The gradient test statistic for outlier detection in generalized estimating equations

Felipe Osorio[a,*], Ángelo Gárate[b], Cibele M. Russo[c]

[a] *Departamento de Matemática, Universidad Técnica Federico Santa Maria, Chile*
[b] *Departamento de Estadística, Pontificia Universidad Católica de Chile*
[c] *Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. Brazil*

Abstract

We develop diagnostic tools for estimating equations, useful for the analysis of data with longitudinal structure. The gradient statistic introduced by Terrell [Comp. Sci.Stat. 34: 206-215, 2002] is used to propose a case deletion measure, as well as a statistic for the detection of outlying observations using a mean-shift outlier model. The proposed methodology is illustrated with an example.

Keywords: Cook's distance, Gradient statistic, Mean shift outlier model

1. Introduction

The assessment of influence on the parameter estimates in statistical models has been an important concern of various researchers in the last decades. Currently, there is considerable interest in developing diagnostic measures for general frameworks. The case-deletion methodology, which consists of studying the impact on the parameter estimates after dropping individual observations, is probably the most employed technique to detect influential or outlying observations. A related formulation that allows the extension of some of these ideas to more general settings is the mean-shift outlier model (Cook and Weisberg, 1982; Wei and Shih, 1994). Some few authors have developed measures to carry out influence diagnostic in generalized estimating equations (GEE) mainly using the case-deletion approach (Preisser and Qaqish, 1996; Preisser and Perin, 2007; Venezuela et al., 2007; Preisser et al., 2008). This work considers two approaches to outlier detection in estimating functions which have not been completely explored. Specifically, we use the gradient-type statistic introduced by Lemonte (2016) (see also Lemonte, 2013) to assess the presence of outlying observation by applying a mean-shift outlier model and by extending case deletion measure proposed for Enea and Plaia (2017). One of the main

[*] Corresponding author.

Email addresses: `felipe.osorios@usm.cl` (Felipe Osorio), `afgarate@uc.cl` (Ángelo Gárate), `cibele@icmc.usp.br` (Cibele M. Russo)

[1] Address for correspondence: Departamento de Matemática, Universidad Técnica Federico Santa Maria, Avenida España 1680, Valparaíso, Chile.

[2] ORCID iD: 0000-0002-4675-5201 (Felipe Osorio), 0000-0003-1356-0245 (Cibele M. Russo)

[3] This preprint has not undergone improvements or corrections. The Version of record of this article is published in *Statistics & Probability Letters*, and is available online at, doi: `10.1016/j.spl.2024.110087`

advantages of the test statistic proposed by Terrell (2002) is that it is defined in terms of a bilinear form, requires little computational effort and is asymptotically equivalent to the score statistic. From the diagnostic perspective, it is important to have several alternatives to assess the influence of observations and to develop relevant diagnostics measures, which can identify observations that might otherwise go unnoticed.

The remainder of the paper unfolds as follows. In Section 2 we review the definition of gradient-type statistic for hypothesis testing in the framework of estimation functions and present the generalized estimating equation method for the analysis of longitudinal data. Diagnostic measures by using case elimination techniques and the mean-shift outlier model are discussed in Section 3. The methodology is illustrated in Section 4 considering a dataset previously analyzed using influence diagnostic procedures and robust methods. Some concluding remarks are given in Section 5. Supplementary Material includes details about the gradient statistic for inference functions, proofs of the theoretical results and the analysis of an additional real dataset.

2. Backgroud

Let $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \dots, \boldsymbol{Y}_n^\top)^\top$ denote the data vector such that $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ are independent for $i = 1, \dots, n$, and consider the class of additive inference functions, defined as:

$$\boldsymbol{\Psi}_n(\boldsymbol{\theta}; \boldsymbol{Y}) = \sum_{i=1}^n \boldsymbol{\Psi}_i(\boldsymbol{\theta}; \boldsymbol{Y}_i), \tag{1}$$

