Influence diagnostics for ridge regression using the Kullback-Leibler divergence

Alonso Ogueda[1] and Felipe Osorio[*]

[1]Department of Mathematical Sciences, George Mason University, 4400 University Dr. Fairfax, VA 22030, USA.

*Corresponding author(s). E-mail(s): faosorios.stat@gmail.com;
Contributing authors: aogueda@gmu.edu;

Abstract

The identification of anomalous observations provides insight into which aspects of the modeling process may be vulnerable. Thus, appropriate diagnostic measures can be developed to prevent certain types of outlying observations from going undetected. This paper proposes an approach to assess the influence diagnostics in ridge regression based on the Kullback-Leibler divergence. To quantify the impact of observations on the ridge estimator two main procedures are explored. Namely, a case-deletion method and the local influence technique considering several perturbation schemes. We provide tractable expressions to assessing the influence of individual observations as well as the derivatives required to characterize the local curvature. The developed measures correspond to a combination of the leverages and the volume of the confidence ellipsoid, which allows an interesting characterization of the detected observations. To evaluate the performance of the proposed methodology, we consider the analysis of two real datasets and performed a comparison with several methods for outlier detection and assessing influence in ridge regression. In such numerical examples, the proposed measures are successful in identifying observations that are not detected by the traditional techniques.*

Keywords: Collinearity, Influence diagnostics, Kullback-Leibler divergence, Regression diagnostics, Ridge estimator

MSC Classification: 62J07 , 62J20 , 94A17

---

*This preprint has not undergone improvements or corrections. The version of record of this article is published in *Statistical Papers*, and is available online at https://doi.org/10.1007/s00362-025-01701-1

1 Introduction

A biased estimation procedure that has been quite successful to overcome the effects of collinearity in regression corresponds to ridge regression (Hoerl and Kennard, 1970). However, a number of papers have shown that the ridge estimator is extremely sensitive to the presence of extreme observations (see for instance Walker and Birch, 1988; Billor and Loynes, 1999; Shi and Wang, 1999; Emami and Emami, 2016). The techniques used for the development of procedures for outliers detection and assessing the influence of observations in ridge regression have been quite diverse. For instance, diagnostic measures based on case-deletion procedures were proposed by Walker and Birch (1988) and more recently in Emami and Emami (2016), whereas the properties of statistical leverage were studied by Steece (1986). On the other hand, Billor and Loynes (1999) and Shi and Wang (1999) developed the local influence in ridge regression based on a pseudo-likelihood for an augmented model and considering the approach of generalized influence function proposed by Shi (1997), respectively. One aspect that should be emphasized is that techniques based on case-deletion, or global influence, have been criticized because they tend to suffer from masking and swamping effects. This problem arises when it is desired to evaluate the joint effect of multiple observations. Indeed, masking has been described as the effect that occurs when groups of observations may not be identified as outliers due to the presence of observations that are individually extreme, while the swamping effect appears when observations are incorrectly labeled as outliers (Meloun and Militký, 2001; Chatterjee and Hadi, 1988). For a formalization of these concepts, see Davies and Gather (1993). This has led to the development of procedures where, instead of eliminating observations, they concentrate on the perturbation of observations or of certain relevant aspects of the model and evaluate their influence on some statistic of interest (see, for instance Pregibon, 1981). This diagnostic procedure was introduced by Cook (1986), is known as the local influence method and has proven to be an extremely flexible tool for determining influential observations that has gained considerable popularity in the statistical literature.

The literature on the use of divergence measures as an input for influence diagnostics is quite limited. To the best of our knowledge, Johnson and Geisser (1983) and Johnson (1985) were the first to use the Kullback-Leibler divergence to propose measures of influence in linear regression and logistic regression, respectively. Whereas, Geisser (1996) in the discussion of the paper by Cook (1986), was who first proposed to consider the Kullback-Leibler divergence to perform the local influence diagnostics from a Bayesian perspective, an idea that was later formalized and applied in linear regression by Shi and Wei (1995). This approach has been successfully used for the assessment of Bayesian local influence in growth curve models by Pan et al. (1996, 1999) and Pan and Fung (2000). More recently, García-Heras et al. (2006) and Muñoz-García et al. (2006) have considered a general class of divergence measures to propose influence measures based on case-deletion.

The main aim of this paper is to develop the influence diagnostics considering the case-deletion technique as well as the local influence procedure based on the Kullback-Leibler divergence which is a diagnostic measure that incorporates two aspects, i.e., the prediction matrix, and the volume of the confidence ellipsoid for the ridge regression framework. Thus, the proposed methodology complements traditional diagnostic

procedures. Additionally, such developments may lead to interesting extensions of the proposed technique to more complex models such as the one introduced in Emami (2018) or to carry out diagnostic analyses in elliptically contoured regression models (Liu, 2000; Galea et al., 2003). In Section 2 we describe some case-deletion diagnostics in ridge regression and present details of the local influence procedure. Section 3 is devoted to provide the main results associated with the Kullback-Leibler divergence-based influence diagnostics. Section 4 presents the analysis of two real datasets commonly used in the literature and reports the results of a simulation study to evaluate the performance of the proposed methodology on finite samples. Finally, Section 5 presents concluding remarks and perspectives for future work. For comparison, the local influence based on the penalized likelihood displacement is presented in the Appendix. Additionally, we examine whether each perturbation scheme is appropriate in the sense outlined by Zhu et al. (2007).

2 Background and definition

Consider the linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{Y}$ is an $n\times 1$ vector of observations, $\boldsymbol{X}$ is an $n\times p$ model matrix with $\mathrm{rk}(\boldsymbol{X}) = p$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random disturbances following a multivariate normal distribution $\mathsf{N}_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$, where $\boldsymbol{I}_n$ denotes the identity matrix of dimension n. In presence of collinearity, the least squares (LS) estimator $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$ becomes unstable (see for instance Stewart, 1987; Belsley, 1991) and several alternatives have been proposed in the literature to reduce its harmful effects. A useful procedure to combat collinearity is the ridge estimator (Hoerl and Kennard, 1970) which is defined as

$$\widehat{\boldsymbol{\beta}}_\lambda = (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{Y}, \qquad \lambda > 0, \tag{2}$$

where λ is the shrinkage parameter, also known as ridge parameter. For the appropriate selection of λ, Golub et al. (1979) suggested the generalized cross-validation (GCV) criterion, which consists in the minimization of the objective function

$$\mathrm{GCV}(\lambda) = \frac{1}{n}\frac{\sum_{i=1}^n (Y_i - \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_\lambda)^2}{\{\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{H}(\lambda))/n\}^2} = \frac{\|(\boldsymbol{I}_n - \boldsymbol{H}(\lambda))\boldsymbol{Y}\|^2/n}{\{\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{H}(\lambda))/n\}^2},$$

where $\boldsymbol{H}(\lambda) = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top$ is called prediction matrix, and $\widehat{Y}_i(\lambda) = \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}_\lambda$ denotes the ith predicted value, for $i = 1,\ldots,n$. It is straightforward to note that the vector of predicted values $\widehat{\boldsymbol{Y}}_\lambda = (\widehat{Y}_1(\lambda),\ldots,\widehat{Y}_n(\lambda))^\top$ assumes the form $\widehat{\boldsymbol{Y}}_\lambda = \boldsymbol{H}(\lambda)\boldsymbol{Y}$. Several techniques exist in literature to estimate the shrinkage parameter λ. For examples, see Hoerl et al. (1975); Kibria (2022), among others.

2.1 Influence diagnostics

The pioneering work of Cook (1977) has led to the development of measures for influence diagnostics in various extensions of the linear regression model. The technique relies on identifying influential observations by studying their effect on some key aspects of the model once such observations are removed from the dataset. In fact, based on the empirical influence curve, Cook and Weisberg (1980) introduced a family of measures of influence in regression models. The Cook's distance for $\widehat{\boldsymbol{\beta}}$ is defined as

$$D_i(\boldsymbol{M}, c) = \frac{(\mathrm{SIC}_i)^\top \boldsymbol{M} (\mathrm{SIC}_i)}{c},$$

where $\mathrm{SIC}_i = (n-1)(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}(i))$, with $\widehat{\boldsymbol{\beta}}(i)$ denoting the least squares estimator of $\boldsymbol{\beta}$ after the ith case is deleted, $\boldsymbol{M}$ is a $p \times p$ positive semidefinite matrix, and $c > 0$ is a scalar. For ridge regression, Walker and Birch (1988) proposed two versions for the Cook distance,

$$D_i(\boldsymbol{M}, c) = \frac{(\widehat{\boldsymbol{\beta}}_\lambda - \widehat{\boldsymbol{\beta}}_\lambda(i))^\top \boldsymbol{M} (\widehat{\boldsymbol{\beta}}_\lambda - \widehat{\boldsymbol{\beta}}_\lambda(i))}{c},$$

say D_i^* and D_i^{**} using, $\boldsymbol{M} = \boldsymbol{X}^\top \boldsymbol{X}$, $\boldsymbol{M} = (\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I})(\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I})$, and $c = ps^2$, respectively, where $s^2 = \|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2/(n-p)$.

Within the likelihood framework, Cook (1986) proposed to assess the influence of extreme observations on the maximum likelihood estimates by considering the curvature of the likelihood displacement,

$$LD(\boldsymbol{\omega}) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))\},$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ denote the maximum likelihood estimates based on the postulated and perturbated models, which are defined as $\mathcal{P} = \{g(\boldsymbol{x}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ and,

$$\mathcal{P}_\omega = \{g(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\omega}) : \boldsymbol{\theta} \in \Theta, \boldsymbol{\omega} \in \Omega\},$$

respectively, with $\boldsymbol{\omega}$ being a q-dimensional perturbation vector that is restricted to some open subset $\Omega \subset \mathbb{R}^q$. In addition, it is assumed that there is a null perturbation, $\boldsymbol{\omega}_0$, satisfying $\mathcal{P}_{\omega_0} = \mathcal{P}$. In general, we may be interested in assessing the influence on objective functions other than the likelihood displacement, for additional details see Wu and Luo (1993). Let $f(\boldsymbol{\omega})$ be a measure of influence. Thus, the main aim of the local influence approach is to analyze the curvature of the curves passing through the influence surface $\varphi(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f(\boldsymbol{\omega}))^\top$ at the critical point $\boldsymbol{\omega}_0$. The idea is to find the direction associated to the largest normal curvature. This direction may evidence those observations that have considerable influence on the objective function under small perturbations on the postulated model and/or the data.

Consider $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + \varepsilon \boldsymbol{h}$, where $\boldsymbol{h}$ is a unitary direction ($\|\boldsymbol{h}\| = 1$) and $\varepsilon \in \mathbb{R}$. As discussed in Cook (1986) the local behavior of the influence function $f(\boldsymbol{\omega})$ around

$\varepsilon = 0$ for a direction $\boldsymbol{h}$ can be characterized by the normal curvature,

$$C_{f,h} = \frac{\boldsymbol{h}^\top \boldsymbol{F}_f \boldsymbol{h}}{(1 + \nabla_f^\top \nabla_f)\boldsymbol{h}^\top (\boldsymbol{I} + \nabla_f \nabla_f^\top)\boldsymbol{h}}, \tag{3}$$

where $\nabla_f = \partial f(\boldsymbol{\omega})/\partial \boldsymbol{\omega}\big|_{\omega=\omega_0}$ and $\boldsymbol{F}_f = \partial^2 f(\boldsymbol{\omega})/\partial \boldsymbol{\omega}\partial \boldsymbol{\omega}^\top\big|_{\omega=\omega_0}$. It is well known that $C_{f,h}$ is not invariant under uniform changes in scale (see Fung and Kwan, 1997). Thus, Poon and Poon (1999) proposed the conformal normal curvature, which is a scale-invariant influence measure and is given by

$$B_{f,h} = \frac{\boldsymbol{h}^\top \boldsymbol{F}_f \boldsymbol{h}}{\|\boldsymbol{F}_f\|_M \boldsymbol{h}^\top (\boldsymbol{I} + \nabla_f \nabla_f^\top)\boldsymbol{h}}, \tag{4}$$

where $\|\cdot\|_M$ denotes some matrix norm such as $\|\boldsymbol{F}_f\|_M = (\mathrm{tr}(\boldsymbol{F}_f^\top \boldsymbol{F}_f))^{1/2}$. An interesting property of the conformal curvature is that $0 \leq |B_{f,h}| \leq 1$. According to matrix theory, the local maximum curvature and the corresponding directions are associated with the generalized eigenvalue-eigenvector solution of the equation

$$\|\boldsymbol{F}_f - \lambda \boldsymbol{K}_f\| = 0, \tag{5}$$

where $\boldsymbol{K}_f$ is defined as $(1 + \nabla_f^\top \nabla_f)(\boldsymbol{I} + \nabla_f \nabla_f^\top)$ or $\|\boldsymbol{F}_f\|_M(\boldsymbol{I} + \nabla_f \nabla_f^\top)$ for the normal or conformal curvature, respectively. The direction of maximum curvature $\boldsymbol{h}_{\max}$ is determinated by the eigenvector associated with the largest eigenvalue associated with the solution of (5). Such direction is used to identify which observations are locally influential. It should be noted that, important simplifications in the computation of the curvature matrices described above occur when $\nabla_f = \boldsymbol{0}$. This holds, for example, when the influence function $f(\boldsymbol{\omega})$ is the likelihood displacement $LD(\boldsymbol{\omega})$.

