Article

# Total Least-Squares Collocation: An Optimal Estimation Technique for the EIV-Model with Prior Information 

Burkhard Schaffrin<br>Geodetic Science Program, School of Earth Sciences, The Ohio State University, Columbus, OH 43210, USA; schaffrin.1@osu.edu

Received: 3 May 2020; Accepted: 1 June 2020; Published: 13 June 2020


#### Abstract

In regression analysis, oftentimes a linear (or linearized) Gauss-Markov Model (GMM) is used to describe the relationship between certain unknown parameters and measurements taken to learn about them. As soon as there are more than enough data collected to determine a unique solution for the parameters, an estimation technique needs to be applied such as 'Least-Squares adjustment', for instance, which turns out to be optimal under a wide range of criteria. In this context, the matrix connecting the parameters with the observations is considered fully known, and the parameter vector is considered fully unknown. This, however, is not always the reality. Therefore, two modifications of the GMM have been considered, in particular. First, 'stochastic prior information' (p. i.) was added on the parameters, thereby creating the - still linear - Random Effects Model (REM) where the optimal determination of the parameters (random effects) is based on 'Least Squares collocation', showing higher precision as long as the p. i. was adequate (Wallace test). Secondly, the coefficient matrix was allowed to contain observed elements, thus leading to the - now nonlinear -Errors-In-Variables (EIV) Model. If not using iterative linearization, the optimal estimates for the parameters would be obtained by 'Total Least Squares adjustment' and with generally lower, but perhaps more realistic precision. Here the two concepts are combined, thus leading to the (nonlinear) 'EIV-Model with p. i.', where an optimal estimation (resp. prediction) technique is developed under the name of 'Total Least-Squares collocation'. At this stage, however, the covariance matrix of the data matrix - in vector form - is still being assumed to show a Kronecker product structure.


Keywords: Errors-In-Variables Model; Total Least-Squares; prior information; collocation vs. adjustment

## 1. Introduction

Over the last 50 years or so, the (linearized) Gauss-Markov Model (GMM) as standard model for the estimation of parameters from collected observation [1,2] has been refined in a number of ways. Two of these will be considered in more detail, namely

- the GMM after strengthening the parameters through the introduction of "stochastic prior information". The relevant model will be the "Random Effects Model (REM)", and the resulting estimation technique has become known as "least-squares collocation" [3,4].
- the GMM after weakening the coefficient matrix through the replacement of fixed entries by observed data, resulting in the (nonlinear) Errors-In-Variables (EIV) Model. When nonlinear normal equations are formed and subsequently solved by iteration, the resulting estimation technique has been termed "Total Least-Squares (TLS) estimation" [5-7]. The alternative approach, based on iteratively linearizing the EIV-Model, will lead to identical estimates of the parameters [8].

After a compact review and comparison of several key formulas (parameter estimates, residuals, variance component estimate, etc.) within the above three models, a new model will be introduced that allows the strengthening of the parameters and the weakening of the coefficient matrix at the same time. The corresponding estimation technique will be called "TLS collocation" and follows essentially the outline that had first been presented by this author in June 2009 at the Intl. Workshop on Matrices and Statistics in Smolenice Castle (Slovakia); cf. Schaffrin [9].

Since then, further computational progress has been made, e.g. Snow and Schaffrin [10]; but several open questions remain that need to be addressed elsewhere. These include issues related to a rigorous error propagation of the "TLS collocation" results; progress may be achieved here along similar lines as in Snow and Schaffrin [11], but is beyond the scope of the present paper.

For a general overview of the participating models, the following diagram (Figure 1) may be helpful.


Figure 1. Model Diagram.
The most informative model defines the top position, and the least informative model is at the bottom. The new model can be formed at the intermediate level (like the GMM), but belongs to the "nonlinear world" where nonlinear normal equations need to be formed and subsequently solved by iteration.

## 2. A Compact Review of the "Linear World"

### 2.1. The (linearized) Gauss-Markov Model (GMM)

Let the Gauss-Markov Model be defined by

$$
\begin{equation*}
y=\underset{n \times m}{A} \xi+e, q:=\operatorname{rk} A \leq \min \{m, n\}, e \sim\left(0, \sigma_{0}^{2} I_{n}\right), \tag{1}
\end{equation*}
$$

possibly after linearization and homogeneization; cf. Koch [2], or Grafarend and Schaffrin [12] among many others.

Here, $y$ denotes the $n \times 1$ vector of (incremental) observations,
$A$ the $n \times m$ matrix of coefficients (given),
$\xi \quad$ the $m \times 1$ vector of (incremental) parameters (unknown),
$e \quad$ the $n \times 1$ vector of random errors (unknown) with expectation $E\{e\}=0$ while the dispersion matrix $D\{e\}=\sigma_{0}^{2} I_{n}$ is split into the (unknown) factor
$\sigma_{0}^{2} \quad$ as variance component (unit-free) and
$I_{n} \quad$ as (homogeneized) $n \times n$ symmetric and positive-definite "cofactor matrix" whose inverse is better known as "weight matrix" $P$; here $P=I_{n}$ for the sake of simplicity.