where $\boldsymbol{\Psi}_i : \Theta \to \mathbb{R}^p$ are independent functions with $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. Next, we will remove the dependence on $\boldsymbol{Y}$ in our notation. We also assume that the conditions established by Yuan and Jennrich (1998) are satisfied, which allows us to guarantee that the sequence of roots $\{\widehat{\boldsymbol{\theta}}_n\}_{n\geq 1}$ of the estimation equation $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \boldsymbol{0}$ is consistent and asymptotically normal. That is, it is assumed that there is a unique $\boldsymbol{\theta}_0 \in \Theta$ such that $\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$. Moreover, $n^{-1/2}\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0) \xrightarrow{D} \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{V}(\boldsymbol{\theta}_0))$, and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{G}^{-1}(\boldsymbol{\theta}_0))$, where $\boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{S}^\top(\boldsymbol{\theta})\boldsymbol{V}^{-1}(\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{\theta})$ denotes the Godambe information matrix, with $\boldsymbol{V}(\boldsymbol{\theta}) = \mathsf{E}\{\boldsymbol{\Psi}_n(\boldsymbol{\theta})\boldsymbol{\Psi}_n^\top(\boldsymbol{\theta})\}$ and $\boldsymbol{S}(\boldsymbol{\theta}) = \mathsf{E}\{-\partial\boldsymbol{\Psi}_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top\}$ being the variability and sensitivity matrices, respectively. Thus, the gradient-type statistic for testing hypotheses such as $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ in the context of inference functions (see Lemonte, 2016, Chapter 5), assumes the form:

$$T_n = \boldsymbol{\Psi}_n^\top(\boldsymbol{\theta}_0)\boldsymbol{V}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{S}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \tag{2}$$

which asymptotically follows a chi-squared distribution with p degrees of freedom under the null hypothesis.

Liang and Zeger (1986) proposed a marginal approach to model repeated measurements with longitudinal structure. Specifically, they considered that the expectation $\mathsf{E}(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i$ is related to p explanatory variables through the linear predictor $\boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta}$ with $\boldsymbol{X}_i$ being an $n_i \times p$ model matrix and $\boldsymbol{\eta}_i = g(\boldsymbol{\mu}_i)$, for some monotone and continuously differentiable link function $g(\cdot)$. It is assumed that the first two moments of the marginal distribution are given by $\mathsf{E}(Y_{ij}) = \mu_{ij}$, and $\mathsf{var}(Y_{ij}) = \phi^{-1}V(\mu_{ij})$, for $i = 1, \dots, n$; $j = 1, \dots, n_i$, where $V(\mu)$ is the variance function and ϕ is a scale parameter. The estimation of the coefficients vector $\boldsymbol{\beta} \in \mathbb{R}^p$ is carried out by solving the equation

$$\boldsymbol{\Psi}_n(\boldsymbol{\beta}) = \phi \sum_{i=1}^n \Big(\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}^\top}\Big)^\top \{\boldsymbol{A}_i^{1/2}\boldsymbol{R}_i(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}\}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}, \tag{3}$$

The local model matrix $\boldsymbol{F}_i(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}^\top$ can be written as $\boldsymbol{F}_i = \boldsymbol{D}_i^{-1}\boldsymbol{X}_i$, with $\boldsymbol{D}_i = \partial\boldsymbol{\eta}_i/\partial\boldsymbol{\mu}_i^\top$. Let $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) = \phi^{-1}\boldsymbol{A}_i^{1/2}\boldsymbol{R}_i(\boldsymbol{\alpha})\boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \operatorname{diag}(V(\mu_{i1}),\dots,V(\mu_{in_i}))$, and $\boldsymbol{R}_i(\boldsymbol{\alpha})$ is an $n_i \times n_i$ working correlation matrix associated with the ith experimental unit whose element (j,j') represents the assumed correlation between Y_{ij} and $Y_{ij'}$ for a parameter of association $\boldsymbol{\alpha}$. Additional details regarding the gradient test and estimation in GEE are described in Appendices A and B of the Supplementary Material, respectively.

3. Diagnostic measures

Next, we describe two procedures to identify outlying observations. First, we develop diagnostic measures considering a mean-shift outlier model. Subsequently, we present an approach of case deletion using a distance based on the gradient-type statistic.