An alternative approach to perform the local influence diagnostics was proposed by Shi (1997) who generalized the local influence defined in Cook (1986) by considering the generalized influence function

$$\mathrm{GIF}(\boldsymbol{T}, \boldsymbol{h}) = \lim_{\varepsilon \to 0} \frac{\boldsymbol{T}(\boldsymbol{\omega}_0 + \varepsilon \boldsymbol{h}) - \boldsymbol{T}(\boldsymbol{\omega}_0)}{\varepsilon},$$

where $\boldsymbol{T} \in \mathbb{R}^p$ represents some statistic of interest. To determine the effect of a small perturbation on $\boldsymbol{T}$, Shi (1997) suggested using the generalized Cook distance, defined as

$$GD_{T,h} = \frac{\{\mathrm{GIF}(\boldsymbol{T}, \boldsymbol{h})\}^\top \boldsymbol{M} \{\mathrm{GIF}(\boldsymbol{T}, \boldsymbol{h})\}}{c},$$

where $\boldsymbol{M}$ is a $p \times p$ positive semidefinite matrix, and $c > 0$ is a scalar. As with the curvatures in (3) or (4) that direction related to the largest local change in $\boldsymbol{T}$ can be used as a diagnostic tool. This approach has been applied by Shi and Wang (1999) to assess the local influence in ridge regression using the ridge estimator as the statistic of interest. The main motivation for its study comes from the fact that this technique allows studying the sensitivity of several aspects of the model under different perturbation schemes.

3 Kullback-Leibler based influence measures

Consider that $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are two estimators of $\boldsymbol{\beta}$. We can measure the discrepancy between $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ using the Kullback-Leibler divergence between their associated density functions. Indeed, let g_1 and g_2 be the densities of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, respectively. Then, the Kullback-Leibler divergence (also known as relative entropy) between $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ is given by

$$I(\boldsymbol{T}_1 : \boldsymbol{T}_2) = \int \log\left(\frac{g_1(\boldsymbol{u})}{g_2(\boldsymbol{u})}\right) \mathrm{d}\, g_1(\boldsymbol{u}).$$

Note that $I(\boldsymbol{T}_1 : \boldsymbol{T}_2)$ is well defined if the support of $\boldsymbol{T}_1$ is contained in the support of $\boldsymbol{T}_2$. Moreover $I(\boldsymbol{T}_1, \boldsymbol{T}_2) \geq 0$ with equality if and only if $g_1 = g_2$ almost everywhere. In general $I(\boldsymbol{T}_1 : \boldsymbol{T}_2)$ is not symmetric, therefore is not a distance function but a directed divergence or pseudo-distance measure. In Jeffreys (1946) was introduced a divergence function that avoids the asymmetry problem, although it does not satisfy the triangular inequality. Hence, also corresponds to a pseudo-distance. Suppose now that $\boldsymbol{T}_1 \sim \mathsf{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\boldsymbol{T}_2 \sim \mathsf{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$. Then it is easy to notice that (see for instance Ullah, 1996),

$$I(\boldsymbol{T}_1 : \boldsymbol{T}_2) = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2}(\operatorname{tr} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1} - p) - \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|}. \tag{6}$$

Following some ideas in Pan et al. (1996, 1999) and Pan and Fung (2000), we develop diagnostic procedures based on case elimination techniques as well as local influence considering the function in (6) as a measure of influence.

3.1 Case-deletion procedure

Consider the model given in (1), under the assumption of normality it follows that the least squares estimator has distribution $\widehat{\boldsymbol{\beta}} \sim \mathsf{N}_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1})$. Based on the relationship between the ridge estimator with the LS estimator, we have

$$\widehat{\boldsymbol{\beta}}_\lambda = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}} \sim \mathsf{N}_p(\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{S}_\lambda^{-1}),$$

where $\boldsymbol{S}_\lambda = \boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I}_p$. Following Wei and Shih (1994), we can assess the influence of the ith observation, using the case-deletion model which is defined as:

$$\boldsymbol{Y}_{(i)} = \boldsymbol{X}_{(i)}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{(i)}, \qquad \boldsymbol{\epsilon}_{(i)} \sim \mathsf{N}_{n-1}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n-1}), \tag{7}$$

where $\boldsymbol{Y}_{(i)}$, $\boldsymbol{X}_{(i)}$ and $\boldsymbol{\epsilon}_{(i)}$ denote respectively, the response vector, model matrix and the vector of random disturbances in the model given in (1) once the ith observation has been removed. It is straightforward to notice that the ridge estimator for model (7), that is, without considering the ith observation, say $\widehat{\boldsymbol{\beta}}_\lambda(i)$, satisfies

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_\lambda(i) &= (\boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)} + \lambda \boldsymbol{I}_p)^{-1}\boldsymbol{X}_{(i)}^\top \boldsymbol{Y}_{(i)} \\ &\sim \mathsf{N}_p(\boldsymbol{S}_\lambda^{-1}(i)\boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)}\boldsymbol{\beta}, \sigma^2 \boldsymbol{S}_\lambda^{-1}(i)\boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)}\boldsymbol{S}_\lambda^{-1}(i)),\end{aligned}$$

with $\boldsymbol{S}_\lambda(i) = \boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)} + \lambda \boldsymbol{I}_p$. Evidently, $\boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)} = \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{x}_i \boldsymbol{x}_i^\top$, which leads to write

$$\boldsymbol{S}_\lambda(i) = \boldsymbol{X}^\top \boldsymbol{X} + \lambda \boldsymbol{I}_p - \boldsymbol{x}_i \boldsymbol{x}_i^\top = \boldsymbol{S}_\lambda - \boldsymbol{x}_i \boldsymbol{x}_i^\top.$$

As an alternative to the techniques for case-deletion diagnostics in ridge regression outlined in Section 2.1, we will consider the Kullback-Leibler divergence between $\widehat{\boldsymbol{\beta}}_\lambda$ and $\widehat{\boldsymbol{\beta}}_\lambda(i)$. This allows us to define a measure to assess the influence of the ith observation on the ridge estimator, as:

$$\begin{aligned} KL_i = I(\widehat{\boldsymbol{\beta}}_\lambda : \widehat{\boldsymbol{\beta}}_\lambda(i)) &= \frac{1}{2} \boldsymbol{\delta}^\top \operatorname{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i)) \boldsymbol{\delta} + \frac{1}{2} \operatorname{tr} \operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda) \operatorname{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i)) \\ &\quad - \frac{1}{2} \log \frac{|\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)|}{|\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(i))|} - \frac{p}{2}, \end{aligned} \tag{8}$$

where $\boldsymbol{\delta} = \mathrm{E}(\widehat{\boldsymbol{\beta}}_\lambda) - \mathrm{E}(\widehat{\boldsymbol{\beta}}_\lambda(i))$. The next proposition gives a computationally attractive expression for KL_i.

Proposition 1. *Let $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ be the singular value decomposition (SVD) of the model matrix $\boldsymbol{X}$ with $\boldsymbol{U} \in \mathbb{R}^{n \times p}$ such that $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_p$, $\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_p)$, where $d_1 \geq \cdots \geq d_p > 0$, are the singular values of $\boldsymbol{X}$, and $\boldsymbol{V}$ is an orthogonal matrix. Then*

$$KL_i = \frac{1}{2}\Big(\frac{\lambda^2 q_i^2(\boldsymbol{\alpha})}{\sigma^2} + 1 + \boldsymbol{u}_i^\top \boldsymbol{\Delta}^2 \boldsymbol{u}_i\Big)\Big(\frac{h_{ii}}{1 - h_{ii}}\Big) + \frac{h_{ii}(\lambda)}{1 - h_{ii}} + \frac{1}{2} \log\Big(\frac{1 - h_{ii}}{(1 - h_{ii}(\lambda))^2}\Big),$$

for $i = 1, \ldots, n$, where $q_i(\boldsymbol{\alpha}) = \boldsymbol{u}_i^\top \boldsymbol{D}^{-1} \boldsymbol{\Delta} \boldsymbol{\alpha}$, $\boldsymbol{\Delta} = (\boldsymbol{D}^2 + \lambda \boldsymbol{I})^{-1} \boldsymbol{D}^2$, $\boldsymbol{\alpha} = \boldsymbol{V}^\top \boldsymbol{\beta}$ and $\boldsymbol{u}_i$ denotes the ith row of the matrix $\boldsymbol{U}$. Moreover, the diagonal elements of $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$ and $\boldsymbol{H}(\lambda)$, can be written as

$$h_{ii} = \boldsymbol{u}_i^\top \boldsymbol{u}_i, \qquad h_{ii}(\lambda) = \boldsymbol{u}_i^\top \boldsymbol{\Delta} \boldsymbol{u}_i.$$

Proof. See Appendix A of the supplementary material. □

Remark 1. *Because KL_i depends on $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$. We must obtain a sample version of KL_i by considering estimates for $\boldsymbol{\beta}$ and σ^2. Among the different alternatives to characterize the ridge estimator, we will consider that (2) can be seen as the penalized maximum likelihood (PML) estimator based on the penalized log-likelihood function, defined as,*

$$\begin{aligned} \ell_\lambda(\boldsymbol{\beta}, \sigma^2) &= \ell(\boldsymbol{\beta}, \sigma^2) - \frac{\lambda}{2\sigma^2} \|\boldsymbol{\beta}\|^2 \\ &= -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\big(\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|^2\big), \end{aligned}$$

where $\ell(\boldsymbol{\beta}, \sigma^2)$ denote the log-likelihood function based on the model in (1). It should be noted that the PML estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$ is given by $\widehat{\boldsymbol{\theta}}_\lambda = (\widehat{\boldsymbol{\beta}}_\lambda^\top, \widehat{\sigma}_\lambda^2)^\top$ with

$\widehat{\boldsymbol{\beta}}_\lambda$ *being defined in* (2)*, whereas*

$$\widehat{\sigma}^2_\lambda = \frac{1}{n}(\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda\|^2).$$

Remark 2. *As highlighted in Pan and Fung (2000), using the Kullback-Leibler divergence is one approach to measure the difference between the densities associated with* $\widehat{\boldsymbol{\beta}}_\lambda$ *and* $\widehat{\boldsymbol{\beta}}_\lambda(i)$*. Indeed,* $KL_i^* = I(\widehat{\boldsymbol{\beta}}_\lambda(i) : \widehat{\boldsymbol{\beta}}_\lambda)$ *can be calculated in much the same fashion as* KL_i*. In addition, this allows to obtain other measures useful for diagnostic purposes, such as Jeffreys' divergence, given by*

$$J_i = I(\widehat{\boldsymbol{\beta}}_\lambda : \widehat{\boldsymbol{\beta}}_\lambda(i)) + I(\widehat{\boldsymbol{\beta}}_\lambda(i) : \widehat{\boldsymbol{\beta}}_\lambda).$$

However, this work is focused on performing case-deletion diagnostics in ridge regression based on the use of (8)*.*

3.2 Local influence procedure

In the following, we consider three perturbation schemes. Namely, the variance perturbation on model defined in Equation (1), the response perturbation scheme, and the explanatory variable perturbation. Each scheme was applied on the Kullback-Leibler divergence as an influence function. The proofs of Propositions 2-4 are deferred to Appendix B of the supplementary material.

3.2.1 Perturbation of variances

Let us consider the perturbed model,

$$\begin{aligned}\mathcal{P}_\omega = \{&\mathsf{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{W}^{-1}) : \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 > 0, \boldsymbol{W} = \operatorname{diag}(\boldsymbol{\omega}),\\ &\text{with } \boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top, \boldsymbol{\omega} \in \mathbb{R}^n_+\}.\end{aligned} \tag{9}$$

In this case we have that the null perturbation vector is given by $\boldsymbol{\omega}_0 = \mathbf{1}_n$. It is easy to notice that the ridge estimator under the perturbed model takes the form,

$$\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}) = (\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X} + \lambda\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{Y},$$

hence,

$$\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}) \sim \mathsf{N}_p(\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})),$$

with $\boldsymbol{S}_\lambda(\boldsymbol{\omega}) = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X} + \lambda\boldsymbol{I}_p$. Furthermore, we have that $\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}_0) = \widehat{\boldsymbol{\beta}}_\lambda$. That is, we propose to use the following influence function

$$\begin{aligned}KL(\boldsymbol{\omega}) &= I(\widehat{\boldsymbol{\beta}}_\lambda : \widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}))\\ &= \frac{1}{2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\operatorname{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}))\boldsymbol{\delta}(\boldsymbol{\omega}) + \frac{1}{2}\operatorname{tr}\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)\operatorname{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}))\\ &\quad - \frac{1}{2}\log\frac{|\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)|}{|\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}))|} - \frac{p}{2}.\end{aligned}$$

where $\boldsymbol{\delta}(\boldsymbol{\omega}) = (\boldsymbol{M}_\lambda - \boldsymbol{M}_\lambda(\boldsymbol{\omega}))\boldsymbol{\beta}$, with $\boldsymbol{M}_\lambda = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}$ and $\boldsymbol{M}_\lambda(\boldsymbol{\omega}) = \boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}$. Thus, we can write $KL(\boldsymbol{\omega})$ as:

$$\begin{aligned} KL(\boldsymbol{\omega}) &= \frac{1}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega}) - \frac{p}{2} \\ &+ \frac{1}{2}\operatorname{tr}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top \\ &+ \log|\boldsymbol{S}_\lambda| - \frac{1}{2}\log|\boldsymbol{X}^\top\boldsymbol{X}| - \log|\boldsymbol{S}_\lambda(\boldsymbol{\omega})| + \frac{1}{2}\log|\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}|. \end{aligned}$$

The following proposition characterizes the local curvature considering the perturbed model $\mathcal{P}_\omega$ defined in Equation (9).

