Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Diagnostics in elliptical regression models with stochastic restrictions applied to econometrics

Víctor Leiva[ab], Shuangzhe Liu[c], Lei Shi[d] & Francisco José A. Cysneiros[e]

[a] Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Viña del Mar, Chile

[b] Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile

[c] Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, Australia

[d] Statistics and Mathematics College, Yunnan University of Finance and Economics, Kunming, People's Republic of China

[e] Department of Statistics, Universidade Federal de Pernambuco, Recife, Brazil

Published online: 28 Aug 2015.

Journal of Applied Statistics, 2015
http://dx.doi.org/10.1080/02664763.2015.1072140

Diagnostics in elliptical regression models with stochastic restrictions applied to econometrics

Víctor Leiva[a,b*], Shuangzhe Liu[c], Lei Shi[d] and Francisco José A. Cysneiros[e]

[a] *Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Viña del Mar, Chile;* [b] *Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile;* [c] *Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, Australia;* [d] *Statistics and Mathematics College, Yunnan University of Finance and Economics, Kunming, People's Republic of China;* [e] *Department of Statistics, Universidade Federal de Pernambuco, Recife, Brazil*

(*Received 23 June 2014; accepted 9 July 2015*)

We propose an influence diagnostic methodology for linear regression models with stochastic restrictions and errors following elliptically contoured distributions. We study how a perturbation may impact on the mixed estimation procedure of parameters in the model. Normal curvatures and slopes for assessing influence under usual schemes are derived, including perturbations of case-weight, response variable, and explanatory variable. Simulations are conducted to evaluate the performance of the proposed methodology. An example with real-world economy data is presented as an illustration.

Keywords: computational statistics; elliptically contoured distributions; generalized least squares; local influence method; maximum-likelihood method; mixed estimation

1. Introduction

Durbin [13] was the pioneer to use sample data and auxiliary information in regression models simultaneously. If a set of stochastic equations is available as auxiliary (prior) information on the regression coefficients, a mixed estimation method can be used [46,49,50]. The method assumes that the prior information is as important as the sample data for the response and explanatory variables, called 'response' and 'covariable' hereafter. Therefore, both sample data and prior information are treated as two equally important equations for the model and are simultaneously considered in the estimation procedure. During the last few decades, the mixed estimation method has been studied and applied to various situations in statistics and econometrics. For recent developments on the topic, see [20, 21,27,37,43,56]. Studies with non-stochastic

*Corresponding author. Email: victorleivasanchez@gmail.com

restrictions were conducted by Cysneiros and Paula [8], Liu and Neudecker [35] and Paula and Cysneiros [40].

Elliptically contoured (EC) distributions are a family of models that contain the normal case as one of its members, among a number of non-normal members, such as the logistic, power exponential and Student-t (called simply t hereafter) cases. The t distribution is more general than the normal distribution and can flexibly deal with several situations involving heavy tails. For more details about EC distributions, see [9–12,15,17–19,23,28–31,33,34,53,54].

Influence diagnostics are an important step to be considered in data modeling. It is often carried out after the parameters are estimated. It allows the stability of the estimation procedure to be assessed. Diagnostic methods have been used in normal linear regression models [5,7]. Cook [6] introduced the local influence method to evaluate the effect of perturbations in the model and/or data on the maximum-likelihood (ML) estimates. The method has played a significant role in regression diagnostics. After Cook [6], who applied his method to the linear regression model under normality, a number of researchers have studied diagnostics for statistical models with further conditions [1,2,16,17,22,24–26,39,41,45,47,55]. For models with restrictions, see [4,23,37,39,48]. All of these last studies are relevant to some restriction, which needs to be imposed by the problem or the data. Influence diagnostics can be studied by means of regression models with stochastic restrictions. Different additional conditions in the model may lead to different diagnostics and identify different influential observations. To our knowledge, influence diagnostics in regression models under the normal distribution with stochastic restrictions was studied by Liu *et al.* [32], but no studies have been reported for EC distributions, which are more general and widely applicable.

The objectives of this paper are to derive and apply tools of influence diagnostics for EC regression models with stochastic linear restrictions. We adapt a likelihood-based approach to produce a local influence analysis, implement it by a computational routine in the statistical software R and apply it to a real-world example from economy.

Section 2 presents EC regression models with stochastic linear restrictions and the estimation of their parameters. Section 3 proposes a methodology based on the local influence method using matrices for the curvatures and vectors for the slopes under different schemes of perturbation. Section 4 conducts two simulation studies to evaluate the performance of the proposed methodology. Section 5 develops an example with economy data as illustration. Section 6 provides some conclusions and possible future studies.

2. Modeling and estimation

2.1 *Modeling*

As in [43, Section 5.10], we consider the linear regression model given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y} = (Y_1, \ldots, Y_n)^\top$ is an $n \times 1$ response vector, $\boldsymbol{X}$ is an $n \times p$ known design matrix of rank p, containing the values of p covariables, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients to be estimated. Note that $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with expectation $\mathrm{E}[\boldsymbol{\varepsilon}] = \boldsymbol{0}$ (null vector) and scale matrix $\phi \boldsymbol{I}_n$. Here, ϕ is a scale parameter and $\boldsymbol{I}_n$ the $n \times n$ identity matrix. We assume that auxiliary information is available, which often may be expressed by a linear restriction of the type

$$\boldsymbol{r} = \boldsymbol{R}\boldsymbol{\beta} + \boldsymbol{v}, \tag{2}$$

where $\boldsymbol{r}$ is a $k \times 1$ vector, $\boldsymbol{R}$ is a $k \times p$ matrix of constants of rank k ($k \le p$), and $\boldsymbol{v}$ is a $k \times 1$ vector of errors with expectation $\mathrm{E}[\boldsymbol{v}] = \boldsymbol{0}$ and scale matrix $\phi \boldsymbol{V}$. Here, $\boldsymbol{V}$ is a $k \times k$ positive definite matrix of fixed values. Note that $\boldsymbol{r}$ may be interpreted as a random vector with expectation $\mathrm{E}[\boldsymbol{r}] = \boldsymbol{R}\boldsymbol{\beta}$. After having the prior information, we assume $\boldsymbol{r}$ to be known (i.e. it is an observed value of the random vector). Then, all the expectations are conditional on $\boldsymbol{r}$ as, for example, $\mathrm{E}[\hat{\boldsymbol{\beta}} \mid \boldsymbol{r}]$. Furthermore, $\boldsymbol{R}$ and $\boldsymbol{V}$ are assumed to be known, and both random errors are assumed to be uncorrelated, that is, $\mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\nu}^\top] = \boldsymbol{0}$. Thus, we can formulate Equations (1) and (2) as one model, with the $[p+1] \times 1$ parameter vector partitioned as $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$, by

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{e} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{e}, \tag{3}$$

where

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{r} \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{R} \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\nu} \end{pmatrix},$$

with $\boldsymbol{z}$, $\boldsymbol{Z}$ and $\boldsymbol{e}$ being $m \times 1$, $m \times p$ and $m \times 1$ vectors, respectively, for $m = n + k$. Note the term $\boldsymbol{e} = \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{\beta}$ with errors of the model has $\mathrm{E}[\boldsymbol{e}] = \boldsymbol{0}$ and the $m \times m$ scale matrix is given by

$$\boldsymbol{\Sigma}^\star = \begin{pmatrix} \phi \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \phi \boldsymbol{V} \end{pmatrix} = \phi \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V} \end{pmatrix} = \phi \boldsymbol{\Sigma}.$$

Representation provided in Equation (3) combines the model and restrictions. When sample data and prior information for the model are not equally important, Schaffrin and Toutenburg [46] indicated that weighted mixed regression may well be useful. We consider sample data and prior information as equally important. Therefore, no obvious difference can be made between the model and restrictions.

2.2 *EC distributions*

Let $\boldsymbol{z}$ be an $m \times 1$ vector following an EC distribution with location vector $\boldsymbol{\mu} \in \mathbb{R}^m$, scale matrix $\boldsymbol{\Sigma}^\star \in \mathbb{R}^{m \times m}$ of $\mathrm{rank}(\boldsymbol{\Sigma}^\star) = m$ and kernel g, denoted by $\boldsymbol{z} \sim \mathrm{EC}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}^\star; g)$. The density of $\boldsymbol{z}$ is given by

$$f(\boldsymbol{z}) = c|\boldsymbol{\Sigma}^\star|^{-1/2} g([\boldsymbol{z} - \boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{\star^{-1}} [\boldsymbol{z} - \boldsymbol{\mu}]), \quad \boldsymbol{z} \in \mathbb{R}^m, \tag{4}$$

where c is a normalizing constant, such that $\int_0^{+\infty} u^{-1/2} g(u)\,\mathrm{d}u = 1/c$. Consider $G(u) = g'(u)/g(u)$, where $g'(u)$ is the derivative of $g(u)$ given in Equation (4) with respect to $u > 0$. If g is continuous and decreasing, then its maximum u_g exists and is finite and positive. If g is differentiable, then u_g is the solution to $G(u) + m/[2u] = 0$. In several EC distributions, the kernel g depends on an additional shape parameter that we denoted by ν, which allows us to control the kurtosis of the distribution. It is known that $u_g = m$ for the normal and t with ν degrees of freedom distributions, the last one denoted by $t(\nu)$. We have $G(u) = -1/2$, $G'(u) = 0$ and $G(u) = -[\nu + m]/[2(\nu + u)]$, $G'(u) = [\nu + m]/[2(\nu + u)^2]$ for the normal and $t(\nu)$ distributions, respectively. For $g(u)$, $G(u)$ and $G'(u)$ of several EC distributions, see [17,44]. Note that, if $m = 1$, we have the case of symmetric distributions instead of EC distributions [15].

