ELSEVIER
Linear Algebra and its Applications 348 (2002) 23–39

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Identification of influential observations on total least squares estimates

Baibing Li [a,*], Bart De Moor [b,1]

[a] *Centre for Process Analytics and Control Technology, Merz Court, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK*

[b] *ESAT/SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001, Leuven-Hevenlee, Belgium*

Received 1 August 2000; accepted 30 October 2001

Submitted by P. Van Dooren

Abstract

It is known that total least squares (TLS) estimates are very sensitive to outliers. Therefore, identification of outliers is important for exploring appropriate model structures and determining reliable TLS estimates of parameters. In this paper, we investigate sensitivities of TLS estimates as observation data are perturbed, and then, based on perturbation theory of matrices, we develop identification indices for detecting observations that highly influence the TLS estimates. Finally, numerical examples are given to illustrate the proposed detection method. © 2002 Elsevier Science Inc. All rights reserved.

AMS classification: 15A18; 62F35; 93E24

Keywords: Outlier; Perturbation theory; Regression diagnostics; Sensitivity analysis; Total least squares estimate

1. Introduction

The objective of this paper is to develop identification indices for detecting influential observations on total least squares (TLS) estimates. To do this, we will exploit some results from matrix perturbation theory.

* Corresponding author. Tel.: +44-191-2226233; fax: +44-191-2225748.
E-mail addresses: baibing.li@ncl.ac.uk (B. Li), bart.demoor@esat.kulueven.ac.be (B. De Moor).

[1] Tel.: +32-16-321715; fax: +32-16-321970.

PII: S0024-3795(01)00562-6

TLS plays an important role in systems modeling and signal processing [1,2]. TLS solves for the unknown vector $\mathbf{x}$ in the over-determined set of equations

$$\mathbf{Ax} \approx \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{m\times n}, \ \ \mathbf{b} \in \mathbb{R}^{m\times 1}, \ \ m > n,$$

when both the data matrix $\mathbf{A}$ and the data vector $\mathbf{b}$ are subject to errors. It is clear that the assumptions of TLS differ from the ones of ordinary least squares (OLS), where only the data vector $\mathbf{b}$ is assumed to be subject to errors.

One of the major problems for the TLS technique is its higher sensitivity to outliers, or more generally, influential observations, than OLS. When influential observations are present, the accuracy of TLS estimates is strongly affected [1].

In general, there are two typical approaches to deal with influential observations on estimates. The first one is to develop robust estimation methods. This approach is based on the belief that the model structure itself is perfect, while something is wrong for a few of the observation data, for instance, mis-recordings of them. Hence, the idea of this approach is to decrease the effects of those data on our estimates. Many robust methods have been proposed in the last two decades (see [3,4] for overviews). Recently, Li and De Moor [5] developed an adaptive robust estimation procedure that can automatically select an appropriate estimate from an estimate family ranging from L_2 estimates to L_1 estimates, depending on how serious the outliers of a data set are. For TLS, a robust approach termed total least norm L_p estimate was proposed to improve the TLS when outliers were present [2].

The second approach does not assume a priori a perfectly established model structure. On the contrary, it is believed that exclusive use of robust methods can obscure important problems, and carefully constructed models in combination with appropriate diagnostic methods provide a useful basis for thorough statistical analysis [6]. In fact, influential observations can often provide useful information. Sometimes, identified influential observations can help analysts to reconsider their models, including questioning certain assumptions of linearity or homogeneity of variances, and from this gain some insight for future modeling.

The key issue of the second approach is to develop appropriate identification indices as diagnostic tools that can be used to identify influential observations. This is crucial for higher dimensional problems since unlike the two-dimensional case, it is no longer possible to visually inspect model validations and outliers through scatter plots of the observed data. Cook and Weisberg [7] gave an overview and practical examples on how to identify and handle influential observations for the OLS regression diagnosis. Recently Hadi and Nyquist [8] considered diagnostic problems of multivariate data based on the Frechet distance. Li and De Moor [9], and Shi [10] investigated identification of influential observations in principal component analysis. In the researches of [9,10], identification indices were developed to detect the observations that have strong influences on principal components, i.e. eigenvalues and their associated eigenvectors of a covariance matrix.

In this paper, we consider the identification of influential observations on TLS estimates. From the point of view of regression analysis, it is an extension of OLS diagnostic methods [6] to the TLS situation, while from the point of view of matrix theory, it is an extension of the perturbation analysis of normalized eigenvectors [9,10] to TLS-scaled eigenvectors.

Section 2 is devoted to the problem formulation. The main results are given in Section 3. Finally, examples are examined in Section 4.

2. Problem formulation

2.1. A brief summary of TLS estimation

Consider an over-determined set of m linear equations for an unknown vector $\mathbf{x} \in \mathbb{R}^{n\times 1}$

$$\mathbf{Ax} \approx \mathbf{b}, \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m]^{\mathrm{T}} \in \mathbb{R}^{m\times n}$ $(m > n)$, $\mathbf{a}_i = [a_{i1}, \dots, a_{in}]^{\mathrm{T}} \in \mathbb{R}^{n\times 1}$, and $\mathbf{b} = [b_1, \dots, b_m]^{\mathrm{T}} \in \mathbb{R}^{m\times 1}$. Denote the ith observation as $\mathbf{z}_i = [\mathbf{a}_i^{\mathrm{T}}, b_i]^{\mathrm{T}}$ and let $\mathbf{Z} = [\mathbf{A}, \mathbf{b}] = [\mathbf{z}_1, \dots, \mathbf{z}_m]^{\mathrm{T}}$.

As mentioned before, TLS considers situations where both the vector $\mathbf{b}$ and the matrix $\mathbf{A}$ are subject to errors. Mathematically, TLS seeks for a solution satisfying (see [1]):

$$\min_{[\hat{\mathbf{A}},\hat{\mathbf{b}}]\in\mathbb{R}^{m\times(n+1)}} \|[\mathbf{A}, \mathbf{b}] - [\hat{\mathbf{A}}, \hat{\mathbf{b}}]\|_{\mathrm{F}} \tag{2}$$

$$\text{subject to} \quad \hat{\mathbf{b}} \in R(\hat{\mathbf{A}}),$$

where $\|\mathbf{B}\|_{\mathrm{F}}$ and $R(\mathbf{B})$ denote the Frobenius norm and the range of a matrix $\mathbf{B}$, respectively.

Let the singular value decomposition of $[\mathbf{A}, \mathbf{b}]$ be

$$[\mathbf{A}, \mathbf{b}] = \mathbf{\Psi\Sigma\Phi}^{\mathrm{T}} \quad \text{with } \mathbf{\Sigma} = \mathrm{diag}\{\sigma_1, \dots, \sigma_{n+1}\} \in \mathbb{R}^{m\times(n+1)}$$

and let $\boldsymbol{\varphi}_{n+1}$ be the last column of $\mathbf{\Phi}$. If

$$\sigma_n > \sigma_{n+1} \quad \text{and} \quad \varphi_{n+1,n+1} \neq 0, \tag{3}$$

where $\varphi_{n+1,n+1}$ is the last element of $\boldsymbol{\varphi}_{n+1}$, then it can be proved that [1]

$$[\hat{\mathbf{A}}, \hat{\mathbf{b}}] = \mathbf{\Psi}\hat{\mathbf{\Sigma}}\mathbf{\Phi}^{\mathrm{T}} \quad \text{and} \quad \hat{\mathbf{\Sigma}} = \mathrm{diag}\ \{\sigma_1, \dots, \sigma_n, 0\}$$

solve the TLS problem (2) and the TLS solution,

$$\hat{\mathbf{x}} = (-1/\varphi_{n+1,n+1})[\varphi_{1,n+1}, \dots, \varphi_{n,n+1}]^{\mathrm{T}}, \tag{4}$$

exists and is the unique solution to $\hat{\mathbf{A}}\mathbf{x} = \hat{\mathbf{b}}$.