Now, it is well known that the Least-Squares Solution (LESS) is based on the principle

$$
\begin{equation*}
e^{\mathrm{T}} e=\min \quad \text { s.t. } e=y-A \xi \tag{2}
\end{equation*}
$$

which leads to the "normal equations"

$$
\begin{equation*}
N \hat{\xi}=c \text { for } \quad[N, c]:=A^{\mathrm{T}}[A, y] \tag{3}
\end{equation*}
$$

Depending on rk $N$, the rank of the matrix $N$, the LESS may turn out uniquely as

$$
\begin{equation*}
\hat{\xi}_{\text {LESS }}=N^{-1} c \quad \text { iff } \quad \text { rk } N=\operatorname{rk} A=q=m ; \tag{4}
\end{equation*}
$$

or it may belong to a solution (hyper)space that can be characterized by certain generalized inverses of $N$, namely

$$
\begin{align*}
\hat{\xi}_{\text {LESS }} & \in\left\{N^{-} c \mid N N^{-} N=N\right\}= \\
& =\left\{N_{r s}^{-} c \mid N N_{r s}^{-} N=N, N_{r s}^{-} N N_{r s}^{-}=N_{r s}^{-}=\left(N_{r s}^{-}\right)^{\mathrm{T}}\right\} ; \tag{5}
\end{align*}
$$

where $N_{r s}^{-}$denotes an arbitrary reflexive symmetric $g$-inverse of $N$ (including the "pseudo-inverse" $N^{+}$),

$$
\begin{equation*}
\text { iff } \operatorname{rk} N=\operatorname{rk} A=q<m \tag{6}
\end{equation*}
$$

In the latter case (6), any LESS will be biased with

$$
\begin{equation*}
\operatorname{bias}\left(\hat{\xi}_{\mathrm{LESS}}\right)=E\left\{\hat{\xi}_{\mathrm{LESS}}-\xi\right\} \in\left\{-\left(I_{m}-N^{-} N\right) \xi\right\}=\left\{-\left(I_{m}-N_{r s}^{-} N\right) \xi\right\} \tag{7}
\end{equation*}
$$

its dispersion matrix will be

$$
\begin{equation*}
D\left\{\hat{\xi}_{\mathrm{LESS}}\right\}=\sigma_{0}^{2} N^{-} N\left(N^{-}\right)^{\mathrm{T}} \in\left\{\sigma_{0}^{2} N_{r s}^{-} \mid N N_{r s}^{-} N=N, N_{r s}^{-} N N_{r s}^{-}=N_{r s}^{-}=\left(N_{r s}^{-}\right)^{\mathrm{T}}\right\} \tag{8}
\end{equation*}
$$

and its Mean Squared Error (MSE) matrix, therefore,

$$
\begin{equation*}
\operatorname{MSE}\left\{\hat{\xi}_{\mathrm{LESS}}\right\} \in\left\{\sigma_{0}^{2}\left[N_{r s}^{-}+\left(I_{m}-N_{r s}^{-} N\right)\left(\xi \sigma_{0}^{-2} \xi^{\mathrm{T}}\right)\left(I_{m}-N_{r s}^{-} N\right)^{\mathrm{T}}\right]\right\} . \tag{9}
\end{equation*}
$$

In contrast to the choice of LESS in case (6), the residual vector will be unique; for any $\hat{\xi}_{\text {LESS }}$ :

$$
\begin{equation*}
\tilde{e}_{\text {LESS }}=y-A \hat{\xi}_{\text {LESS }} \sim\left(0, \sigma_{0}^{2}\left(I_{n}-A N_{r s}^{-} A^{\mathrm{T}}\right)=D\{y\}-D\left\{A \hat{\xi}_{\text {LESS }}\right\}\right) \tag{10}
\end{equation*}
$$

Hence, the optimal variance component estimate will also be unique:

$$
\begin{align*}
\hat{\sigma}_{0}^{2} & =(n-q)^{-1} \tilde{e}_{\text {LESS }}^{\mathrm{T}} \cdot \tilde{e}_{\text {LESS }}=(n-q)^{-1}\left(y^{\mathrm{T}} y-c^{\mathrm{T}} \hat{\xi}_{\text {LESS }}\right),  \tag{11}\\
E\left\{\hat{\sigma}_{0}^{2}\right\} & =\sigma_{0}^{2}, D\left\{\hat{\sigma}_{0}^{2}\right\} \approx 2\left(\sigma_{0}^{2}\right)^{2} /(n-q) \text { under quasi-normality } . \tag{12}
\end{align*}
$$

For the corresponding formulas in the full-rank case (4), the reflexive symmetric g-inverse $N_{r s}^{-}$ simply needs to be replaced by the regular inverse $N^{-1}$, thereby showing that the LESS turns into an unbiased estimate of $\xi$ while its MSE-matrix coincides with its dispersion matrix (accuracy precision).