2.3 *Estimation*

We can use the ML method for estimating the parameters of the model formulated in Equation (3) along the lines of the works by Billor and Loynes [4] and Labra *et al.* [23] employed for ridge regression. To derive the ML estimator of the parameter $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\beta}}^\top, \hat{\phi}]^\top$, we assume the model error follows an EC distribution. Specifically, we assume for the model formulated by

Equation (3) that $\boldsymbol{e} = \boldsymbol{z} - \boldsymbol{\mu} = \boldsymbol{z} - \boldsymbol{Z}\boldsymbol{\beta} \sim \mathrm{EC}_m(\boldsymbol{0}, \phi\boldsymbol{\Sigma}; g)$. Thus, from Equation (4), to obtain $\hat{\boldsymbol{\theta}}$, we consider the log-likelihood function (ignoring the constant term) for $\boldsymbol{\theta}$ given by

$$\ell(\boldsymbol{\theta}) = -\frac{m}{2}\log(\phi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) + \log\left(g\left(\frac{1}{\phi}Q\right)\right), \tag{5}$$

where $Q = Q(\boldsymbol{\beta}) = \boldsymbol{e}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$. The score vector obtained from the derivatives of Equation (5) partitioned accordingly to $\boldsymbol{\theta}$ is

$$\dot{\boldsymbol{\ell}}(\boldsymbol{\theta}) = (\dot{\boldsymbol{\ell}}_1(\boldsymbol{\theta})^\top, \dot{\ell}_2(\boldsymbol{\theta}))^\top, \tag{6}$$

where

$$\dot{\boldsymbol{\ell}}_1(\boldsymbol{\theta}) = -\frac{2}{\phi}G\boldsymbol{Z}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}, \quad \dot{\ell}_2(\boldsymbol{\theta}) = -\frac{m}{2\phi} - \frac{1}{\phi^2}GQ,$$

with $G = G(u_g)$ and u_g maximizing the function $t(u) = u^{m/2}g(u)$. As known, elements $\dot{\boldsymbol{\ell}}_1(\boldsymbol{\theta})$ and $\dot{\ell}_2(\boldsymbol{\theta})$ expressed in Equation (6) must be equated to zero to find a solution for $\boldsymbol{\theta}$, establishing the ML estimators of $\boldsymbol{\beta}$ and ϕ as

$$\hat{\boldsymbol{\beta}} = [\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{R}^\top\boldsymbol{V}^{-1}\boldsymbol{R}]^{-1}[\boldsymbol{X}^\top\boldsymbol{y} + \boldsymbol{R}^\top\boldsymbol{V}^{-1}\boldsymbol{r}], \tag{7}$$

$$\hat{\phi} = \frac{1}{u_g}\hat{Q} = \frac{1}{u_g}[(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}})^\top\boldsymbol{V}^{-1}(\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}})], \tag{8}$$

where $\hat{Q} = Q(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{e}}^\top\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{e}}$, with $\hat{\boldsymbol{e}} = \boldsymbol{z} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}$.

We can also use the method of generalized least squares (GLS) for estimating the parameter vector $\boldsymbol{\beta}$ defined by Equation (3), which coincides with the ML estimator $\hat{\boldsymbol{\beta}}$ provided in Equation (7) [43, Section 5.10]. However, different from $\hat{\phi}$ in Equation (8), the GLS estimator of ϕ is given by

$$\tilde{\phi} = \frac{1}{m-p}[(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}})^\top\boldsymbol{V}^{-1}(\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}})].$$

2.4 *Fisher information matrix*

Consider the model formulated in Equation (3) and $\boldsymbol{\theta}$ partitioned accordingly. Then, the corresponding Hessian matrix obtained from the derivatives of Equation (5) is given by

$$\ddot{\boldsymbol{\ell}}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\boldsymbol{\ell}}_{11}(\boldsymbol{\theta}) & \ddot{\boldsymbol{\ell}}_{12}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{\ell}}_{21}(\boldsymbol{\theta}) & \ddot{\ell}_{22}(\boldsymbol{\theta}) \end{pmatrix}, \tag{9}$$

where $\ddot{\boldsymbol{\ell}}_{11}(\boldsymbol{\theta}), \ddot{\boldsymbol{\ell}}_{12}(\boldsymbol{\theta}), \ddot{\boldsymbol{\ell}}_{21}(\boldsymbol{\theta}), \ddot{\ell}_{22}(\boldsymbol{\theta})$ are detailed in the appendix. Thus, the observed Hessian matrix $\ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\theta}})$ is obtained evaluating Equation (9) at the ML estimates of $\boldsymbol{\beta}$ and ϕ, getting

$$\ddot{\boldsymbol{\ell}}_{11}(\hat{\boldsymbol{\theta}}) = \frac{2}{\hat{\phi}}G\boldsymbol{Z}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}, \quad \ddot{\boldsymbol{\ell}}_{12}(\hat{\boldsymbol{\theta}}) = \ddot{\boldsymbol{\ell}}_{21}(\hat{\boldsymbol{\theta}})^\top = \boldsymbol{0}, \quad \ddot{\ell}_{22}(\hat{\boldsymbol{\theta}}) = \frac{m}{2\hat{\phi}^2} + \frac{2}{\hat{\phi}^3}G\hat{Q} + \frac{1}{\hat{\phi}^4}G'\hat{Q}^2,$$

where $G' = G'(u_g)$ and $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$ are as given in Equations (7) and (8), respectively. In practice, we can approximate the expected Fisher information matrix by its observed version throughout the observed Hessian matrix $-\ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\theta}})$; see [14] for details about the accuracy of the ML estimator obtained from observed vs. expected Fisher information matrices.

3. Local influence

3.1 *The local influence method*

Let $\ell(\boldsymbol{\theta})$ be the log-likelihood function for the model of parameter $\boldsymbol{\theta}$ formulated in Equation (3), called the non-perturbed model. Let $\boldsymbol{w} = (w_1, \ldots, w_m)^\top$ be an $m \times 1$ vector of perturbations in the model with $\boldsymbol{w} \in \Omega$, where Ω is an open set of relevant perturbations. Let $\ell(\boldsymbol{\theta} \mid \boldsymbol{w})$ be the log-likelihood function of the model perturbed with $\boldsymbol{w}$, called the perturbed model, and $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ be the ML estimate of $\boldsymbol{\theta}$ obtained from $\ell(\boldsymbol{\theta} \mid \boldsymbol{w})$. In addition, let $\boldsymbol{w}_0 \in \Omega$ be an $m \times 1$ non-perturbed vector with $\boldsymbol{w}_0 = (0, \ldots, 0)^\top$, or $\boldsymbol{w}_0 = (1, \ldots, 1)^\top$, or a third choice, depending on the context, such that $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta} \mid \boldsymbol{w}_0)$. Suppose that $\ell(\boldsymbol{\theta} \mid \boldsymbol{w})$ is twice continuously differentiable in a neighborhood of $(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0)$. We are interested in comparing the parameter estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ by using the local influence method. We investigate how the inference is affected by the perturbation.

We consider two approaches to the local influence method: one proposed by Cook [6] and the other one by Tsai [52] and Billor and Loynes [3,4]. We refer to them as Cook and alternative approaches, respectively.

3.2 *The Cook approach*

In this approach, the likelihood displacement (LD) is chosen to be

$$\mathrm{LD}(\boldsymbol{w}) = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{w}})], \tag{10}$$

which is used to assess the influence of the perturbation $\boldsymbol{w}$. Large values of $\mathrm{LD}(\boldsymbol{w})$ in Equation (10) indicate that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ differ considerably in relation to the contours of the log-likelihood function based on the non-perturbed model, $\ell(\boldsymbol{\theta})$. The approach studies the local behavior of the influence graph $a(\boldsymbol{w}) = (\boldsymbol{w}^\top, \mathrm{LD}(\boldsymbol{w}))^\top$ around $\boldsymbol{w}_0$. Cook [6] suggested to investigate the direction of maximum curvature of the surface $a(\boldsymbol{w})$. For $\mathrm{LD}(\boldsymbol{w})$ in Equation (10), this curvature is $C_{\max} = \max_{\|\boldsymbol{d}\|=1} C_{\boldsymbol{d}}$, where $C_{\boldsymbol{d}} = 2|\boldsymbol{d}^\top \boldsymbol{F} \boldsymbol{d}|$, with $\boldsymbol{F}$ being an $m \times m$ matrix and $\boldsymbol{d}$ the unit-length direction vector. To find $C_{\max}$ and the direction vector $\boldsymbol{d}_{\max}$, we need to calculate

$$\boldsymbol{F} = -\boldsymbol{\Delta}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0)^\top \ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\theta}})^{-1} \boldsymbol{\Delta}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0). \tag{11}$$