Proposition 2. *For the perturbed model given in (9), we have that the gradient and Hessian matrix of $KL(\boldsymbol{\omega})$ evaluated at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $(\boldsymbol{\beta}^\top, \sigma^2)^\top = (\widehat{\boldsymbol{\beta}}_\lambda^\top, \widehat{\sigma}_\lambda^2)^\top$, adopt the form*

$$\begin{aligned} \nabla_{KL} &= \frac{\partial KL(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\Big|_{\omega=\omega_0,\theta=\widehat{\theta}_\lambda} = \mathbf{0}, \\ \boldsymbol{F}_{KL} &= \frac{\partial^2 KL(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top}\Big|_{\omega=\omega_0,\theta=\widehat{\theta}_\lambda} = \boldsymbol{B}_n^\top\Big\{\boldsymbol{H}(\lambda)\otimes\boldsymbol{H}(\lambda) - \frac{1}{\widehat{\sigma}_\lambda^2}\boldsymbol{H}\otimes\boldsymbol{H}(\lambda)\boldsymbol{C}\boldsymbol{H}(\lambda) \\ &\qquad + \frac{1}{2}\boldsymbol{H}\otimes(\boldsymbol{H} - 4\boldsymbol{H}(\lambda) + \boldsymbol{H}^2(\lambda))\Big\}\boldsymbol{B}_n, \end{aligned}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{C} = \boldsymbol{Y}\boldsymbol{Y}^\top + \widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top$ and $\boldsymbol{B}_n$ is known as the transition matrix (see Nel, 1980) which satisfies $\operatorname{vec}\boldsymbol{W} = \boldsymbol{B}_n\boldsymbol{\omega}$.

3.2.2 Response perturbation

Let $\boldsymbol{Y}(\boldsymbol{\omega}) = \boldsymbol{Y} + \boldsymbol{\omega}$ be response shifts, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ and the null perturbation vector is given by $\boldsymbol{\omega}_0 = \mathbf{0}$. Thus, the perturbed version of the ridge estimator, satisfies

$$\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}) = (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}(\boldsymbol{\omega}) = \widehat{\boldsymbol{\beta}}_\lambda + (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{\omega}.$$

It is straightforward to verify that $\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega})) = \operatorname{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)$, which allows us to write

$$\begin{aligned} KL(\boldsymbol{\omega}) &= I(\widehat{\boldsymbol{\beta}}_\lambda : \widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega})) \\ &= \frac{1}{2}\boldsymbol{\omega}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\operatorname{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda)(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{\omega} \\ &= \frac{1}{2}\boldsymbol{\omega}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\omega}. \end{aligned}$$

Explicit formulas for ∇_{KL} and $\boldsymbol{F}_{KL}$ required to evaluate the local curvature are provided in the following proposition.

Proposition 3. *For the response perturbation* $\boldsymbol{Y}(\boldsymbol{\omega}) = \boldsymbol{Y} + \boldsymbol{\omega}$*, we have that the gradient and Hessian matrix of* $KL(\boldsymbol{\omega})$ *evaluated at* $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ *and* $(\boldsymbol{\beta}^\top, \sigma^2)^\top = (\widehat{\boldsymbol{\beta}}_\lambda^\top, \widehat{\sigma}_\lambda^2)^\top$*, are given by*

$$\nabla_{KL} = \frac{\partial KL(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\Big|_{\omega=\omega_0, \theta=\widehat{\theta}_\lambda} = \mathbf{0},$$
$$\boldsymbol{F}_{KL} = \frac{\partial^2 KL(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top}\Big|_{\omega=\omega_0, \theta=\widehat{\theta}_\lambda} = \boldsymbol{H},$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$.

It is interesting to note that the curvature matrix associated with this perturbation scheme is fully characterized by the projection matrix $\boldsymbol{H}$. That is, from this perspective, the leverage matrix associated with least squares estimation may still contain information relevant for diagnostics in ridge regression. It is well known that a projection matrix has all its non-zero eigenvalues equal to unity. In fact, the singular value decomposition of the matrix $\boldsymbol{X}$ used in Proposition 1 allows to write $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{U}^\top$, which leads to a very efficient procedure to obtain these p eigenvectors. From a diagnostic perspective, our interest is in examining the magnitude of the elements of the $\boldsymbol{U}$ matrix to reveal those observations that have a strong impact when the response perturbation scheme is considered.

3.2.3 Perturbation of the explanatory variables

Our interest is on perturbing a particular continuous explanatory variable. Thus, we assume the following perturbed model

$$\mathcal{P}_\omega = \{\mathsf{N}_n(\boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{\beta}, \sigma^2\boldsymbol{I}) : \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 > 0, \boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top \in \mathbb{R}^n\}, \tag{10}$$

where $\boldsymbol{X}(\boldsymbol{\omega}) = \boldsymbol{X} + a\boldsymbol{\omega}\boldsymbol{c}_t^\top$ denotes the perturbed model matrix, $a > 0$ is a scale factor and $\boldsymbol{c}_t$ is a $p \times 1$ vector with 1 at the tth position and zero elsewhere. In this case one has $\boldsymbol{\omega}_0 = \mathbf{0}$. Under this perturbation scheme we have that

$$\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}) = (\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}) + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{Y}.$$

Under the assumption of $\boldsymbol{Y} \sim \mathsf{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$, it follows that

$$\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}) \sim \mathsf{N}_p(\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}(\boldsymbol{\omega})).$$

Under this perturbation scheme the influence function $KL(\boldsymbol{\omega}) = I(\widehat{\boldsymbol{\beta}}_\lambda : \widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}))$ adopts the form,

$$KL(\boldsymbol{\omega}) = \frac{1}{2\sigma^2}\boldsymbol{\delta}^\top(\boldsymbol{\omega})\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{\delta}(\boldsymbol{\omega}) - \frac{p}{2}$$
$$+ \frac{1}{2}\operatorname{tr}\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})(\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega}))^{-1}\boldsymbol{S}_\lambda(\boldsymbol{\omega})\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top$$

$$+\log|\boldsymbol{S}_\lambda| - \frac{1}{2}\log|\boldsymbol{X}^\top\boldsymbol{X}| - \log|\boldsymbol{S}_\lambda(\boldsymbol{\omega})| + \frac{1}{2}\log|\boldsymbol{X}^\top(\boldsymbol{\omega})\boldsymbol{X}(\boldsymbol{\omega})|.$$

Therefore, the elements necessary to characterize the local curvature associated with the perturbed model in (10) are given in the following proposition.

Proposition 4. *Considering the perturbed model defined in Equation* (10)*, we have that* ∇_{KL} *and* $\boldsymbol{F}_{KL}$ *evaluated at* $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ *and* $(\boldsymbol{\beta}^\top, \sigma^2)^\top = (\widehat{\boldsymbol{\beta}}_\lambda^\top, \widehat{\sigma}_\lambda^2)^\top$*, are given by*

$$\begin{aligned}
\nabla_{KL} &= \frac{\partial KL(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\Big|_{\omega=\omega_0,\theta=\widehat{\theta}_\lambda} = \mathbf{0},\\
\boldsymbol{F}_{KL} &= \frac{\partial^2 KL(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top}\Big|_{\omega=\omega_0,\theta=\widehat{\theta}_\lambda}\\
&= \frac{a^2}{\widehat{\sigma}_\lambda^2}\Big\{r_{tt}\widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top - 2(r_{tt}\boldsymbol{H}(\lambda) + \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{E}_t\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top)\widehat{\boldsymbol{Y}}_\lambda\widehat{\boldsymbol{Y}}_\lambda^\top(\boldsymbol{I}_n - \tfrac{1}{2}\boldsymbol{H}(\lambda))\Big\}\\
&\quad + a^2\big\{(l_{tt} + m_t^2/\widehat{\sigma}_\lambda^2 + r_{tt}^2)\boldsymbol{H} - 2r_{tt}\boldsymbol{H}(\lambda)(\boldsymbol{I}_n - \tfrac{1}{2}\boldsymbol{H}(\lambda)) + 2\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}\boldsymbol{E}_t\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\\
&\quad + \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{E}_t\big[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top - 4\boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top(\boldsymbol{I}_n - \tfrac{1}{2}\boldsymbol{H}(\lambda))\big]\big\},
\end{aligned}$$

where $\boldsymbol{R} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}$, $\boldsymbol{L} = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{S}_\lambda^{-1}$, $\boldsymbol{m} = \boldsymbol{S}_\lambda^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{Y}}_\lambda$. *Thus,* $r_{tt} = \boldsymbol{c}_t^\top\boldsymbol{R}\boldsymbol{c}_t$ *and* $l_{tt} = \boldsymbol{c}_t^\top\boldsymbol{L}\boldsymbol{c}_t$, $m_t = \boldsymbol{c}_t^\top\boldsymbol{m}$, *and* $\boldsymbol{E}_t = \boldsymbol{c}_t\boldsymbol{c}_t^\top$.

It is interesting to note that the first derivative of $KL(\boldsymbol{\omega})$ for all the perturbation schemes under consideration satisfies $\nabla_{KL} = \mathbf{0}$, which simplifies the computation of the curvature matrix associated with (4). This result is expected, because $KL(\boldsymbol{\omega})$ attains its local minimum at $\boldsymbol{\omega}_0$.

4 Numerical experiments

In this section we illustrate the proposed methodology through the analysis of two real datasets previously studied in literature. In addition, we present a simulation study to evaluate the performance of the influence diagnostics based on the Kullback-Leibler divergence. Datasets and R codes for the influence diagnostic procedures described in the previous section are available on github.[1] In our analyses we also have used routines available in the india (Osorio, 2023) and fastmatrix (Osorio and Ogueda, 2024) packages.

4.1 Monte Carlo study

We report our findings from a Monte Carlo simulation study, which was designed to evaluate the performance of the influence diagnostic procedure described in the previous section on small samples. Our experiment is based on the simulation study reported by Hadi and Nyquist (1993) who proposed the construction of a model matrix $n \times p$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{(i)} \\ \boldsymbol{x}_i^\top \end{pmatrix}, \tag{11}$$

[1] URL: https://github.com/faosorios/KL_influence

Table 1 Outlier detection percentage using different influence measures, $n = 20$.

k	Scenario	d	case deletion		$LD(\boldsymbol{\omega})$		$KL(\boldsymbol{\omega})$	
			D_i^{**}	KL_i	variance	response	variance	response
1	I	1	7.3	2.4	3.6	2.8	2.7	91.7
		5	53.1	100.0	0.0	0.0	21.2	100.0
		10	73.9	100.0	0.0	0.2	25.7	100.0
	II	1	17.0	59.6	3.0	1.4	4.0	100.0
		5	62.9	100.0	0.0	0.0	5.8	100.0
		10	82.5	100.0	0.0	0.0	3.6	100.0
	III	1	13.4	16.3	3.4	3.2	5.0	100.0
		5	60.7	100.0	0.0	0.8	29.8	100.0
		10	79.5	100.0	0.0	0.6	35.1	100.0
2	I	1	19.4	98.7	0.7	0.3	10.4	100.0
		5	56.1	100.0	0.0	0.0	20.8	100.0
		10	78.2	100.0	0.0	0.0	24.2	100.0
	II	1	28.3	100.0	0.3	0.0	10.6	100.0
		5	71.4	100.0	0.0	0.0	4.8	100.0
		10	82.8	100.0	0.0	0.0	1.1	100.0
	III	1	26.2	99.9	0.6	0.2	11.2	100.0
		5	68.3	100.0	0.0	0.0	29.2	100.0
		10	87.1	100.0	0.0	0.1	30.0	100.0
3	I	1	22.5	100.0	0.0	0.0	10.7	100.0
		5	50.8	100.0	0.0	0.0	24.6	100.0
		10	75.1	100.0	0.0	0.1	30.4	100.0
	II	1	28.2	100.0	0.0	0.0	13.2	100.0
		5	68.8	100.0	0.0	0.0	6.3	100.0
		10	84.8	100.0	0.0	0.0	2.3	100.0
	III	1	30.7	100.0	0.0	0.0	12.9	100.0
		5	65.4	100.0	0.0	0.0	32.6	100.0
		10	88.1	100.0	0.0	0.3	38.3	100.0

according to the following procedure: First, we need to create the matrix $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3)$, where $\boldsymbol{Z}_j \sim \mathsf{N}_m(\boldsymbol{0}, \boldsymbol{I}_m)$ are independent vectors, for $j = 1, 2, 3$, with $m = n - 1$, and then consider

$$\boldsymbol{W}_1 = \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{\eta}_1, \qquad \boldsymbol{W}_2 = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\eta}_2,$$

where $\boldsymbol{a}$ y $\boldsymbol{b}$ are vectors of constants $p \times 1$, whereas $\boldsymbol{\eta}_1 \sim \mathsf{N}_m(\boldsymbol{0}, \sigma^2(k)\boldsymbol{I}_m)$ and $\boldsymbol{\eta}_2 \sim \mathsf{N}_m(\boldsymbol{0}, \sigma^2(k)\boldsymbol{I}_m)$ such that $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are independent, with $\sigma^2(k) = 10^{-k}$. Let,

$$\boldsymbol{X}_{(i)} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3, \boldsymbol{W}_1, \boldsymbol{W}_2), \tag{12}$$

and obtain the eigenvectors $\boldsymbol{\Psi}$ of $\boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)}$. The next step is the construction of the vector:

$$\boldsymbol{x}_i = d\boldsymbol{\Psi}\boldsymbol{\tau},$$

where d is a scalar and $\boldsymbol{\tau}$ is a unitary vector $p \times 1$. Finally, the vector $\boldsymbol{x}_i^\top$ is appended to $\boldsymbol{X}_{(i)}$ as the last row of the matrix in (11). As mentioned by Hadi and Nyquist (1993), the conditioning of the matrix $\boldsymbol{X}_{(i)}$ defined in (12) is characterized by the