3.1. Outlier detection through the mean-shift outlier model

A general approach to detect outliers in regression models is the *mean-shift outlier model* (see Cook and Weisberg, 1982, Sec. 2.2.2). It has been demonstrated that this is equivalent to the assessment of influence by case-deletion in linear and non-linear models when the response belongs to the exponential family (Wei and Shih, 1994). Thus, an approach to identify outlying observations in GEE is the mean-shift outlier model, defined by $g(\boldsymbol{\mu}_j) = \boldsymbol{\eta}_j$, with

$$\boldsymbol{\eta}_j = \begin{cases} \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{\gamma}_i, & j = i, \\ \boldsymbol{X}_j\boldsymbol{\beta}, & j \neq i, \end{cases} \tag{4}$$

for $j = 1,\dots,n$, where $\boldsymbol{B}_i$ is an $n_i \times m_i$ known matrix and $\boldsymbol{\gamma}_i$ is an $m_i \times 1$ parameter vector. This formulation allows us to identify outlying observations from an hypothesis testing perspective. In fact, in order to detect whether the ith subject is an outlier, we can test the hypotheses

$$H_0 : \boldsymbol{\gamma}_i = \boldsymbol{0} \qquad \text{against} \qquad H_1 : \boldsymbol{\gamma}_i \neq \boldsymbol{0}. \tag{5}$$

Pardo and Hobza (2014) proposed to assess the hypotheses in (5) based on the generalized Wald and score-type statistics defined by Rotnitzky and Jewell (1990). Now, we suggest to test (5) considering the gradient-type statistic described in Lemma A.2 from the Supplementary Material with $\boldsymbol{\delta}_i = (\boldsymbol{\gamma}_i^\top, \boldsymbol{\beta}^\top)^\top$ being the parameter of interest. In this context, we have that the variability and sensitivity matrices for $\boldsymbol{\delta}_i$ are respectively given by:

$$\boldsymbol{V}(\boldsymbol{\delta}_i) = \phi^2 \begin{pmatrix} \boldsymbol{B}_i^\top\boldsymbol{W}_i\boldsymbol{\Lambda}_i\boldsymbol{W}_i\boldsymbol{B}_i & \boldsymbol{B}_i^\top\boldsymbol{W}_i\boldsymbol{\Lambda}_i\boldsymbol{W}_i\boldsymbol{X}_i \\ \boldsymbol{X}_i^\top\boldsymbol{W}_i\boldsymbol{\Lambda}_i\boldsymbol{W}_i\boldsymbol{B}_i & \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}\boldsymbol{X} \end{pmatrix},$$

$$\boldsymbol{S}(\boldsymbol{\delta}_i) = \phi \begin{pmatrix} \boldsymbol{B}_i^\top\boldsymbol{W}_i\boldsymbol{B}_i & \boldsymbol{B}_i^\top\boldsymbol{W}_i\boldsymbol{X}_i \\ \boldsymbol{X}_i^\top\boldsymbol{W}_i\boldsymbol{B}_i & \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X} \end{pmatrix},$$

where $\boldsymbol{X} = (\boldsymbol{X}_1^\top,\dots,\boldsymbol{X}_n^\top)^\top$ and $\boldsymbol{W} = \bigoplus_{i=1}^n \boldsymbol{W}_i$ being a block diagonal matrix, with $\boldsymbol{W}_i = \boldsymbol{D}_i^{-1}\boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i^{-1}(\boldsymbol{\alpha})\boldsymbol{A}_i^{-1/2}\boldsymbol{D}_i^{-1}$, and $\boldsymbol{\Lambda} = \bigoplus_{i=1}^n \boldsymbol{\Lambda}_i$ with $\boldsymbol{\Lambda}_i = \boldsymbol{D}_i \operatorname{Cov}(\boldsymbol{Y}_i)\boldsymbol{D}_i$. These expressions enables the computation of score-type and gradient-type statistics for testing the hypothesis defined in (5), respectively, as

