Total Least-Squares Adjustment with Prior Information vs. the Penalized Least-Squares Approach to EIV-Models

Burkhard Schaffrin[1,3] and Kyle Snow[1,2]

[1]Geodetic Science Program, School of Earth Sciences, The Ohio State University, Columbus, Ohio, USA
[2]Topcon Positioning Systems, Inc., Columbus, Ohio, USA
[3]Corresponding author: Burkhard Schaffrin, e-mail: schaffrin.1@osu.edu

Abstract

In the framework of a Gauss-Markov Model (GMM), it is well known that the Bayesian estimator, based on prior information, seems to coincide with the penalized least-squares estimator, based on Tykhonov regularization. However, their Mean Squared Error (MSE) matrices will be different, giving rise to a criterion that can determine the superiority of one over the other with respect to their MSE-risk.

Here, a similar comparison will be undertaken within the framework of Errors-In-Variables (EIV) Models. An attempt will be made to determine the MSE-risk for both the Bayesian Total Least-Squares (TLS) estimator and the penalized TLS estimator, which should result in a generalized criterion for the superiority of one over the other.

Key Words: EIV-Model, MSE-risk comparison, penalized TLS approach, prior information, TLS collocation

0. Introduction

It has long been known that Tykhonov regularization (i.e., the penalized Least-Squares approach) within a Gauss-Markov Model (GMM) leads to numerically identical estimates for the unknown parameters as the standard (weighted) Least-Squares approach within a Random Effects Model (REM), where the fixed parameters have been replaced by random effects with known expectation and dispersion; more details may be found in the books by Rao et al. (2008), Engl et al. (1996), or Grafarend and Schaffrin (1993), among others. It has been pointed out, however, by Schaffrin (2008) that the formal MSE-risk turns out differently for both estimators. In fact, a criterion was provided to decide under which condition one of the estimators will be superior over the other; cf. Schaffrin (2008, Theorem 1).

In this contribution, a similar comparison will be undertaken between the penalized Total Least-Squares (TLS) approach within an Errors-In-Variables (EIV) Model and TLS Collocation within an EIV-Model with prior information. For this purpose, the EIV-Model will be interpreted as a nonlinear Gauss-Helmert Model (GHM) that, after linearization, allows the derivation of MSE-risk formulas in first-order approximation.

Starting with a brief review of *penalized TLS adjustment* according to Schaffrin and Snow (2010) in chapter 1, a similar review will follow to explain *TLS Collocation* in accordance with Schaffrin (2009) and Snow and Schaffrin (2012) in chapter 2. Finally, chapter 3 will provide the key theorem on *MSE-risk* comparisons before chapter 4 concludes this contribution.

1. A brief review of the penalized TLS adjustment

Let the Errors-In-Variables (EIV) Model be defined as

$$\boldsymbol{y} = \underbrace{(\boldsymbol{X} - \boldsymbol{E}_X)}_{k\times m} \cdot \boldsymbol{\beta}_\mu + \boldsymbol{e}_y, \ \operatorname{rk} \boldsymbol{X} = m < k, \tag{1.1a}$$

$$\begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_X := \operatorname{vec} \boldsymbol{E}_X \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{km} \end{bmatrix} = \sigma_0^2 \boldsymbol{I}_n \right), \tag{1.1b}$$

where

$[\boldsymbol{y}, \boldsymbol{X}]$	denotes a $k \times (m+1)$ matrix of observations,
$[\boldsymbol{e}_y, \boldsymbol{E}_X]$	a $k \times (m+1)$ matrix of random errors,
$\boldsymbol{\beta}_\mu$	an $m \times 1$ vector of (unknown) parameters, and
σ_0^2	the (unknown) variance component in front of the so-called "cofactor matrix," here the identity matrix $\boldsymbol{I}_n$ of size $n \times n$ with $n := k(m+1)$.