Table 2 Outlier detection percentage using different influence measures, $n = 50$.

k	Scenario	d	case deletion		$LD(\boldsymbol{\omega})$		$KL(\boldsymbol{\omega})$	
			D_i^{**}	KL_i	variance	response	variance	response
1	I	1	4.8	0.1	3.8	3.8	2.8	97.5
		5	56.7	100.0	1.0	0.0	48.3	100.0
		10	74.9	100.0	0.0	0.0	59.2	100.0
	II	1	18.1	57.4	3.7	2.5	9.4	100.0
		5	68.6	100.0	0.0	0.0	27.6	100.0
		10	84.4	100.0	0.0	0.0	31.3	100.0
	III	1	10.8	6.4	3.5	3.7	8.5	100.0
		5	63.9	100.0	0.5	0.1	65.0	100.0
		10	82.3	100.0	0.0	0.0	81.8	100.0
2	I	1	29.8	99.6	1.5	1.2	18.2	100.0
		5	64.1	100.0	0.0	0.0	47.3	100.0
		10	80.8	100.0	0.0	0.0	50.1	100.0
	II	1	38.5	100.0	0.9	0.2	18.1	100.0
		5	77.0	100.0	0.0	0.0	24.0	100.0
		10	87.3	100.0	0.0	0.0	23.3	100.0
	III	1	32.6	100.0	1.2	0.8	23.1	100.0
		5	77.7	100.0	0.0	0.0	58.9	100.0
		10	92.5	100.0	0.0	0.0	71.3	100.0
3	I	1	37.8	100.0	0.0	0.0	20.9	100.0
		5	66.2	100.0	0.0	0.0	49.9	100.0
		10	86.4	100.0	0.0	0.0	52.7	100.0
	II	1	40.6	100.0	0.0	0.0	18.5	100.0
		5	80.1	100.0	0.0	0.0	24.6	100.0
		10	93.1	100.0	0.0	0.0	23.9	100.0
	III	1	40.5	100.0	0.0	0.0	23.8	100.0
		5	82.4	100.0	0.0	0.0	60.7	100.0
		10	96.8	100.0	0.0	0.0	67.8	100.0

scalar k and the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. In fact, for $k = 0$ and $\boldsymbol{a} = \boldsymbol{b} = (0, 0, 0)^\top$ the matrix $\boldsymbol{X}_{(i)}$ is well-conditioned. Additionally, the parameter d determines the leverage of $\boldsymbol{x}_i$, and as d increases it becomes more collinearity-influential, whereas $\boldsymbol{\tau}$ determines the position of $\boldsymbol{x}_i$ relative to the eigenvectors $\boldsymbol{\Psi}$ (see Hadi and Nyquist, 1993).

Next, we consider the model,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{13}$$

where the model matrix has been generated from the procedure described above. We use $\boldsymbol{\beta} = (1, 1, 1, 1, 1)^\top$ and $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\boldsymbol{0}, \phi \boldsymbol{I}_n)$ with $\phi = 1$. Following Hadi and Nyquist (1993), we consider $\boldsymbol{a} = (1, 1, 0)^\top$ and $\boldsymbol{b} = (0, 0, 1)^\top$ to create two sets of collinearities and the scenarios:

$$\text{I)}\ \boldsymbol{\tau} = \frac{1}{\sqrt{5}}(1, 1, 1, 1, 1)^\top, \quad \text{II)}\ \boldsymbol{\tau} = \frac{1}{\sqrt{2}}(0, 0, 0, 1, 1)^\top, \quad \text{III)}\ \boldsymbol{\tau} = \frac{1}{\sqrt{2}}(1, 0, 0, 0, 1)^\top.$$

Table 3 Outlier detection percentage using different influence measures, $n = 100$.

k	Scenario	d	case deletion		$LD(\boldsymbol{\omega})$		$KL(\boldsymbol{\omega})$	
			D_i^{**}	KL_i	variance	response	variance	response
1	I	1	3.8	0.0	3.7	3.8	4.0	99.8
		5	60.6	100.0	3.4	0.3	65.8	100.0
		10	77.1	100.0	0.0	0.0	75.6	100.0
	II	1	15.7	74.6	4.6	3.2	18.0	100.0
		5	72.0	100.0	0.2	0.0	54.4	100.0
		10	85.3	100.0	0.0	0.0	67.7	100.0
	III	1	8.4	2.4	5.3	5.1	14.5	100.0
		5	67.0	100.0	2.2	0.1	91.0	100.0
		10	83.8	100.0	0.0	0.0	97.0	100.0
2	I	1	34.5	99.9	3.0	2.3	27.8	100.0
		5	73.6	100.0	0.0	0.0	62.4	100.0
		10	84.6	100.0	0.0	0.0	69.0	100.0
	II	1	42.4	100.0	2.6	0.7	31.4	100.0
		5	80.5	100.0	0.0	0.0	48.7	100.0
		10	91.6	100.0	0.0	0.0	55.0	100.0
	III	1	37.6	100.0	3.1	2.0	34.5	100.0
		5	80.9	100.0	0.0	0.0	84.3	100.0
		10	92.8	100.0	0.0	0.0	88.3	100.0
3	I	1	47.2	100.0	0.2	0.0	30.7	100.0
		5	77.0	100.0	0.0	0.0	58.7	100.0
		10	90.4	100.0	0.0	0.0	58.4	100.0
	II	1	52.7	100.0	0.0	0.0	24.9	100.0
		5	87.8	100.0	0.0	0.0	46.0	100.0
		10	95.8	100.0	0.0	0.0	52.6	100.0
	III	1	49.3	100.0	0.0	0.0	37.4	100.0
		5	88.4	100.0	0.0	0.0	72.7	100.0
		10	98.3	100.0	0.0	0.0	75.0	100.0

In our simulation study we will also consider the following values for d and k, $d = 1, 5, 10$ and $k = 1, 2, 3$, respectively. 1 000 datasets with sample size of $n = 20, 50$ and 100 were created from model (13).

Note that the synthetic collinearity-influential point corresponds to the last observation in the matrix defined in (11). Let $\boldsymbol{h} = |\boldsymbol{h}_{\max}|$, thus we detect this extreme point if h_n is greater than the following threshold $\overline{h} + 2\,\mathrm{sd}(\boldsymbol{h})$, where $\overline{h}$ and $\mathrm{sd}(\boldsymbol{h})$ are the average and standard deviation of $\boldsymbol{h}$, respectively. We emphasize that this benchmark will be used in our analysis with real datasets from the following section. Tables 1, 2 and 3 contain the collinearity-influential detection percentages computed using this threshold for different values of k and d considering several influence measures.

As expected, the detection percentages improve as d increases. Our findings suggest that, methods based on the case elimination technique using the Kullback-Leibler divergence as well as local influence under the response perturbation scheme are quite efficient for detecting of the collinearity-influential observation. We should stress that the results based on the evaluation of local influence based on the likelihood displacement, $LD(\boldsymbol{\omega})$ (see Appendix A), are disappointing, whereas the Cook's distance, D_i^{**},

and the local influence procedure under the scheme of perturbation of variances are more conservative in detecting the extreme observation.

4.2 Real-life examples

4.2.1 Portland cement data

We will consider the experimental study on heat emission during the hardening process of 13 Portland cement samples introduced by Woods et al. (1932) who related the heat emission after 180 days of curing, measured in calories per gram of cement as a function of four predictors corresponding to the percentages of the following four compounds: tricalcium aluminate, tricalcium silicate, tetracalcium aluminate ferrite and dicalcium silicate. This dataset has been extensively analyzed to illustrate the harmful effects of collinearity (see, for example, Kowalski, 1990; Kaçıranlar et al., 1999; Lukman et al., 2019, as well as the references therein). In particular, Hadi (1988) and Wang and Nyquist (1991) used this dataset for the detection of collinearity-influential observations in the model given in (1). Following Gorman and Toman (1966) we considered a linear regression model with intercept, that is,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i, \tag{14}$$

for $i = 1, \ldots, 13$. This model is called an inhomogeneous linear model by Kaçıranlar et al. (1999), and whose design matrix $\boldsymbol{X}$ has scaled condition number (see Belsley, 1991), $\kappa(\boldsymbol{X}) = 249.578$ suggesting the presence of rather severe collinearity. We consider ridge estimation, choosing the shrinkage parameter by generalized cross-validation. Table 4 presents the estimation results for the model given in (14). We also report the selected ridge parameter, as well as the effective degrees of freedom, $\mathsf{edf} = \operatorname{tr} \boldsymbol{H}(\lambda)$, and the determinant of the estimated covariance matrix for $\widehat{\boldsymbol{\beta}}_\lambda$.

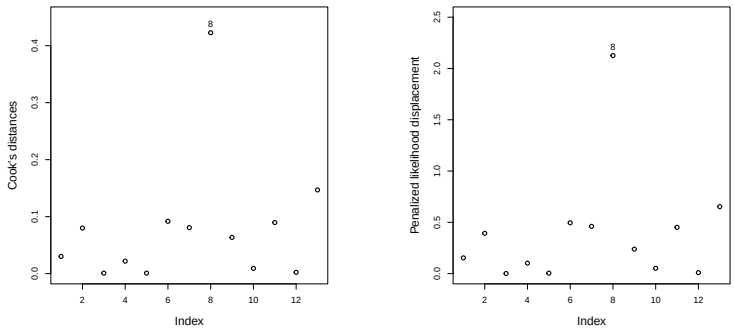
Figure 1 displays several diagnostic measures. Specifically, the Cook's distance, the penalized likelihood displacement Cook et al. (1988), given by (see Appendix D of the supplementary material):

$$\begin{aligned} LD_i(\boldsymbol{\beta}|\sigma^2) &= 2\{\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda, \widehat{\sigma}^2_\lambda) - \max_{\sigma^2} \ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda(i), \sigma^2)\} \\ &= n \log\Big(\frac{\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda(i)\|^2}{\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}_\lambda\|^2}\Big), \end{aligned}$$

and the direction of largest curvature $\boldsymbol{h}_{\max}$ associated with the variance perturbation scheme (see Figure 1 (e)), using the penalized likelihood displacement,

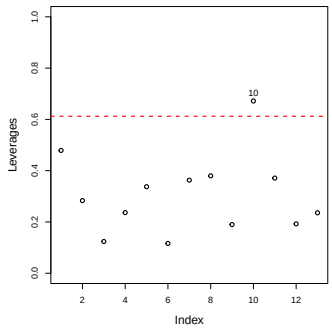
$$LD(\boldsymbol{\omega}) = 2\{\ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda, \widehat{\sigma}^2_\lambda) - \ell_\lambda(\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega}), \widehat{\sigma}^2_\lambda(\boldsymbol{\omega}))\},$$

reveal that observation 8 has a strong impact on the estimation of the regression coefficients (for details on the derivation of the curvature matrix in this case see Appendix A), whereas the index plot of $\boldsymbol{h}_{\max}$ for $LD(\boldsymbol{\omega})$ under the response perturbation allows us to identify observation 6. In addition, observation 10 has a pronounced leverage

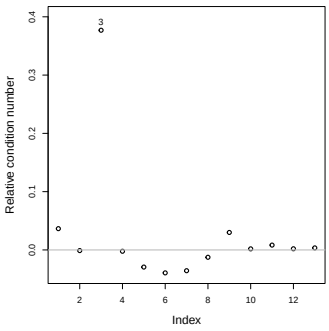


(a)

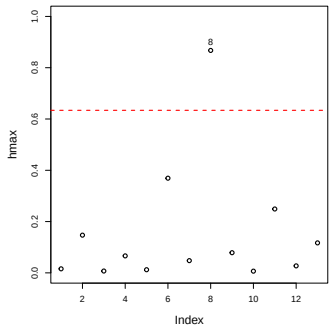
(b)



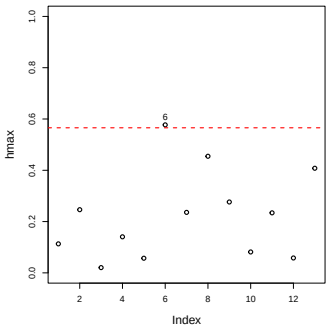
(c)



(d)



(e)



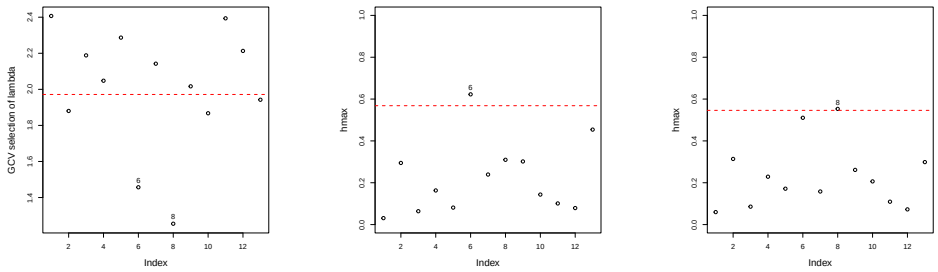
(f)

Fig. 1 Influence measures for Portland cement data: (a) Cook's distances, D_i^{**}, (b) penalized likelihood displacement, $LD_i(\boldsymbol{\beta}|\sigma^2)$, (c) index plot of leverages, $h_{ii}(\lambda)$, (d) index plot of relative condition index. Index plot of $\boldsymbol{h}_{\max}$ based on $LD(\boldsymbol{\omega})$ under (e) perturbation of variances, (f) and response perturbation.