$$R_i = \frac{1}{n}\widetilde{\phi}^{-2}\boldsymbol{\Psi}_1^\top(\widetilde{\boldsymbol{\delta}}_i)(\boldsymbol{B}_i^\top\widetilde{\boldsymbol{W}}_i\widetilde{\boldsymbol{P}}_i\widetilde{\boldsymbol{\Lambda}}_i\widetilde{\boldsymbol{W}}_i\boldsymbol{B}_i)^{-1}\boldsymbol{\Psi}_1(\widetilde{\boldsymbol{\delta}}_i),$$
$$T_i = \widetilde{\phi}^{-1}\boldsymbol{\Psi}_1^\top(\widetilde{\boldsymbol{\delta}}_i)(\boldsymbol{B}_i^\top\widetilde{\boldsymbol{W}}_i\widetilde{\boldsymbol{P}}_i\widetilde{\boldsymbol{\Lambda}}_i\widetilde{\boldsymbol{W}}_i\boldsymbol{B}_i)^{-1}\boldsymbol{B}_i^\top\widetilde{\boldsymbol{W}}_i\widetilde{\boldsymbol{P}}_i\boldsymbol{B}_i\widehat{\gamma}_i,$$

where $\boldsymbol{P}_i = \boldsymbol{I} - \boldsymbol{\Lambda}_i \boldsymbol{W}_i \boldsymbol{X}_i (\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top \boldsymbol{W}_i$ and $\boldsymbol{\Psi}_1(\boldsymbol{\delta}_i) = \phi \boldsymbol{B}_i^\top \boldsymbol{W}_i \boldsymbol{D}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$ must be evaluated at $\widetilde{\boldsymbol{\delta}}_i = (\boldsymbol{0}, \widehat{\boldsymbol{\beta}}^\top)^\top$, i.e., the estimate obtained under the null hypothesis $H_0 : \boldsymbol{\gamma}_i = \boldsymbol{0}$. Here $\widehat{\boldsymbol{\delta}}_i = (\widehat{\boldsymbol{\gamma}}_i^\top, \widehat{\boldsymbol{\beta}}_*^\top)^\top$ represent the estimates for the model defined in (4). Hence, we reject $H_0 : \boldsymbol{\gamma}_i = \boldsymbol{0}$ by comparing the values of R_i or T_i with the percentile $100(1-\alpha)\%$ of the chi-square distribution with m_i degrees of freedom. One-step approximations for $\widehat{\boldsymbol{\gamma}}_i$ and $\widehat{\boldsymbol{\beta}}_*$ are given in Proposition C.1 of the electronic supplementary material.

3.2. A case deletion measure

An interesting perspective for the detection of atypical observations based on case deletion techniques is provided by Enea and Plaia (2017), who proposed to use a gradient distance as an influence measure. Let $\boldsymbol{\Psi}_{(i)}(\boldsymbol{\theta}) = \sum_{j\neq i}^{n} \boldsymbol{\Psi}_j(\boldsymbol{\theta})$ be the inference function given in (1) when the ith function is discarded. Therefore, the gradient-type statistic defined in (2), together with the first-order Taylor approximation of $\boldsymbol{\Psi}_{(i)}(\boldsymbol{\theta})$ around $\widehat{\boldsymbol{\theta}}$, leads to the following one-step approximation:

$$\widehat{\boldsymbol{\theta}}_{(i)}^{1} = \widehat{\boldsymbol{\theta}} + \Big\{ - \frac{\partial \boldsymbol{\Psi}_{(i)}(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^\top} \Big\}^{-1} \boldsymbol{\Psi}_{(i)}(\widehat{\boldsymbol{\theta}}),$$

where $\widehat{\boldsymbol{\theta}}_{(i)}$ denotes the estimate of $\boldsymbol{\theta}$ based on $\boldsymbol{\Psi}_{(i)}(\boldsymbol{\theta})$. Following the same argument used by Jørgensen and Knudsen (2004), we can write $\widehat{\boldsymbol{\theta}}_{(i)}^{1} - \widehat{\boldsymbol{\theta}} = \boldsymbol{S}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{\Psi}_{(i)}(\widehat{\boldsymbol{\theta}})$, with $\boldsymbol{S}(\widehat{\boldsymbol{\theta}}) = \mathsf{E}\{-\partial \boldsymbol{\Psi}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top\}\big|_{\theta=\widehat{\theta}}$. Thus, pre-multiplying by $\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{S}(\widehat{\boldsymbol{\theta}})$ and noting that $\boldsymbol{\Psi}_{(i)}(\widehat{\boldsymbol{\theta}}) = -\boldsymbol{\Psi}_i(\widehat{\boldsymbol{\theta}})$ we can define the *gradient distance* as:

$$TD_i = \boldsymbol{\Psi}_i^\top(\widehat{\boldsymbol{\theta}})\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{S}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(i)}). \tag{6}$$

