ON MIXED TOTAL LEAST SQUARES FOR PARAMETER ESTIMATION

Cuiping Guo\textsuperscript{a}, Junhuan Peng\textsuperscript{b} and Chuantao Li\textsuperscript{a}

\textsuperscript{a}School of Science, China University of Geosciences (Beijing), Beijing, 100083 P. R. China
\textsuperscript{b}School of Land Science and Technology, China University of Geosciences (Beijing), Beijing, 100083, P. R. China

Abstract

A mixed weighted total least squares (MWTLS) method is presented for an errors-in-variables (EIV) model with some fixed columns in the coefficient matrix. An iterative algorithm is derived based on the proposed MWTLS method. Compared with the classical WTLS method, the method represented in this paper improves the computing speed of the estimated parameter, while the fixed columns of the coefficient matrix keeps unperturbed. Furthermore, a simulated example is carried out to demonstrate the performance of the proposed algorithm.

Keywords: errors-in-variables, total least squares, mixed weighted total least squares, parameter estimation.
1. Introduction

Consider the linear model

\[ Y = A\beta + \epsilon, \]

\[ \epsilon \sim (0, \sigma^2 Q), \quad (1) \]

where \( A \in R^{n \times p} \) is the coefficient matrix with \( n > m \), and \( Y \in R^p \) is the observation vector. It is often assumed in the least square (LS) problem that, the coefficient matrix is exactly known, and only the observation vector has measurement errors. In practice, however, we may encounter the situation in which some or all the elements of the coefficient matrix \( A \) are not given deterministically, and they are measured or derived from measurements with random errors. Ignoring such errors in the coefficient matrix usually results in biased estimates of the parameter vector \( \beta \), which may cause difficulties and complications in conducting statistical analysis. Models with a random coefficient matrix \( A \) have been well known as errors-in-variables (EIV) models (see Fuller [7]; Cheng and Van Ness [5]; Carroll et al. [4]; Buonaccorsi [3]). An appropriate approach to solve EIV models is the total least squares (TLS) method. Now, there are many researches about the total least squares in algorithms such as the singular value decomposition (SVD) algorithm (Golub and Van Loan [8]) and the algorithm based on the Lagrange function (Schaffrin et al. [16, 17]; Fang [6]). For more information about the methodology of TLS, one can refer to Huffel et al. [10, 11]. Recently, TLS has attained much attention. In many applications, such as remote sensing and geodetic datum transformation, data are collected by different instruments with different precisions. Then the ordinary TLS approach is generalized to so called weighted TLS (WTLS) approach where both the observation vector and coefficient matrix have different weight matrices (Schaffrin et al. [16]; Jazaeri et al. [12]).
In practice, some of the columns of the coefficient matrix may be known exactly. For example, in linear regression with measured independent variables, the elements of the first column of the coefficient matrix is often equal to unity and are surely not random. In system identification, when inputs and outputs of a system are measured with error, and the system is modelled as a transfer function, the coefficient matrix of the function may have exactly known columns. An EIV model with exactly known columns is called a mixed EIV (MEIV) model. Obviously, the computational cost of the MEIV model will increase dramatically according to the traditional solutions of EIV model. For the MEIV model, Golub et al. [9] developed the mixed LS-TLS method by computing a QR factorization of the known columns and then solved a TLS problem of reduced dimension. And later, this method was extended by Huffel et al. [10, 11] to the multi-dimensional case.

In survey engineering, Akyilmaz [1] performed mixed LS-TLS method to solve coordinate transformation problems. Amiri-Simkooei et al. [2] presented the weighted TLS formulation for a MEIV model based on the standard least squares. In this paper, we will provide the mixed weighted total least squared (MWTLS) estimators for the MEIV models by Lagarange-multiplier method. The structure of this paper is as follows. Section 2 presents MWTLS estimators for MEIV models. A simulation study is included in Section 3, and the paper is concluded in Section 4.

2. Weighted Total Least Squared Estimators for the MEIV Model

Let the EIV model be defined as follows:

$$y = (A - E_A)\xi + e_y,$$

$$\begin{bmatrix} e_y \\ e_A \end{bmatrix} \sim \begin{bmatrix} 0 \\ \sigma_\delta^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} \end{bmatrix}, \quad P_y = Q_y^{-1}, \quad P_A = Q_A^{-1},$$

(2)
where $y$ and $e_y$ are the $n \times 1$ observation vector and the corresponding random error vector in (2), respectively. Matrix $A$ and $E_A$ are full column-rank $n \times p$ stochastic and the corresponding random error matrix, respectively. Vector $\xi$ is the unknown parameter vector by $n \times 1$. Denote $e_A = \text{vec}(E_A)$ (‘vec’ denotes the operator that stacks one column of a matrix underneath the previous one). The symbol $\sigma_0^2$ denotes the unknown variance component. Matrices $Q_y$ and $P_y$ are the cofactor matrix and the weight matrix of the observation vector $y$, and $Q_A$ and $P_A$ are the cofactor matrix and the weight matrix of the matrix $e_A$.