The "vec" operator stacks each column of a matrix underneath the previous one. After introducing the new symbols

$$\boldsymbol{Y} := \operatorname{vec}\left[\boldsymbol{y}, \boldsymbol{X}\right], \ \boldsymbol{e} := \operatorname{vec}\left[\boldsymbol{e}_y, \boldsymbol{E}_X\right], \ \boldsymbol{\Xi} := \boldsymbol{\beta}_\mu, \tag{1.2}$$

the above EIV-Model can be interpreted as a *nonlinear Gauss-Helmert Model (GHM)*

$$\boldsymbol{b}(\underbrace{\boldsymbol{Y} - \boldsymbol{e}}_{n\times 1}, \underbrace{\boldsymbol{\Xi}}_{m\times 1}) = \boldsymbol{0}, \ \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 \boldsymbol{P}^{-1}), \tag{1.3}$$

where $\boldsymbol{b} : \mathbb{R}^{m+n} \to \mathbb{R}^{m+r}$ denotes a given multivariate nonlinear function $(r < n)$.

Obviously, in the case of the EIV-Model (1.1a) and (1.1b), it holds that:

$$\boldsymbol{b}(\boldsymbol{Y} - \boldsymbol{e}, \boldsymbol{\Xi}) := \boldsymbol{y} - \boldsymbol{X} \cdot \boldsymbol{\beta}_\mu - \left[\boldsymbol{I}_k, -(\boldsymbol{\beta}_\mu^T \otimes \boldsymbol{I}_k)\right] \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_X \end{bmatrix} = \boldsymbol{0}, \tag{1.4a}$$

$$\boldsymbol{P} := \boldsymbol{I}_n, \ r = n - \operatorname{rk} \boldsymbol{X} = n - m \ (\text{"redundancy"}). \tag{1.4b}$$

Here, $\otimes$ denotes the Kronecker-Zehfuss product of matrices, defined by

$$\boldsymbol{G} \otimes \boldsymbol{H} := \left[g_{ij} \cdot \boldsymbol{H}\right] \ \text{if} \ \boldsymbol{G} = [g_{ij}]. \tag{1.5}$$

The model (1.4a) and (1.4b) is better known in the form

$$\boldsymbol{y} := \underset{k\times m}{\boldsymbol{X}} \cdot \boldsymbol{\beta}_\mu + \underset{k\times n}{\boldsymbol{B}}\, \boldsymbol{e}, \quad \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 \boldsymbol{I}_n), \tag{1.6a}$$

where the $k \times n$ matrix $\boldsymbol{B}$ is defined as

$$\boldsymbol{B} := \left[\boldsymbol{I}_k, -(\boldsymbol{\beta}_\mu^T \otimes \boldsymbol{I}_k)\right] = \boldsymbol{B}(\boldsymbol{\beta}_\mu), \tag{1.6b}$$

thus depending on the parameter vector $\boldsymbol{\beta}_\mu$. Therefore, the standard Least-Squares solution of (1.6a) and (1.6b), which corresponds to the TLS solution of (1.1a) and (1.1b), can be generated by *iterative linearization* via the "normal equations" (for $j \in \mathbb{N}_0$)

$$\boxed{\left(\boldsymbol{A}^{(j)}\right)^T \left[\boldsymbol{B}^{(j)} \left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1} \boldsymbol{A}^{(j)} \cdot \hat{\boldsymbol{\xi}}_{j+1} = \left(\boldsymbol{A}^{(j)}\right)^T \left[\boldsymbol{B}^{(j)} \left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1} \boldsymbol{w}_j}, \tag{1.7a}$$

where

$$\boldsymbol{\xi}_{j+1} := \boldsymbol{\Xi} - \boldsymbol{\Xi}^{(j)} = \boldsymbol{\beta}_\mu - \boldsymbol{\beta}_\mu^{(j)}, \quad \boldsymbol{\mu}_j := \boldsymbol{Y} - \tilde{\boldsymbol{e}}^{(j)}, \tag{1.7b}$$

$$\boldsymbol{A}^{(j)} := \left.\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\Xi}^T}\right|_{\boldsymbol{\mu}_j, \boldsymbol{\beta}_\mu^{(j)}} = \boldsymbol{X} - \tilde{\boldsymbol{E}}_X^{(j)}, \quad \boldsymbol{B}^{(j)} = \boldsymbol{B}(\boldsymbol{\beta}_\mu^{(j)}), \tag{1.7c}$$