Evidently, these results provide the framework to develop diagnostic measures for assessing the effect of the ith subject (or cluster) on the estimate of $\boldsymbol{\beta}$ based on the estimating equation given in (3). Proposition C.1 from the Supplementary Material provides an alternative way to define a version of Cook's distance for the context of GEE, which is defined as (see Cook and Weisberg, 1982) $D_i = (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})^\top \boldsymbol{M}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})$, for $i = 1, \dots, n$, where $\boldsymbol{M}$ is a $p \times p$ semipositive definite matrix. Using Equation (C.2) of Supplementary Material and following the recommendations of Vens and Ziegler (2012), we can choose $\boldsymbol{M}$ as the inverse of the empirical estimator of covariance matrix for $\widehat{\boldsymbol{\beta}}$ (see Appendix B of electronic supplementary material), which leads us to define a one-step approximation of the Cook's distance as:

$$\begin{aligned} D_i^1 &= (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}^1)^\top \boldsymbol{X}^\top \widehat{\boldsymbol{W}}\boldsymbol{X}(\boldsymbol{X}^\top \widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \widehat{\boldsymbol{W}}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}^1) \\ &= \{\widehat{\gamma}_i^1\}^\top \boldsymbol{B}_i^\top \widehat{\boldsymbol{W}}_i \boldsymbol{X}_i (\boldsymbol{X}^\top \widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top \widehat{\boldsymbol{W}}_i \boldsymbol{B}_i \widehat{\gamma}_i^1. \end{aligned} \tag{7}$$

This provides an alternative to the influence measures proposed by Preisser and Qaqish (1999) and Venezuela et al. (2007). In fact, by choosing $\boldsymbol{M}$ as the inverse of the model-based estimator of the covariance matrix for $\widehat{\boldsymbol{\beta}}$, leads to the proposal of Venezuela et al. (2007), that is

$$\begin{aligned} D_{i,\mathsf{VBS}}^1 &= \widehat{\phi}\,(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}^1)^\top \boldsymbol{X}^\top \widehat{\boldsymbol{W}}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}^1) \\ &= \widehat{\phi}\,\{\widehat{\gamma}_i^1\}^\top \boldsymbol{B}_i^\top \widehat{\boldsymbol{W}}_i \boldsymbol{X}_i (\boldsymbol{X}^\top \widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top \widehat{\boldsymbol{W}}_i \boldsymbol{B}_i \widehat{\gamma}_i^1, \end{aligned} \tag{8}$$

which allow us to manipulate any subset of observations by using the matrix $\boldsymbol{B}_i$.

4. Example: GUIDE study data

We revisit a dataset from the Guidelines for Urinary Incontinence Discussion and Evaluation (GUIDE) study, introduced by Preisser and Qaqish (1999) and which aims to assess the impact of urinary incontinence on the lives of elderly patients over 76 years of age. The response is binary, indicating whether the individual perceives that his or her daily routine is affected by accidental urine leakage. There are 137 elderly patients from 38 medical practices (i.e. cluster). Observations are unbalanced, ranging from 1 to 8 patients per practice. Five regressors are available, gender (sex), age (age), daily leaking accidents (accidents), severity of leaking (severe) and number of times using the toilet daily (toilet). A logistic link function was considered for the marginal model as follows:

$$\operatorname{logit}(\mu_{ij}) = \beta_0 + \beta_1 \,\text{sex} + \beta_2 \,\text{age} + \beta_3 \,\text{accidents} + \beta_4 \,\text{severe} + \beta_5 \,\text{toilet}, \tag{9}$$