It is easy to check that $\boldsymbol{\varphi}_{n+1}$ is an eigenvector of the matrix $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ associated with the smallest eigenvalue σ_{n+1}^2, i.e.

$$\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\varphi_{n+1} = \sigma_{n+1}^2\boldsymbol{\varphi}_{n+1}. \tag{5}$$

Hence, the TLS estimate $\hat{\mathbf{x}}$ is an eigenvector of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ corresponding to the smallest eigenvalue after multiplying a TLS scaling factor $-1/\varphi_{n+1,n+1}$ and crossing out the last element, -1.

Finally, it should be noted that from (5) we can easily derive a closed-form expression of the TLS estimate

$$\hat{\mathbf{x}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} - \sigma_{n+1}^2\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}, \tag{4$'$}$$

where $\mathbf{I}$ is an identity matrix of appropriate dimensions.

2.2. Sensitivity of TLS estimates to perturbations

Suppose $\boldsymbol{\omega}$ is a perturbation vector restricted to an open subset $\Omega \subset \mathbb{R}^{m\times 1}$. Denote the perturbed version of the data matrix $\mathbf{Z} = [\mathbf{A}, \mathbf{b}]$ as $\mathbf{Z}(\boldsymbol{\omega}) = [\mathbf{A}(\boldsymbol{\omega}), \mathbf{b}(\boldsymbol{\omega})]$, and the perturbed version of the TLS estimate $\hat{\mathbf{x}}$ as $\hat{\mathbf{x}}(\boldsymbol{\omega})$ which is given by $(4)'$ for the data matrix $\mathbf{Z}(\boldsymbol{\omega})$:

$$\hat{\mathbf{x}}(\boldsymbol{\omega}) = \left[\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega}) - \sigma_{n+1}^2(\boldsymbol{\omega})\mathbf{I}\right]^{-1}\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{b}(\boldsymbol{\omega}),$$

where $\sigma_{n+1}^2(\boldsymbol{\omega})$ is the smallest eigenvalue of $\mathbf{Z}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{Z}(\boldsymbol{\omega})$.

It is assumed that there exists a null perturbation point $\boldsymbol{\omega}_0 \in \Omega$ such that $\hat{\mathbf{x}}(\boldsymbol{\omega}_0) = \hat{\mathbf{x}}$. To investigate behaviors of the TLS estimate $\hat{\mathbf{x}}$ under small perturbations around the null perturbation point, consider a pencil of straight lines through the null perturbation point $\boldsymbol{\omega}_0$ (see [6,10]):

$$\boldsymbol{\omega}(\varepsilon) = \boldsymbol{\omega}_0 + \varepsilon\mathbf{h}, \tag{6}$$

where $\varepsilon \in \mathbb{R}^1$ and $\boldsymbol{\omega}(\varepsilon) = [\omega_1(\varepsilon), \ldots, \omega_m(\varepsilon)]^{\mathrm{T}}$. $\mathbf{h} = [h_1, \ldots, h_m]^{\mathrm{T}} \in \mathbb{R}^{m\times 1}$ is a vector of unit length.

Following the results in [6,10], if $\hat{\mathbf{x}}(\boldsymbol{\omega})$ is differentiable, we consider the rate of change of the TLS estimate $\hat{\mathbf{x}}$ in the direction $\mathbf{h}$ under small perturbations:

$$\begin{aligned} R(\hat{\mathbf{x}}; \mathbf{h}) &= \lim_{\varepsilon\to 0}[\hat{\mathbf{x}}(\boldsymbol{\omega}_0 + \varepsilon\mathbf{h}) - \hat{\mathbf{x}}(\boldsymbol{\omega}_0)]/\varepsilon \\ &= [\mathrm{d}\hat{\mathbf{x}}(\boldsymbol{\omega}_0 + \varepsilon\mathbf{h})/\mathrm{d}\varepsilon]_{\varepsilon=0} = [\partial\hat{\mathbf{x}}(\boldsymbol{\omega})/\partial\boldsymbol{\omega}]_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}\mathbf{h}, \end{aligned} \tag{7}$$

where $[\partial\hat{\mathbf{x}}(\boldsymbol{\omega})/\partial\boldsymbol{\omega}]_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}$ is an $n \times m$ matrix.

The rate of change characterizes how sensitive the estimate $\hat{\mathbf{x}}$ is under small perturbations in the direction $\mathbf{h}$. The sensitivity, $S(\hat{\mathbf{x}})$, of a TLS estimate $\hat{\mathbf{x}}$ is defined to be the supremum of a norm of the rate of change:

$$S(\hat{\mathbf{x}}) = \sup_{\|\mathbf{h}\|=1} \left\|R(\hat{\mathbf{x}}; \mathbf{h})\right\|_{\mathbf{M}}^2 = \sup_{\mathbf{h}\neq 0}\left\{\mathbf{h}^{\mathrm{T}}(\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P})\mathbf{h}\right\} / \left\{\mathbf{h}^{\mathrm{T}}\mathbf{h}\right\} = \|\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}\|_2^2, \tag{8}$$

where $\mathbf{M} > 0$ (positive definite) is a given weighting matrix. $\mathbf{P} = [\partial\hat{\mathbf{x}}(\boldsymbol{\omega})/\partial\boldsymbol{\omega}]_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}$ is termed the sensitivity matrix. $S(\hat{\mathbf{x}})$ gives the maximum magnitude of the rates of change of a TLS estimate $\hat{\mathbf{x}}$. The direction of the perturbation that maximizes (8) is given by $\mathbf{h}^* = \arg\max_{\mathbf{h}\neq 0}\{\mathbf{h}^{\mathrm{T}}(\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P})\mathbf{h}\}/\{\mathbf{h}^{\mathrm{T}}\mathbf{h}\}$ which is equal to the eigenvector of $\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}$ corresponding to the largest eigenvalue.

The direction $\mathbf{h}^*$ indicates how to perturb the postulated model to attain the greatest rate of change. It can be used for diagnostic purposes [6]. Each of its elements reflects the contribution of the corresponding observation to the sensitivity $S(\hat{\mathbf{x}})$ of the TLS estimate $\hat{\mathbf{x}}$. Suppose that the ith element of $\mathbf{h}^*$ is found to be relatively large. This indicates that perturbations for the ith observation may lead to substantial changes in the results of the analysis and thus the ith observation is relatively influential. The elements of the vector $\mathbf{h}^*$ are called identification indices.

3. Identification indices of TLS estimates

In this section, we develop identification indices of TLS estimates. We consider only the situation where the TLS estimate exists and is unique. This is insured by the conditions (3).