Note that $\boldsymbol{\Delta}(\boldsymbol{\theta}, \boldsymbol{w})$ is a $p \times m$ matrix for the perturbed model obtained from Equation (3) and given by

$$\boldsymbol{\Delta}(\boldsymbol{\theta}, \boldsymbol{w}) = \frac{\partial^2 \ell(\boldsymbol{\theta} \mid \boldsymbol{w})}{\partial \boldsymbol{\theta} \partial \boldsymbol{w}^\top} = \begin{pmatrix} \dfrac{\partial^2 \ell(\boldsymbol{\theta} \mid \boldsymbol{w})}{\partial \boldsymbol{\beta} \partial \boldsymbol{w}^\top} \\ \dfrac{\partial^2 \ell(\boldsymbol{\theta} \mid \boldsymbol{w})}{\partial \phi \partial \boldsymbol{w}^\top} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta}_1(\boldsymbol{\theta}, \boldsymbol{w}) \\ \boldsymbol{\Delta}_2(\boldsymbol{\theta}, \boldsymbol{w}) \end{pmatrix}. \tag{12}$$

Note also that Equation (12) must be evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0$. Then, $\boldsymbol{d}_{\max}$ is a unit-length eigenvector associated with the largest absolute eigenvalue $C_{\max}$ of $\boldsymbol{F}$ given in Equation (11). Large absolute values of the elements of $\boldsymbol{d}_{\max}$ reveal that the observations are likely to be influential.

3.3 *An alternative approach*

In this approach, the LD is now chosen to be

$$\mathrm{LD}^*(\boldsymbol{w}) = -2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} \mid \boldsymbol{w})]. \tag{13}$$

Note that $\mathrm{LD}(\boldsymbol{w})$ in the Cook approach and $\mathrm{LD}^*(\boldsymbol{w})$ in Equation (13) are different. $\mathrm{LD}(\boldsymbol{w})$ has the first derivative, with respect to $\boldsymbol{w}$ evaluated at $\boldsymbol{w}_0$, equal to zero. Cook [6] used normal curvature (second-order derivative) to evaluate the local influence. However, the first derivative of $\mathrm{LD}^*(\boldsymbol{w})$ does not vanish. Therefore, the slope can be used to study the local change. Note that $\mathrm{LD}^*(\boldsymbol{w})$ considers the log-likelihood function based on the perturbed model and a negative sign multiplying it to assure its positive value. We have the surface $a^*(\boldsymbol{w}) = (\boldsymbol{w}^\top, \mathrm{LD}^*(\boldsymbol{w}))^\top$ and its first derivative (slope) $S_{\boldsymbol{d}}$ in the direction $\boldsymbol{d}$ does not vanish, except in trivial cases. Thus, it provides valuable information about the local behavior of $\mathrm{LD}^*(\boldsymbol{w})$ in Equation (13). The maximum slope $S_{\max}$ and the corresponding direction vector $\boldsymbol{d}_{\max}$ are useful in detecting local influence. Because the alternative approach only involves first derivatives and not second derivatives, it is easier to use than the Cook approach. In fact, the maximum slope for $S_{\boldsymbol{d}}$ is given by $S_{\max} = 2|\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{w}_0)|$, where

$$\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{w}) = \frac{\partial \ell(\boldsymbol{\theta} \mid \boldsymbol{w})}{\partial \boldsymbol{w}^\top}, \tag{14}$$

which must be evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0$.

3.4 *Normal curvatures and slopes*

We determine curvatures and slopes for the Cook and alternative approaches, respectively, in the model formulated in Equation (3) and its perturbed version. For the Cook approach, we compute the observed information matrix $-\ddot{\boldsymbol{\ell}}(\hat{\boldsymbol{\theta}})$, the derivative matrix $\boldsymbol{\Delta}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0)$, and then the eigenvector associated with the largest absolute eigenvalue of $\boldsymbol{F}$ given in Equation (11) as our local diagnostic. For the alternative approach, we calculate the derivative vector $\boldsymbol{h}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0)$ defined in Equation (14). We use one model perturbation scheme and two data perturbation schemes. The model perturbation scheme is equivalent to the case-weight scheme considered in [17], addressing the multivariate model, but not the univariate model studied in [19]. Details of the derivatives obtained from Equation (5) used for establishing the curvatures and slopes are provided in the appendix.

Case-weight perturbation (c)

Let $\boldsymbol{w}_0 = (1, \ldots, 1)^\top$ be the non-perturbed vector, such that $\ell(\boldsymbol{\theta} \mid \boldsymbol{w}_0) = \ell(\boldsymbol{\theta})$. Also, let $\boldsymbol{W} = \mathrm{diag}(\boldsymbol{w})$ be an $n \times n$ matrix containing the perturbations $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ in its diagonal. The w_is are positive and denote the weight corresponding to the case i. Using $\phi \boldsymbol{W}^{-1}$ instead of $\phi \boldsymbol{I}_n$ in formulation given in Equation (3), we obtain the relevant part of the log-likelihood function for the perturbed model as

$$\ell_{\mathrm{c}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = -\frac{m}{2}\log(\phi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}_{\boldsymbol{w}}|) + \log\left(g\left(\frac{1}{\phi} Q_{w_{\mathrm{c}}}\right)\right), \tag{15}$$

where

$$Q_{w_{\mathrm{c}}} = \boldsymbol{e}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1} \boldsymbol{e}, \quad \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1} = \begin{pmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}^{-1} \end{pmatrix}.$$

The Cook approach. We specify $\boldsymbol{\Delta}(\boldsymbol{\theta}, \boldsymbol{w})$ given in Equation (12) taking the derivatives of $\ell_{\mathrm{c}}(\boldsymbol{\theta} \mid \boldsymbol{w})$ established in Equation (15) with respect to $\boldsymbol{\beta}$ and ϕ, and then with respect to $\boldsymbol{w}$, evaluating them at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0 = (1, \ldots, 1)^\top$. Thus, according to Equations (12) and (15), we obtain

$$\boldsymbol{\Delta}_{1_{\mathrm{c}}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = -\frac{2}{\hat{\phi}} G[\hat{\boldsymbol{\varepsilon}}^\top \otimes \boldsymbol{X}^\top]\boldsymbol{J} = -\frac{2}{\hat{\phi}} G \boldsymbol{X}^\top \mathrm{D}(\hat{\boldsymbol{\varepsilon}}),$$

$$\boldsymbol{\Delta}_{2_{\mathrm{c}}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = -\left[\frac{1}{\hat{\phi}^3} G' \hat{Q} + \frac{1}{\hat{\phi}^2} G\right][\hat{\boldsymbol{\varepsilon}}^\top \otimes \hat{\boldsymbol{\varepsilon}}^\top]\boldsymbol{J} = -\frac{1}{\hat{\phi}^3}[G'\hat{Q} + \hat{\phi} G]\hat{\boldsymbol{\varepsilon}}^\top \mathrm{D}(\hat{\boldsymbol{\varepsilon}}),$$

where $\otimes$ is the Kronecker product, $\boldsymbol{J} = (\boldsymbol{i}_1 \otimes \boldsymbol{i}_1, \dots, \boldsymbol{i}_n \otimes \boldsymbol{i}_n)$ the $n^2 \times n$ selection matrix, with $\boldsymbol{i}_t$, for $t = 1, \dots, n$, being the tth column of $\boldsymbol{I}_n = (\boldsymbol{i}_1, \dots, \boldsymbol{i}_n)$, and $\mathrm{D}(\hat{\boldsymbol{\varepsilon}}) = \mathrm{diag}(\hat{\boldsymbol{\varepsilon}}_1, \dots, \hat{\boldsymbol{\varepsilon}}_n)$ an $n \times n$ diagonal matrix. For more details about $\boldsymbol{J}$ and its statistical applications, see [38].

Alternative approach. We specify $\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{w})$ given in Equation (14) taking derivatives of $\ell_{\mathrm{c}}(\boldsymbol{\theta} \mid \boldsymbol{w})$ established in Equation (15) with respect to $\boldsymbol{w}$ and evaluating it at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0 = (1, \dots, 1)^\top$. Thus, according to Equations (14) and (15), we obtain

$$\boldsymbol{h}_{\mathrm{c}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = \left[\frac{1}{2}\mathrm{vec}^\top(\boldsymbol{I}_n) + \frac{1}{\hat{\phi}} G(\hat{\boldsymbol{\varepsilon}}^\top \otimes \hat{\boldsymbol{\varepsilon}}^\top)\right]\boldsymbol{J} = \frac{1}{2}\mathbf{1}_n^\top + \frac{1}{\hat{\phi}} G\hat{\boldsymbol{\varepsilon}}^\top \mathrm{D}(\hat{\boldsymbol{\varepsilon}}),$$

where $\mathbf{1}_n^\top = (1, \dots, 1)$ is an $1 \times n$ vector.