where μ_{ij} denotes, for jth patient in ith cluster, the probability of being bothered. In addition, the exchangeable working correlation structure is assumed. The GUIDE dataset and `R` code to perform the analysis described in this section are available at github[4]. Several papers have used the GUIDE data for robust estimation or diagnostic analysis in GEE. For example, Preisser and Qaqish (1999) and Qu and Song (2004) found that the harmful effect of patients 8, 19, 42, 44 and 88 on the parameter estimates can be mitigated by considering robust estimation procedures, whereas Preisser and Garcia (2005) and Hammill and Preisser (2006) detected patients 8, 44 and 122 and clusters 27, 41, 107 and 156 as outliers. More recently, Jung (2008) using the local influence approach studied the role of patients 8 and 44. The supplementary material contains all the tables and figures listed below.

Model given in (9) was fitted using GEE. Table D.1 presents the results of the fit for the entire dataset as well as for subsets where certain observations have been removed. It should be stressed that patients 8 and 44 produce a change in the inference associated with the sex and toilet variables, respectively, whereas observation 64 does not produce inferential changes but increases the p-value of the sex and age variables. It is interesting to note that observations 64 and 88 have a strong impact on the empirical variance estimator, while patients 8, 44 and 88 exert a large effect on the model-based variance estimator. Details on how to study the role of outliers on the covariance matrices (or equivalently the confidence ellipsoids) associated with $\widehat{\boldsymbol{\beta}}$ can be found in Pardo and Alonso (2012).

Figure D.1 (a) presents the cluster-level gradient distance TD_i, for $i = 1, \ldots, 38$. Thus, the gradient distance allows us to identify clusters 27, 41, 107, 156 and 235 as medical practices with a strong influence on the fitting results. It should be noted that patients 8, 44 and 88 are in clusters 27, 107 and 156, respectively. Figure D.1 (b) shows an interesting aspect, the one-step approximation of TD_i given in Equation (6) presents a strong agreement with the bilinear form distance obtained by deleting the ith cluster and fully iterating the fitting procedure until reach convergence.

Now, consider that we are interested on detecting which observations within a cluster are outliers. Thus, we can employ $\boldsymbol{B}_i = (0, \ldots, 1, \ldots, 0)^\top$ an $n_i \times 1$ vector with one at the jth position and zero elsewhere, in which case the null hypothesis $H_0 : \gamma_{ij} = 0$ is rejected if T_{ij} or R_{ij} exceeds the quantile $1 - \alpha$ of the chi-square distribution with one

[4] `https://github.com/faosorios/outlier_GEE`

degree of freedom. This also allows us to obtain versions of the Cook's distance using Equations (7) and (8), say D_{ij} and $D_{ij,\mathsf{VBS}}$, oriented to determine the effect of the jth observation in the ith cluster on the parameter estimates. For the GUIDE dataset, the one-step approximation D^1_{ij} (see Figure D.2 (a)) reveals observations 8, 44 and 64 as influential. To the best of our knowledge, previously patient 64 had not been detected in the GUIDE dataset, who was bothered although reporting a very low frequency of toileting and leaking accidents. This patient exerts a strong impact on the estimation of the regression coefficients, and on the empirical estimator of the covariance of the coefficients estimates. We stress that using the $D^1_{ij,\mathsf{VBS}}$ distances proposed by Venezuela et al. (2007) it is not possible to identify observation 64 as influential. Figure D.3 presents gradient-type and score-type statistics to assess outlying observations. We can note that both statistics coincide in identifying patient 86 as an outlier. Indeed, this is a female patient who is bothered by the high number of leaking accidents, while reporting low severity and toileting. In addition, observation 86 has effect on the estimation of the correlation parameter α and exerts a slight change on the estimation of $\mathrm{Cov}_{\mathsf{Emp}}(\widehat{\boldsymbol{\beta}})$. We also note that in previous works using this dataset this observation had not been identified.