Table 4 Parameter estimates, standard errors (in parenthesis) and percentage change in parameter estimates for Portland cement data.

Parameter	Full	Removed observations			
	data	3	6	8	10
β_0	0.085	0.084	0.089	0.084	0.089
	(0.040)	(0.026)	(0.045)	(0.051)	(0.042)
	—	-1.88%	3.88%	-1.72%	3.98%
β_1	2.165	2.164	2.133	2.247	2.135
	(0.170)	(0.179)	(0.145)	(0.144)	(0.279)
	—	-0.08%	-1.52%	3.76%	-1.42%
β_2	1.159	1.160	1.154	1.119	1.164
	(0.044)	(0.047)	(0.037)	(0.040)	(0.058)
	—	0.09%	-0.41%	-3.38%	0.44%
β_3	0.738	0.736	0.738	0.903	0.726
	(0.146)	(0.154)	(0.124)	(0.138)	(0.176)
	—	-0.35%	-0.07%	22.31%	-1.68%
β_4	0.490	0.490	0.493	0.482	0.492
	(0.038)	(0.040)	(0.032)	(0.032)	(0.044)
	—	0.06%	0.78%	-1.50%	0.56%
σ^2	5.090	5.634	3.600	3.477	5.446
	—	10.69%	-29.28%	-31.69%	6.98%
λ	1.972	2.188	1.457	1.255	1.867
	—	10.99%	-26.11%	-36.36%	-5.29%
edf	3.979	3.977	3.985	3.985	3.960
	—	-0.07%	0.14%	0.15%	-0.49%
[1]det(Cov($\widehat{\boldsymbol{\beta}}_\lambda$))	4.504	3.332	1.654	2.587	19.630
	—	-26.01%	-63.28%	-42.56%	335.88%

[1] values multiplied by 10^{14}.



(a) (b) (c)

Fig. 2 Influence measures for Portland cement data: (a) selected ridge parameter using GCV when the ith observation is deleted, Index plot of $\boldsymbol{h}_{\max}$ based on GCV$(\lambda, \boldsymbol{\omega})$ under (b) perturbation of variances, and (c) response perturbation.

(see Figure 1 (c), the benchmark in this case is $2 \cdot \mathsf{edf}/n$ which corresponds to twice the average of the $h_{ii}(\lambda)$'s), whereas using the relative condition index (Hadi, 1988) $\gamma_i = (\kappa_{(i)} - \kappa)/\kappa$, for $i = 1, \ldots, n$, where $\kappa = \kappa(\boldsymbol{X})$ and $\kappa_{(i)} = \kappa(\boldsymbol{X}_{(i)})$, identifies the

3rd case as an observation that exerts a strong effect on the conditioning of the design matrix, i.e., eliminating observation 3 increases its condition number.

Furthermore, Figure 2 shows that observations 6 and 8 also have some effect on the selection of the ridge parameter. The panel in (a) presents the λ estimates based on the GCV criterion when the ith observation was removed from the dataset, in this case the segmented line represents the selected ridge parameter considering the full data, which is reported in Table 4. Following Thomas (1991) (see also Osorio, 2016) by perturbing the generalized cross-validation criterion, we can evaluate those observations that have a strong influence on the selection of the ridge parameter. It is possible to show that the direction of greatest local change on GCV is given by $\boldsymbol{h}_{\max} \propto \partial\widehat{\lambda}(\boldsymbol{\omega})/\partial\boldsymbol{\omega}\big|_{\omega=\omega_0}$. Furthermore,

$$\frac{\partial\widehat{\lambda}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\Big|_{\omega=\omega_0} = \Big\{ -\Big(\frac{\partial^2 \operatorname{GCV}(\lambda,\boldsymbol{\omega})}{\partial\lambda^2}\Big)^{-1} \frac{\partial^2 \operatorname{GCV}(\lambda,\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial\lambda} \Big\}\Big|_{\omega=\omega_0},$$

that is, $\boldsymbol{h}_{\max} \propto \partial^2 \operatorname{GCV}(\lambda,\boldsymbol{\omega})/\partial\boldsymbol{\omega}\partial\lambda\big|_{\omega=\omega_0,\lambda=\widehat{\lambda}}$. For the perturbation scheme defined in Section 3.2.1, we have that the perturbed GCV criterion adopts the form,

$$\operatorname{GCV}(\lambda,\boldsymbol{\omega}) = \frac{\|(\boldsymbol{I}_n - \boldsymbol{H}(\lambda,\boldsymbol{\omega}))\boldsymbol{Y}\|^2/n}{\{\operatorname{tr}(\boldsymbol{I}_n - \boldsymbol{H}(\lambda,\boldsymbol{\omega}))/n\}^2},$$

where $\boldsymbol{H}(\lambda,\boldsymbol{\omega}) = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X} + \lambda\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{W}$, with $\boldsymbol{W} = \operatorname{diag}(\omega_1,\ldots,\omega_n)$ and $\boldsymbol{\omega}_0 = \boldsymbol{1}_n$. Under the response perturbation scheme $\boldsymbol{Y}(\boldsymbol{\omega}) = \boldsymbol{Y} + \boldsymbol{\omega}$, where $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)$ with $\boldsymbol{\omega}_0 = \boldsymbol{0}$, the perturbed GCV criterion is given by

$$\operatorname{GCV}(\lambda,\boldsymbol{\omega}) = \frac{\|(\boldsymbol{I}_n - \boldsymbol{H}(\lambda))\boldsymbol{Y}(\boldsymbol{\omega})\|^2/n}{\{\operatorname{tr}(\boldsymbol{I}_n - \boldsymbol{H}(\lambda))/n\}^2}.$$

Closed-form expressions for $\boldsymbol{h}_{\max}$ follow directly from Thomas (1991) (see also the discussion in Osorio, 2016). In our case, we can see from Figure 2 (b,d) that observations 6 and 8 have a strong influence on the choice of λ. Moreover, from Table 4, we note that the elimination of these observations exert a 26.1% and 36.4% decrease in the selection of this parameter, respectively.

Using case-deletion and local influence methods based on Kullback-Leibler divergence (see Figure 3 and Figure 4, respectively), allow the identification of observation 10 as strongly influential. Indeed, this observation not only has a high leverage, but, as can be seen from Table 4, it also has a strong impact on the covariance of the ridge estimator. It is interesting to note that KL_i or the influence function, $KL(\boldsymbol{\omega})$ for the perturbation of variances are monotonic functions of the ratio of the determinants of the covariance matrices between the estimators $\widehat{\boldsymbol{\beta}}_\lambda$ and $\widehat{\boldsymbol{\beta}}_\lambda(i)$ or $\widehat{\boldsymbol{\beta}}_\lambda(\boldsymbol{\omega})$, respectively. That is, the measures proposed in our work are related to the COVRATIO statistic given in Belsley et al. (1980). It should be stressed that no discrepant observations were detected by perturbing each of the explanatory variables. Additionally, the index plot of the magnitude of the eigenvectors $\boldsymbol{U}$ associated with the response perturbation highlights observation 3 which has been identified as a collinearity-influential point

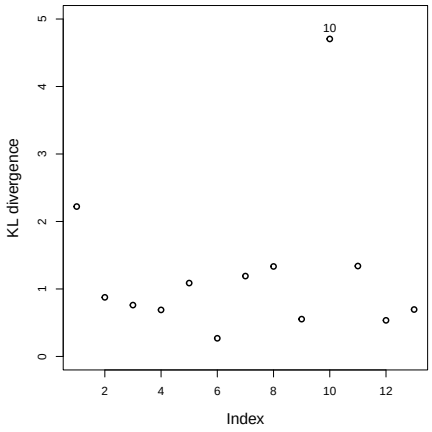
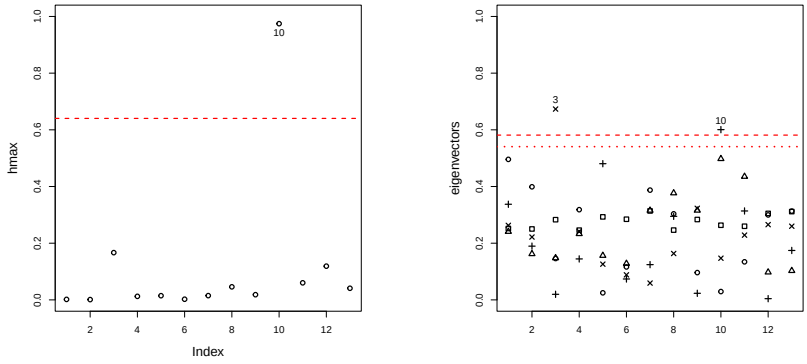


Fig. 3 Index plot of KL_i for Portland cement data.



(a) (b)

Fig. 4 Index plot of $\boldsymbol{h}_{\max}$ based on $KL(\boldsymbol{\omega})$ under (a) perturbation of variances, and (b) response perturbation, Portland cement data.

(see, for instance Hadi, 1988). It should be noted that in this case we have calculated the threshold described in Section 4.1 for each of the columns of $\boldsymbol{U}$. In Figure 4 (b), we have plotted only those thresholds that allowed us to identify influential observations. Finally, it should be noted that, to the best of our knowledge, observation 10 had not previously been detected as influential using traditional diagnostic techniques.

4.2.2 Aerial biomass

In Meloun and Militký (2001), the dataset introduced by Linthurst (1979) was considered to exemplify the use of various diagnostic techniques in linear regression. The

main aim of the study was to characterize those variables that influence the aerial biomass production of the marsh grass Spartina alterniflora using 45 soil samples from 5 random sites. In this work we consider the model analyzed in Meloun and Militký (2001), who assumed that the response variable, Y biomass aerial (g/m^2), is related to five physicochemical properties of the substrate: x_1 salinity (‰), x_2 acidity as measured in water pH, x_3 potassium (ppm), x_4 sodium (ppm), and x_5 zinc (ppm).

Table 5 Parameter estimates, standard errors (in parenthesis) and percentage change in parameter estimates for Aerial biomass data.