### 2.2. The Random Effects Model (REM)

Now, additional prior information (p. i.) is introduced to strengthen the parameter vector within the GMM (1). Based on Harville's [13] notation of a $m \times 1$ vector $\underset{\sim}{0}$ of so-called "pseudo-observations", the new $m \times 1$ vector

$$
\begin{gather*}
x:=\xi-\underset{\sim}{0} \text { of "random effects" (unknown) is formed with }  \tag{13}\\
E\{x\}=\beta_{0} \text { as given } m \times 1 \text { vector of } \mathrm{p} . \mathrm{i} \text {. and } \tag{14}
\end{gather*}
$$

$D\{x\}=\sigma_{0}^{2} Q_{0}$ as $n \times n$ positive-(semi)definite dispersion matrix of the $\mathrm{p} . \mathrm{i}$.
Consequently, the Random Effects Model (REM) can be stated as

$$
\begin{gather*}
\underset{\sim}{y}:=y-A \cdot \underset{\sim}{0}=\underset{n \times m}{A} x+e, q:=\operatorname{rk} A \leq \min \{m, n\}, \underset{m \times 1}{\beta_{0}}:=x+e_{0}  \tag{16}\\
{\left[\begin{array}{c}
e \\
e_{0}
\end{array}\right] \sim\left(\left[\begin{array}{l}
0 \\
0
\end{array}\right], \sigma_{0}^{2}\left[\begin{array}{cc}
I_{n} & 0 \\
0 & Q_{0}
\end{array}\right]\right), \underset{m \times m}{Q_{0}} \text { symmetric and nnd (non-negative definite). }} \tag{17}
\end{gather*}
$$

If $P_{0}:=Q_{0}^{-1}$ exists, the estimation/prediction of $x$ may now be based on the principle

$$
\begin{equation*}
e^{\mathrm{T}} e+e_{0}^{\mathrm{T}} P_{0} e_{0}=\min \quad \text { s.t. } e=\underset{\sim}{y}-A x, e_{0}=\beta_{0}-x \tag{18}
\end{equation*}
$$

which leads to the "normal equations" for the "Least-Squares Collocation (LSC)" solution

$$
\begin{equation*}
\left(N+P_{0}\right) \cdot \tilde{x}=\underset{\sim}{c}+P_{0} \cdot \beta_{0} \tag{19}
\end{equation*}
$$

which is unique even in the rank-deficient case (6).
If $Q_{0}$ is singular, but $e_{0} \in \mathfrak{R}\left(Q_{0}\right)$ with probability 1 , then there must exist an $m \times 1$ vector $v^{0}$ with

$$
\begin{equation*}
Q_{0} v^{0}=-e_{0}=x-\beta_{0} \text { with probability } 1 \text { (a.s. } \hat{=} \text { almost surely). } \tag{20}
\end{equation*}
$$

Thus, the principle (18) may be equivalently replaced by

$$
\begin{equation*}
e^{\mathrm{T}} e+\left(v^{0}\right)^{\mathrm{T}} Q_{0} v^{0}=\min \text { s.t. } e=\underset{\sim}{y}-A x \text { and }(20) \tag{21}
\end{equation*}
$$

which generates the LSC solution uniquely from the "modified normal equations"

$$
\begin{equation*}
\left(I_{m}+Q_{0} N\right) \cdot \tilde{x}_{\mathrm{LSC}}=\beta_{0}+Q_{0} \underset{\sim}{c} \text { for } \underset{\sim}{c}:=A^{\mathrm{T}} \underset{\sim}{y} \tag{22}
\end{equation*}
$$

or, alternatively, via the "update formula"

$$
\begin{equation*}
\tilde{x}_{\mathrm{LSC}}=\beta_{0}+Q_{0}\left(I_{m}+N Q_{0}\right)^{-1}\left(\underset{\sim}{c}-N \beta_{0}\right) \tag{23}
\end{equation*}
$$

which exhibits the "weak (local) unbiasedness" of $\tilde{x}_{\text {LSC }}$ via

$$
\begin{equation*}
E\left\{\tilde{x}_{\mathrm{LSC}}\right\}=\beta_{0}+Q_{0}\left(I_{m}+N Q_{0}\right)^{-1}\left(E\{\underset{\sim}{c}\}-N \beta_{0}\right)=\beta_{0}+0=E\{x\} . \tag{24}
\end{equation*}
$$

Consequently, the MSE-matrix of $\tilde{x}_{\text {LSC }}$ can be obtained from

$$
\begin{equation*}
\operatorname{MSE}\left\{\tilde{x}_{\mathrm{LSC}}\right\}=D\left\{\tilde{x}_{\mathrm{LSC}}-x\right\}=\sigma_{0}^{2} Q_{0}\left(I_{m}+N Q_{0}\right)^{-1}= \tag{25}
\end{equation*}
$$

$$
\begin{equation*}
=\sigma_{0}^{2}\left(N+P_{0}\right)^{-1} \text { if } P_{0}:=Q_{0}^{-1} \text { exists, } \tag{26}
\end{equation*}
$$

whereas the dispersion matrix of $\tilde{x}_{\text {LSC }}$ itself is of no relevance here. It holds:

$$
\begin{equation*}
D\left\{\tilde{x}_{\mathrm{LSC}}\right\}=D\{x\}-D\left\{x-\tilde{x}_{\mathrm{LSC}}\right\}=\sigma_{0}^{2} Q_{0}-\sigma_{0}^{2} Q_{0}\left(I_{m}+N Q_{0}\right)^{-1} \tag{27}
\end{equation*}
$$