Perturbation of the response (r)

We replace $\boldsymbol{y}$ by $\boldsymbol{y} + \boldsymbol{w}$, where $\boldsymbol{w} = (w_1, \dots, w_n)^\top$ is an $n \times 1$ vector with $w_i > 0$ denoting the perturbation corresponding to the case i. Here, w_i can be proportional to the standard deviation (SD) of the response and $\boldsymbol{w}_0 = (0, \dots, 0)^\top$ is the non-perturbed vector. Hence, the relevant part of the log-likelihood function is given by

$$\ell_{\mathrm{r}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = -\frac{m}{2}\log(\phi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) + \log\left(g\left(\frac{1}{\phi}Q_{\boldsymbol{w}_{\mathrm{r}}}\right)\right), \tag{16}$$

where

$$Q_{\boldsymbol{w}_{\mathrm{r}}} = \boldsymbol{e}_{\boldsymbol{w}_{\mathrm{r}}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_{\boldsymbol{w}_{\mathrm{r}}}, \quad \boldsymbol{e}_{\boldsymbol{w}_{\mathrm{r}}} = \begin{pmatrix} \boldsymbol{y} + \boldsymbol{w} \\ \boldsymbol{r} \end{pmatrix} - \boldsymbol{Z}\boldsymbol{\beta}.$$

The Cook approach. We specify $\boldsymbol{\Delta}(\boldsymbol{\theta}, \boldsymbol{w})$ given in Equation (12) by taking the corresponding derivatives of $\ell_{\mathrm{r}}(\boldsymbol{\theta} \mid \boldsymbol{w})$ established in Equation (16) and evaluating them at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0 = (0, \dots, 0)^\top$. Thus, from Equations (12) and (16), we obtain

$$\boldsymbol{\Delta}_{1_{\mathrm{r}}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = -\frac{2}{\hat{\phi}} GX^\top, \quad \boldsymbol{\Delta}_{2_{\mathrm{r}}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = -\frac{2}{\hat{\phi}^3} G'\hat{Q}\hat{\boldsymbol{\varepsilon}}^\top - \frac{2}{\hat{\phi}^2} G\hat{\boldsymbol{\varepsilon}}^\top = -\frac{2}{\hat{\phi}^3}[G'\hat{Q} + \hat{\phi}G]\hat{\boldsymbol{\varepsilon}}^\top.$$

Alternative approach. We specify $\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{w})$ given in Equation (14) by taking the derivatives of $\ell_{\mathrm{r}}(\boldsymbol{\theta} \mid \boldsymbol{w})$ established in Equation (16), evaluating it at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0 = (0, \dots, 0)^\top$. Hence, from Equations (14) and (16), we obtain

$$\boldsymbol{h}_{\mathrm{r}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = \frac{2}{\hat{\phi}} G\hat{\boldsymbol{\varepsilon}}^\top.$$

Perturbation of a covariable (e)

We replace $\boldsymbol{X}$ by $\boldsymbol{X} + \boldsymbol{W}$, where $\boldsymbol{W} = \mathrm{diag}(\boldsymbol{w})$ is an $n \times n$ matrix containing the perturbations $\boldsymbol{w} = (w_1, \dots, w_n)^\top$ in its diagonal. Here, w_i can be proportional to the SD of the covariable and $\boldsymbol{w}_0 = (0, \dots, 0)^\top$ is the non-perturbed vector. Hence, the relevant part of the log-likelihood function is given by

$$\ell_{\mathrm{e}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = -\frac{m}{2}\log(\phi) - \frac{1}{2}\log(|\boldsymbol{\Sigma}|) + \log\left(g\left(\frac{1}{\phi}Q_{\boldsymbol{w}_{\mathrm{e}}}\right)\right), \tag{17}$$

where

$$Q_{\boldsymbol{w}_{\mathrm{e}}} = \boldsymbol{e}_{\boldsymbol{w}_{\mathrm{e}}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_{\boldsymbol{w}_{\mathrm{e}}}, \quad \boldsymbol{e}_{\boldsymbol{w}_{\mathrm{e}}} = \boldsymbol{z} - \boldsymbol{Z}_{\boldsymbol{w}}\boldsymbol{\beta}, \quad \boldsymbol{Z}_{\boldsymbol{w}} = \begin{pmatrix} \boldsymbol{X} + \boldsymbol{W} \\ \boldsymbol{R} \end{pmatrix}.$$

The Cook approach. We specify $\boldsymbol{\Delta}(\boldsymbol{\theta}, \boldsymbol{w})$ given in Equation (12) by taking the derivatives of $\ell_{\mathrm{e}}(\boldsymbol{\theta} \mid \boldsymbol{w})$ established in Equation (17) and evaluating them at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0 = (0, \dots, 0)^\top$.

Thus, from Equations (12) and (17), we obtain

$$\boldsymbol{\Delta}_{1_{\rm e}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = -\frac{2}{\hat{\phi}} G[(\hat{\boldsymbol{\varepsilon}}^\top \otimes \boldsymbol{I}_p)\boldsymbol{K} - \boldsymbol{X}^\top(\hat{\boldsymbol{\beta}}^\top \otimes \boldsymbol{I}_n)] = -\frac{2}{\hat{\phi}} G[\boldsymbol{I}_p \otimes \hat{\boldsymbol{\varepsilon}}^\top - \hat{\boldsymbol{\beta}}^\top \otimes \boldsymbol{X}^\top],$$

$$\boldsymbol{\Delta}_{2_{\rm e}}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = \frac{2}{\hat{\phi}^3} G'\hat{Q}[\hat{\boldsymbol{\beta}}^\top \otimes \hat{\boldsymbol{\varepsilon}}^\top] + \frac{2}{\hat{\phi}^2} G[\hat{\boldsymbol{\beta}}^\top \otimes \hat{\boldsymbol{\varepsilon}}^\top] = \frac{2}{\hat{\phi}^3}[G'\hat{Q} + \hat{\phi}G][\hat{\boldsymbol{\beta}}^\top \otimes \hat{\boldsymbol{\varepsilon}}^\top],$$

where $\boldsymbol{K}$ is the $np \times np$ commutation matrix as used in [36].

Alternative approach. We specify $\boldsymbol{h}(\boldsymbol{\theta}, \boldsymbol{w})$ given in Equation (14) by taking the derivative of $\ell_{\rm e}(\boldsymbol{\theta} \mid \boldsymbol{w})$ established in Equation (17) and evaluating it at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{w} = \boldsymbol{w}_0$. Hence, from Equations (14) and (17), we obtain

$$\boldsymbol{h}_{\rm e}(\hat{\boldsymbol{\theta}}, \boldsymbol{w}_0) = -\frac{2}{\hat{\phi}} G[\hat{\boldsymbol{\beta}}^\top \otimes \hat{\boldsymbol{\varepsilon}}^\top].$$

4. Simulation

4.1 *Computational framework*

A computational framework to analyze data using the proposed diagnostic methodology has been implemented in R, a non-commercial open source software for statistics and graphs (see www.r-project.org and [42]).

4.2 *Model for simulation*

For our simulation studies, we consider the model formulated in Equation (3) now specified as

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \tag{18}$$

with stochastic restrictions $-0.7 = \beta_1 + v_1, 1 = \beta_2 + v_2$, and the vector of errors following a three-variate EC distribution detailed as

$$\boldsymbol{e} = \begin{pmatrix} \varepsilon \\ v_1 \\ v_2 \end{pmatrix} \sim \text{EC}_3\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \phi\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.0225 & -0.0100 \\ 0 & -0.0100 & 0.0225 \end{pmatrix}; g\right).$$

The choice of the model parameters and data for the above-described simulation studies are based on a real-word scenario provided in the example of Section 5.

4.3 *Simulation study I*

First, in the specification given in Equation (18) with model errors following a normal distribution, consider $y_{17} = 2.21880$ and its perturbations $y_{17}^* = y_{17} + d$, with d varying between -10 and 10 by 1 (see [44] for a similar study). Figure 1 shows the behavior of the estimators of β_1 (left) and of its standard error $\text{SE}(\hat{\beta}_1)$ (right) by the relative change $\text{RC}_d = |[\hat{\theta}_i - \hat{\theta}_i^d]/\hat{\theta}_i| \times 100$, where $\theta_1 = \beta_0$, $\theta_2 = \beta_1$, $\theta_3 = \beta_2$, $\theta_4 = \phi$, and $\hat{\theta}_i^d$ is the estimate of θ_i with observed response y_{17}^*, for $i = 1, 2, 3, 4$. The figure shows that, as the absolute value of d increases, the estimator of β_1 and its estimated SE are more sensitive to detecting atypical observations. A similar behavior is observed for other covariables. Then, the proposed diagnostic methodology under normality should identify atypical observations, which is analyzed in simulation study II and the application

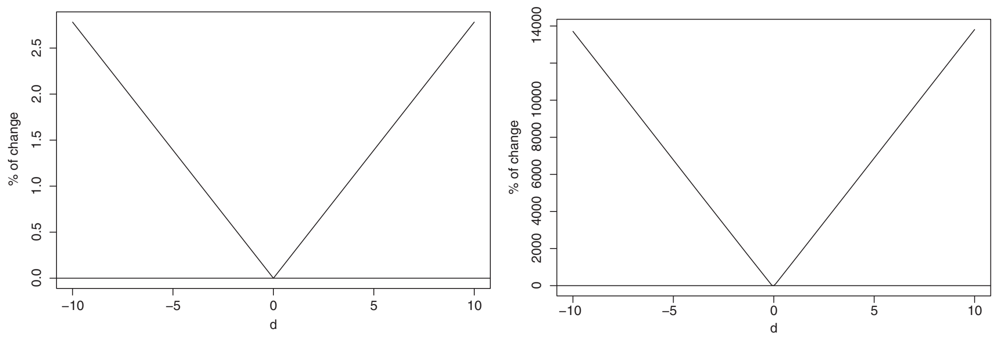

Figure 1. Impact of atypical data for the model specified in Equation (18) on $\hat{\beta}_1$ (left) and $\widehat{\mathrm{SE}}(\hat{\beta}_1)$ (right).

section. However, it is well known that, in the case of EC models, no differences among the estimates of their parameters nor among the corresponding estimated SE under normal and $t(\nu)$ error assumptions exist, even when restrictions are present. Only differences between estimators of the scale parameter and of its asymptotic SE are detected (see theoretical details in Section 2.3 and empirical evidence in Section 5.2). Then, under the specification and perturbation used to construct Figure 1, but with a $t(\nu)$ error assumption, similar results are obtained.