Parameter	Full	Removed observations				
	data	5	7	11	12	14
β_0	121.907	91.522	94.177	337.617	528.834	99.278
	(92.004)	(70.039)	(72.055)	(261.619)	(383.275)	(76.651)
	—	-24.92%	-22.75%	176.94%	333.80%	-18.56%
β_1	-8.800	-9.138	-7.481	-12.920	-23.619	-14.804
	(9.147)	(9.566)	(9.338)	(10.301)	(11.281)	(9.144)
	—	3.85%	-14.98%	46.83%	168.41%	68.23%
β_2	364.885	372.445	360.907	355.588	368.315	384.519
	(44.692)	(48.196)	(46.432)	(47.796)	(48.691)	(43.334)
	—	2.07%	-1.09%	-2.55%	0.94%	5.38%
β_3	-0.177	-0.329	-0.129	-0.198	-0.114	-0.173
	(0.315)	(0.485)	(0.342)	(0.317)	(0.300)	(0.301)
	—	85.14%	-27.12%	11.33%	-35.66%	-2.52%
β_4	-0.014	-0.007	-0.016	-0.014	-0.014	-0.012
	(0.014)	(0.022)	(0.015)	(0.014)	(0.013)	(0.013)
	—	-49.72%	10.68%	-4.581	-0.22%	-13.83%
β_5	-9.097	-8.242	-9.302	-11.341	-11.604	-5.841
	(6.821)	(7.028)	(6.978)	(7.375)	(7.565)	(6.637)
	—	-9.40%	2.26%	24.67%	27.57%	-35.79%
[1]σ^2	1.436	1.475	1.474	1.433	1.254	1.314
	—	2.72%	2.66%	-0.20%	-12.63%	-8.45%
λ	1.220	1.601	1.506	0.346	0.193	1.414
	—	31.20%	23.48%	-71.60%	-84.17%	15.92%
edf	5.060	5.029	5.033	5.205	5.345	5.046
	—	-0.61%	-0.55%	2.87%	5.64%	-0.28%
[1]det(Cov($\widehat{\boldsymbol{\beta}}_\lambda$))	0.144	0.248	0.121	1.278	1.510	0.072
	—	72.44%	-15.69%	788.57%	950.10%	-49.92%

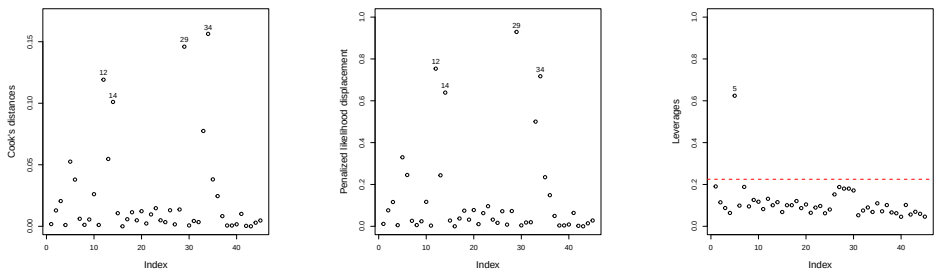
[1] values multiplied by 10^5.

Following Meloun and Militký (2001), we fit a linear regression model with intercept. The scaled condition number is given by $\kappa(\boldsymbol{X}) = 58.977$ which is an indication of strong collinearity, a result that agrees with that presented in Section 11.4.3 of Rawlings et al. (1998). The fit using ridge regression is presented in the second column of Table 5, whereas the ordinary least squares fit can be consulted in Meloun and Militký (2001). From Figure 5 we note that the Cook's distance, the penalized likelihood displacement allow us to identify observations 12, 14, 29 and 34 as influential on the regression coefficients. Such observations have also been detected by Meloun and Militký (2001) using diagnostic techniques for LS estimation.

Table 6 Parameter estimates, standard errors (in parenthesis) and percentage change in parameter estimates for Aerial biomass data.

Parameter	Full	Removed observations				
	data	15	29	30	33	34
β_0	121.907	77.666	468.516	125.931	293.648	49.173
	(92.004)	(61.222)	(346.402)	(95.240)	(226.974)	(49.228)
	—	-36.29%	284.32%	3.30%	140.88%	-59.66%
β_1	-8.800	-8.563	-16.559	-8.958	-7.521	-1.986
	(9.147)	(9.286)	(10.584)	(9.262)	(9.682)	(8.233)
	—	-2.69%	88.18%	1.80%	-14.53%	-77.43%
β_2	364.885	366.748	378.537	366.769	337.685	350.156
	(44.692)	(44.872)	(48.927)	(47.030)	(45.292)	(39.907)
	—	0.51%	3.74%	0.52%	-7.45%	-4.04%
β_3	-0.177	-0.149	-0.280	-0.183	-0.206	-0.163
	(0.315)	(0.320)	(0.300)	(0.320)	(0.297)	(0.284)
	—	-16.08%	57.93%	3.08%	15.85%	-8.27%
β_4	-0.014	-0.015	-0.013	-0.014	-0.014	-0.019
	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)
	—	2.75%	-9.48%	-0.38%	-0.39%	36.75%
β_5	-9.097	-8.702	-14.371	-9.270	-13.730	-10.067
	(6.821)	(6.864)	(7.449)	(6.968)	(6.938)	(6.127)
	—	-4.34%	57.98%	1.90%	50.94%	10.67%
$^1\sigma^2$	1.436	1.475	1.281	1.466	1.274	1.173
	—	2.73%	-10.76%	2.14%	-11.27%	-18.27%
λ	1.220	1.800	0.228	1.185	0.402	2.100
	—	47.56%	-81.30%	-2.88%	-67.09%	72.15%
edf	5.060	5.021	5.308	5.061	5.199	5.014
	—	-0.77%	4.90%	0.01%	2.74%	-0.91%
$^1\det(\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_\lambda))$	0.144	0.076	1.452	0.207	0.543	0.016
	—	-47.12%	909.50%	43.60%	277.36%	-89.13%

[1] values multiplied by 10^5.



(a) (b) (c)

Fig. 5 Influence measures for Aerial biomass data: (a) Cook's distances, D_i^{**}, (b) penalized likelihood displacement, $LD_i(\boldsymbol{\beta}|\sigma^2)$, and (c) index plot of leverages, $h_{ii}(\lambda)$.

As can be seen from Figure 5 (c) observation 5 is identified as a leverage, moreover is labeled as influential using the Kullback-Leibler divergence (see Figure 7). Observations 12, 14, 33 and 34 are identified as extreme by the local influence procedure

based on $LD(\boldsymbol{\omega})$ (see Figures 6 (a) and (b)). In addition, the direction of the greatest local change on the GCV criterion depicted in Figures 6 (c) and (d) allows us to note that observations 7, 11, 12, 15 and 34 have a strong impact on the selection of the ridge parameter (compare with Tables 5 and 6).

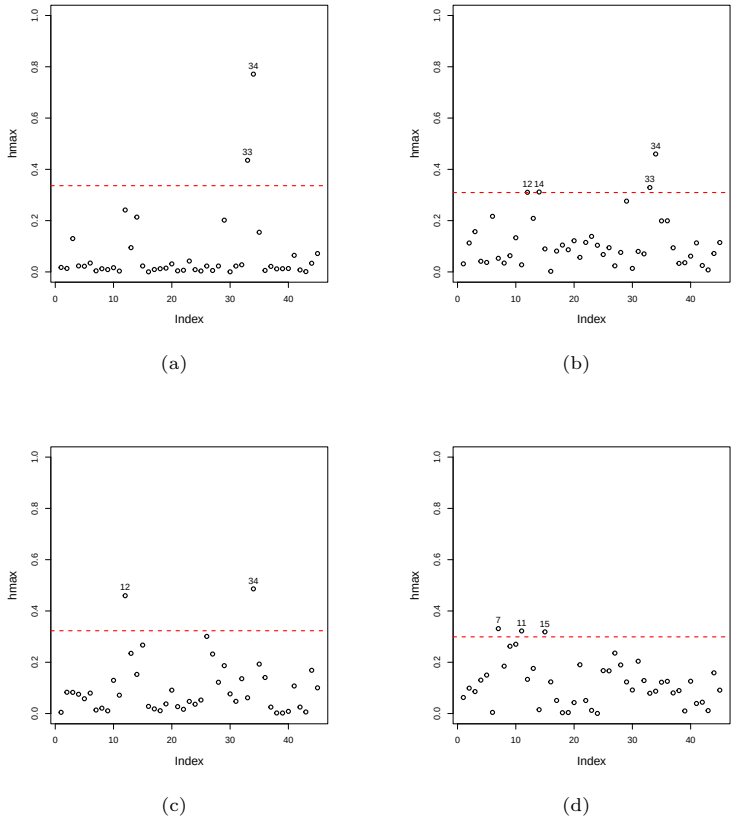


Fig. 6 Influence measures for Aerial biomass data: Index plot of $\boldsymbol{h}_{\max}$ based on $LD(\boldsymbol{\omega})$ under (a) perturbation of variances, (b) response perturbation. Index plot of $\boldsymbol{h}_{\max}$ based on $\mathrm{GCV}(\lambda, \boldsymbol{\omega})$ under (c) perturbation of variances, and (d) response perturbation.

Assessment of local influence based on Kullback-Leibler divergence identifies observations 29 and 30 as influential (see Figure 8 (a)). The brief confirmatory analysis shown in Tables 5 and 6 present the percentages of change when some selected observations are removed from the dataset. This allows us to verify the role played by each of these extreme data. For example, it can be noted that removing observation 29 increases the determinant of the covariance matrix of the estimated coefficients by 909.50%. It is remarkable to note that although observation 30 individually exerts an increase of 43.60% on the determinant of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)$. When both observations are removed, i.e., 29 and 30, this relative change rises to 1903.26%, which allows us to

realize the extreme nature of these cases. Therefore, it is interesting to note that observation 30 is not identified by traditional diagnostic methods. In fact, using traditional diagnostic methods, observation 29 hides the influence of observation 30, whereas the local influence technique is able to reveal their joint effect on the covariance of the regression coefficients. Furthermore, it is interesting to note from Figure 8 (b) that the response perturbation scheme on $KL(\boldsymbol{\omega})$ leads to identify observation 5 as well as observations 7, 11, and 15 as strongly influential. In fact, removing the 11th observation increases the covariance of the coefficients by 788.57%. This result is consistent with the observations identified in Figure 6 (b).

Figure 9 presents the influence plot considering the perturbation of explanatory variables. We can notice that observations 27, 28 and 29 (Figure 9 (a)) have impact on x_1, whereas cases 32, 34 and 35 exert a strong influence on x_3 and x_4 (see Figure 9 (b), (c)), observations that, as reported in the previous figures, affect different aspects of the proposed model. It is interesting to note that when x_5 is perturbed, observations 37 and 44 are detected. To the best of our knowledge, these cases have not been detected by other diagnostic methods.

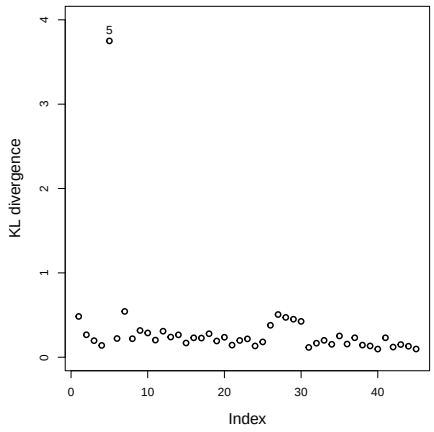
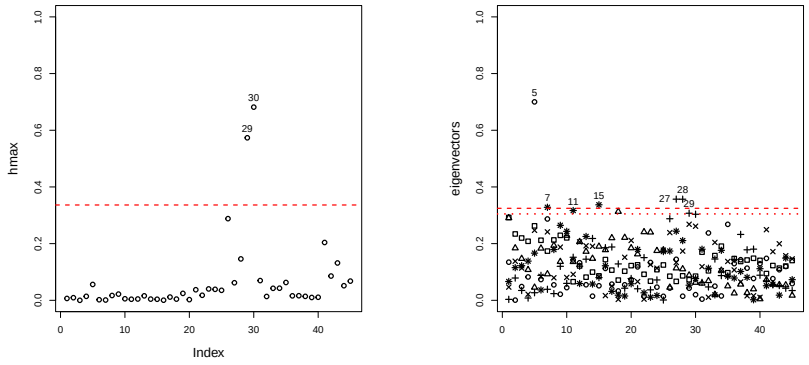


Fig. 7 Index plot of KL_i for Aerial biomass data.

5 Concluding remarks

In this paper we have explored two perspectives to perform the influence diagnostics based on the Kullback-Leibler divergence considering the distributions associated to the estimators for the case-deletion model as well as the local influence analysis under different perturbation schemes. Previous experience using divergence measures to perform diagnostic analyses has only considered techniques based on case-deletion. To the best of our knowledge, the assessment of local influence based on divergence measures has been studied only from a Bayesian perspective, while the approach proposed



(a) (b)

Fig. 8 Index plot of $\boldsymbol{h}_{\max}$ based on $KL(\boldsymbol{\omega})$ under (a) perturbation of variances, and (b) response perturbation, Aerial biomass data.

in this paper has not been considered previously. It is noteworthy that the measures introduced in this work have allowed the identification of observations that were not determined using traditional diagnostic methods and thus represent a valuable complement to traditional procedures.