$$\boldsymbol{w}_j := \boldsymbol{b}(\boldsymbol{\mu}_j, \boldsymbol{\beta}_\mu^{(j)}) + \boldsymbol{B}^{(j)}(\boldsymbol{Y} - \boldsymbol{\mu}_j) \approx \boldsymbol{b}(\boldsymbol{Y}, \boldsymbol{\beta}_\mu^{(j)}) = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_\mu^{(j)}, \tag{1.7d}$$

and

$$\boxed{\tilde{\boldsymbol{e}}^{(j+1)} = \left(\boldsymbol{B}^{(j)}\right)^T \left[\boldsymbol{B}^{(j)} \left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1} \left(\boldsymbol{w}_j - \boldsymbol{A}^{(j)} \hat{\boldsymbol{\xi}}_{j+1}\right)}. \tag{1.7e}$$

After convergence, indicated by $\|\hat{\boldsymbol{\xi}}_{j+1}\| < \delta$ for a chosen threshold δ, the Mean Squared Error (MSE) matrix and the dispersion matrix of $\hat{\boldsymbol{\Xi}} = \hat{\boldsymbol{\beta}}_\mu = \boldsymbol{\beta}_\mu^{(j)} + \hat{\boldsymbol{\xi}}_{j+1}$ coincide in first-order approximation as:

$$D\{\hat{\boldsymbol{\Xi}}\} = \sigma_0^2 \Big[\big(\boldsymbol{A}^{(j)}\big)^T \Big[\boldsymbol{B}^{(j)} \big(\boldsymbol{B}^{(j)}\big)^T\Big]^{-1} \boldsymbol{A}^{(j)} \Big]^{-1} = \\ = \Big[(\boldsymbol{X} - \tilde{\boldsymbol{E}})^T (\boldsymbol{X} - \tilde{\boldsymbol{E}}) \Big]^{-1} \cdot \sigma_0^2 (1 + \hat{\boldsymbol{\Xi}}^T \hat{\boldsymbol{\Xi}}) \approx \mathrm{MSE}\{\hat{\boldsymbol{\Xi}}\}, \tag{1.8a}$$

with the variance component estimate best computed through:

$$\hat{\sigma}_0^2 (1 + \hat{\boldsymbol{\Xi}}^T \hat{\boldsymbol{\Xi}}) = r^{-1} \cdot \boldsymbol{w}_j^T (\boldsymbol{w}_j - \boldsymbol{A}^{(j)} \hat{\boldsymbol{\xi}}_{j+1}) = \tag{1.8b}$$

$$= r^{-1} \cdot (\boldsymbol{y} - \boldsymbol{X} \cdot \hat{\boldsymbol{\Xi}})^T (\boldsymbol{y} - \boldsymbol{X} \cdot \hat{\boldsymbol{\Xi}}). \tag{1.8c}$$

In the *penalized TLS approach*, the original objective function

$$\boldsymbol{e}^T \boldsymbol{e} = \min. \text{ subject to (1.6a) and (1.6b)} \tag{1.9}$$

is now replaced by

$$\boldsymbol{e}^T \boldsymbol{e} + \lambda \cdot \boldsymbol{\beta}_\mu^T \boldsymbol{R} \boldsymbol{\beta}_\mu = \min. \text{ subject to (1.6a) and (1.6b),} \tag{1.10}$$

where λ is a chosen positive number and $\boldsymbol{R}$ a chosen positive-definite $m \times m$ matrix. Following Schaffrin and Snow (2010), the "modified normal equations" now read (for $j \in \mathbb{N}_0$)

$$\Big[\lambda\big(1 + (\boldsymbol{\beta}_\mu^{(j)})^T \boldsymbol{\beta}_\mu^{(j)}\big) \cdot \boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)\Big] \cdot \hat{\hat{\boldsymbol{\xi}}}_{j+1} = \\ = \Big[\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)^T \boldsymbol{w}_j - \lambda\big(1 + (\boldsymbol{\beta}_\mu^{(j)})^T \boldsymbol{\beta}_\mu^{(j)}\big) \cdot \boldsymbol{R} \cdot \boldsymbol{\beta}_\mu^{(j)}\Big] \tag{1.11a}$$

with

$$\tilde{\tilde{\boldsymbol{e}}}^{(j+1)} = \big(\boldsymbol{B}^{(j)}\big)^T \Big[\boldsymbol{B}^{(j)} \big(\boldsymbol{B}^{(j)}\big)^T\Big]^{-1} \big(\boldsymbol{w}_j - \boldsymbol{A}^{(j)} \hat{\hat{\boldsymbol{\xi}}}_{j+1}\big). \tag{1.11b}$$

