Chilean Journal of Statistics
Vol. 1, No. 2, September 2010, 17–34

Special Issue "In Memory of Pilar Loreto Iglesias Zuazola"
Research Paper

Influential observations in the independent Student-t measurement error model with weak nondifferential error

Ignacio Vidal[1,*] and Luis M. Castro[2]

[1]Instituto de Matemática y Física, Universidad de Talca, Talca, Chile,
[2]Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas,
Universidad de Concepción, Concepción, Chile.

(Received: 23 March 2010 · Accepted in final form: 11 August 2010)

Abstract

As in regression analysis, inference in measurement error models (MEM) can be strongly modified by the inclusion or deletion of a small set of observations. Such observations are called influential data. In this work, we present different influence measures based on the Bayes risk and the q-divergence. These measures quantify the influence of a small subset of the data on the posterior distribution for the structural parameters of the independent Student-t MEM with weak nondifferential error. The advantage of the influence measures presented in this work is that we can compute them for any subset of data by using only one sample drawn from the posterior distribution. The samples from the posterior distributions are obtained through Gibbs sampler algorithm, assuming specific proper prior distributions. The Bayesian identifiability of the independent Student-t MEM with weak nondifferential error is also discussed. Finally, the results are illustrated with applications on two well-known real data sets.

Keywords: Bayesian analysis · Influential observations · Measurement error models · MCMC methods · Student-t distribution.

Mathematics Subject Classification: Primary 62F15 · Secondary 62J99.

1. Introduction

A simple measurement error model (MEM) is given by

$$\mathbf{y} = \boldsymbol{\Sigma}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

and

$$\mathbf{x} = \boldsymbol{\xi} + \mathbf{u}\,, \tag{2}$$

where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\top$, $\boldsymbol{\Sigma}$ is the $n \times 2$ matrix $(\mathbb{1}_n\ \boldsymbol{\xi})$, with $\mathbb{1}_n$ being a vector of n ones and $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. Here, $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{x} = (x_1, \ldots, x_n)^\top$ are the data and

*Corresponding author. Ignacio Vidal. Instituto de Matemática y Física, Universidad de Talca, Casillas 747 ó 721, Talca, Chile. Email: ividal@inst-mat.utalca.cl

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ and $\mathbf{u} = (u_1, \ldots, u_n)^\top$ are the error terms. A normal distribution usually is assumed for errors, for example, $\boldsymbol{\epsilon}|\,\phi_1 \sim \mathrm{N}_n\left(\mathbf{0}, \phi_1^{-1}\mathbf{I}_n\right)$ and $\mathbf{u}|\,\phi_2 \sim \mathrm{N}_n\left(\mathbf{0}, \phi_2^{-1}\mathbf{I}_n\right)$, where $\boldsymbol{\epsilon}$ and $\mathbf{u}$ are conditionally independent given the precision parameters ϕ_1 and ϕ_2, i.e., $(\boldsymbol{\epsilon} \perp \mathbf{u})|\,(\phi_1, \phi_2)$, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Hereafter, $\mathbf{w}\,|\mathbf{v}$ denotes a random variable $\mathbf{w}$ given a random variable $\mathbf{v}$. This model has been broadly studied by Fuller (1987) and Cheng and Van Ness (1999). Equations (1) and (2) are suitable to describe experiments where the variables $\boldsymbol{\xi}$ and $\mathbf{y}$ have a linear relationship and $\boldsymbol{\xi}$ is observed through $\mathbf{x}$. The books of Fuller (1987), Cheng and Van Ness (1999), Gustafson (2004) and Carrol et al. (2006) present several applications of MEM.

If ξ_i, for $i = 1, \ldots, n$, are assumed fixed, the model given by Equations (1) and (2) is called functional. However, if the elements of $\boldsymbol{\xi}$ are random variables, it is called a structural model. A well-known structural MEM is when the ξ_i are normally and independently distributed, i.e., $\boldsymbol{\xi}|\,\phi_3 \sim \mathrm{N}_n\left(\mu\mathbb{1}_n, \phi_3^{-1}\mathbf{I}_n\right)$, where ϕ_3 is the precision parameter. Thus, the normal structural MEM is given by

$$y_i = \beta_0 + \beta_1 \xi_i + \epsilon_i \tag{3}$$

and

$$x_i = \xi_i + u_i,$$

with

$$(\epsilon_i, u_i, \xi_i)^\top |\boldsymbol{\phi}, \mu \overset{\text{i.i.d.}}{\sim} \mathrm{N}_3\left((0, 0, \mu)^\top,\ \operatorname{diag}^{-1}(\boldsymbol{\phi})\right), \quad i = 1, \ldots, n, \tag{4}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)^\top$. Applications and classical statistical analysis of the structural normal MEM given by Equations (3) and (4) can be found in Fuller (1987) and Cheng and Van Ness (1999). Since explicit formulas of the posterior distributions can not be obtained, Bayesian inference of Equations (3) and (4) have been mainly made by using Markov chain Monte Carlo simulations. For Bayesian analysis of MEM and its applications see, e.g., Lindley and El-Sayyad (1968), Zellner (1971, Chapter 5), Villegas (1972), Florens et al. (1974), Bolfarine and Cordani (1993), Richardson and Gilks (1993), Dellaportas and Stephens (1995), Richardson (1996), Aoki et al. (2003), Bolfarine and Arellano-Valle (2005) and Bolfarine and Lachos (2006).