As mentioned before, Li and De Moor [9], and Shi [10] recently developed identification indices to detect those observations that have a strong influence on normalized principal components, i.e. eigenvectors of unit length, and their associated eigenvalues of a covariance matrix. Those results are closely related to a TLS estimate that can be either expressed as a TLS-scaled eigenvector of the matrix $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ given by (4), or characterized as a vector-valued function of the smallest eigenvalue, $\hat{\mathbf{x}} = \mathbf{f}(\hat{\sigma}^2_{n+1})$ given by (4)$'$. In this paper, the second expression, $\hat{\mathbf{x}} = \mathbf{f}(\hat{\sigma}^2_{n+1})$ given by (4)$'$, is adopted as the starting point for influential analysis.

Let $\boldsymbol{\omega}(0) = [\omega_1(0), \ldots, \omega_m(0)]^{\mathrm{T}} = \boldsymbol{\omega}_0 = \mathbf{1}$ be the null perturbation point, where $\mathbf{1}$ is an $m \times 1$ vector of ones. For small perturbations $\boldsymbol{\omega}(\varepsilon) = \boldsymbol{\omega}_0 + \varepsilon\mathbf{h}$ given by (6), consider the following perturbation scheme where perturbed data are constituted by "true" data plus perturbation increments:

$$\mathbf{z}_i(\boldsymbol{\omega}) = \omega_i(\varepsilon)\mathbf{z}_i = \mathbf{z}_i + \varepsilon h_i \mathbf{z}_i \quad \text{for } i = 1, \ldots, m \tag{9}$$

and at the null perturbation point, $\mathbf{z}_i(\boldsymbol{\omega}_0) = \mathbf{z}_i$. The perturbed version of the observation matrix $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ is then given by

$$\begin{aligned}\mathbf{Z}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{Z}(\boldsymbol{\omega}) &= \begin{bmatrix}\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega}) & \mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{b}(\boldsymbol{\omega})\\ \mathbf{b}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega}) & \mathbf{b}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{b}(\boldsymbol{\omega})\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{A}^{\mathrm{T}}\mathbf{A} & \mathbf{A}^{\mathrm{T}}\mathbf{b}\\ \mathbf{b}^{\mathrm{T}}\mathbf{A} & \mathbf{b}^{\mathrm{T}}\mathbf{b}\end{bmatrix} + 2\varepsilon\begin{bmatrix}\sum_{i=1}^m h_i\mathbf{a}_i\mathbf{a}_i^{\mathrm{T}} & \sum_{i=1}^m h_i\mathbf{a}_i b_i\\ \sum_{i=1}^m h_i\mathbf{a}_i^{\mathrm{T}} b_i & \sum_{i=1}^m h_i b_i^2\end{bmatrix}\\ &\quad + \mathrm{O}(\varepsilon^2).\end{aligned} \tag{10}$$

Then, we obtain from (10)

$$\mathrm{d}[\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega})]/\mathrm{d}\varepsilon|_{\varepsilon=0} = 2\sum_{i=1}^m h_i\mathbf{a}_i\mathbf{a}_i^{\mathrm{T}},$$

$$\mathrm{d}[\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{b}(\boldsymbol{\omega})]/\mathrm{d}\varepsilon|_{\varepsilon=0} = 2\sum_{i=1}^{m} h_i \mathbf{a}_i b_i. \tag{11}$$

Define $\mathbf{E} = \operatorname{diag}\{e_1, \ldots, e_m\}$ and $\mathbf{e} = [e_1, \ldots, e_m]^{\mathrm{T}}$, where $e_i = b_i - \mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}}$ is the residual $(j = 1, \ldots, m)$.

Lemma 1. *Let* $\sigma_{n+1}^2(\boldsymbol{\omega})$ *be the smallest eigenvalue of* $\mathbf{Z}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{Z}(\boldsymbol{\omega})$, *and* $[\hat{\mathbf{x}}^{\mathrm{T}}, -1]^{\mathrm{T}}$ *the eigenvector of* $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ *associated with the smallest eigenvalue. Then for perturbation scheme* (9)

$$\mathrm{d}\sigma_{n+1}^2(\boldsymbol{\omega})/\mathrm{d}\varepsilon|_{\varepsilon=0} = 2\left(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}}\right)^{-1} \mathbf{e}^{\mathrm{T}}\mathbf{E}\mathbf{h}.$$

The proof is given in Appendix A.

Lemma 2. *Suppose* $\mathbf{B}(\varepsilon)$, $\mathbf{C}(\varepsilon)$, *and* $\mathbf{G}(\varepsilon)$ *are matrices of appropriate dimensions of which the entries are differentiable functions of the scalar* ε, *and* $\mathbf{G}(\varepsilon)$ *is nonsingular. Then*

$$\mathrm{d}[\mathbf{B}(\varepsilon)\mathbf{C}(\varepsilon)]/\mathrm{d}\varepsilon = [\mathrm{d}\mathbf{B}(\varepsilon)/\mathrm{d}\varepsilon]\mathbf{C}(\varepsilon) + \mathbf{B}(\varepsilon)[\mathrm{d}\mathbf{C}(\varepsilon)/\mathrm{d}\varepsilon]$$

and

$$\mathrm{d}[\mathbf{G}^{-1}(\varepsilon)]/\mathrm{d}\varepsilon = -\mathbf{G}^{-1}(\varepsilon)[\mathrm{d}\mathbf{G}(\varepsilon)/\mathrm{d}\varepsilon]\mathbf{G}^{-1}(\varepsilon).$$

The proof of Lemma 2 is trivial.

Let $\mathbf{T}(\boldsymbol{\omega}) = \mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega}) - \sigma_{n+1}^2(\boldsymbol{\omega})\mathbf{I}$ and $\mathbf{T} = \mathbf{A}^{\mathrm{T}}\mathbf{A} - \sigma_{n+1}^2\mathbf{I}$. Then $\mathbf{T}(\boldsymbol{\omega}_0) = \mathbf{T}$. From Lemmas 1 and 2 we have:

Theorem 1. *For a given weighting matrix* $\mathbf{M} > 0$, *the sensitivity matrix* $\mathbf{P}$ *of the TLS estimate* $\hat{\mathbf{x}}$ *for problem* (1) *is given by*

$$\mathbf{P} = 2\mathbf{T}^{-1}\,[(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}\mathbf{e}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}]\mathbf{E}. \tag{12}$$

Proof. From Lemmas 1 and 2, and noting Eq. (11)

$$\begin{aligned}
&\mathrm{d}\mathbf{T}^{-1}(\boldsymbol{\omega})/\mathrm{d}\varepsilon|_{\varepsilon=0}\\
&\quad= \Big[-\mathbf{T}^{-1}(\boldsymbol{\omega})\left\{\mathrm{d}\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{A}(\boldsymbol{\omega})/\mathrm{d}\varepsilon - (\mathrm{d}\sigma_{n+1}^2(\boldsymbol{\omega})/\mathrm{d}\varepsilon)\mathbf{I}\right\}\mathbf{T}^{-1}(\boldsymbol{\omega})\Big]_{\varepsilon=0}\\
&\quad= 2\mathbf{T}^{-1}\left[(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\mathbf{e}^{\mathrm{T}}\mathbf{E}\mathbf{h} - \sum_{i=1}^{m} h_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}}\right]\mathbf{T}^{-1}.
\end{aligned}$$