After convergence, where $\|\hat{\hat{\boldsymbol{\xi}}}_{j+1}\| < \delta$ for a chosen threshold δ, the *bias vector* results in first-order approximation as

$$\boldsymbol{\beta} \approx -\Big[\boldsymbol{I}_m + \lambda^{-1} (1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}})^{-1} \cdot \boldsymbol{R}^{-1} \cdot \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\Big]^{-1} \cdot \boldsymbol{\Xi}, \tag{1.12a}$$

with

$$\hat{\hat{\boldsymbol{\Xi}}} = \hat{\hat{\boldsymbol{\beta}}}_\mu = \boldsymbol{\beta}_\mu^{(j)} + \hat{\hat{\boldsymbol{\xi}}}_{j+1}, \tag{1.12b}$$

and the *dispersion matrix* of $\hat{\hat{\boldsymbol{\Xi}}}$ as

$$D\{\hat{\hat{\boldsymbol{\Xi}}}\} \approx \sigma_0^2 (1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}) \cdot \Big[\lambda\big(1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}\big) \cdot \boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\Big]^{-1} \cdot \\ \cdot \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big) \cdot \Big[\lambda\big(1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}\big) \cdot \boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\Big]^{-1}, \tag{1.12c}$$

from which the MSE matrix of $\hat{\hat{\boldsymbol{\Xi}}}$ can be computed via

$$MSE\{\hat{\hat{\boldsymbol{\Xi}}}\} = D\{\hat{\hat{\boldsymbol{\Xi}}}\} + \boldsymbol{\beta}\boldsymbol{\beta}^T \approx \\ \approx \sigma_0^2 (1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}) \cdot \Big[\lambda\big(1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}\big) \cdot \boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\Big]^{-1} \cdot \\ \cdot \Big[\lambda\big(1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}\big) \cdot \boldsymbol{R}\boldsymbol{\Xi} \cdot (\lambda/\sigma_0^2) \cdot \boldsymbol{\Xi}^T \boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\Big] \cdot \\ \cdot \Big[\lambda\big(1 + \hat{\hat{\boldsymbol{\Xi}}}^T \hat{\hat{\boldsymbol{\Xi}}}\big) \cdot \boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\Big]^{-1}. \tag{1.12d}$$

This formula will become the basis for the MSE-risk comparison in chapter 3.

2. TLS Collocation: The Least-Squares approach to the EIV-Model with prior information

In this chapter, *stochastic prior information* is added to the EIV-Model (1.1a) and (1.1b) so that it now shows a random effects vector $\boldsymbol{b}_\mu$ in lieu of the original parameter vector $\boldsymbol{\beta}_\mu$:

$$\boldsymbol{y} = \underbrace{(\boldsymbol{X} - \boldsymbol{E}_X)}_{k\times m}\boldsymbol{b}_\mu + \boldsymbol{e}_y, \quad \boldsymbol{\beta}_0 = \boldsymbol{b}_\mu + \boldsymbol{e}_0, \tag{2.1a}$$

$$\begin{bmatrix}\boldsymbol{e}_y\\ \boldsymbol{e}_X\\ \boldsymbol{e}_0\end{bmatrix} \sim \left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2\left[\begin{array}{cc|c}\boldsymbol{I}_k & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_{km} & \boldsymbol{0}\\ \hline \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Q}_0\end{array}\right]\right), \tag{2.1b}$$

where $\boldsymbol{\beta}_0$ is a given $m \times 1$ vector of prior information, and
$\boldsymbol{e}_0$ is an $m \times 1$ vector of random errors with
$\boldsymbol{Q}_0$ as a nonnegative-definite (nnd) cofactor matrix.