Without loss of generality, we assume that the first $p - t$ columns of the coefficient matrix do not contain errors, but the last $t$ columns of the coefficient matrix do not contain errors. The MEIV model be defined as follows:

$$ y = A_1 \beta_1 + (A_2 - E_2) \beta_2 + e_y, $$

$$ \begin{bmatrix} e_y \\ e_2 \end{bmatrix} \sim \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_2^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_2 \end{bmatrix}. $$

(3)

In (3), $y$ and $e_y$ are the $n \times 1$ observation vector and the corresponding random error vector, respectively. Matrices $A_1$ is the $n \times (p - t)$ matrix, $A_2$ is the $n \times t$ matrix, $E_2$ is the corresponding $n \times t$ matrix of random errors, $\beta_1$ is the $p - t$ vector of unknown parameters, $\beta_2$ is the $t$ vector of unknown parameters, $e_2 = \text{vec}(E_2)$ (‘vec’ denotes the operator that stacks one column of a matrix underneath the previous one). The symbol $\sigma_0^2$ denotes the unknown variance component and $Q_y$ is the cofactor matrix of the observation vector $y$, and $Q_2$ is the cofactor matrix of $E_2$. 
Using $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$ (Lütkepohl [13]), the model in (3) can be rewritten as

$$y = A_1\beta_1 + A_2\beta_2 - (\beta_2^T \otimes I_n )e_2 + e_y,$$

where $\otimes$ denotes the Kronecker product of matrices.

The weighted total least squares principle is to minimize the objective function

$$S = e_y^T Q_y^{-1}e_y + e_2^T Q_2^{-1}e_2.$$

By employing the equivalent target function in accordance with Lagrange, we have

$$\Phi(e_y, e_2, \lambda, \xi) = e_y^T Q_y^{-1}e_y + e_2^T Q_2^{-1}e_2 + 2\lambda^T [y - A_1\hat{\beta}_1 - A_2\hat{\beta}_2 + (\beta_2^T \otimes I_n )e_2 - e_y].$$

Then the necessary Euler-Lagrange conditions are derived, namely,

$$\frac{1}{2} \frac{\partial \Phi}{\partial e_y} \bigg|_{\tilde{e}_y, \tilde{e}_2, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = Q_y^{-1}\tilde{e}_y - \hat{\lambda} = 0,$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial e_2} \bigg|_{\tilde{e}_y, \tilde{e}_2, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = Q_2^{-1}\tilde{e}_2 + \left(\hat{\beta}_2^T \otimes I_n \right)\hat{\lambda} = 0,$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \lambda} \bigg|_{\tilde{e}_y, \tilde{e}_2, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = y - A_1\hat{\beta}_1 - A_2\hat{\beta}_2 + \left(\hat{\beta}_2^T \otimes I_n \right)\tilde{e}_2 - \tilde{e}_y = 0,$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \hat{\beta}_1} \bigg|_{\tilde{e}_y, \tilde{e}_2, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = -A_1^T \hat{\lambda} = 0,$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \hat{\beta}_2} \bigg|_{\tilde{e}_y, \tilde{e}_2, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2} = -A_2^T \hat{\lambda} + \tilde{E}_2^T \hat{\lambda} = 0,$$

where tildas indicate “predicted” vectors, and hats indicate “estimated” ones. Now, \tilde{e}_y and \tilde{e}_2 can be expressed in terms of \hat{\lambda} by using (5) and (6). This leads to
\[
\begin{align*}
\tilde{e}_y &= Q_y \hat{\lambda}, \\
\tilde{e}_2 &= -Q_2 \left( \hat{\beta}_2 \otimes I_n \right) \hat{\lambda},
\end{align*}
\] (10)

and after inserting this into (7), we obtain
\[
\hat{\lambda} = \left[ Q_y + \left( \hat{\beta}_2^T \otimes I_n \right) Q_2 \left( \hat{\beta}_2 \otimes I_n \right) \right]^{-1} \left( y - A_1 \hat{\beta}_1 - A_2 \hat{\beta}_2 \right). \\
\] (12)