Then, we obtain from (4)′, (11), and Lemmas 1 and 2

$$R(\hat{\mathbf{x}}; \mathbf{h}) = \left\{[\mathrm{d}\mathbf{T}^{-1}(\boldsymbol{\omega})/\mathrm{d}\varepsilon]\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{b}(\boldsymbol{\omega}) + \mathbf{T}^{-1}(\boldsymbol{\omega})\mathrm{d}[\mathbf{A}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{b}(\boldsymbol{\omega})]/\mathrm{d}\varepsilon\right\}_{\varepsilon=0}$$

$$
\begin{aligned}
&= 2\mathbf{T}^{-1}\left[(1+\hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\mathbf{e}^{\mathrm{T}}\mathbf{E}\mathbf{h} - \sum_{i=1}^{m} h_i \mathbf{a}_i \mathbf{a}_i^{\mathrm{T}}\right]\mathbf{T}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \\
&\quad + 2\,\mathbf{T}^{-1}\sum_{i=1}^{m} h_i \mathbf{a}_i b_i \\
&= 2\mathbf{T}^{-1}[(1+\hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}\mathbf{e}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}]\mathbf{E}\mathbf{h},
\end{aligned}
$$

which yields $\mathbf{P} = 2\mathbf{T}^{-1}[(1+\hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}\mathbf{e}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}]\mathbf{E}$. □

For the sensitivity matrix $\mathbf{P}$ obtained by Theorem 1, the vector of identification indices of influential observations is a vector of unit length given by

$$
\mathbf{h}^* = \arg\max_{\mathbf{h}\neq 0} \left\{\mathbf{h}^{\mathrm{T}}(\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P})\mathbf{h}\right\}/\left\{\mathbf{h}^{\mathrm{T}}\mathbf{h}\right\}, \tag{13}
$$

which is a normalized eigenvector of the matrix $\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}$ associated with the largest eigenvalue. Particularly, when the unknown x is univariate, i.e. $n = 1$ and $\mathbf{A} = [a_1, \ldots, a_m]^{\mathrm{T}} \in \mathbb{R}^{m\times 1}$ in (1), it is easy to check that

$$
\mathbf{h}^* = c[e_1 a_1 + e_1^2\hat{x}/(1+\hat{x}^2), \ldots, e_m a_m + e_m^2\hat{x}/(1+\hat{x}^2)]^{\mathrm{T}}, \tag{14}
$$

where c is a normalization scalar to ensure that $\mathbf{h}^*$ has unit length.

It can be seen from (14) that the impact of an influential observation on the TLS estimate depends on both the residual e_i and its value $\mathbf{a_i}$, i.e. whether it is an outlier and/or whether it is a high leverage point.

It is of interest to compare the above results of TLS with its counterparts for OLS. From Eq. (29) of [6], for OLS estimates with known error variance, the vector of identification indices for a univariate explanatory variable is given by

$$
\mathbf{h}^*_{\mathrm{OLS}} = c[e_1\, a_1, \ldots,\ e_m a_m]^{\mathrm{T}}, \tag{14$'$}
$$

where c is a normalization scalar to ensure that $\mathbf{h}^*_{\mathrm{OLS}}$ has unit length.

It can be can seen that the ith element of the vector of identification indices of TLS (14) is a quadratic function of the residual e_i, while its counterpart of the OLS is a linear function of e_i. This means that an observation with a large residual such that its quadratic term dominates the identification index has a much stronger influence on a TLS estimate than on an OLS estimate for the same magnitude of residual. This quantitatively verifies the empirical conclusion in [1, p. 268] that if outliers are large, TLS estimates are much more sensitive to outliers than OLS estimates.

On the other hand, however, for any observation, if its quadratic term does not dominate the identification index of the TLS solution, the linear and quadratic terms may cancel out each other, resulting in an even smaller index than its counterpart of the OLS. A numerical example was given in [9] for the eigenvalues of a covariance matrix, where the effects of the linear and quadratic terms cancelled out each other such that a significantly outlying observation point had a small influence on the eigenvalue. This indicates that for these circumstances, the TLS solution may even be superior to the OLS solution in terms of sensitivity to outliers.

The above results can be extended to the situation where there is an intercept term in the equation systems (1), i.e.

$$[\mathbf{1}, \mathbf{A}][\alpha, \mathbf{x}^{\mathrm{T}}]^{\mathrm{T}} \approx \mathbf{b}$$

According to [1], the TLS estimate of $[\alpha, \mathbf{x}^{\mathrm{T}}]^{\mathrm{T}}$ is given by $[\hat{\alpha}, \hat{\mathbf{x}}^{\mathrm{T}}]^{\mathrm{T}} = [\bar{\mathbf{b}} - \bar{\mathbf{A}}\hat{\mathbf{x}}, \hat{\mathbf{x}}^{\mathrm{T}}]^{\mathrm{T}}$, where $\bar{\mathbf{b}}$ and $\bar{\mathbf{A}}$ are arithmetic means of the columns of $\mathbf{b}$ and $\mathbf{A}$, respectively, and $\hat{\mathbf{x}}$ is the solution of the following equations without intercept term:

$$\mathbf{A}_{\mathrm{c}}\mathbf{x} \approx \mathbf{b}_{\mathrm{c}}, \tag{15}$$

where $\mathbf{A}_{\mathrm{c}} = \mathbf{JA}$, $\mathbf{b}_{\mathrm{c}} = \mathbf{Jb}$ and $\mathbf{J} = \mathbf{I} - \mathbf{11}^{\mathrm{T}}/m$.

Define $\mathbf{E} = \mathrm{diag}\{e_1, \dots, e_m\}$ and $\mathbf{e} = [e_1, \dots, e_m]^{\mathrm{T}}$, where $e_i = b_i - \hat{\alpha} - \mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}}$ is the residual $(j = 1, \dots, m)$. Let $\mathbf{T}_{\mathrm{c}} = \mathbf{A}_{\mathrm{c}}^{\mathrm{T}}\mathbf{A}_{\mathrm{c}} - \tau_{n+1}^2\mathbf{I}$, where τ_{n+1}^2 is the smallest eigenvalue of $\mathbf{Z}_{\mathrm{c}}^{\mathrm{T}}\mathbf{Z}_{\mathrm{c}}$ and $\mathbf{Z}_{\mathrm{c}} = [\mathbf{A}_{\mathrm{c}}, \mathbf{b}_{\mathrm{c}}]$. Then similar to Theorem 1, we have:

Theorem 1′. *Given a weighting matrix* $\mathbf{M} > 0$, *the sensitivity matrix* $\mathbf{P}_{\mathrm{c}}$ *of the TLS estimate* $\hat{\mathbf{x}}$ *for the TLS problem* (15) *is given by*

$$\mathbf{P}_{\mathrm{c}} = \mathbf{T}_{\mathrm{c}}^{-1}\left[2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}(\mathbf{e} + \hat{\alpha}\mathbf{1})^{\mathrm{T}}\mathbf{E} + \mathbf{A}^{\mathrm{T}}\mathbf{E} + \mathbf{A}_{\mathrm{c}}^{\mathrm{T}}(\mathbf{E} + \hat{\alpha}\mathbf{I})\right].$$

The proof is given in Appendix B.