4.4 *Simulation study II*

Second, again in specification given in Equation (18) with a normal error assumption, consider the perturbations on the data set expressed as $y^*_{17} = y_{17} + 4s_y$ and $x^*_{1,17} = x_{1,17} + 4s_x$, where s_y, s_x are the SDs of the response and the covariable X_1, respectively. Figure 2 displays index plots of the Cook (left) and alternative (right) approaches for perturbations of case-weight (first panel), response (second panel) and covariable X_1 (third panel). From the figure, note that, as expected, case #17 is revealed as influential in all perturbation schemes. Then, the diagnostic measures proposed in the paper are able to detect outlying observations. However, due to reasons mentioned in simulation study I, they show similar results when normal and $t(\nu)$ error assumptions are considered in EC regression models with stochastic restrictions. Differently, when symmetric regression models are used, a more noticeable difference between the parameter estimates for the normal and $t(\nu)$ error assumptions should be detected. Nevertheless, the derivation of diagnostic measures for symmetric regression models with stochastic restrictions is beyond the objective of the present paper. Therefore, we will study that topic in a future work.

5. Example with real-world data

5.1 *Data description*

We illustrate our results with data taken from [51], corresponding to the demand for textile products in the Netherlands during the period 1923–1939. The response (Y) is the logarithm of the per capita textile consumption per year, which is obtained dividing the money value of textile consumption per family households by pN, where p is the retail price index of clothing for the city of Amsterdam and N is accordingly the size of the population of the Netherlands. Two covariables are considered (X_1, X_2). Here, X_1 is the logarithm of the deflated price index of clothing, which is obtained as the ratio p/γ. In addition, X_2 is the logarithm of real per capita income, which is calculated dividing the monetary value of household income by γN, where

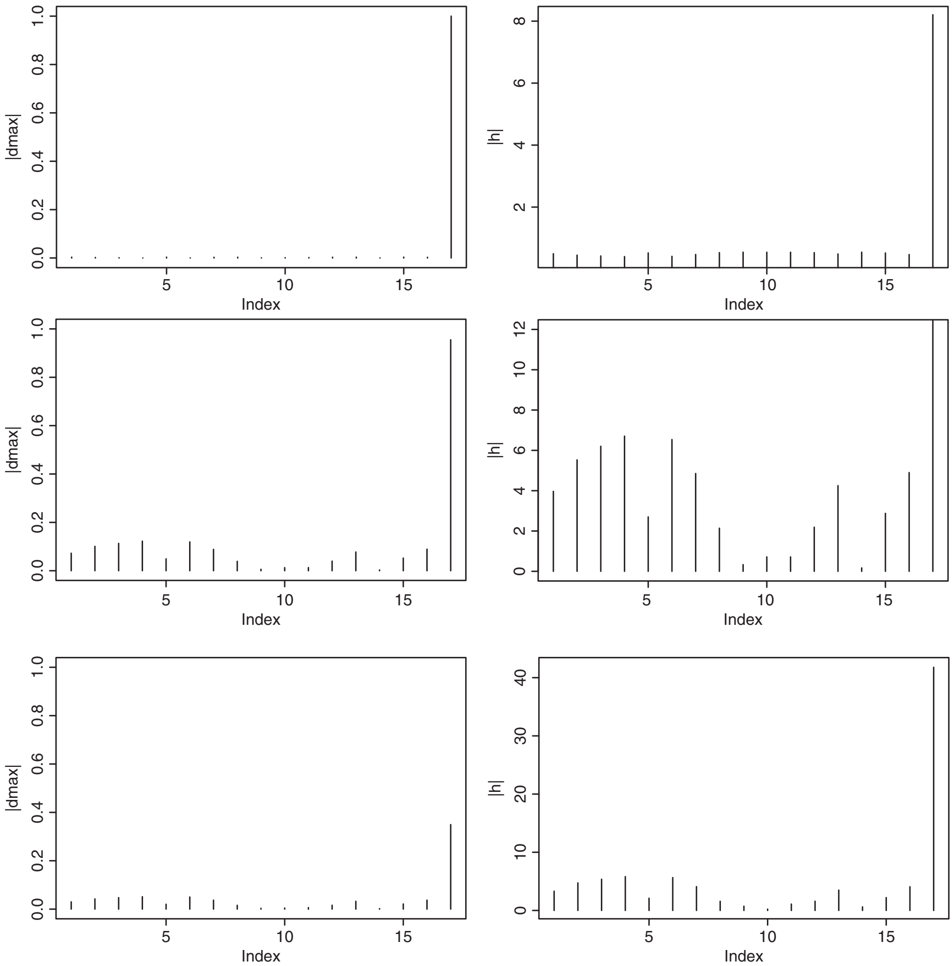

Figure 2. Index plots of influence with the Cook (left) and alternative (right) approaches under perturbation schemes of case-weight, response and covariable (first, second and third panels, respectively) for perturbed textile data.

γ is the general retail price index. The data are displayed in Table 1. In [49, p. 407], information about stochastic restrictions for variables X_1 and X_2 is provided. Specifically, for the price elasticity a prior estimate of -0.7 is taken, and for the income elasticity an estimate of 1 is considered. In both cases, an SE of 0.15 is assumed. Also, if the prior beliefs overestimate the income sensitivity, the same should happen for the price sensitivity and vice versa, at least on average. Given the negative sign of the price elasticity, this indicates that the prior cross-moment is negative.

5.2 *Modeling and estimation*

Our analysis is based on the model given in Equation (1), which in this example is specified as

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \tag{19}$$

Table 1. Logarithms of per capita textile consumption (Y), deflated price index of clothing (X_1) and per capita income (X_2) taken from [51] for the indicated year.

Index	Year	Y	X_1	X_2
1	1923	1.99651	2.00432	1.98543
2	1924	1.99564	2.00043	1.99167
3	1925	2.00000	2.00000	2.00000
4	1926	2.04766	1.95713	2.02078
5	1927	2.08707	1.93702	2.02078
6	1928	2.07041	1.95279	2.03941
7	1929	2.08314	1.95713	2.04454
8	1930	2.13354	1.91803	2.05038
9	1931	2.18808	1.84572	2.03862
10	1932	2.18639	1.81558	2.02243
11	1933	2.20003	1.78746	2.00732
12	1934	2.14799	1.79588	1.97955
13	1935	2.13418	1.80346	1.98408
14	1936	2.22531	1.72099	1.98945
15	1937	2.18837	1.77597	2.01030
16	1938	2.17319	1.77452	2.00689
17	1939	2.21880	1.78746	2.01620

where Y is the response related to consumption and x_1, x_2 are the values of the two covariables X_1, X_2 corresponding to textile demand; β_0, β_1 and β_2 are the model coefficients to be estimated and ε is the term of the model error, which we assume to follow an EC distribution, in particular, a normal distribution or a $t(4)$ distribution. The corresponding stochastic restrictions, imposed a priori based on the economic arguments mentioned at the end of Section 5.1, are defined as

$$\boldsymbol{r} = \begin{pmatrix} -0.7 \\ 1 \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{pmatrix} 0.0225 & -0.0100 \\ -0.0100 & 0.0225 \end{pmatrix}.$$

Based on the results presented in Section 3, we estimate the parameters of the model formulated in Equation (19). We assume the $t(4)$ distribution for the model error. As a comparison, we also provide estimates for the normal distribution. The corresponding ML estimates (with the estimated asymptotic SE of the respective estimators in parentheses) are $\hat{\beta}_0 = 1.4211(0.0063)$, $\hat{\beta}_1 = -0.7004(0.0024)$ and $\hat{\beta}_2 = 1.0002(0.0024)$, whereas the ML estimates of ϕ is $\hat{\phi} = 0.00002$, with estimated asymptotic SE of 0.00008 for the $t(4)$ distribution and of 0.0002 for the normal distribution. The GLS estimate of ϕ is $\tilde{\phi} = 0.00003$, with its estimated asymptotic SE equal to that from the ML estimator. Therefore, as mentioned in simulation study I, no differences among the regression parameter estimates under normal and $t(4)$ error assumptions exist. Only differences between scale parameter estimates and estimates of the asymptotic SE are observed.