Let
\[
Q_1 = Q_y + \left( \hat{\beta}_2^T \otimes I_n \right) Q_2 \left( \hat{\beta}_2 \otimes I_n \right),
\] (13)

where \( Q_1 \) is invertible. We readily obtain
\[
\hat{\lambda} = Q_1^{-1} \left( y - A_1 \hat{\beta}_1 - A_2 \hat{\beta}_2 \right). \\
\] (14)

Inserting (14) into (8), we get
\[
\hat{\beta}_1 = \left( A_1^T Q_1^{-1} A_1 \right)^{-1} A_1^T Q_1^{-1} \left( y - A_2 \hat{\beta}_2 \right). \\
\] (15)

Let
\[
Q_3 = A_2^T Q_1^{-1} \left[ I_n - A_1 \left( A_1^T Q_1^{-1} A_1 \right)^{-1} A_1^T Q_1^{-1} \right],
\]

and using
\[
\tilde{E}_2^T \hat{\lambda} = \left( I_1 \otimes \hat{\lambda}^T \right) \bar{e}_2, \\
\] (16)

we obtain from (9) that
\[
\hat{\beta}_2 = (Q_3 A_2)^{-1} \left[ Q_3 y - \left( I_1 \otimes \hat{\lambda}^T \right) \bar{e}_2 \right]. \\
\] (17)

Inserting (17) into (15), the closed-form expression of the estimated parameter vector \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) can be derived as
\[
\begin{align*}
\hat{\beta}_1 &= \left( A_1^T Q_1^{-1} A_1 \right)^{-1} A_1^T Q_1^{-1} \left[ I - A_2 (Q_3 A_2)^{-1} \right] y + A_2 (Q_3 A_2)^{-1} \left( I_1 \otimes \hat{\lambda}^T \right) \bar{e}_2, \\
\hat{\beta}_2 &= (Q_3 A_2)^{-1} \left[ Q_3 y - \left( I_1 \otimes \hat{\lambda}^T \right) \bar{e}_2 \right].
\end{align*}
\] (18)
In the analogy with the standard least squares, the estimator of the variance component $\sigma^2$ is given as

$$\hat{\sigma}^2 = \frac{e_y^T Q_y^{-1} e_y + e_2^T Q_2^{-1} e_2}{n - p}. \quad (19)$$

Substitution of $\tilde{e}_y$ from (10) and $e_2$ from Equation (11) into (19), yields

$$\hat{\sigma}^2 = \frac{\lambda^T \left[ Q_y + (\hat{\beta}_2^T \otimes I_n) Q_2 (\hat{\beta}_2 \otimes I_n) \right] \lambda}{n - p}. \quad (20)$$

From (13), $\hat{\sigma}^2$ can be obtained as

$$\hat{\sigma}^2 = \frac{\lambda^T (y - A_1 \hat{\beta}_1 - A_2 \hat{\beta}_2)}{n - p}. \quad (21)$$

After giving an initial value of the parameter vector, we compute the solution with (18) iteratively. Note that, in each iteration, the matrices and vectors containing the parameter vector and the error matrix should be updated. Therefore, the MWTLS procedure for parameter estimation is summarized as follows:

**MWTLS Algorithm:**

**Step 1.** Set $\begin{pmatrix} \hat{\beta}_1^0 \\ \hat{\beta}_2^0 \end{pmatrix} = \left( (A_1, A_2)^T P_y \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \right)^{-1} (A_1, A_2)^T y$.

**Step 2.** For $i \in \mathbb{N}$, compute

$$Q_y^i = Q_y + (\hat{\beta}_2^i)^T (\hat{\beta}_2^i \otimes I_n) Q_2 (\hat{\beta}_2 \otimes I_n), \quad \lambda^i = \left( Q_y^i \right)^{-1} \left( y - A_1 \hat{\beta}_1^i - A_2 \hat{\beta}_2^i \right), \quad \tilde{e}_2^i = -Q_2 (\hat{\beta}_2^i \otimes I_n) \lambda^i, \quad \hat{\beta}_2^i = \left( Q_2 A_2 \right)^{-1} \left[ Q_2 y - (I \otimes (\lambda^i)^T) \tilde{e}_2^i \right].$$

**Step 3.** Stop when $\| \hat{\beta}_1^i - \hat{\beta}_1^{i-1} \| < \varepsilon$ and $\| \hat{\beta}_2^i - \hat{\beta}_2^{i-1} \| < \varepsilon$ for a chosen threshold $\varepsilon$. 

3. Numerical Examples

In order to demonstrate the performance of Algorithm 1, in this section, the algorithm will be applied to a straight line fitting problem representing the MEIV problem, compared with the general WLS algorithm and the algorithm proposed by Fang [6]. The observed data and their corresponding weights are listed in Table 1. We try to estimate the slope $a$ and the intercept $b$ of the regression line

$$y_i - e_{y_i} = a \cdot (x_i - e_{x_i}) + b,$$

using the MWTLS algorithm above.