Again, both residual vectors are uniquely determined from

$$
\begin{gather*}
\left(\tilde{e}_{0}\right)_{\mathrm{LSC}}=\beta_{0}-\tilde{x}_{\mathrm{LSC}}=-Q_{0} \hat{v}_{\mathrm{LSC}}^{0} \text { for }  \tag{28}\\
\hat{v}_{\mathrm{LSC}}^{0}:=\left(I_{m}+N Q_{0}\right)^{-1}\left(\underset{\sim}{c}-N \beta_{0}\right), \text { and }  \tag{29}\\
\tilde{e}_{\mathrm{LSC}}=\underset{\sim}{y}-A \tilde{x}_{\mathrm{LSC}}=\left[I_{m}-A Q_{0}\left(I_{m}+N Q_{0}\right)^{-1} A^{\mathrm{T}}\right]\left(\underset{\sim}{y}-A \beta_{0}\right) \tag{30}
\end{gather*}
$$

with the control formula

$$
\begin{equation*}
A^{\mathrm{T}} \tilde{\mathcal{e}}_{\text {LSC }}=\hat{\mathcal{v}}_{\mathrm{LSC}}^{0} . \tag{31}
\end{equation*}
$$

An optimal estimate of the variance component may now be obtained from

$$
\left.\begin{array}{rl}
\left(\hat{\sigma}_{0}^{2}\right)_{\mathrm{LSC}} & =n^{-1} \cdot\left(\tilde{e}_{\mathrm{LSC}}^{\mathrm{T}} \cdot \tilde{e}_{\mathrm{LSC}}+\left(\hat{v}_{\mathrm{LSC}}^{0}\right)^{\mathrm{T}} Q_{0} \hat{v}_{\mathrm{LSC}}^{0}\right)= \\
& =n^{-1} \cdot\left(\underset{\sim}{y^{\mathrm{T}}} \underset{\sim}{y}-{\underset{\sim}{c}}^{\mathrm{c}} \tilde{x}_{\mathrm{LSC}}-\beta_{0}^{\mathrm{T}} \hat{v}_{\mathrm{LSC}}^{0}\right. \tag{32}
\end{array}\right) .
$$

## 3. An Extension into the "Nonlinear World"

### 3.1. The Errors-In-Variables (EIV) Model

In this scenario, the Gauss-Markov Model (GMM) is further weakened by allowing some or all of the entries in the coefficient matrix $A$ to be observed. So, after introducing a corresponding $n \times m$ matrix of unknown random errors $E_{A}$, the original GMM (1) turns into the EIV-Model

$$
\begin{equation*}
y=\underbrace{\left(A-E_{A}\right)}_{n \times m} \cdot \xi+e, q:=\operatorname{rk} A=m<n, e \sim\left(0, \sigma_{0}^{2} I_{n}\right), \tag{33}
\end{equation*}
$$

with the vectorized form of $E_{A}$ being characterized through

$$
\begin{equation*}
\underset{n m \times 1}{e_{A}}:=\operatorname{vec} E_{A} \sim\left(0, \sigma_{0}^{2}\left(I_{m} \otimes I_{n}\right)=\sigma_{0}^{2} I_{m n}\right), C\left\{e, e_{A}\right\}=0 \text { (assumed) } . \tag{34}
\end{equation*}
$$

Here, the vec operation transform a matrix into a vector by stacking all columns underneath each other while $\otimes$ denotes the Kronecker-Zehfuss product of matrices, defined by

$$
\begin{equation*}
\underset{p \times q}{G} \otimes \underset{r \times s}{H}=\left[g_{i j} \cdot H\right]_{p r \times q s} \text { if } G=\left[g_{i j}\right] \text {. } \tag{35}
\end{equation*}
$$

In particular, the following key formula holds true:

$$
\begin{equation*}
\operatorname{vec}\left(A B C^{\mathrm{T}}\right)=(C \otimes A) \cdot \operatorname{vec} B \tag{36}
\end{equation*}
$$

for matrices of suitable size. Note that, in (34), the choice $Q_{A}:=I_{m n}$ is a very special one. In general, $Q_{A}$ may turn out singular whenever some parts of the matrix $A$ remain unobserved (i.e., nonrandom).

In any case, thanks to the term

$$
\begin{equation*}
E_{A} \cdot \xi=\left(\xi^{\mathrm{T}} \otimes I_{n}\right) e_{A}, \tag{37}
\end{equation*}
$$

the model (33) needs to be treated in the "nonlinear world" even though the vector $\xi$ may contain only incremental parameters. From now on, $A$ is assumed to have full column rank, rk $A=: q=m$.