5.3 *Influence diagnostics*

Figure 3 displays the index plots of the Cook and alternative approaches under perturbations of case-weight, response and covariable. From the figure, note that the 17th year (1939) is revealed as an influential case, in different degrees, under the three perturbation schemes in the two approaches. However, the alternative approach is easier to calculate than the Cook approach. Note that the year 1939 coincides with the beginning of World War II. The proposed diagnostic methodology has been useful for detecting potentially influential cases on the ML estimates.

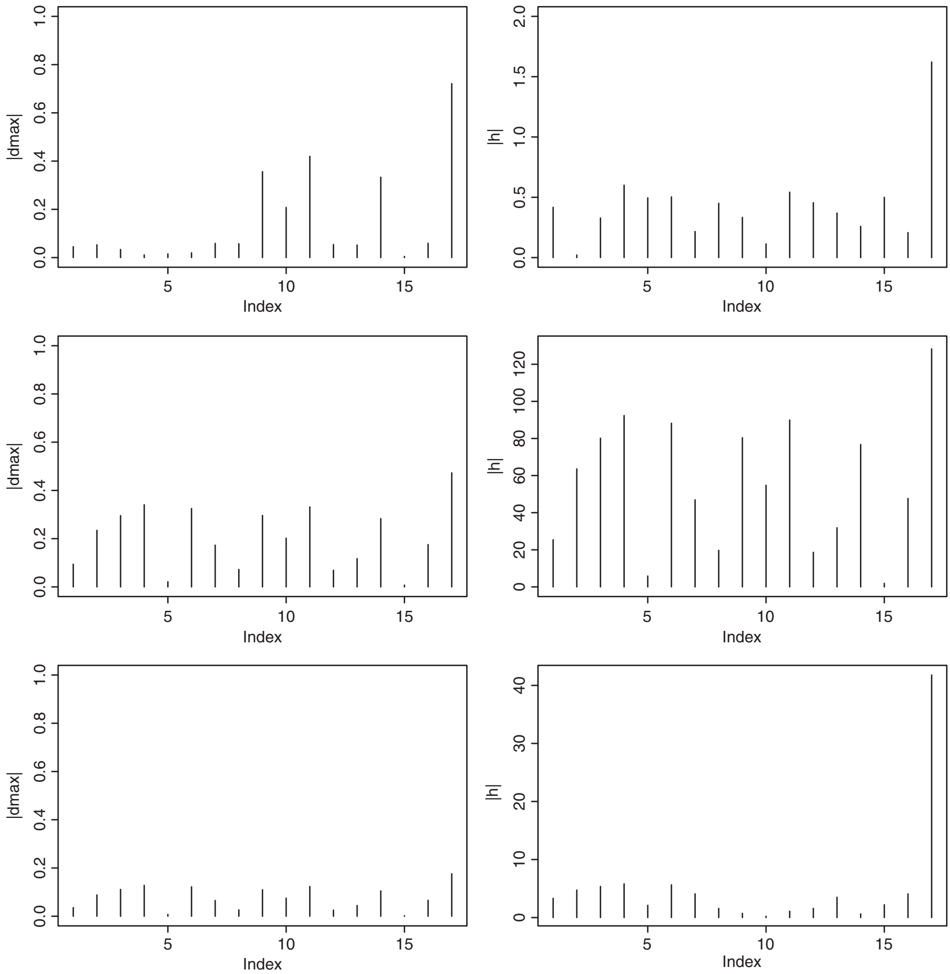

Figure 3. Index plots of influence with the Cook (left) and alternative (right) approaches under perturbation schemes ofcase-weight, response and X_1 covariable (first, second and third panels, respectively).

6. Conclusions and future studies

We proposed a methodology based on influence diagnostics for the linear regression model with stochastic restrictions and errors following an EC distribution. For this model with restrictions, we studied how a perturbation may impact on the mixed estimation procedure of its parameters. In addition, we derived expressions for curvatures and slopes of two approaches to the local influence method with three perturbation schemes: case-weight, response variable and covariable. Then, local influence assessment was carried out based on these two approaches. The performance of the proposed methodology was studied by simulation and illustrated with an example from real-world economy data. Our empirical studies with simulated and real data demonstrated that the proposed diagnostic methodology is useful for detecting potentially influential cases on the ML estimates. However, because no differences between the parameter estimates nor between the corresponding estimated SEs under normal and t error assumptions exist, diagnostics based on both assumptions are not distinguishable. For diagnostics in linear regression models with stochastic restrictions and normal t error assumptions to be distinguishable, we are considering a future study with the assumption of errors being symmetrically distributed rather than elliptically distributed.

Acknowledgments

The authors wish to thank the Editor-in-Chief, Professor Robert Aykroyd and three anonymous referees for their constructive comments on an earlier version of this manuscript. The research was initiated and conducted by the authors who met in pairs at Yunnan University of Finance and Economics (China), University of Canberra (Australia), Universidad de Valparaiso (Chile) and Universidade Federal de Pernambuco (Brazil).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research of F.J.A. Cysneiros and V. Leiva was supported by FONDECYT [1120879] (Chile) and Capes, CNPq and FACEPE (Brazil) grants. The research of L. Shi was supported by Natural Science Foundation [grant number 11161053] (China)

References

[1] A.C. Atkinson, *Econometric applications of the forward search in regression: Robustness, diagnostics, and graphics*, Econometric Rev. 28 (2009), pp. 21–39.
[2] M. Barros, M. Galea, M. Gonzalez, and V. Leiva, *Influence diagnostics in the Tobit censored response model*, Stat. Methods Appl. 19 (2010), pp. 379–397.
[3] N. Billor and R.M. Loynes, *Local influence: A new approach*, Comm. Statist. Theory Methods 22 (1993), pp. 1595–1611.
[4] N. Billor and R.M. Loynes, *An application of the local influence approach to ridge regression*, J. Amer. Statist. Assoc. 2 (1999), pp. 177–183.
[5] S. Chatterjee and A.S. Hadi, *Sensitivity Analysis in Linear Regression*, Wiley, New York, 1988.
[6] D. Cook, *Assessment of local influence*, J. R. Stat. Soc. B 48 (1986), pp. 133–169.
[7] R.D. Cook and S. Weisberg, *Residuals and Influence in Regression*, Chapman & Hall, London, 1982.
[8] F.J.A. Cysneiros and G.A. Paula, *Restricted methods in symmetrical linear regression models*, Comput. Statist. Data Anal. 49 (2005), pp. 689–708.
[9] F.J.A. Cysneiros and L.H. Vanegas, *Residuals and their statistical properties in symmetrical nonlinear models*, Statist. Probab. Lett. 78, (2008), pp. 3269–3273.
[10] J.A. Díaz-García, M. Galea Rojas, and V. Leiva-Sánchez, *Influence diagnostics for multivariate elliptical multivariate linear regression models*, Comm. Statist. Theory Methods 32 (2003), pp. 625–641.
[11] J.A. Díaz-García and V. Leiva-Sánchez, *Doubly non-central t and $\mathcal{F}$ distribution obtained under singular and non-singular elliptic distributions*, Comm. Statist Theory Methods 32 (2003), pp. 11–32.
[12] J.A. Díaz-García, V. Leiva-Sánchez, and M. Galea, *Singular elliptic distribution: density and applications*, Comm. Statist. Theory Methods 31 (2002), pp. 665–681.
[13] J. Durbin, *A note on regression when there is extraneous information about one of the coefficients*, J. Amer. Statist. Assoc. 48 (1953), pp. 799–808.
[14] B. Efron and D.V. Hinkley, *Assessing the accuracy of the maximum likelihood estimator: Observed vs. expected Fisher information*, Biometrika 65 (1978), pp. 457–483.
[15] K.T. Fang, S. Kotz, and K.W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman, London, 1990.
[16] M. Galea, V. Leiva-Sánchez, and G.A. Paula, *Influence diagnostics in log-Birnbaum–Saunders regression models*, J. Appl. Stat. 31 (2004), pp. 1049–1064.
[17] M. Galea, G.A. Paula, and H. Bolfarine, *Local influence in elliptical linear regression models*, Statistician 46 (1997), pp. 71–79.
[18] M. Galea, G.A. Paula, and F.J.A. Cysneiros, *On diagnostic in symmetrical nonlinear models*, Statist. Papers 73 (2005), pp. 459–467.
[19] M. Galea, G.A. Paula, and M. Uribe-Opazo, *On influence diagnostic in univariate elliptical linear regression models*, Statist. Papers 44 (2003), pp. 23–45.