**Table 1.** Observed data $(x_i, y_i)$ and corresponding weights, taken from Neri et al. [14]

<table>
<thead>
<tr>
<th>$i$</th>
<th>$x_i$</th>
<th>$W_{x_i}$</th>
<th>$y_i$</th>
<th>$W_{y_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>1000.0</td>
<td>5.9</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
<td>1000.0</td>
<td>5.4</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>1.8</td>
<td>500.0</td>
<td>4.4</td>
<td>4.0</td>
</tr>
<tr>
<td>4</td>
<td>2.6</td>
<td>800.0</td>
<td>4.6</td>
<td>8.0</td>
</tr>
<tr>
<td>5</td>
<td>3.3</td>
<td>200.0</td>
<td>3.5</td>
<td>20.0</td>
</tr>
<tr>
<td>6</td>
<td>4.4</td>
<td>80.0</td>
<td>3.7</td>
<td>20.0</td>
</tr>
<tr>
<td>7</td>
<td>5.2</td>
<td>60.0</td>
<td>2.8</td>
<td>70.0</td>
</tr>
<tr>
<td>8</td>
<td>6.1</td>
<td>20.0</td>
<td>2.8</td>
<td>70.0</td>
</tr>
<tr>
<td>9</td>
<td>6.5</td>
<td>1.8</td>
<td>2.4</td>
<td>100.0</td>
</tr>
<tr>
<td>10</td>
<td>7.4</td>
<td>1.0</td>
<td>1.5</td>
<td>500.0</td>
</tr>
</tbody>
</table>
Modifying (22) as follows:

\[
\begin{pmatrix}
y_1 \\ y_2 \\ \vdots \\ y_n
\end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} b + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} e_{x_1} \\ e_{x_2} \\ \vdots \\ e_{x_n} \end{pmatrix} a,
\]

(23)

where (23) is a typical MEIV model and * means that there is no result about the estimator of the variance component in Neri's paper.

**Table 2.** Results of straight line fit to observed data of Table 1

<table>
<thead>
<tr>
<th>Parameter estimate</th>
<th>Exact solution (Neri et al.)</th>
<th>WLS (Schaffrin et al.)</th>
<th>WTLS (Schaffrin et al.)</th>
<th>MTLS (this paper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>(-0.480533407)</td>
<td>(-0.610812956)</td>
<td>(-0.480533407)</td>
<td>(-0.480533407)</td>
</tr>
<tr>
<td>(b)</td>
<td>(5.47991022)</td>
<td>(6.100109317)</td>
<td>(5.479910224)</td>
<td>(5.479910224)</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>*</td>
<td>(4.29315094)</td>
<td>(1.48329415)</td>
<td>(1.48329415)</td>
</tr>
<tr>
<td>Cpu time</td>
<td>*</td>
<td>(0.0468)</td>
<td>(0.0624)</td>
<td>(\approx 0)</td>
</tr>
</tbody>
</table>

The threshold \(\epsilon = 10^{-10}\) has been chosen such as to allow a comparison of our MTLS to the “exact solution” reported by Neri et al. [14] and the solution reported by Schaffrin and Wieser [17]. In fact, the results indicate that the estimated line parameters are the same and coincide with the exact solutions as reported by Neri et al. [14] and Schaffrin and Wieser [17]. Therefore, it is possible to find exact results for a regression line analysis affected by errors, without requiring any kind of approximation. Compared with WTLS, the MTLS algorithm presented in this paper consumes little time, therefore, the computation speed is accelerated. In fact, if the coefficient matrix has more fixed columns, the MTLS algorithm can present its superiority over the TLS algorithm.
4. Conclusion

For a MEIV model, considering the random errors may exist in both of the observation vector and the coefficient matrix, and the coefficient matrix have some fixed columns, the MWTLS algorithm is constructed in this paper. A numerical example is carried out to demonstrate the performance of the MWTLS compared with WTLS. When all the columns of the coefficient matrix are fixed, the MTLS solution reduces to the ordinary LS estimate. When all the columns of the coefficient matrix are random, the MTLS solution becomes the TLS solution. So, it will have a more wide range of application.

Acknowledgement

The research is supported by the Fundamental Research Funds for Central Universities (No. 2652015193), Undergraduate Teaching-Reform Project (No. JGYB201529), Graduate Teaching-Reform Project of China of University of Geosciences (Beijing), NSFC (41374016, 41104025 and 41330634), and SKLGED2013-4-8-e.

References