The vector of identification indices is then given by

$$\mathbf{h}_{\mathrm{c}}^* = \arg\max\nolimits_{\mathbf{h}\neq 0}\left\{\mathbf{h}^{\mathrm{T}}(\mathbf{P}_{\mathrm{c}}^{\mathrm{T}}\mathbf{M}\mathbf{P}_{\mathrm{c}})\mathbf{h}\right\} / \left\{\mathbf{h}^{\mathrm{T}}\mathbf{h}\right\},$$

i.e. an eigenvector of the matrix $\mathbf{P}_{\mathrm{c}}^{\mathrm{T}}\mathbf{M}\mathbf{P}_{\mathrm{c}}$ associated with the largest eigenvalue. In most of cases, the parameters of interest are the “slope” parameter vector $\mathbf{x}$. In those situations, we can use $\mathbf{h}_{\mathrm{c}}^*$ for model explorations.

If in some cases the intercept α is also of interest, we can similarly obtain

$$R(\hat{\alpha}; \mathbf{h}) = \mathrm{d}[\bar{\mathbf{b}}(\varepsilon) - \bar{\mathbf{A}}(\varepsilon)\hat{\mathbf{x}}(\varepsilon)]/\mathrm{d}\varepsilon|_{\varepsilon=0} = (\mathbf{e}^{\mathrm{T}}/m - \bar{\mathbf{A}}\mathbf{P}_{\mathrm{c}})\mathbf{h}.$$

Then the vector of identification indices for $\hat{\alpha}$ is given by

$$\mathbf{h}_{\alpha}^* = \arg\max_{h\neq 0}|R(\hat{\alpha}; \mathbf{h})| = c(\mathbf{e}^{\mathrm{T}}/m - \bar{\mathbf{A}}\mathbf{P}_{\mathrm{c}})^{\mathrm{T}},$$

where c is a scalar to ensure that $\mathbf{h}_{\alpha}^*$ has unit length.

Finally, we consider the invariance of the vector of identification indices $\mathbf{h}^*$ under column orthogonal transformations.

Theorem 2. *The vector of identification indices* $\mathbf{h}^*$ *given by* (13) *is invariant under column orthogonal transformations of the data matrix* $\mathbf{A}$ *if* $\mathbf{M} = \mathbf{I}$ *or* $\mathbf{M} = \mathbf{A}^{\mathrm{T}}\,\mathbf{A}$.

The proof is given in Appendix C.

From Theorem 2, the vector of identification indices $\mathbf{h}^*$ is invariant when orthogonal transformations of the coordinate system are made for the explanatory variables in (1). In addition, Theorem 2 gives two often used weighting matrices. The choice of the weighting matrix **M** reflects specific interests in applications. According to [11], it is likely that both of them would give approximately the same information.

4. Numerical examples

In this section, two numerical examples are examined to illustrate the developed identification indices.

Example 1. Consider a set of artificial data given in Table 1 and the following system model:

$$y = r_1 t_1 + r_2 t_2,$$

where r_1 and r_2 are parameters to be estimated.

Fig. 1 gives scatter plots of explanatory variables t_i $(i = 1, 2)$ versus response variable y. It seems that there exists a linear relationship between the explanatory variables and the response variable, and there do not exist any anomalous data in Fig. 1. We then obtain the TLS estimate $[\hat{r}_1, \hat{r}_2]^{\mathrm{T}}$ of the parameter vector $[r_1, r_2]^{\mathrm{T}}$ by using the data in Table 1.

Fig. 2 gives the diagnostic plot of identification indices $\mathbf{h}^*$ for the weighting matrix $\mathbf{M} = \mathbf{A}^{\mathrm{T}}\ \mathbf{A}$. Since the absolute value of the 26th element of $\mathbf{h}^*$ is significantly larger than others, it is suggested that observation 26 has a very strong influence on the TLS estimate $[\hat{r}_1, \hat{r}_2]^{\mathrm{T}}$.

Table 1
A set of artificial data

Obs	t_1	t_2	y	Obs	t_1	t_2	y
1	5.0201	−1.1716	4.1729	16	2.9824	−0.7835	−2.5195
2	−0.8679	−1.3365	0.4369	17	−1.9843	1.9214	5.0180
3	−4.1001	0.5409	−3.9756	18	−2.3285	0.7073	0.2280
4	3.7196	−1.0706	−0.0041	19	0.4660	0.4492	4.7791
5	2.3391	−3.1700	−1.4151	20	−0.9890	0.8738	5.2420
6	1.2468	−1.3758	−3.9933	21	1.1382	3.0059	1.7552
7	−0.5670	−4.0769	−3.0176	22	2.3931	1.8138	4.1303
8	2.6972	0.5523	7.4431	23	−2.5760	−2.6618	−6.5651
9	−4.4617	2.0661	0.8322	24	0.9415	1.7316	6.8336
10	−4.3151	−0.2339	−3.4556	25	−1.3638	−0.0617	−4.7384
11	−0.7117	2.7923	−5.4102	26	−3.1540	−2.8607	4.1049
12	0.6919	−0.3793	2.4953	27	−1.3087	−0.1137	1.2912
13	0.9216	−2.8373	−4.2514	28	−3.4907	−1.8876	−0.2924
14	−1.5165	−1.2533	−2.1991	29	0.9726	−1.9629	2.3269
15	4.7572	−0.7279	1.3625	30	−1.1689	3.2024	0.6291

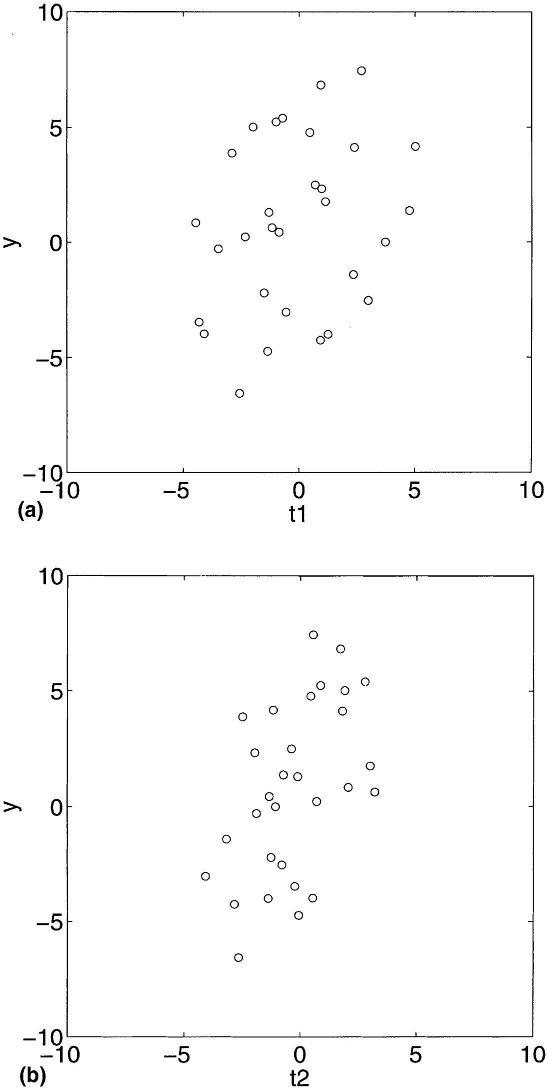

Fig. 1. (a) Scatter plot: t_1 versus y; (b) scatter plot: t_2 versus y.