[20] J. Gross, *Linear Regression*, Springer, Berlin, 2003.
[21] M.H.J. Gruber, *Regression Estimators: A Comparative Study*, J Hopkins University Press, Baltimore, 2010.
[22] J. Haslett and D. Dillane, *Application of 'delete = replace' to deletion diagnostics for variance component estimation in the linear mixed model*, J. R. Stat. Soc. B 66 (2004), pp. 131–143.
[23] F.V. Labra, R. Aoki, and F. Rojas, *An application of the local influence diagnostics to ridge regression under elliptical model*, Comm. Statist. Theory Methods 36 (2007), pp. 767–779.
[24] V. Leiva, E. Rojas, M. Galea, and A. Sanhueza, *Diagnostics in Birnbaum–Saunders accelerated life models with an application to fatigue data*, Appl. Stoch. Models Bus. Ind. 30 (2014), pp. 115–131.
[25] V. Leiva, M. Santos-Neto, F.J.A. Cysneiros, and M. Barros, *Birnbaum–Saunders statistical modelling: A new approach*, Stat. Model. 14 (2014), pp. 21–48.
[26] V. Leiva, H. Saulo, J. Leao, and C. Marchant, *A family of autoregressive conditional duration models applied to financial data*, Comput. Statist. Data Anal. 79 (2014), pp. 175–191.
[27] Y. Li and H. Yang, *A new stochastic mixed ridge estimator in linear regression model*, Statist. Papers 51 (2010), pp. 315–323.
[28] S. Liu, *On local influence for elliptical linear models*, Statist. Papers 41 (2000), pp. 211–224.
[29] S. Liu, *Local influence in multivariate elliptical linear regression models*, Linear Algebra Appl. 354 (2002), pp. 159–174.
[30] S. Liu, *On diagnostics in conditionally heteroskedastic time series models under elliptical distributions*, J. Appl. Probab. 41 (2004), pp. 393–405.
[31] S. Liu, *On influence diagnostics in multivariate regression models under elliptical distributions*, in *Contributions to Probability and Statistics: Applications and Challenges*, P. Brown, S. Liu, and D. Sharma, eds., World Scientific, Singapore, 2006, pp. 129–141.
[32] S. Liu, S.E. Ahmed, and L.Y. Ma, *Influence diagnostics in the linear regression model with linear stochastic restrictions*, Pakistan J. Statist. 25 (2009), pp. 647–662.
[33] S. Liu and C.C. Heyde, *On estimation in conditional heteroskedastic time series models under non-normal distributions*, Statist. Papers 49 (2008), pp. 455–469.
[34] S. Liu, C.C. Heyde, and W.-K. Wong, *Moment matrices in conditional heteroskedastic models under elliptical distributions with applications in AR-ARCH models*, Statist. Papers 52 (2011), pp. 621–632.
[35] S. Liu and H. Neudecker, *Local sensitivity of the restricted least squares estimator in the linear model*, Statist. Papers 48 (2007), p. 525.
[36] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester, 1999.
[37] S. Najariana, M. Arashia, and G.B.M. Kibria, *A simulation study on some restricted ridge regression estimators*, Comm. Statist. Simulation Comput. 42 (2013), pp. 871–890.
[38] H. Neudecker, S. Liu, and W. Polasek, *The Hadamard product and some of its applications in statistics*, Statistics 26 (1995), pp. 365–373.
[39] G.A. Paula, *Assessing local influence in restricted regression models*, Comput. Statist. Data Anal. 16 (1993), pp. 63–79.
[40] G.A. Paula and F.J.A. Cysneiros, *Local influence under parameter constraints*, Comm. Statist. Theory Methods 39 (2010), pp. 1212–1228.
[41] G.A. Paula, V. Leiva, M. Barros, and S. Liu, *Robust statistical modeling using the Birnbaum–Saunders-t distribution applied to insurance*, Appl. Stoch. Models Bus. Ind. 28 (2012), pp. 16–34.
[42] R Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. Available at www.R-project.org.
[43] C.R. Rao, H. Toutenburg, and S.C. Heumann, *Linear Models and Generalizations*, Springer, Berlin, 2008.
[44] M. Riquelme, V. Leiva, M. Galea, and A. Sanhueza, *Influence diagnostics on the coefficient of variation of elliptically contoured distributions*, J. Appl. Stat. 38 (2011), pp. 513–532.
[45] L. Santana, F. Vilca, and V. Leiva, *Influence analysis in skew-Birnbaum–Saunders regression models and applications*, J. Appl. Stat. 38 (2011), pp. 1633–1649.
[46] B. Schaffrin and H. Toutenburg, *Weighted mixed regression*, ZAMM J. Appl. Math. Mech/Z. Angew. Math. Mech. 70 (1990), pp. 735–738.
[47] L. Shi, *Local influence in principal components analysis*, Biometrika 84 (1997), pp. 175–186.
[48] L. Shi and X. Wang, *Local influence in ridge regression*, Comput. Statist. Data Anal. 31 (1999), pp. 341–353.
[49] H. Theil, *On the use of incomplete prior information in regression analysis*, J. Amer. Statist. Assoc. 58 (1963), pp. 401–414.
[50] H. Theil and A.S. Goldberger, *On pure and mixed statistical estimation in economics*, Int. Econ. Rev. 2 (1961), pp. 65–78.
[51] H. Theil and A.L. Nagar, *Testing the independence of regression disturbances*, J. Amer. Statist. Assoc. 56 (1961), pp. 793–806.

[52] C.L. Tsai, *Discussion of assessment of local influence by RD Cook*, J. R. Stat. Soc. B 48 (1986), p. 165.

[53] L.H. Vanegas and F.J.A. Cysneiros, *Assessment of diagnostic procedures in symmetrical nonlinear regression models*, Comput. Statist. Data Anal. 54 (2010), pp. 1002–1016.

[54] L.H. Vanegas, L.M. Rondon, and F.J.A. Cysneiros, *Assessing robustness of inference in symmetrical nonlinear regression models*, Comm. Statist. Theory Methods 42 (2013), pp. 1692–1711.

[55] C. Villegas, G.A. Paula, and V. Leiva, *Birnbaum–Saunders mixed models for censored reliability data analysis*, IEEE Trans. Reliab. 60 (2011), pp. 748–758.

[56] J. Xu and H. Yang, *Performances of the positive-rule Stein-type ridge estimator in a linear regression model with spherically symmetric error distributions*, Comm. Statist. Theory Methods 42 (2013), pp. 543–560.

Appendix

We detail our formulas for derivatives used in the paper. For these formulas, we recall that $G(u) = g'(u)/g(u)$, where $g'(u)$ is the derivative of the kernel $g(u)$ given in Equation (4) with respect to $u > 0$, and $G = G(u_g)$, with u_g maximizing $t(u) = u^{m/2}g(u)$. According to the model expressed in Equation (3), we also recall

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{R} \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{v} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V} \end{pmatrix},$$

where $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$, $\boldsymbol{v} = \boldsymbol{r} - \boldsymbol{R\beta}$, $Q = \boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{e}$. In the following, $\boldsymbol{Z}_{\boldsymbol{w}}$, $\boldsymbol{e}_{\boldsymbol{w}}$, $\boldsymbol{\varepsilon}_{\boldsymbol{w}}$, $\boldsymbol{\Sigma}_{\boldsymbol{w}}$, $Q_{\boldsymbol{w}}$ and $G_{\boldsymbol{w}}$ denote their adaptations for each perturbation scheme.

A.1. *Score vector*

To establish the elements $\dot{\boldsymbol{\ell}}_1$ and $\dot{\ell}_2$ of the score vector in Equation (6), we take the derivatives of $\ell(\boldsymbol{\theta}) = \ell$ given in Equation (5) with respect to $\boldsymbol{\beta}$ and ϕ obtaining, after rearranging terms,

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = -\frac{2}{\phi} G\boldsymbol{Z}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{e} = \dot{\boldsymbol{\ell}}_1(\boldsymbol{\theta}), \quad \frac{\partial \ell}{\partial \phi} = -\frac{m}{2\phi} - \frac{1}{\phi^2} GQ = \dot{\ell}_2(\boldsymbol{\theta}). \tag{A1}$$

A.2. *Hessian matrix*

To establish the elements $\dot{\boldsymbol{\ell}}_{11}$, $\dot{\boldsymbol{\ell}}_{12} = \dot{\boldsymbol{\ell}}_{21}^\top$, and $\dot{\ell}_{22}$ of the Hessian matrix in Equation (9), we take the derivatives of the score vector given in (A1) obtaining, after rearranging terms,

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^\top} = \frac{2}{\phi} G\boldsymbol{Z}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \frac{4}{\phi^2} G'\boldsymbol{Z}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{e}\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Z} = \dot{\boldsymbol{\ell}}_{11}(\boldsymbol{\theta}),$$

$$\frac{\partial^2 \ell}{\partial \phi \partial \boldsymbol{\beta}^\top} = \frac{2}{\phi^2} G\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + \frac{2}{\phi^3} G'Q\boldsymbol{e}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Z} = \dot{\boldsymbol{\ell}}_{21}(\boldsymbol{\theta}),$$

$$\frac{\partial^2 \ell}{\partial \phi^2} = \frac{m}{2\phi^2} + \frac{2}{\phi^3} GQ + \frac{1}{\phi^4} G'Q^2 = \dot{\ell}_{22}(\boldsymbol{\theta}).$$