To simply the formulas, $\boldsymbol{\beta}_0 := \boldsymbol{0}$ is set in the following.

Obviously, in its linearized form corresponding to (1.6a) and (1.6b), the model (2.1a) and (2.1b) can be summarized as:

$$\boldsymbol{y} = \underset{k\times m}{\boldsymbol{X}} \cdot \boldsymbol{b}_\mu + \underset{k\times n}{\boldsymbol{B}}\,\boldsymbol{e}, \quad \boldsymbol{0} = \boldsymbol{b}_\mu + \boldsymbol{e}_0, \tag{2.2a}$$

$$\begin{bmatrix}\boldsymbol{e}\\ \boldsymbol{e}_0\end{bmatrix} \sim \left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2\begin{bmatrix}\boldsymbol{I}_n & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{Q}_0\end{bmatrix}\right), \tag{2.2b}$$

where the $k \times n$ matrix $\boldsymbol{B}$ now depends on $\boldsymbol{b}_\mu$:

$$\boldsymbol{B} := \left[\boldsymbol{I}_n, -(\boldsymbol{b}_\mu^T \otimes \boldsymbol{I}_k)\right] = \boldsymbol{B}(\boldsymbol{b}_\mu). \tag{2.2c}$$

Following the approach by Snow and Schaffrin (2012), who applied weighted TLS adjustment to model (2.2a) to (2.2c), the predicted vector of random effects may be iteratively computed (for $j \in \mathbb{N}_0$) via

$$\begin{aligned}\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j+1)} &= -\tilde{\tilde{\boldsymbol{e}}}_0^{(j+1)} = \left[(\boldsymbol{A}^{(j)})^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1}\boldsymbol{A}^{(j)} + \boldsymbol{Q}_0^{-1}\right]^{-1}\cdot\\ &\quad\cdot(\boldsymbol{A}^{(j)})^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1}\left(\boldsymbol{y} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\cdot\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j)}\right) = \quad \text{if } \boldsymbol{Q}_0^{-1} \text{ exists,} \qquad (2.3a)\\ &= \left[\boldsymbol{I}_m + \boldsymbol{Q}_0(\boldsymbol{A}^{(j)})^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1}\boldsymbol{A}^{(j)}\right]^{-1}\cdot\\ &\quad\cdot\boldsymbol{Q}_0(\boldsymbol{A}^{(j)})^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1}\left(\boldsymbol{y} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\cdot\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j)}\right) = \qquad (2.3b)\\ &= \boldsymbol{Q}_0(\boldsymbol{A}^{(j)})^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T + \boldsymbol{A}^{(j)}\boldsymbol{Q}_0(\boldsymbol{A}^{(j)})^T\right]^{-1}\left(\boldsymbol{y} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\cdot\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j)}\right) \qquad (2.3c)\\ &\quad \text{if } \boldsymbol{Q}_0 \text{ is possibly singular,}\end{aligned}$$

where, analogous to (1.7c), the identity $\boldsymbol{A}^{(j)} = \boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}$ has been used. The iteration procedure also permits computation of the residual vector, $\tilde{\tilde{\boldsymbol{e}}} := \text{vec}[\tilde{\tilde{\boldsymbol{e}}}_y, \tilde{\tilde{\boldsymbol{E}}}_X]$, as

$$\begin{aligned}\tilde{\tilde{\boldsymbol{e}}}^{(j+1)} &= \left(\boldsymbol{B}^{(j)}\right)^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1}\cdot\left[\boldsymbol{y} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\cdot\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j)} - \boldsymbol{A}^{(j)}\cdot\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j+1)}\right] = \qquad (2.3d)\\ &= \left(\boldsymbol{B}^{(j)}\right)^T\left[\boldsymbol{B}^{(j)}\left(\boldsymbol{B}^{(j)}\right)^T\right]^{-1}\cdot\left[\left(\boldsymbol{y} - \boldsymbol{X}\cdot\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j+1)}\right) + \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\left(\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j+1)} - \tilde{\tilde{\boldsymbol{b}}}_\mu^{(j)}\right)\right]. \qquad (2.3e)\end{aligned}$$

