker H (1988) Matrix differential calculus with applications in statistics and econometrics. Wiley, New York

Mann ME, Emanuel KA (2006) Atlantic hurricane trends linked to climate change. EOS Trans AGU 87:233–241

Markovsky I, van Huffel S (2007) Overview of total least squares methods. Signal Proc 87:2283–2302

Marquardt DW (1963) An algorithm for least squares estimation of nonlinear parameters. SIAM J Appl Math 11:431–441

Patterson HD, Thompson R (1975) Maximum likelihood estimation of components of variance. In: Proceedings of the 8th international biometric conference, pp 197–207

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572

Pincus R (1974) Estimability of parameters of the covariance matrix and variance components. Math Operationsforsch Stat 5:245–248

Pope AJ (1972) Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. In: Proceedings of the 38th annual meeting, American society of photogrammetry, Washington, pp 449–471

Pukelsheim F (1976) Estimating variance components in linear models. J Multivar Anal 6:626–629

Pukelsheim F, Styan GPH (1978) Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model, Technical Report No. 125, Department of Statistics, Stanford University, California

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

Rao CR (1971b) Minimum variance quadratic unbiased estimation of variance components. J Multivar Anal 1:445–456

Rao CR, Kleffe J (1988) Estimation of variance components and applications. north-holland, Amsterdam

Ratkowsky DA (1983) Nonlinear regression modeling: a practical unified approach. Marcel Dekker, New York

Schaffrin B (1983) Varianz-Kovarianz-Komponenten-Schätzung bei der Ausgleichung heterogener Wiederholungsmessungen, C282. Deutsche Geodätische Kommission, München

Schaffrin B, Wieser A (2008) On weighted total least-squares adjustment for linear regression. J Geod 82:415–421

Schaffrin B, Felus YA (2009) An algorithmic approach to the total least-squares problem with linear and quadratic constraints. Stud Geophys Geod 53:1–16

Schaffrin B, Snow K (2010) Total least-squares regularization of Tykhonov type and an ancient racetrack in Corinth. Linear Algebra Appl 432:2061–2076

Searle SR (1971) Linear models. Wiley, New York

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York

Seber G, Wild C (1989) Nonlinear regression. Wiley, New York

Snow K (2012) Topics in total least-squares adjustment within the errors-in-variables model: Singular cofactor matrices and prior information, Technical Report No. 502, Geodetic Science, The Ohio State University, Columbus

Teunissen P (1989a) First and second moments of non-linear least-squares estimators. Bull Géod 63:253–262

Teunissen P (1989b) Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity. Lecture Notes Earth Sciences, vol 29. Springer, Berlin, pp 241–264

Teunissen P, Amiri-Simkooei AR (2008) Least-squares variance component estimation. J Geod 82:65–82

van Huffel S, Vandewalle J (1989) Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error. SIAM J Matrix Anal Appl 10:2577–2596

van Huffel S, Vandewalle J (1991) The total least squares problem: computational aspects and analysis. SIAM, Philadelphia

Wang N, Davidian M (1996) A note on covariate measurement error in nonlinear mixed effects models. Biometrika 83:801–812

Wang N, Lin X, Gutierrez R, Carroll RJ (1998) Bias analysis and SIMEX approach in generalized linear mixed measurement error models. J Am Stat Assoc 93:249–261

Wulff SS, Birkes D (2005) Minimum variance unbiased invariant estimation of variance components under normality. Statistics 39:53–65

Xu PL (2009) Iterative generalized cross-validation for fusing heteroscedastic data of inverse ill-posed problems. Geophys J Int 179:182–200

Xu PL, Shimada S (2000) Least squares estimation in multiplicative noise models. Commun Stat B29:83–96

Xu PL, Shen YZ, Fukuda Y, Liu YM (2006) Variance component estimation in inverse ill-posed linear models. J Geod 80:69–81

Xu PL, Liu YM, Shen YZ, Fukuda Y (2007a) Estimability analysis of variance and covariance components. J Geod 81:593–602

Xu PL, Liu YM, Shen YZ, Fukuda Y (2007b) Reply to the comments by K.R. Koch and J. Kusche on Xu et al. (2006) Variance component estimation in linear inverse ill-posed models, J Geod 80:69–8. J Geod 81:633–635

Xu PL, Liu JN, Shi C (2012) Total least squares adjustment in partial errors-in-variables models: algorithm and statistical analysis. J Geod 86:661–675