To investigate what happened for the data in Table 1, we adopt a new coordinate system which is a counterclockwise rotation of $\pi/6$ from the old coordination system:

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}.$$

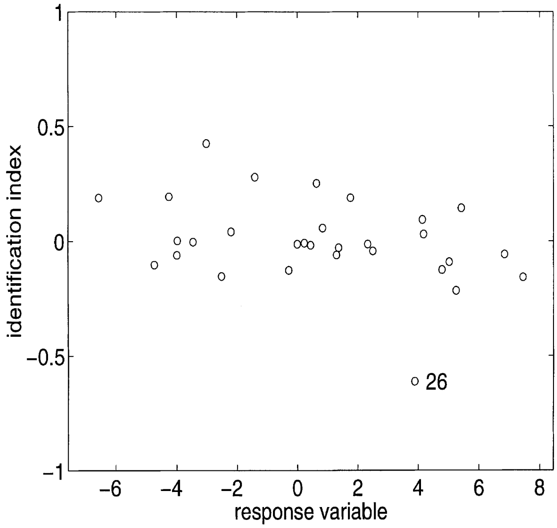

Fig. 2. Diagnostic plot of $\mathbf{h}^*$.

Fig. 3 gives the scatter plot of explanatory variable s_2 versus response variable y. It can be seen that observation 26 is an anomalous point.

This example shows that analysis on TLS estimates for higher dimensional problems can be very complicated. Simple scatter plots for response variable versus each of explanatory variables may not reveal anomalies in observation data. For those situations, a diagnostic plot of the identification indices $\mathbf{h}^*$ is very helpful

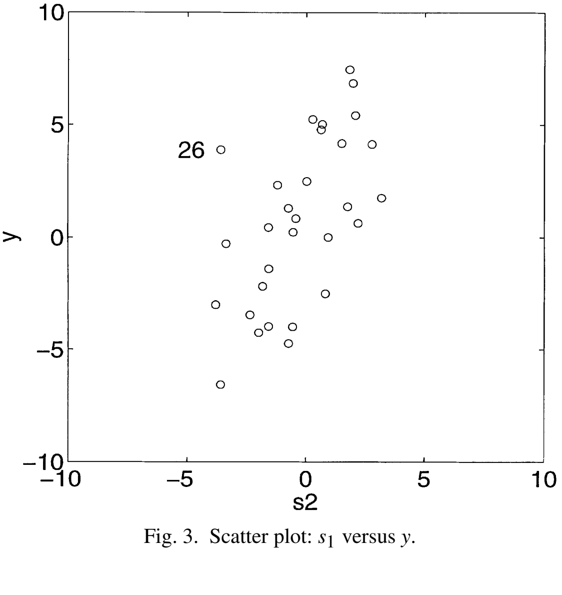

Fig. 3. Scatter plot: s_1 versus y.

for detecting those observations that have larger contributions to the rates of change of TLS estimates.

In practice, further investigations based on background knowledge are often needed to look for explanation of such outliers and then make a decision how to treat them. Sometimes a diagnostic plot of the identification indices $\mathbf{h}^*$ exhibits some systematic characteristics. Such information may provide useful suggestions to further model explorations.

Example 2. Consider the following system model

$$y = f(t_1, t_2; r_1, r_2, r_3) = r_1 t_1 + r_2 t_2 + r_3 t_1 \, t_2 \tag{16}$$

with "true" parameters $r_1 = 2$, $r_2 = 1$ and $r_3 = 0.05$. Fig. 4 gives a plot of the function (16).

Suppose the "true" data set is constructed as

$$\{(s_1, s_2, s_3) \,|\, s_1 = 0, 2, 4, 6, 8; \;\; s_2 = 0, 2, 4, 6, 8; \;\; s_3 = f(s_1, s_2\,; 2, 1, 0.05)\}$$

while the observed data set, given in Table 2, is generated as the "true" data plus noises which are mutually independent random variables with normal distribution of zero mean and unit variance.

In general, modeling higher dimensional data is not easy without a priori information on model structures. One often used approach is to inspect scatter plots for each of the explanatory variables versus the response variable, and adopt a linear model as a starting point if possible. Fig. 6 gives the scatter plots of t_i $(i = 1, 2)$ versus y

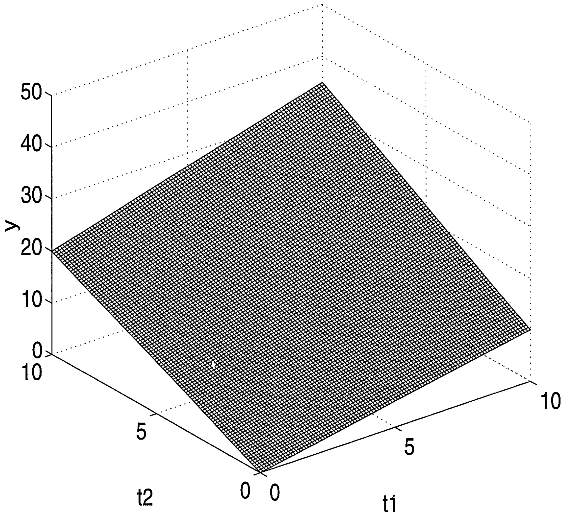

Fig. 4. Plot of function $y = 2t_1 + t_2 + 0.05t_1t_2$.

Table 2
Observed data for system (16)

Obs	t_1	t_2	y	Obs	t_1	t_2	y
1	−0.3609	0.5536	−1.5564	14	3.1074	6.2787	16.4542
2	−0.2067	1.5744	4.4938	15	5.6035	8.5743	21.9207
3	−0.8709	4.0798	7.4784	16	5.8486	0.3158	7.3437
4	−1.4139	5.6157	11.5421	17	3.7622	3.2929	10.2215
5	−0.2915	7.6988	14.4114	18	6.0025	4.8846	15.7825
6	3.0943	1.3242	1.8735	19	4.3858	4.4963	20.3736
7	1.2628	2.2137	5.7995	20	5.0895	6.3687	24.0409
8	2.0649	2.2420	12.0867	21	7.6024	−1.1613	6.8902
9	2.3274	6.7160	16.1986	22	8.2907	0.0898	14.1148
10	−0.0647	7.2564	18.9762	23	8.6653	3.7249	17.5770
11	4.5278	−0.5532	4.2983	24	7.0920	4.9563	22.7735
12	2.7734	1.8103	8.0983	25	8.9015	9.2785	27.0715
13	4.9570	3.4666	11.8989				

for the data in Table 2. It seems that there exists a strong linear relationship between them. We then start modeling from the following linear structure:

$$y = r_1 t_1 + r_2 t_2. \tag{17}$$

From Fig. 4 we can see that the "true" system model (16) in the area $(t_1, t_2) \in [0, 10] \times [0, 10]$ is quite close to a plane except for the area where both t_1 and t_2 are relatively large. The mis-specified system model, (17), leads to over-estimations of

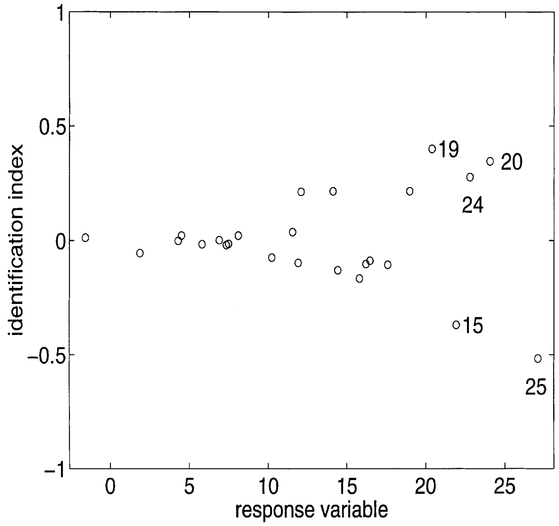

Fig. 5. Diagnostic plot of $\mathbf{h}^*$.