5. Concluding remarks

This work provides an alternative to the diagnostic procedures proposed by Preisser and Qaqish (1996) and Venezuela et al. (2007) who used case deletion techniques, as well as to the developments reported by Wei and Fung (1999) and Pardo and Hobza (2014) who proposed methods for outlier identification using the mean-shift outlier model. Explicit expressions were obtained for the gradient-type statistic T_i (and the score-type statistic, R_i) associated with the hypothesis in Equation (5). It is worth noting that the one-step estimators developed in this work have allowed us to develop one-step approximations for the Cook's distance, proposal that maintains the simplicity of the results reported by Preisser and Qaqish (1996) and Venezuela et al. (2007). Additionally, our findings have allowed us to extend the gradient distance introduced by Enea and Plaia (2017) to the general context of inference functions with particular emphasis on GEE.

Data availability

We have added links to data/codes in the manuscript.

Acknowledgments

We would like to thank two referees for the helpful comments and suggestions which have helped to improve the presentation of the paper. This work was supported by Comisión Nacional de Investigación Científica y Tecnológica, FONDECYT grant 1140580 and UTFSM grant PI_LI_19_11. Felipe Osorio also acknowledge financial support from the regional MATH-AmSud program, grant number 20-MATH-03. Ángelo Gárate was supported by the National Scholarship Program of Chile, ANID-Chile. The authors are indebted to Agustín Uribe for providing preliminary computational support.

References

Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression.* New York: Chapman & Hall.

Enea, M., Plaia, A. (2017). A gradient-based deletion diagnostic measure for generalized linear mixed models. *Communications in Statistics - Theory and Methods* 46, 1972-1982.

Hammill, B.G., Preisser, J.S. (2006). A SAS/IML program for GEE and regression diagnostics. *Computational Statistics and Data Analysis* 51, 1197-1212.

Jørgensen, B., Knudsen, S.J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics* 31, 93-114.

Jung, K. (2008). Local influence in generalized estimating equations. *Scandinavian Journal of Statistics* 35, 286-294.

Lemonte, A.J. (2013). On the gradient statistic under model misspecification. *Statistics & Probability Letters* 380, 390-398.

Lemonte, A.J. (2016). *The Gradient Test: Another Likelihood-based Test.* London: Academic Press.

Liang, K.-Y., Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

Pardo, M.C., Alonso, R. (2012). Influence measures based on the volume of confidence ellipsoids for GEE. *Biometrical Journal* 54, 552-567.

Pardo, M.C., Hobza, T. (2014). Outlier detection method in GEEs. *Biometrical Journal* 56, 838-850.

Preisser, J.S., Garcia, D.I. (2005). Alternative computational formulae for generalized linear model diagnostics: Identifying influential observations with SAS software. *Computational Statistics & Data Analysis* 48, 755-764.

Preisser, J.S., Perin, J. (2007). Deletion diagnostics for marginal mean and correlation model parameters in estimating equations. *Statistics and Computing* 17, 381-393.

Preisser, J.S., Qaqish, B.F. (1996). Deletion diagnostics for generalized estimating equations. *Biometrika* 83, 551-562.

Preisser, J.S., Qaqish, B.F. (1999). Robust regression for clustered data with application to binary responses. *Biometrics* 55, 574-579.

Preisser, J.S., Qaqish, B.F., Perin, J. (2008). A note on deletion diagnostics for estimating equations. *Biometrika* 95, 509-513.

Qu, A., Song, P. (2004). Assessing robustness of generalised estimating equations and quadratic inference functions. *Biometrika* 91, 447-459.

Rotnitzky, A., Jewell, N.P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for clustered data. *Biometrika* 77, 485-497.

Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics* 34, 206-215.

Venezuela, M.K. Botter, D.A., Sandoval, M.C. (2007). Diagnostic techniques in generalized estimating equations. *Journal of Statistical Computation and Simulation* 77, 879-888.

Vens, M., Ziegler, A. (2012). Generalized estimating equations and regression diagnostics for longitudinal controlled clinical trials: A case study. *Computational Statistics & Data Analysis* 56, 1232-1242.

Wei, W.H., Fung, W.K. (1999). The mean-shift outlier model in general weighted regression and its applications. *Computational Statistics & Data Analysis* 30, 429-441.

Wei, B.C., Shih, J.Q. (1994). On statistical models for regression diagnostics. *Annals of the Institute of Statistical Mathematics* 46, 267-278.

Yuan, K.H., Jennrich, R.I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis* 65, 245-260.