Following Schaffrin and Wieser [7], for instance, the "Total-Least Squares Solution (TLSS)" can be based on the principle

$$
\begin{equation*}
e^{\mathrm{T}} e+e_{A}^{\mathrm{T}} e_{A}=\min \text { s.t. }(33-34) \tag{38}
\end{equation*}
$$

and the Lagrange function

$$
\begin{align*}
\Phi\left(e, e_{A}, \xi, \lambda\right): & =e^{\mathrm{T}} e+e_{A}^{\mathrm{T}} e_{A}+2 \lambda^{\mathrm{T}}\left(y-A \xi-e+E_{A} \xi\right)=  \tag{39}\\
& =e^{\mathrm{T}} e+e_{A}^{\mathrm{T}} e_{A}+2 \lambda^{\mathrm{T}}\left[y-A \xi-e+\left(\xi^{\mathrm{T}} \otimes I_{n}\right) e_{A}\right]
\end{align*}
$$

which needs to be made stationary. The necessary Euler-Lagrange conditions then read:

$$
\begin{gather*}
\frac{1}{2} \frac{\partial \Phi}{\partial e}=\tilde{e}-\hat{\lambda} \doteq 0 \Rightarrow \hat{\lambda}=\tilde{e}  \tag{40}\\
\frac{1}{2} \frac{\partial \Phi}{\partial e_{A}}=\tilde{e}_{A}+\left(\hat{\xi}_{\mathrm{TLS}} \otimes I_{n}\right) \hat{\lambda} \doteq 0 \Rightarrow \tilde{E}_{A}=-\hat{\lambda} \hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}}  \tag{41}\\
\frac{1}{2} \frac{\partial \Phi}{\partial \xi}=-\left(A-\tilde{E}_{A}\right)^{\mathrm{T}} \hat{\lambda} \doteq 0 \Rightarrow A^{\mathrm{T}} \hat{\lambda}=\tilde{E}_{A}^{\mathrm{T}} \hat{\lambda}=-\hat{\xi}_{\mathrm{TLS}} \cdot\left(\hat{\lambda}^{\mathrm{T}} \hat{\lambda}\right)  \tag{42}\\
\frac{1}{2} \frac{\partial \Phi}{\partial \lambda}=y-A \hat{\xi}_{\mathrm{TLS}}-\tilde{e}+\tilde{E}_{A} \hat{\xi}_{\mathrm{TLS}} \doteq 0 \Rightarrow y-A \hat{\xi}_{\mathrm{TLS}}=\hat{\lambda}\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)  \tag{43}\\
\Rightarrow c-N \hat{\xi}_{\mathrm{TLS}}:=A^{\mathrm{T}}\left(y-A \hat{\xi}_{\mathrm{TLS}}\right)=A^{\mathrm{T}} \hat{\lambda}\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)=-\hat{\xi}_{\mathrm{TLS}} \cdot \hat{v}_{\mathrm{TLS}} \tag{44}
\end{gather*}
$$

for

$$
\begin{gather*}
\hat{v}_{\mathrm{TLS}}:=\left(\hat{\lambda}^{\mathrm{T}} \hat{\lambda}\right) \cdot\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)=\hat{\lambda}^{\mathrm{T}}\left(y-A \hat{\xi}_{\mathrm{TLS}}\right)=  \tag{45}\\
=\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)^{-1} \cdot\left(y-A \hat{\xi}_{\mathrm{TLS}}\right)^{\mathrm{T}}\left(y-A \hat{\xi}_{\mathrm{TLS}}\right)=  \tag{46}\\
=\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)^{-1}\left[y^{\mathrm{T}}\left(y-A \hat{\xi}_{\mathrm{TLS}}\right)-\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}}\left(c-N \hat{\xi}_{\mathrm{TLS}}\right)\right] \geq 0  \tag{47}\\
\Rightarrow\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right) \cdot \hat{v}_{\mathrm{TLS}}=y^{\mathrm{T}} y-c^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}+\left(\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right) \cdot \hat{v}_{\mathrm{TLS}} \\
\Rightarrow \hat{v}_{\mathrm{TLS}}=y^{\mathrm{T}} y-c^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}} \tag{48}
\end{gather*}
$$

which needs to be solved in connection with the "modified normal equations" from (44), namely

$$
\begin{equation*}
\left(N-\hat{v}_{\mathrm{TLS}} I_{m}\right) \hat{\xi}_{\mathrm{TLS}}=c \text {. } \tag{49}
\end{equation*}
$$

Due to the nonlinear nature of $\hat{\xi}_{\text {TLS }}$, it is not so easy to determine if it is an unbiased estimate, or how its MSE-matrix may exactly look like. First attempts of a rigorous error propagation have recently been undertaken by Amiri-Simkooei et al. [14] and by Schaffrin and Snow [15], but are beyond the scope of this paper.