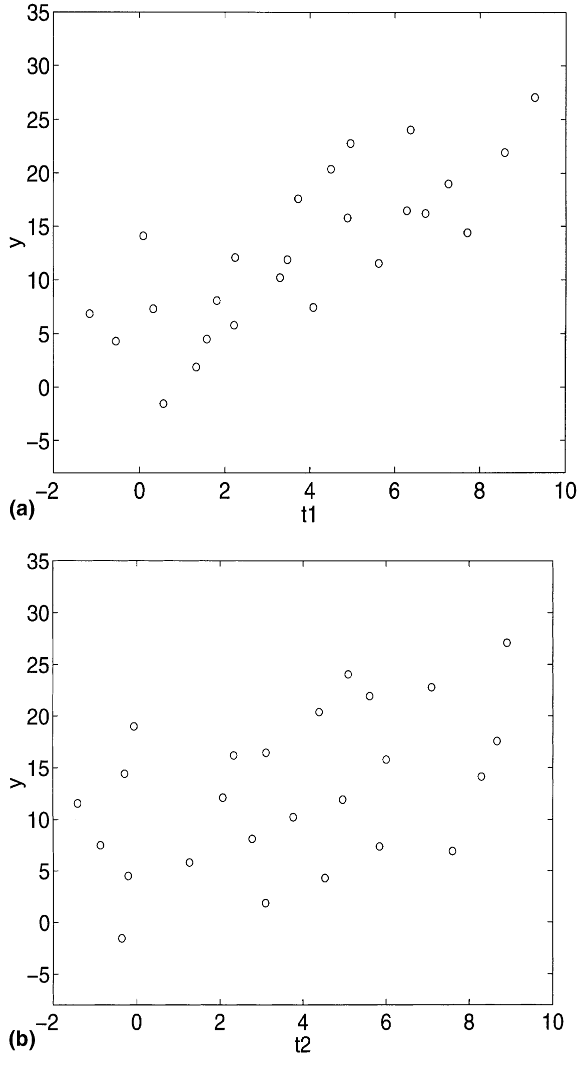

Fig. 6. (a) The scatter plot of t_1 versus y. (b) The scatter plot of t_2 versus y.

$\hat{r}_1$ and $\hat{r}_2$ since the neglected term $r_3t_1t_2$ has potentials to pull the plane (17) up when $r_3t_1t_2 > 0$.

The TLS estimates of parameters for (17) are given as $[\hat{r}_1, \hat{r}_2] = [2.2183, 1.2387]$ which confirm that they over-estimate the "true" values [2, 1]. Fig. 5 gives a diagnostic plot for the mis-specified model (17) for the weighting matrix $\mathbf{M} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$. It can be seen from Fig. 5 that the absolute values of the identification indices $\mathbf{h}^*$ increase gradually as the observation values of the response variable increase. In particular, observations 15, 19, 20, 24, and 25 have relatively large absolute values in the identification indices. This means that the larger the values of the response variable, the worse the fitness of the model (17) is. Due to this systematic trend, it is suggested that there may be something wrong in the specification of the model structure (17). The positions of the observations 15, 19, 20, 24, and 25 in the sample space in combination with the systematic trend in Fig. 5, may provide clues for exploring the next version of the model.

In general, however, there may exist several possible reasons for a single symptom in diagnostic problems. For instance, for the diagnostic plot given by Fig. 5, the problem may be caused by heterogeneous variances of response variables, or as in this example, by missing some important "explanatory" terms. Further careful analysis is then necessary based on background knowledge and data analysis. If, as in this example, there is no a priori knowledge for the model structure and no evidence of heterogeneous variances in data measurements, it is natural to try some second-order models in the next modeling step.

Appendix A. Proof of Lemma 1

Lemma A.1 [12]. *Let* $\mathbf{B}$ *and* $\mathbf{C}$ *be symmetric matrices*, λ *a simple eigenvalue of* $\mathbf{B}$ *and* ξ *an associated eigenvector of unit length. Let* $\mathbf{B}$ *be perturbed to* $\mathbf{B}(\varepsilon) = \mathbf{B} + \varepsilon\mathbf{C} + \mathrm{O}(\varepsilon^2)$, *and assume that the corresponding perturbation of* λ *is* $\lambda(\varepsilon) = \lambda + \varepsilon\mu + \mathrm{O}(\varepsilon^2)$. *Then* $\mu = \boldsymbol{\xi}^{\mathrm{T}}\mathbf{C}\boldsymbol{\xi}$.

Proof of Lemma 1. Noting that the eigenvector of unit length associated with the smallest eigenvalue σ_{n+1}^2 of the matrix $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ is $(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1/2}[\hat{\mathbf{x}}^{\mathrm{T}}, -1]^{\mathrm{T}}$, and from Lemma A.1, the perturbed version of the smallest eigenvalue of $\mathbf{Z}^{\mathrm{T}}(\boldsymbol{\omega})\mathbf{Z}(\boldsymbol{\omega})$ is given by

$$\sigma_{n+1}^2(\boldsymbol{\omega}) = \sigma_{n+1}^2 + \varepsilon\mu + \mathrm{O}(\varepsilon^2),$$

where from (10),

$$\begin{aligned}\mu &= 2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\begin{bmatrix}\hat{\mathbf{x}}^{\mathrm{T}} & -1\end{bmatrix}\begin{bmatrix}\sum_{i=1}^m h_i\mathbf{a}_i\mathbf{a}_i^{\mathrm{T}} & \sum_{i=1}^m h_i\mathbf{a}_i b_i \\ \sum_{i=1}^m h_i\mathbf{a}_i^{\mathrm{T}} b_i & \sum_{i=1}^m h_i b_i^2\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}} \\ -1\end{bmatrix} \\ &= 2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\mathbf{e}^{\mathrm{T}}\mathbf{E}\mathbf{h}.\end{aligned}$$

Hence, we obtain $\mathrm{d}\sigma_{n+1}^2(\boldsymbol{\omega})/\mathrm{d}\varepsilon|_{\varepsilon=0} = \mu = 2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\mathbf{e}^{\mathrm{T}}\mathbf{E}\mathbf{h}$. $\square$

Appendix B. Proof of Theorem 1′

Let $\mathbf{Z}_{\mathrm{c}}(\boldsymbol{\omega})$ be a perturbed version of $\mathbf{Z}_{\mathrm{c}} = [\mathbf{A}_{\mathrm{c}}, \mathbf{b}_{\mathrm{c}}]$. For the perturbation scheme (9),

$$\mathbf{Z}_{\mathrm{c}}^{\mathrm{T}}(\omega)\mathbf{Z}_{\mathrm{c}}(\omega) = \mathbf{Z}^{\mathrm{T}}\mathbf{J}\mathbf{Z} + \varepsilon\mathbf{U}_{\mathrm{c}} + \mathrm{O}(\varepsilon^2),$$

where $\mathbf{U}_{\mathrm{c}} = \mathrm{diag}\,\{h_1, \ldots, h_m\}\,\mathbf{Z}^{\mathrm{T}}\mathbf{J}\mathbf{Z} + (\mathbf{J}\mathbf{Z})^{\mathrm{T}}\mathrm{diag}\,\{h_1, \ldots, h_m\}\mathbf{Z}$.