The solution $\tilde{\tilde{\boldsymbol{b}}}_\mu$, after convergence where $\|\tilde{\tilde{\boldsymbol{b}}}_\mu^{(j+1)} - \tilde{\tilde{\boldsymbol{b}}}_\mu^{(j)}\| < \delta$ for a chosen threshold δ, is known to be "*weakly unbiased*" in the sense that

$$E\{\tilde{\tilde{\boldsymbol{b}}}_\mu\} = \boldsymbol{0} = \boldsymbol{\beta}_0 = E\{\boldsymbol{b}_\mu\} \tag{2.4}$$

holds for the given prior information vector $\boldsymbol{\beta}_0$. Thus, the MSE matrix of $\tilde{\tilde{\boldsymbol{b}}}_\mu$ and the dispersion matrix of $(\tilde{\tilde{\boldsymbol{b}}}_\mu - \boldsymbol{b}_\mu)$ will coincide in first-order approximation, namely, if $\boldsymbol{Q}_0^{-1}$ exists:

$$D\{\tilde{\tilde{\boldsymbol{b}}}_\mu - \boldsymbol{b}_\mu\} = \sigma_0^2\big[\boldsymbol{Q}_0^{-1} + (\boldsymbol{A}^{(j)})^T\big[\boldsymbol{B}^{(j)}\big(\boldsymbol{B}^{(j)}\big)^T\big]^{-1}\boldsymbol{A}^{(j)}\big]^{-1} =$$
$$= \sigma_0^2\big(1 + \tilde{\tilde{\boldsymbol{b}}}_\mu^T\tilde{\tilde{\boldsymbol{b}}}_\mu\big)\big[\big(1 + \tilde{\tilde{\boldsymbol{b}}}_\mu^T\tilde{\tilde{\boldsymbol{b}}}_\mu\big)\boldsymbol{Q}_0^{-1} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\big]^{-1} \approx \mathrm{MSE}\{\tilde{\tilde{\boldsymbol{b}}}_\mu\}, \quad (2.5a)$$

with the variance component estimate best computed via

$$\hat{\hat{\sigma}}_0^2\big(1 + \tilde{\tilde{\boldsymbol{b}}}_\mu^T\tilde{\tilde{\boldsymbol{b}}}_\mu\big) = n^{-1}\cdot\big(\boldsymbol{y} - \boldsymbol{X}\tilde{\tilde{\boldsymbol{b}}}_\mu\big)^T\cdot$$
$$\cdot\big[\boldsymbol{I}_m + \big(1 + \tilde{\tilde{\boldsymbol{b}}}_\mu^T\tilde{\tilde{\boldsymbol{b}}}_\mu\big)^{-1}\cdot\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\boldsymbol{Q}_0\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T\big]\big(\boldsymbol{y} - \boldsymbol{X}\tilde{\tilde{\boldsymbol{b}}}_\mu\big). \quad (2.5b)$$

In the following chapter, formulas (2.3a) to (2.3c) need to be compared with (1.11a), and – above all – a criterion ought to be found that explains when $\tilde{\tilde{\boldsymbol{b}}}_\mu$ is superior over $\hat{\hat{\boldsymbol{\Xi}}} = \hat{\hat{\boldsymbol{\beta}}}_\mu$, and vice versa.

3. A comparison of MSE-risks

In order to compare $\tilde{\tilde{\boldsymbol{b}}}_\mu$ directly with $\hat{\hat{\boldsymbol{\Xi}}} = \hat{\hat{\boldsymbol{\beta}}}_\mu = \boldsymbol{\beta}_\mu^{(j)} + \hat{\hat{\boldsymbol{\xi}}}_{j+1}$, formula (1.11a) needs to be further transformed into