As noted in Remark 2, other diagnostic measures based on the Kullback-Leibler divergence can be proposed. For example, in our context of ridge regression, Jeffreys' divergence adopts the form

$$J_i = \frac{1}{2}\boldsymbol{\delta}^\top\{\mathrm{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda) + \mathrm{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i))\}\boldsymbol{\delta} - p \\ + \frac{1}{2}\,\mathrm{tr}\{\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)\,\mathrm{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda(i)) + \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(i))\,\mathrm{Cov}^{-1}(\widehat{\boldsymbol{\beta}}_\lambda)\}.$$

However, one aspect in which this measure differs with (8) is the term that depends on the ratio of determinants $|\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_\lambda)|/|\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_\lambda(i))|$, which is related to the COVRATIO statistic. Exploring the connections that may exist between the diagnostic measures developed in this work with existing measures following a perspective such as that adopted by Schall and Dunne (1992) deserves to be addressed. Along these lines it could be of particular interest to adapt these types of diagnostic techniques based on divergence measures for the detection of collinearity-influential data points. Several studies (Mason and Gunst, 1985; Hadi, 1988; Walker, 1989; Hadi and Wells, 1990; Wang and Nyquist, 1991) have warned that certain extreme observations can exert a strong effect on the eigenstructure that originates from the model matrix. The main conclusion of such studies is that observations with high leverage are precisely those that can strongly affect the conditioning of the model matrix. In particular, our numerical experiments provide some evidence that the proposed methodology is capable of identifying collinearity-influential observations. This kind of development is noteworthy, since in fact Hadi and Velleman (1987) points out in the discussion of

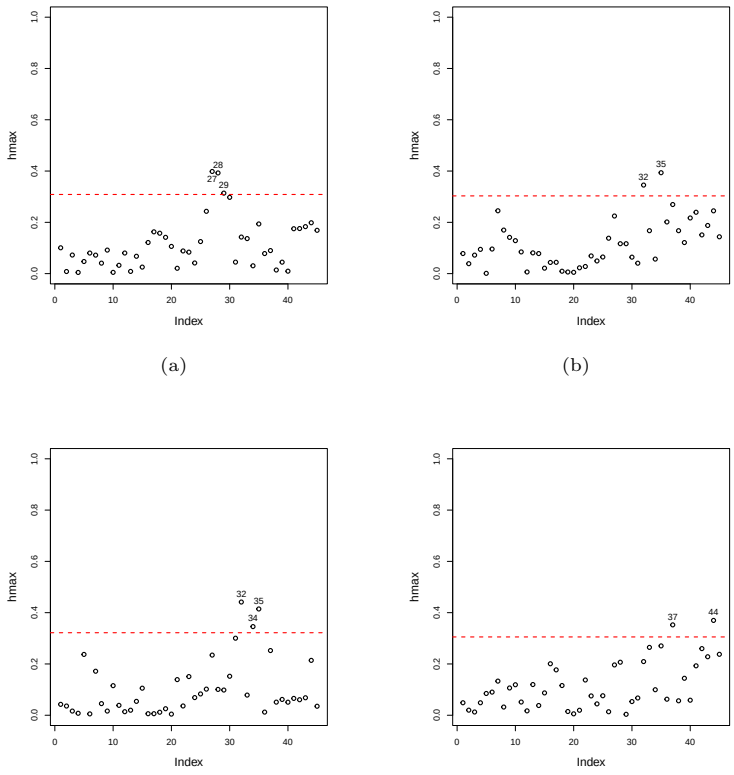


Fig. 9 Index plot of $\boldsymbol{h}_{\max}$ based on $KL(\boldsymbol{\omega})$ under perturbation of explanatory variables, for (a) x_1 salinity, (b) x_3 potassium, (c) x_4 sodium, and (d) x_5 zinc.

Stewart (1987) that measures to detect the presence of collinearity can be affected by observations with high leverage as well as collinearity-influential observations.

An interesting aspect pointed out to us by one of the referees corresponds to the evaluation of the local influence on the choice of the ridge parameter as developed in Shi and Wang (1999). In this paper, we have addressed the effect of atypical observations on the selection of λ considering the GCV criterion. Furthermore, a very relevant aspect of our developments has been to study whether each of the considered perturbation schemes are appropriate according to the perspective outlined in Zhu et al. (2007).

We should stress that the threshold we used in the local influence plots is quite simple. Our aim is to improve the benchmark choice following the methodology proposed by Shi and Huang (2011) and use the detection strategy described by Aoki et al. (2022, 2023). These developments will also be included in the india package (Osorio, 2023) for R.

Data availability

The replication files related to this article are available online at https://github.com/faosorios/KL_influence

Acknowledgments

This work was written while the second author was at Universidad Técnica Federico Santa María. The support of the computing infrastructure of the Applied Laboratory of Spatial Analysis UTFSM - ALSA (MT7018) is acknowledged. The authors thank Ronny Vallejos and Michael Karkulik for their insightful discussions. Additionally, the authors thank an associate editor and two anonymous reviewers for their thorough review of the manuscript, which helped us to enhance the clarity and quality of this work.

ORCID

Alonso Ogueda https://orcid.org/0000-0001-5411-1484
Felipe Osorio https://orcid.org/0000-0002-4675-5201

Appendix A Local influence based on the penalized likelihood for ridge regression

In this appendix we derive the differentials $\mathrm{d}^2_\theta\, \ell_\lambda(\boldsymbol{\theta})$ and $\mathrm{d}^2_{\theta\omega}\, \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega})$ for the perturbation schemes introduced in Section 3.2. The necessary curvature matrices

$$\boldsymbol{F}_{LD} = \boldsymbol{\Delta}^\top(\boldsymbol{\omega}_0)\{-\ddot{\ell}_\lambda(\widehat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}(\boldsymbol{\omega}_0),$$

where

$$\ddot{\ell}_\lambda(\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 \ell_\lambda(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\theta=\widehat{\theta}_\lambda}, \qquad \boldsymbol{\Delta}(\boldsymbol{\omega}_0) = \frac{\partial^2 \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}\Big|_{\theta=\widehat{\theta}_\lambda, \omega=\omega_0},$$

are obtained using the differentiation method and by applying some identification theorems discussed in Magnus and Neudecker (2019).

A.1 Observed information matrix

Following Osorio (2016), we consider the penalized likelihood displacement as an influence function, i.e,

$$LD(\boldsymbol{\omega}) = 2\{\ell_\lambda(\widehat{\boldsymbol{\theta}}_\lambda) - \ell_\lambda(\widehat{\boldsymbol{\theta}}_\lambda(\boldsymbol{\omega}))\},$$

where

$$\ell_\lambda(\boldsymbol{\theta}) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\big(\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2\big).$$

Obtaining the first differential of $\ell_\lambda(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ and σ^2, it follows that

$$\mathrm{d}_\beta\, \ell_\lambda(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\{\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) - \lambda\boldsymbol{\beta}\}^\top \mathrm{d}\,\boldsymbol{\beta}$$

$$\mathrm{d}_{\sigma^2}\,\ell_\lambda(\boldsymbol{\theta}) = \Big\{ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\big(\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2\big)\Big\}\,\mathrm{d}\,\sigma^2.$$

Differentiating again in relation to $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$, we arrive at

$$\begin{aligned}
\mathrm{d}^2_{\beta}\,\ell_\lambda(\boldsymbol{\theta}) &= -\frac{1}{\sigma^2}(\mathrm{d}\,\boldsymbol{\beta})^\top(\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I}_p)\,\mathrm{d}\,\boldsymbol{\beta},\\
\mathrm{d}^2_{\sigma^2\beta}\,\ell_\lambda(\boldsymbol{\theta}) &= -\frac{1}{\sigma^4}\,\mathrm{d}\,\sigma^2\{\boldsymbol{X}^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) - \lambda\boldsymbol{\beta}\}^\top\,\mathrm{d}\,\boldsymbol{\beta}\\
\mathrm{d}^2_{\sigma^2}\,\ell_\lambda(\boldsymbol{\theta}) &= \Big\{\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\big(\|\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2\big)\Big\}\,\mathrm{d}\,\sigma^2\,\mathrm{d}\,\sigma^2.
\end{aligned}$$

Noting that $\widehat{\boldsymbol{\beta}}_\lambda = (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$, leads to $\boldsymbol{X}^\top(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda) - \lambda\widehat{\boldsymbol{\beta}}_\lambda = \boldsymbol{0}$, which allows us to write

$$\mathrm{d}^2_{\sigma^2\beta}\,\ell_\lambda(\boldsymbol{\theta})\big|_{\theta=\widehat{\theta}_\lambda} = \boldsymbol{0}.$$

In addition

$$\mathrm{d}^2_{\sigma^2}\,\ell_\lambda(\boldsymbol{\theta})\big|_{\theta=\widehat{\theta}_\lambda} = \Big(\frac{n}{2\widehat{\sigma}^4_\lambda} - \frac{n\widehat{\sigma}^2_\lambda}{\widehat{\sigma}^6_\lambda}\Big)\,\mathrm{d}\,\sigma^2\,\mathrm{d}\,\sigma^2 = -\frac{n}{2\widehat{\sigma}^4_\lambda}\,\mathrm{d}\,\sigma^2\,\mathrm{d}\,\sigma^2.$$

Using the second identification theorem of Magnus and Neudecker (2019), yields

$$-\ddot{\ell}_\lambda(\boldsymbol{\theta})\big|_{\theta=\widehat{\theta}_\lambda} = -\frac{\partial^2\ell_\lambda(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\theta=\widehat{\theta}_\lambda} = \frac{1}{\widehat{\sigma}^2_\lambda}\begin{pmatrix}\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I}_p & \boldsymbol{0}\\ \boldsymbol{0} & \frac{n}{2\widehat{\sigma}^2_\lambda}\end{pmatrix}.$$

A.2 Perturbation schemes

Next, we derive matrix $\boldsymbol{\Delta}(\boldsymbol{\omega}_0)$ for different perturbation schemes. For each case, we obtain the second differential $\mathrm{d}^2_{\theta\omega}\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})$. It is straightforward to note that evaluating these differentials at $\boldsymbol{\theta} = \widehat{\theta}_\lambda$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ leads to $\boldsymbol{\Delta}(\boldsymbol{\omega}_0) = \big(\boldsymbol{\Delta}^\top_\beta(\boldsymbol{\omega}_0), \boldsymbol{\Delta}^\top_{\sigma^2}(\boldsymbol{\omega}_0)\big)^\top$, where

$$\boldsymbol{\Delta}_\beta(\boldsymbol{\omega}_0) = \frac{\partial^2\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\Big|_{\theta=\widehat{\theta}_\lambda,\omega=\omega_0}, \qquad \boldsymbol{\Delta}_{\sigma^2}(\boldsymbol{\omega}_0) = \frac{\partial^2\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})}{\partial\sigma^2\partial\boldsymbol{\omega}^\top}\Big|_{\theta=\widehat{\theta}_\lambda,\omega=\omega_0},$$

A.2.1 Perturbation of variances

For the perturbation scheme defined in Equation (9), the perturbed log-likelihood function is given by

$$\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega}) = -\frac{n}{2}\log 2\pi\sigma^2 + \frac{1}{2}\log|\boldsymbol{W}| - \frac{1}{2\sigma^2}\big\{(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|^2\big\}.$$

Thus, the second differential of $\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})$ in relation to $\boldsymbol{\beta}$ and σ^2, assume the form

$$\mathrm{d}^2_{\beta\omega}\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega}) = \frac{1}{\sigma^2}(\mathrm{d}\,\boldsymbol{\beta})^\top\boldsymbol{X}^\top\,\mathrm{diag}(\boldsymbol{\epsilon})\,\mathrm{d}\,\boldsymbol{\omega},$$

$$\mathrm{d}^2_{\sigma^2\omega}\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega}) = \frac{1}{2\sigma^4}\,\mathrm{d}\,\sigma^2\boldsymbol{\epsilon}^\top\,\mathrm{diag}(\boldsymbol{\epsilon})\,\mathrm{d}\,\boldsymbol{\omega}.$$

The above allows us to write,

$$\boldsymbol{\Delta}_\beta(\boldsymbol{\omega}_0) = \frac{1}{\widehat{\sigma}^2_\lambda}\boldsymbol{X}^\top\,\mathrm{diag}(\boldsymbol{e}_\lambda), \qquad \boldsymbol{\Delta}_{\sigma^2}(\boldsymbol{\omega}_0) = \frac{1}{\widehat{\sigma}^4_\lambda}\boldsymbol{e}_\lambda^\top\,\mathrm{diag}(\boldsymbol{e}_\lambda),$$

where $\boldsymbol{e}_\lambda = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda$.

Obtaining the second differential with relation to $\boldsymbol{\omega}$ and taking expectations we arrive to,

$$\begin{aligned}\mathrm{E}\{-\mathrm{d}^2_\omega\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})\} &= \frac{1}{2}\,\mathrm{tr}\,\boldsymbol{W}^{-1}(\mathrm{d}\,\boldsymbol{W})\boldsymbol{W}^{-1}\,\mathrm{d}\,\boldsymbol{W}\\ &= \frac{1}{2}(\mathrm{d}\,\mathrm{vec}\,\boldsymbol{W})^\top(\boldsymbol{W}^{-1}\otimes\boldsymbol{W}^{-1})\,\mathrm{d}\,\mathrm{vec}\,\boldsymbol{W}.\end{aligned}$$

Evaluating at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ $(= \mathbf{1}_n)$, and using the second identification theorem stated in Magnus and Neudecker (2019), it follows that

$$\boldsymbol{G}(\boldsymbol{\omega}_0) = \mathrm{E}\left\{-\frac{\partial^2\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top}\right\}\Big|_{\omega=\omega_0} = \tfrac{1}{2}\,\boldsymbol{I}_n.$$

This allows to verify that this perturbation scheme is appropriate under the terms defined by Zhu et al. (2007).