Instead, both the optimal residual vector $\tilde{e}_{\text {TLS }}$ and the optimal residual matrix $\left(\tilde{E}_{A}\right)_{\text {TLS }}$ are readily available through (40) and (43) as

$$
\begin{equation*}
\tilde{e}_{\mathrm{TLS}}=\hat{\lambda}=\left(y-A \hat{\xi}_{\mathrm{TLS}}\right) \cdot\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)^{-1} \tag{50}
\end{equation*}
$$

and through (41) as

$$
\begin{equation*}
\left(\tilde{E}_{A}\right)_{\mathrm{TLS}}=-\tilde{e}_{\mathrm{TLS}} \cdot \hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}}=-\left(y-A \hat{\xi}_{\mathrm{TLS}}\right) \cdot\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)^{-1} \cdot \hat{\xi}_{\mathrm{TLS}} . \tag{51}
\end{equation*}
$$

As optimal variance component estimate, it is now proposed to use the formula

$$
\left(\hat{\sigma}_{0}^{2}\right)_{\mathrm{TLS}}=(n-m)^{-1} \cdot\left[\tilde{e}_{\mathrm{TLS}}^{\mathrm{T}} \tilde{e}_{\mathrm{TLS}}+\left(\tilde{e}_{A}\right)_{\mathrm{TLS}}^{\mathrm{T}}\left(\tilde{e}_{A}\right)_{\mathrm{TLS}}\right]=
$$

$$
\begin{equation*}
=(n-m)^{-1} \cdot \hat{\lambda}^{\mathrm{T}} \hat{\lambda}\left(1+\hat{\xi}_{\mathrm{TLS}}^{\mathrm{T}} \hat{\xi}_{\mathrm{TLS}}\right)=\hat{v}_{\mathrm{TLS}} /(n-m) \tag{52}
\end{equation*}
$$

in analogy to the previous estimates (11) and (32).

### 3.2. A New Model: The EIV-Model with Random Effects (EIV-REM)

In the following, the above EIV-Model (33-34) is strengthened by introducing stochastic prior information (p. i.) on the parameters which thereby change their character and become "random effects" as in (13-15). The EIV-REM can, therefore, be stated as

$$
\begin{equation*}
\underset{\sim}{y}=\underbrace{\left(A-E_{A}\right)}_{n \times m} \cdot x+e, q:=\operatorname{rk} A \leq \min \{m, n\}, \beta_{0}=x+e_{0} \text { (given) } \tag{53}
\end{equation*}
$$

with

$$
\left[\begin{array}{c}
e  \tag{54}\\
e_{A}=\operatorname{vec} E_{A} \\
e_{0}
\end{array}\right] \sim\left(\left[\begin{array}{l}
0 \\
0 \\
0
\end{array}\right], \sigma_{0}^{2}\left[\begin{array}{ccc}
I_{n} & 0 & 0 \\
0 & I_{m n} & 0 \\
0 & 0 & Q_{0}
\end{array}\right]\right), Q_{0} \text { symmetric and nnd. }
$$

The first set of formulas will be derived by assuming that the weight matrix $P_{0}:=Q_{0}^{-1}$ exists uniquely for the p. i. Then, the "TLS collocation (TLSC)" may be based on the principle

$$
\begin{equation*}
e^{\mathrm{T}} e+e_{A}^{\mathrm{T}} e_{A}+e_{0}^{\mathrm{T}} P_{0} e_{0}=\min \text { s.t. }(53-54) \tag{55}
\end{equation*}
$$

resp. on the Lagrange function

$$
\begin{equation*}
\Phi\left(e, e_{A}, e_{0}, \lambda\right):=e^{\mathrm{T}} e+e_{A}^{\mathrm{T}} e_{A}+e_{0}^{\mathrm{T}} P_{0} e_{0}+2 \lambda^{\mathrm{T}}[(\underset{\sim}{y}-A \beta_{0}-e+\left(\beta_{0}^{\mathrm{T}} \otimes I_{n}\right) e_{A}+A e_{0} \underbrace{-E_{A} e_{0}}_{=-\left(e_{0}^{\mathrm{T}} \otimes I_{n}\right) \cdot e_{A}}] \tag{56}
\end{equation*}
$$

which needs to be made stationary. The necessary Euler-Lagrange conditions then read:

$$
\begin{gather*}
\frac{1}{2} \frac{\partial \Phi}{\partial e}=\tilde{e}-\hat{\lambda} \doteq 0 \Rightarrow \hat{\lambda}=\tilde{e}  \tag{57}\\
\frac{1}{2} \frac{\partial \Phi}{\partial e_{A}}=\tilde{e}_{A}+\left[\left(\beta_{0}-\tilde{e}_{0}\right) \otimes I_{n}\right] \hat{\lambda} \doteq 0 \Rightarrow \tilde{E}_{A}=-\hat{\lambda} \cdot\left(\beta_{0}-\tilde{e}_{0}\right)^{\mathrm{T}}=:-\hat{\lambda} \cdot \tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}}  \tag{58}\\
\frac{1}{2} \frac{\partial \Phi}{\partial e_{0}}=P_{0} \tilde{e}_{0}+\left(A-\tilde{E}_{A}\right)^{\mathrm{T}} \hat{\lambda} \doteq 0 \Rightarrow A^{\mathrm{T}} \hat{\lambda}=\tilde{E}_{A}^{\mathrm{T}} \hat{\lambda}-P_{0} \tilde{e}_{0} \Rightarrow  \tag{59}\\
\Rightarrow A^{\mathrm{T}} \hat{\lambda}=-\tilde{x}_{\mathrm{TLSC}} \cdot\left(\hat{\lambda}^{\mathrm{T}} \hat{\lambda}\right)+\hat{v}_{\mathrm{TLSC}}^{0} \text { for } \hat{\mathrm{v}}_{\mathrm{TLSC}}^{0}:=-P_{0} \tilde{e}_{0}=P_{0}\left(\beta_{0}-\tilde{x}_{\mathrm{TLSC}}\right)  \tag{60}\\
\frac{1}{2} \frac{\partial \Phi}{\partial \lambda}=\underset{\sim}{y}-A\left(\beta_{0}-\tilde{e}_{0}\right)-\tilde{e}+\tilde{E}_{A}\left(\beta_{0}-\tilde{e}_{0}\right) \doteq 0 \Rightarrow \underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}=\hat{\lambda}\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right) \Rightarrow  \tag{61}\\
\left.\Rightarrow \hat{\lambda}=\underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}\right) \cdot\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)^{-1} . \tag{62}
\end{gather*}
$$

Combining (60) with (62) results in

$$
\begin{equation*}
\left(\underset{\sim}{c}-N \tilde{x}_{\mathrm{TLSC}}\right) \cdot\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)^{-1}=A^{\mathrm{T}} \hat{\lambda}=-\tilde{x}_{\mathrm{TLSC}}\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)^{-2}\left(\underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}\right)^{\mathrm{T}}\left(\underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}\right)+\hat{v}_{\mathrm{TLSC}}^{0} \tag{63}
\end{equation*}
$$

and finally in

$$
\begin{equation*}
\left(N+\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right) P_{0}-\hat{v}_{\mathrm{TLSC}} I_{m}\right) \tilde{x}_{\mathrm{TLSC}}=\underset{\sim}{c}+P_{0} \beta_{0} \cdot\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right) \tag{64}
\end{equation*}
$$

where

$$
\begin{array}{|l}
\hat{v}_{T L S C}:=\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)^{-1}\left(\underset{\sim}{y}-A \tilde{x}_{\text {TLSC }}\right)^{\mathrm{T}}\left(\underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}\right), \text { and }  \tag{65}\\
\hat{v}_{\mathrm{TLSC}}^{0}:=-P_{0}\left(\beta_{0}-\tilde{x}_{\mathrm{TLSC}}\right)=-\tilde{P}_{0}, \text { provided that } P_{0} \text { exists. } \\
\hline
\end{array}
$$

In the more general case of a singular matrix $Q_{0}$, an approach similar to (20) can be followed, leading to the equation system

$$
\begin{equation*}
\left[\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right) \cdot I_{m}+Q_{0} N-\hat{v}_{\mathrm{TLSC}} \cdot Q_{0}\right] \tilde{x}_{\mathrm{TLSC}}=\beta_{0} \cdot\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)+Q_{0} \underset{\sim}{c} \tag{66}
\end{equation*}
$$

that needs to be solved in connection with (65). Obviously,

$$
\begin{array}{|llllllll|}
\hline \tilde{x}_{\text {TLSC }} & \longmapsto & \beta_{0} & \text { if } & Q_{0} & \longmapsto & 0, & \text { and }  \tag{67}\\
\tilde{x}_{\text {TLSC }} & \longmapsto & \tilde{x}_{\text {TLS }} & \text { if } & P_{0} & \longmapsto & 0 . & \\
\hline
\end{array}
$$

Again, it is still unclear if $\tilde{x}_{\text {TLSC }}$ represents an unbiased prediction of the vector $x$ of random effects. Also, very little (if anything) is known about the corresponding MSE-matrix of $\tilde{x}_{\text {TLSC }}$. The answers to these open problems will be left for a future contribution. It is, however, possible to find the respective residual vectors/matrices represented as follows:

$$
\begin{gather*}
\tilde{e}_{\mathrm{TLSC}}=\hat{\lambda}=\left(\underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}\right) \cdot\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)^{-1}  \tag{68}\\
\left(\tilde{E}_{A}\right)_{\mathrm{TLSC}}=-\hat{\lambda} \cdot \tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}}=-\left(\underset{\sim}{y}-A \tilde{x}_{\mathrm{TLSC}}\right) \cdot\left(1+\tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}} \tilde{x}_{\mathrm{TLSC}}\right)^{-1} \cdot \tilde{x}_{\mathrm{TLSC}}^{\mathrm{T}}  \tag{69}\\
\left(\tilde{e}_{0}\right)_{\mathrm{TLSC}}=-Q_{0} \cdot \hat{v}_{\mathrm{TLSC}}^{0}=\beta_{0}-\tilde{x}_{\mathrm{TLSC}}, \tag{70}
\end{gather*}
$$