$$\big[\lambda\big(1 + (\boldsymbol{\beta}_\mu^{(j)})^T\boldsymbol{\beta}_\mu^{(j)}\big)\cdot\boldsymbol{R} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)^T\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)\big]\big(\boldsymbol{\beta}_\mu^{(j)} + \hat{\hat{\boldsymbol{\xi}}}_{j+1}\big) =$$
$$= \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)^T\big[\big(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_\mu^{(j)}\big) + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)\boldsymbol{\beta}_\mu^{(j)}\big] = \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)^T\big(\boldsymbol{y} - \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\cdot\boldsymbol{\beta}_\mu^{(j)}\big), \quad (3.1)$$

where the identity $\boldsymbol{w}_j = \boldsymbol{y} - \big(\boldsymbol{A}^{(j)} + \tilde{\tilde{\boldsymbol{E}}}_X^{(j)}\big)\boldsymbol{\beta}_\mu^{(j)}$ has been used. Equation (3.1) clearly coincides with (2.3a) *numerically* if and only if the condition

$$\boldsymbol{Q}_0^{-1} = \lambda\cdot\boldsymbol{R} \quad (3.2)$$

is fulfilled. Under this assumption, however, formula (1.12d) turns into

$$\mathrm{MSE}\{\hat{\hat{\boldsymbol{\Xi}}} = \hat{\hat{\boldsymbol{\beta}}}_\mu\} = \sigma_0^2(1 + \hat{\hat{\boldsymbol{\Xi}}}^T\hat{\hat{\boldsymbol{\Xi}}})\cdot\big[(1 + \hat{\hat{\boldsymbol{\Xi}}}^T\hat{\hat{\boldsymbol{\Xi}}})\cdot\boldsymbol{Q}_0^{-1} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\big]^{-1}\cdot$$
$$\cdot\big[(1 + \hat{\hat{\boldsymbol{\Xi}}}^T\hat{\hat{\boldsymbol{\Xi}}})\cdot\boldsymbol{Q}_0^{-1}(\boldsymbol{\Xi}\sigma_0^{-2}\boldsymbol{\Xi}^T)\boldsymbol{Q}_0^{-1} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\big]\cdot$$
$$\cdot\big[(1 + \hat{\hat{\boldsymbol{\Xi}}}^T\hat{\hat{\boldsymbol{\Xi}}})\cdot\boldsymbol{Q}_0^{-1} + \big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)^T\big(\boldsymbol{X} - \tilde{\tilde{\boldsymbol{E}}}_X\big)\big]^{-1}, \quad (3.3)$$

with $\tilde{\tilde{\boldsymbol{E}}}_X$ representing the same residual matrix in both formulas.

Hence, (3.3) can now be directly compared with formula (2.5a) for the MSE matrix of $\tilde{\tilde{\boldsymbol{b}}}_\mu$, resulting in the

Theorem. *The Penalized Total Least-Squares solution* $\hat{\hat{\boldsymbol{\Xi}}} = \hat{\hat{\boldsymbol{\beta}}}_\mu$ *with* $\boldsymbol{Q}_0^{-1} := \lambda\cdot\boldsymbol{R}$ *is* superior *to the "Bayesian estimate"* $\tilde{\tilde{\boldsymbol{b}}}_\mu$*, that is based on the "TLS Collocation" solution, whenever the difference between their MSE matrices is, at least, positive-semidefinite:*

$$\mathrm{MSE}\{\tilde{\tilde{\boldsymbol{b}}}_\mu\} - \mathrm{MSE}\{\hat{\hat{\boldsymbol{\beta}}}_\mu\} \geq_L \boldsymbol{0}, \quad (3.4)$$

where $\geq_L$ *denotes Löwner's partial ordering of matrices; see, e.g., Marshall et al. (2009). This is the case if and only if one of the equivalent inequalities*

$$\boldsymbol{\Xi}^T\boldsymbol{Q}_0^{-1}\boldsymbol{\Xi} \leq \sigma_0^2 \quad \textit{or} \quad \boldsymbol{\beta}_\mu^T\boldsymbol{R}\boldsymbol{\beta}_\mu \leq \sigma_0^2/\lambda \quad (3.5a\text{-}b)$$

holds true.