Denote the smallest eigenvalue of $\mathbf{Z}_{\mathrm{c}}^{\mathrm{T}}(\omega)\mathbf{Z}_{\mathrm{c}}(\omega)$ as $\tau_{n+1}^2(\omega)$. From Lemma A.1

$$\tau_{n+1}^2(\omega) = \tau_{n+1}^2 + \varepsilon\theta + \mathrm{O}(\varepsilon^2),$$

where $\theta = (1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}[\hat{\mathbf{x}}^{\mathrm{T}}, -1]\mathbf{U}_{\mathrm{c}}[\hat{\mathbf{x}}^{\mathrm{T}}, -1]^{\mathrm{T}} = 2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}(\mathbf{e} + \hat{\alpha}\mathbf{1})^{\mathrm{T}}\mathbf{E}\mathbf{h}$. Hence we obtain

$$\mathrm{d}\tau_{n+1}^2(\omega)/\mathrm{d}\varepsilon|_{\varepsilon=0} = 2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}(\mathbf{e} + \hat{\alpha}\mathbf{1})^{\mathrm{T}}\mathbf{E}\mathbf{h}.$$

Finally, similar to the proof of Theorem 1, we obtain

$$\begin{aligned} R(\hat{\mathbf{x}}; \mathbf{h}) &= \partial\hat{\mathbf{x}}_{n+1}(\omega)/\partial\omega|_{\omega=\omega_0}\mathbf{h} \\ &= \mathbf{T}_{\mathrm{c}}^{-1}[2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}(\mathbf{e} + \hat{\alpha}\mathbf{1})^{\mathrm{T}}\mathbf{E} + \mathbf{A}^{\mathrm{T}}\mathbf{E} + \mathbf{A}_{\mathrm{c}}^{\mathrm{T}}(\mathbf{E} + \hat{\alpha}\mathbf{I})]\mathbf{h}. \end{aligned}$$

This yields

$$\mathbf{P}_{\mathrm{c}} = \mathbf{T}_{\mathrm{c}}^{-1}[2(1 + \hat{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{x}})^{-1}\hat{\mathbf{x}}(\mathbf{e} + \hat{\alpha}\mathbf{1})^{\mathrm{T}}\mathbf{E} + \mathbf{A}^{\mathrm{T}}\mathbf{E} + \mathbf{A}_{\mathrm{c}}^{\mathrm{T}}(\mathbf{E} + \hat{\alpha}\mathbf{I})]. \qquad \Box$$

Appendix C. Proof of Theorem 2

Let $\mathbf{A}_{\mathbf{Q}}$ denote the data matrix $\mathbf{A}$ after applying the transformation

$$\mathbf{A}_{\mathbf{Q}} = \mathbf{A}\mathbf{Q} \quad \text{for orthogonal matrix } \mathbf{Q}. \tag{C.1}$$

Then for $\mathbf{Z}_{\mathbf{Q}} = [\mathbf{A}_{\mathbf{Q}}, \mathbf{b}]$, we have $\mathbf{Z}_{\mathbf{Q}}^{\mathrm{T}}\mathbf{Z}_{\mathbf{Q}} = \mathrm{diag}\{\mathbf{Q}^{\mathrm{T}}, 1\}\,\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\,\mathrm{diag}\{\mathbf{Q}, 1\}$.

It is clear that for the orthogonal transformation of the orthogonal matrix diag $\{\mathbf{Q},1\}$, the eigenvalues of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ are invariant, and the eigenvector $[\hat{\mathbf{x}}_{\mathbf{Q}}^{\mathrm{T}}, -1]^{\mathrm{T}}$ of $\mathbf{Z}_{\mathbf{Q}}^{\mathrm{T}}\mathbf{Z}_{\mathbf{Q}}$ corresponding to an eigenvector $[\hat{\mathbf{x}}^{\mathrm{T}}, -1]^{\mathrm{T}}$ of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$ satisfies

$$\begin{bmatrix} \hat{\mathbf{x}}_{\mathbf{Q}} \\ -1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^{\mathrm{T}} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \hat{\mathbf{x}} \\ -1 \end{bmatrix} \quad \text{or} \quad \hat{\mathbf{x}}_{\mathbf{Q}} = \mathbf{Q}^{\mathrm{T}}\hat{\mathbf{x}}. \tag{C.2}$$

Then we conclude that $\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}$ is invariant by inserting (C.1) and (C.2) into (12) and noting that both $\mathbf{E}$ and $\mathbf{e}$ are invariant. Hence, $\mathbf{h}^*$ as an eigenvector of $\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}$ is invariant. $\Box$

References

[1] S. Van Huffel, J. Vandewalle, The Total Least Squares Problem: Computational Aspects and Analysis, SIAM, Philadelphia, PA, 1991.

[2] S. Van Huffel, Recent Advances in Total Least Squares Technique and Errors-in-Variables Modeling, SIAM, Philadelphia, PA, 1997.

[3] P.J. Huber, Robust Statistics, Wiley, New York, 1981.

[4] D.C. Hoaglin, F. Mosteller, J.W. Tukey, Understanding Robust and Exploratory Data Analysis, Wiley, New York, 1983.

[5] B. Li, B. De Moor, A family of adaptive robust estimates based on symmetrical generalized logistic distribution, Comm. Statist. Theory Methods 28 (1999) 1293–1310.

[6] R.D. Cook, Assessment of local influence, J. Roy. Statist. Soc. B 48 (1986) 133–169.

[7] R.D. Cook, S. Weisberg, Residuals and Influence in Regression, Chapman & Hall, New York, 1982.

[8] A.S. Hadi, H. Nyquist, Frechet distance as a tool for diagnosing multivariate data, Linear Algebra Appl. 289 (1999) 183–201.

[9] B. Li, B. De Moor, More on local influence on principal component analysis, Comm. Statist. Theory Methods 28 (1999) 2487–2495.

[10] L. Shi, Local influence in principal components analysis, Biometrika 84 (1997) 175–186.

[11] R.D. Cock, Influential observations in linear regression, J. Am. Statist. Assoc. 74 (1979) 169–174.

[12] R. Sibson, Studies in robustness of multidimensional scaling: perturbational analysis of classical scaling, J. Roy. Statist. Soc. B 41 (1979) 217–229.