while a suitable formula for the variance component is suggested as

$$
\begin{equation*}
\left(\hat{\sigma}_{0}^{2}\right)_{\text {TLSC }}=n^{-1} \cdot\left[\hat{v}_{\mathrm{TLSC}}+\left(\hat{v}_{\mathrm{TLSC}}^{0}\right)^{\mathrm{T}} Q_{0}\left(\hat{v}_{\mathrm{TLSC}}^{0}\right)\right] . \tag{71}
\end{equation*}
$$

## 4. Conclusions and Outlook

Key formulas have been developed successfully to optimally determine the parameters and residuals within the new 'EIV-Model with p. i.' (or EIV-REM) which turns out to be more general than the other three models considered here (GMM, REM, EIV-Model). In particular, it is quite obvious that

- EIV-REM becomes the REM if $D\left\{e_{A}\right\}:=0$,
- EIV-REM becomes the EIV-Model if $P_{0}:=0$,
- EIV-REM becomes the GMM if both $P_{0}:=0$ and $D\left\{e_{A}\right\}:=0$.

Hence the new EIV-REM can indeed serve as a universal representative of the whole class of models presented here.

Therefore, in a follow-up paper, it is planned to also cover more general dispersion matrices for $e$ and $e_{A}$ in (54), similarly to the work by Schaffrin et al. [16] for the EIV-Model with singular dispersion matrices for $e_{A}$.

Funding: This research received no external funding.
Conflicts of Interest: The author declares no conflict of interest.

## References

1. Rao, C.R. Estimation of parameters in a linear model. Ann. Statist. 1976, 4, 1023-1037. [CrossRef]
2. Koch, K.-R. Parameter Estimation and Hypothesis Testing in Linear Models; Springer: Berlin/Heidelberg, Germany, 1999; p. 334.
3. Moritz, H. A generalized least-squares model. Studia Geophys. Geodaet. 1970, 14, 353-362. [CrossRef]
4. Schaffrin, B. Model Choice and Adjustment Techniques in the Presence of Prior Information; Technical Report No. 351; The Ohio State University Department of Geodetic Science and Survey: Columbus, OH, USA, 1983; p. 37.
5. Golub, G.H.; van Loan, C.F. An analysis of the Total Least-Squares problem. SIAM J. Numer. Anal. 1980, 17, 883-893. [CrossRef]
6. Van Huffel, S.; Vandewalle, J. The Total Least-Squares Problem: Computational Aspects and Analysis; SIAM: Philadelphia, PA, USA, 1991; p. 300.
7. Schaffrin, B.; Wieser, A. On weighted Total Least-Squares adjustment for linear regression. J. Geod. 2008, 82, 415-421. [CrossRef]
8. Schaffrin, B. Adjusting the Errors-In-Variables model: Linearized Least-Squares vs. nonlinear Total Least-Squares procedures. In VIII Hotine-Marussi Symposium on Mathematical Geodesy; Sneeuw, N., Novák, P., Crespi, M., Sansó, F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 301-307.
9. Schaffrin, B. Total Least-Squares Collocation: The Total Least-Squares approach to EIV-Models with prior information. In Proceedings of the 18th International Workshop on Matrices and Statistics, Smolenice Castle, Slovakia, 23-27 June 2009.
10. Snow, K.; Schaffrin, B. Weighted Total Least-Squares Collocation with geodetic applications. In Proceedings of the SIAM Conference on Applied Linear Algebra, Valencia, Spain, 18-22 June 2012.
11. Schaffrin, B.; Snow, K. Progress towards a rigorous error propagation for Total Least-Squares estimates. J. Appl. Geod. 2020. accepted for publication. [CrossRef]
12. Grafarend, E.; Schaffrin, B. Adjustment Computations in Linear Models; Bibliographisches Institut: Mannheim, Germany; Wiesbaden, Germany, 1993; p. 483. (In German)
13. Harville, D.A. Using ordinary least-squares software to compute combined intra-interblock estimates of treatment contrasts. Am. Stat. 1986, 40, 153-157.
14. Amiri-Simkooei, A.R.; Zangeneh-Nejad, F.; Asgari, J. On the covariance matrix of weighted Total Least-Squares estimates. J. Surv. Eng. 2016, 142. [CrossRef]
15. Schaffrin, B.; Snow, K. Towards a more rigorous error propagation within the Errors-In-Variables Model for applications in geodetic networks. In Proceedings of the 4th Joint International Symposium on Deformation Monitoring (JISDM 2019), Athens, Greece, 15-17 May 2019. electronic proceedings only.
16. Schaffrin, B.; Snow, K.; Neitzel, F. On the Errors-In-Variables Model with singular dispersion matrices. J. Geod. Sci. 2014, 4, 28-36. [CrossRef]