A.3. *Local influence*

Case-weight perturbation

The Cook approach. The derivatives of $\ell_{\mathrm{c}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = \ell_{\boldsymbol{w}_{\mathrm{c}}}$ established in Equation (15) with respect to $\boldsymbol{\beta}$ and ϕ are, after rearranging terms, given by

$$\frac{\partial \ell_{\boldsymbol{w}_{\mathrm{c}}}}{\partial \boldsymbol{\beta}} = -\frac{2}{\phi} G_{\boldsymbol{w}_{\mathrm{c}}} \boldsymbol{Z}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}\boldsymbol{e}, \quad \frac{\partial \ell_{\boldsymbol{w}_{\mathrm{c}}}}{\partial \phi} = -\frac{m}{2\phi} - \frac{1}{\phi^2} G_{\boldsymbol{w}_{\mathrm{c}}} Q_{\boldsymbol{w}_{\mathrm{c}}}. \tag{A2}$$

The derivatives of Equation (A2) with respect to $\boldsymbol{w}$ are given by

$$\frac{\partial^2 \ell_{\boldsymbol{w}_{\mathrm{c}}}}{\partial \boldsymbol{\beta}\partial \boldsymbol{w}^\top} = -\frac{2}{\phi^2} G'_{\boldsymbol{w}_{\mathrm{c}}} \boldsymbol{Z}^\top \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}\boldsymbol{e}[\boldsymbol{\varepsilon}^\top \otimes \boldsymbol{\varepsilon}^\top]\boldsymbol{J} - \frac{2}{\phi} G_{\boldsymbol{w}_{\mathrm{c}}}[\boldsymbol{\varepsilon}^\top \otimes \boldsymbol{X}^\top]\boldsymbol{J} = \boldsymbol{\Delta}_{1_{\mathrm{c}}}(\boldsymbol{\theta}, \boldsymbol{w}),$$

$$\frac{\partial^2 \ell_{\boldsymbol{w}_{\mathrm{c}}}}{\partial \phi \partial \boldsymbol{w}^\top} = -\frac{1}{\phi^3} G'_{\boldsymbol{w}_{\mathrm{c}}} Q_{\boldsymbol{w}_{\mathrm{c}}}[\boldsymbol{\varepsilon}^\top \otimes \boldsymbol{\varepsilon}^\top]\boldsymbol{J} - \frac{1}{\phi^2} G_{\boldsymbol{w}_{\mathrm{c}}}[\boldsymbol{\varepsilon}^\top \otimes \boldsymbol{\varepsilon}^\top]\boldsymbol{J} = \boldsymbol{\Delta}_{2_{\mathrm{c}}}(\boldsymbol{\theta}, \boldsymbol{w}).$$

Alternative approach. The derivative of $\ell_{\mathrm{c}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = \ell_{\boldsymbol{w}_{\mathrm{c}}}$ established in Equation (15) with respect to $\boldsymbol{w}$ is given by

$$\frac{\partial \ell_{\boldsymbol{w}_{\mathrm{r}}}}{\partial \boldsymbol{w}^{\top}} = \frac{1}{2}\mathrm{vec}^{\top}(\boldsymbol{W}^{-1})\boldsymbol{J} + \frac{1}{\phi}G_{\boldsymbol{w}_{\mathrm{c}}}[\boldsymbol{\varepsilon}^{\top} \otimes \boldsymbol{\varepsilon}^{\top}]\boldsymbol{J} = \boldsymbol{h}_{\mathrm{c}}(\boldsymbol{\theta}, \boldsymbol{w}).$$

Perturbation of the response

The Cook approach. The derivatives of $\ell_{\mathrm{r}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = \ell_{\boldsymbol{w}_{\mathrm{r}}}$ established in Equation (16) with respect to $\boldsymbol{\beta}$ and ϕ are, after rearranging terms, given by

$$\frac{\partial \ell_{\boldsymbol{w}_{\mathrm{r}}}}{\partial \boldsymbol{\beta}} = -\frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{Z}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{\boldsymbol{w}_{\mathrm{r}}}, \quad \frac{\partial \ell_{\boldsymbol{w}_{\mathrm{r}}}}{\partial \phi} = -\frac{m}{2\phi} - \frac{1}{\phi^2}G_{\boldsymbol{w}_{\mathrm{r}}}Q_{\boldsymbol{w}_{\mathrm{r}}}. \tag{A3}$$

The derivatives of Equation (A3) with respect to $\boldsymbol{w}$ are given by

$$\frac{\partial^2 \ell_{\boldsymbol{w}_{\mathrm{r}}}}{\partial \boldsymbol{\beta}\partial \boldsymbol{w}^{\top}} = -\frac{4}{\phi^2}G'_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{Z}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{r}}}^{\top} - \frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{X}^{\top} = \boldsymbol{\Delta}_{1_{\mathrm{r}}}(\boldsymbol{\theta}, \boldsymbol{w}),$$

$$\frac{\partial^2 \ell_{\boldsymbol{w}_{\mathrm{r}}}}{\partial \phi\partial \boldsymbol{w}^{\top}} = -\frac{2}{\phi^3}G'_{\boldsymbol{w}_{\mathrm{r}}}Q_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{r}}}^{\top} - \frac{2}{\phi^2}G_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{r}}}^{\top} = \boldsymbol{\Delta}_{2_{\mathrm{r}}}(\boldsymbol{\theta}, \boldsymbol{w}),$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{r}}} = \boldsymbol{y} + \boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta}$.

Alternative approach. The derivative of $\ell_{\mathrm{r}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = \ell_{\boldsymbol{w}_{\mathrm{r}}}$ established in Equation (16) with respect to $\boldsymbol{w}$ is given by

$$\frac{\partial \ell_{\boldsymbol{w}_{\mathrm{r}}}}{\partial \boldsymbol{w}^{\top}} = \frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{r}}}\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{r}}}^{\top} = \boldsymbol{h}_{\mathrm{r}}(\boldsymbol{\theta}, \boldsymbol{w}).$$

Perturbation of a covariable

The Cook approach. The derivatives of $\ell_{\mathrm{e}}(\boldsymbol{\theta} \mid \boldsymbol{w}) = \ell_{\boldsymbol{w}_{\mathrm{e}}}$ established in Equation (17) with respect to $\boldsymbol{\beta}$ and ϕ are, after rearranging terms, given by

$$\frac{\partial \ell_{\boldsymbol{w}_{\mathrm{e}}}}{\partial \boldsymbol{\beta}} = -\frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{e}}}\boldsymbol{Z}_{\boldsymbol{w}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{\boldsymbol{w}_{\mathrm{e}}}, \quad \frac{\partial \ell_{\boldsymbol{w}_{\mathrm{e}}}}{\partial \phi} = -\frac{m}{2\phi} - \frac{1}{\phi^2}G_{\boldsymbol{w}_{\mathrm{e}}}Q_{\boldsymbol{w}_{\mathrm{e}}}. \tag{A4}$$

The derivatives of Equation (A4) with respect to $\boldsymbol{w}$ are given by

$$\frac{\partial^2 \ell_{\boldsymbol{w}_{\mathrm{e}}}}{\partial \boldsymbol{\beta}\partial \boldsymbol{w}^{\top}} = \frac{4}{\phi^2}G'_{\boldsymbol{w}_{\mathrm{e}}}\boldsymbol{Z}_{\boldsymbol{w}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{\boldsymbol{w}_{\mathrm{e}}}[\boldsymbol{\beta}^{\top} \otimes \boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{e}}}^{\top}] - \frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{e}}}[\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{e}}}^{\top} \otimes \boldsymbol{I}_p]\boldsymbol{K} + \frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{e}}}\boldsymbol{X}_{\boldsymbol{w}}^{\top}[\boldsymbol{\beta}^{\top} \otimes \boldsymbol{I}_n] = \boldsymbol{\Delta}_{1_{\mathrm{e}}}(\boldsymbol{\theta}, \boldsymbol{w}),$$

$$\frac{\partial^2 \ell_{\boldsymbol{w}_{\mathrm{e}}}}{\partial \phi\partial \boldsymbol{w}^{\top}} = \frac{2}{\phi^3}G'_{\boldsymbol{w}_{\mathrm{e}}}Q_{\boldsymbol{w}_{\mathrm{e}}}[\boldsymbol{\beta}^{\top} \otimes \boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{e}}}^{\top}] + \frac{2}{\phi^2}G_{\boldsymbol{w}_{\mathrm{e}}}[\boldsymbol{\beta}^{\top} \otimes \boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{e}}}^{\top}] = \boldsymbol{\Delta}_{2_{\mathrm{e}}}(\boldsymbol{\theta}, \boldsymbol{w}),$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{e}}} = \boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{w}}\boldsymbol{\beta}$, with $\boldsymbol{X}_{\boldsymbol{w}} = \boldsymbol{X} + \boldsymbol{W}$.

Alternative approach. The derivative of $\ell(\boldsymbol{\theta} \mid \boldsymbol{w}) = \ell_{\boldsymbol{w}}$ established in Equation (17) with respect to $\boldsymbol{w}$ is given by

$$\frac{\partial \ell_{\boldsymbol{w}_{\mathrm{e}}}}{\partial \boldsymbol{w}^{\top}} = -\frac{2}{\phi}G_{\boldsymbol{w}_{\mathrm{e}}}[\boldsymbol{\beta}^{\top} \otimes \boldsymbol{\varepsilon}_{\boldsymbol{w}_{\mathrm{e}}}^{\top}] = \boldsymbol{h}_{\mathrm{e}}(\boldsymbol{\theta}, \boldsymbol{w}).$$