Proof. From a direct comparison of (2.5a) with (3.3), noting that $\hat{\hat{\boldsymbol{\Xi}}} = \hat{\hat{\boldsymbol{\beta}}} \hat{=} \tilde{\tilde{\boldsymbol{b}}}_\mu$ numerically, it becomes immediately clear that

$$\text{MSE}\{\tilde{\tilde{\boldsymbol{b}}}_\mu\} \geq_L \text{MSE}\{\hat{\hat{\boldsymbol{\beta}}}_\mu\} \iff \boldsymbol{Q}_0^{-1} \geq_L \boldsymbol{Q}_0^{-1}(\Xi\sigma_0^{-2}\Xi^T)\boldsymbol{Q}_0^{-1},$$

which first implies the inequality $\Xi^T\boldsymbol{Q}_0^{-1}\Xi \geq (\Xi^T\boldsymbol{Q}_0^{-1}\Xi)^2/\sigma_0^2$ and, with (3.2) and $\Xi = \boldsymbol{\beta}_\mu \neq \mathbf{0}$, also the inequality $\boldsymbol{\beta}_\mu^T\boldsymbol{R}\boldsymbol{\beta}_\mu \leq \sigma_0^2/\lambda$.

That each of these inequalities is also sufficient for (3.4) to hold true, now follows from a result by Baksalary and Kala (1983). □

Obviously, the inequalities (3.5a-b) become only practical when the quantities $\Xi = \boldsymbol{\beta}_\mu$, σ_0^2, and λ are replaced by certain estimates such as $\tilde{\tilde{\boldsymbol{b}}}_\mu$ from (2.3a) to (2.3c) and $\hat{\hat{\sigma}}_0^2$ from (2.5b). How a proper estimate can be found for the "*Tykhonov parameter*" λ, must be left to a future paper.

4. Conclusions and outlook

In this contribution the Penalized TLS approach within an EIV-Model has been compared with the "Bayesian" TLS approach within an EIV-Model with prior information, also known as "TLS-Collocation." Surprisingly, both results turn out to be *numerically identical* under the mild matrix identity (3.2). They are not fully identical, however, since they have *different MSE-risks* associated with them. In fact, a criterion could be given that allows to decide which solution is superior over the other. This criterion proves to be *essentially the same* as the one developed by Schaffrin (2008) to compare the MSE-risk of the Penalized Least-Squares solution within a Gauss-Markov Model with the "Bayesian" Least-Squares solution within a Random Effects Model, i.e., the "LS-Collocation" solution. A quasi-optimal estimate of "Tykhonov's regularization parameter," needed for the Penalized TLS solution, will be presented in a future paper.

References

Baksalary, J. and Kala, R. (1983). Partial ordering between matrices one of which is of rank one. *Bulletin of the Polish Acadamey of Sciences: Mathematics*, 31:5–7.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht/NL.

Grafarend, E. W. and Schaffrin, B. (1993). *Adjustment Computations in Linear Models (in German)*, Bibliograph Inst., Mannheim etc., Germany.

Marshall, A. W., Olkin, I., and Arnold, B. C. (2009). *Inequalities: Theory of Majorization and Its Applications*, 2nd. edition, Springer, New York, etc.

Rao, C. R., Toutenburg, H., Shalabh, and Heumann, C. (2008). *Linear Models and Generalizations: Least Squares and Alternatives*, 3rd. edition, Springer, Berlin, Germany.

Schaffrin, B. (2008). On Penalized Least-Squares: Its Mean Squared Error and a quasi-optimal weight ratio. In: Shalabh and Heumann, C., editors, *Recent Advances in Linear Models and Related Areas*, pages 313–322. Physica-Verlag, Heidelberg, Germany.

Schaffrin, B. (2009). Total Least-Squares Collocation: The Total-Least Squares approach to EIV-Models with prior information. Presented at the 18th Intl. Workshop on Matrices and Statistics, Smolenice Castle, Slovakia.

Schaffrin, B. and Snow, K. (2010). Total Least-Squares regularization of Tykhonov type and an ancient racetrack in Corinth. *Linear Algebra and its Applications*, 432(8):2061–2076.

Snow, K. and Schaffrin, B. (2012). Weighted Total Least-Squares Collocation with Geodetic Applications. Presented at the 2012 SIAM Conference on Applied Linear Algebra, Valencia, Spain.