A.2.2 Response perturbation

In the case of the perturbation of observed responses, i.e., $\boldsymbol{Y}(\boldsymbol{\omega}) = \boldsymbol{Y} + \boldsymbol{\omega}$ we obtain that the perturbed log-likelihood function becomes

$$\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega}) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left\{\|\boldsymbol{Y}+\boldsymbol{\omega}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2\right\}.$$

Using the differentiation method, we have that

$$\begin{aligned}\mathrm{d}^2_{\beta\omega}\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega}) &= \frac{1}{\sigma^2}(\mathrm{d}\,\boldsymbol{\beta})^\top\boldsymbol{X}^\top\,\mathrm{d}\,\boldsymbol{\omega},\\ \mathrm{d}^2_{\sigma^2\omega}\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega}) &= \frac{1}{2\sigma^4}\,\mathrm{d}\,\sigma^2(\boldsymbol{Y}-\boldsymbol{\omega}-\boldsymbol{X}\boldsymbol{\beta})^\top\,\mathrm{d}\,\boldsymbol{\omega}.\end{aligned}$$

This leads to,

$$\boldsymbol{\Delta}_\beta(\boldsymbol{\omega}_0) = \frac{1}{\widehat{\sigma}^2_\lambda}\boldsymbol{X}^\top, \qquad \boldsymbol{\Delta}_{\sigma^2}(\boldsymbol{\omega}_0) = \frac{1}{\widehat{\sigma}^4_\lambda}\boldsymbol{e}_\lambda^\top,$$

where $\boldsymbol{e}_\lambda = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda$. It is straightforward to verify that

$$\mathrm{E}\{-\mathrm{d}^2_\omega\,\ell_\lambda(\boldsymbol{\theta},\boldsymbol{\omega})\} = \frac{1}{\sigma^2}(\mathrm{d}\,\boldsymbol{\omega})^\top\,\mathrm{d}\,\boldsymbol{\omega}.$$

That is

$$\boldsymbol{G}(\boldsymbol{\omega}_0) = \mathrm{E}\Big\{ -\frac{\partial^2 \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top} \Big\}\Big|_{\omega=\omega_0} = \frac{1}{\sigma^2}\, \boldsymbol{I}_n,$$

from this it follows that the perturbation scheme is the appropriate one.

A.2.3 Perturbation of the explanatory variables

For the perturbed model in (10), the log-likelihood function adopts the form

$$\ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega}) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\big\{\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - a\boldsymbol{\omega}\boldsymbol{c}_t^\top\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2\big\}.$$

It is possible to show that

$$\mathrm{d}^2_{\beta\omega}\, \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega}) = \frac{a}{\sigma^2}(\mathrm{d}\,\boldsymbol{\beta})^\top \big\{\boldsymbol{c}_t(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{\omega})\boldsymbol{\beta})^\top - \boldsymbol{c}_t^\top\boldsymbol{\beta}\boldsymbol{X}(\boldsymbol{\omega})\big\}^\top \mathrm{d}\,\boldsymbol{\omega},$$

$$\mathrm{d}^2_{\sigma^2\omega}\, \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega}) = -\frac{a\boldsymbol{c}_t^\top\boldsymbol{\beta}}{\sigma^4}\,\mathrm{d}\,\sigma^2(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - a\boldsymbol{\omega}\boldsymbol{c}_t^\top\boldsymbol{\beta})^\top\,\mathrm{d}\,\boldsymbol{\omega}.$$

Using the second identification theorem in Magnus and Neudecker (2019), it follows that

$$\boldsymbol{\Delta}_\beta(\boldsymbol{\omega}_0) = \frac{a}{\widehat{\sigma}^2_\lambda}(\boldsymbol{e}_\lambda\boldsymbol{c}_t^\top - \boldsymbol{c}_t^\top\widehat{\boldsymbol{\beta}}_\lambda\boldsymbol{X})^\top, \qquad \boldsymbol{\Delta}_{\sigma^2}(\boldsymbol{\omega}_0) = -\frac{a\boldsymbol{c}_t^\top\widehat{\boldsymbol{\beta}}_\lambda}{\widehat{\sigma}^4_\lambda}\boldsymbol{e}_\lambda^\top,$$

where $\boldsymbol{e}_\lambda = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda$.

To evaluate whether the perturbation scheme is appropriate note that,

$$\mathrm{d}^2_\omega\, \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega}) = -\frac{a^2(\boldsymbol{c}_t^\top\boldsymbol{\beta})}{\sigma^2}(\mathrm{d}\,\boldsymbol{\omega})^\top\,\mathrm{d}\,\boldsymbol{\omega}.$$

It follows that the information matrix with respect to $\boldsymbol{\omega}$ is given by

$$\boldsymbol{G}(\boldsymbol{\omega}_0) = \mathrm{E}\Big\{ -\frac{\partial^2 \ell_\lambda(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top} \Big\}\Big|_{\omega=\omega_0} = \frac{a^2\beta_t^2}{\sigma^2}\, \boldsymbol{I}_n,$$

where $\beta_t = \boldsymbol{c}_t^\top\boldsymbol{\beta}$, and therefore the perturbation scheme is appropriate.

References

Aoki R, Bustamante JPM, Paula GA (2022) Local influence diagnostics with forward search in regression analysis. Statistical Papers 63:1477-1497.

Aoki R, Bustamante JPM, Russo CM, Paula GA (2023) Conformal normal curvature and detection of masked observations in multivariate null intercept measurement error models. Journal of Applied Statistics 51:1545-1569.

Belsley DA (1991) Conditioning Diagnostics: Collinearity and Weak Data in Regression. Wiley, New York.

Belsley DA, Kuh E, Welsch RE (1980) Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley, New York.

Billor N, Loynes RM (1999) An application of the local influence approach to ridge regression. Journal of Applied Statistics 26:177-183.

Chatterjee S, Hadi AS (1988) Sensitivity Analysis in Linear Regression. Wiley, New York.

Cook RD (1977) Detection of influential observations in linear regression. Technometrics 19:15-18.

Cook RD (1986) Assessment of local influence (with discussion). Journal of the Royal Statistical Society, Series B 48:133-169.

Cook RD, Weisberg S (1980) Characterizations of an empirical influence function for detecting influential cases in regression. Technometrics 22:495-508.

Cook RD, Peña D, Weisberg S (1988) The likelihood displacement: A unifying principle for influence measures. Communications in Statistics - Theory & Methods 17:623-640.

Davies L, Gather U (1993) The identification of multiple outliers. Journal of the American Statistical Association 88:782-792.

Emami H (2018) Local influence for Liu estimators in semiparametric linear models. Statistical Papers 59:529-544.

Emami H, Emami M (2016) New influence diagnostics in ridge regression. Journal of Applied Statistics 43:476-489.

Fung W, Kwan C (1997) A note on local influence based on normal curvature. Journal of the Royal Statistical Society, Series B 59:839-843.

Galea M, Paula GA, Uribe-Opazo M (2003) On influence diagnostics in univariate elliptical linear regression models. Statistical Papers 44:23-45.

García-Heras J, Muñoz-García J, Muñoz-Pichardo JM, Pardo L (2006) Influence measures based on Cressie-Read divergence measures in multivariate linear model. Communications in Statistics - Theory and Methods 35:2055-2073.

Geisser S (1996) Discussion of a paper by R.D. Cook. Journal of the Royal Statistical Society, Series B 48:163.

Golub G, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21:215-223.

Gorman JW, Toman RJ (1966) Selection of variables for fitting equations to data. Technometrics 8:27-51.

Hadi AS (1988) Diagnosing collinerity-influential observations. Computational Statistics & Data Analysis 7:143-159.

Hadi AS, Velleman PF (1987) Discussion of a paper by G. W. Stewart. Statistical Science 2:93-98.

Hadi AS, Wells MT (1990) Assessing the effects of multiple rows on the condition number of a matrix. Journal of the American Statistical Association 85:786-792.

Hadi AS, Nyquist H (1993) Further theoretical results and a comparison between two methods for approximating eigenvalues of perturbed covariance matrices. Statistics and Computing 3:113-123.

Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55-67.

Hoerl AE, Kannard RW, Baldwin KF (1975). Ridge regression: Some simulations. Communications in Statistics - Theory and Methods 4:105-123.

Jeffreys H (1946) An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London, Series A 186:453-461.

Johnson W (1985) Influence measures for logistic regression: Another point of view. Biometrika 72:59-65.

Johnson W, Geisser S (1983) A predictive view of the detection and characterization of influential observations in regression. Journal of the American Statistical Association 78:137-144.

Kaçıranlar S, Sakallıoğlu S, Akdeniz F, Styan GPH, Werner HJ (1999) A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on Portland cement. Sankhyā, Series B 61:443-459.

Kibria BMG (2022) More than hundred (100) estimators for estimating the shrinkage parameter in a linear and generalized linear ridge regression models. Journal of Econometrics and Statistics 2:233-252.

Kowalski KG (1990) On the predictive performance of biased regression methods and multiple linear regression. Chemometrics and Intelligent Laboratory Systems 9:177-184.

Linthurst RA (1979). Aeration, nitrogen, pH and salinity as factors affecting Spartina Alterniflora growth and dieback. PhD thesis, North Carolina State University.

Liu S (2000) On local influence for elliptical linear models. Statistical Papers 41:211-224.

Liu S, Trenkler G, Kollo T, von Rosen D, Baksalary OM (2024) Professor Heinz Neudecker and matrix differential calculus. Statistical Papers 65:2605-2639.

Lukman AF, Ayinde K, Binuomote S, Clement OA (2019) Modified ridge-type estimator to combat multicollinearity: Application to chemical data. Journal of Chemometrics 33:e3125.

Magnus JR, Neudecker H (2019) Matrix Differential Calculus with Applications in Statistics and Econometrics (3rd Ed.). Wiley, Chichester.

Mason RL, Gunst RF (1985) Outlier-induced collinearities. Technometrics 27:401-407.

Meloun M, Militký J (2001) Detection of single influential points in OLS regression model building. Analytica Chimica Acta 439:169-191.

Muñoz-García J, Muñoz-Pichardo JM, Pardo L (2006) Cressie and Read power-divergences as influence measures for logistic regression models. Computational Statistics & Data Analysis 50:3199-3221.

Nel D (1980) On matrix differentiation in statistics. South African Statistical Journal 14:137-193.

Osorio F (2016) Influence diagnostics for robust P-splines using scale mixture of normal distributions. Annals of the Institute of Statistical Mathematics 68:589-619.

Osorio F (2023) india: Influence diagnostics in statistical models. R package version 0.1. https://doi.org/10.32614/CRAN.package.india

Osorio F, Ogueda A (2024) fastmatrix: Fast computation of some matrices useful in statistics. R package version 0.5-772. https://doi.org/10.32614/CRAN.package.fastmatrix

Pan JX, Fang KT, Liski EP (1996) Bayesian local influence for the growth curve model with Rao's simple covariance structure. Journal of Multivariate Analysis 58:55-81.

Pan JX, Fang KT, von Rosen D (1999) Bayesian local influence for the growth curve model with unstructured covariance. Biometrical Journal 41:641-658.

Pan JX, Fung WK (2000) Bayesian influence assessment in the growth curve model with unstructured covariance. Annals of the Institute of Statistical Mathematics 52:737-752.

Poon W, Poon Y (1999) Conformal normal curvature and assessment of local influence. Journal of the Royal Statistical Society, Series B 61:51-61.

Pregibon D (1981) Logistic regression diagnostics. The Annals of Statistics 9:705-724.

Rawlings JO, Pantula SG, Dickey DA (1998) Applied Regression Analysis: A Research Tool. 2nd Ed. Springer, New York.

Schall R, Dunne TT (1992) A note on the relationship between parameter collinearity and local influence. Biometrika 79:399-404.

Shi JQ, Wei BC (1995) Bayesian local influence. Mathematica Applicata 8:237-245.

Shi L (1997) Local influence in principal components analysis. Biometrika 84:175-186.

Shi L, Wang X (1999) Local influence in ridge regression. Computational Statistics & Data Analysis 31:341-353.

Shi L, Huang M (2011) Stepwise local influence analysis. Computational Statistics & Data Analysis 55:973-982.

Steece BN (1986) Regression space outliers in ridge regression. Communications in Statistics - Theory & Methods 15:3599-3605.

Stewart GW (1987) Collinearity and least squares regression (with discussion). Statistical Science 2:68-100.

Thomas, W (1991) Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. Journal of the American Statistical Association 86:693-698.

Ullah A (1996) Entropy, divergence and distance measures with econometric applications. Journal of Statistical Planning and Inference 49:137-162.

Walker E (1989) Detection of collinearity-influential observations. Communications in Statistics: Theory & Methods 18:1675-1690.

Walker E, Birch JB (1988) Influence measures in ridge regression. Technometrics 30:221-227.

Wang SG, Nyquist H (1991) Effects on the eigenstructure of a data matrix when deleting an observation. Computational Statistics & Data Analysis 11:179-188.

Wei BC, Shih JQ (1994) On statistical models for regression diagnostics. Annals of the Institute of Statistical Mathematics 46:267-278.

Woods H, Steinour H, Starke H (1932) Effect of composition of portland cement on heat evolving during hardening. Industrial & Engineering Chemistry 24:1207-1214.

Wu X, Luo Z (1993) Second, order approach to local influence. Journal of the Royal Statistical Society, Series B 55:929-936.

Zhu H, Ibrahim JG, Lee S, Zhang H (2007) Perturbation selection and influence measures in local influence analysis. The Annals of Statistics 35:2565-2588.