The model given by Equations (3) and (4) has two types of parameters: $(\boldsymbol{\beta}, \boldsymbol{\phi}, \mu)$, called structural parameters, and $\xi_1, \ldots, \xi_n$, called latent (or incidental) parameters. The objective usually is to make inference about the structural parameters or functions of them. In this work, we assume a Student-t distribution in Equation (4) and propose different measures for assessing the influence of a given subset of observations on the posterior distribution of the structural parameters. The problem of detecting influential observations in MEM has been treated from the classical point of view by Kelly (1984), Wellman and Gunst (1991), Abdullah (1995), Lee and Zhao (1996), Kim (2000), Galea et al. (2002a,b) and Montenegro et al. (2009), among others. However, we only know the works by Quintana et al. (2005) and Vidal et al. (2007) on detection of influential observations in MEM from a Bayesian perspective.

The influence measures proposed in this work are based on the Bayes risk under quadratic loss and the q-divergence between the posterior distributions of the structural parameters. These measures were applied to other kinds of models for assessing the influence of model assumptions; see, e.g., Kempthorne (1986), Peng and Dey (1995) and Arellano-Valle et al. (2000). To compute them, we combine the idea of perturbation function; see, e.g., Kass et al. (1989) and Weiss (1996) and Markov chain Monte Carlo (MCMC) methods.

The rest of the paper is structured as follows. In Section 2, we set the MEM to be analyzed and discuss identifiability. In Section 3, we present some influence measures based on posterior Bayes risk and q-divergence and apply it to the structural Student-t MEM with weak nondifferential error (WNDE). In Section 4, we focus on the implementation of the Gibbs sampler algorithm. In Section 5, we apply the obtained results to two well-known data sets. Finally, in Section 6, we sketch some conclusions and final remarks.

2. The Independent Student-t MEM with WNDE

From Equations (3) and (4), we have that $(\mathbf{y} \perp \mathbf{x})|\,(\boldsymbol{\beta}, \boldsymbol{\xi}, \phi_1, \phi_2)$. In this case, it is said that the MEM has nondifferential error (NDE). In other words, $\mathbf{x}$ does not contain information about $\mathbf{y}$ given $\boldsymbol{\xi}$. For more details about NDE models, see Bolfarine and Arellano-Valle (1998) and Carrol et al. (2006).

It is important to stress that the inference based on the normality assumption for the MEM can be strongly affected by any perturbation in the data. Particularly, it is well known that the normal models are quite vulnerable to the presence of outliers. For that reason, we propose a model than provide results more robust than the normal one by considering a Student-t distribution in Equation (4). Thus, the probability density function (pdf) for $(\boldsymbol{\epsilon}, \mathbf{u}, \boldsymbol{\xi})^\top$ is given by

$$[\boldsymbol{\epsilon}, \mathbf{u}, \boldsymbol{\xi}\,|\boldsymbol{\phi}, \mu, \nu] = \prod_{i=1}^{n} t_3\left((0,0,\mu)^\top, \operatorname{diag}^{-1}(\boldsymbol{\phi}), \nu\right), \tag{5}$$

where $[\mathbf{w}\,|\mathbf{v}]$ denotes the conditional pdf of the distribution of $\mathbf{w}$ given $\mathbf{v}$ and $t_n(\boldsymbol{\eta}, \boldsymbol{\Omega}, \nu)$ denotes the n-variate Student-t distribution with a location parameter $\boldsymbol{\eta}$, a scale parameter $\boldsymbol{\Omega}$ and ν degrees of freedom. Since the elements of a multivariate Student-t distribution are not independent, this structural Student-t MEM has differential error because does not satisfy $(\mathbf{y} \perp \mathbf{x})|\,(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\phi}, \mu, \nu)$. However, the model defined by Equations (1), (2) and (5) satisfies $\mathbb{E}(\mathbf{y}|\,\boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{x}) = \mathbb{E}(\mathbf{y}|\,\boldsymbol{\beta}, \boldsymbol{\xi}) = \boldsymbol{\Sigma}\boldsymbol{\beta}$, whenever $\nu > 1$. In this case, $\mathbb{E}(\mathbf{y}|\,\boldsymbol{\beta}, \boldsymbol{\xi}, \mathbf{x})$ does not depend on $\mathbf{x}$. Therefore, it is said that the MEM given by Equations (1), (2) and (5) has a WNDE. Definition and some characterization resulting from WNDE models can be seen in Bolfarine and Arellano-Valle (1998).

Under the Bayesian approach, we have to specify the prior distributions of the statistical model and, therefore, we have to determine the dimension of the identified parameter vector. This task is relevant since a parameterization of a statistical model must be able to differentiate points on the parameter space that can be distinguished from the observations. If this does not happen, then it is necessary to restrict the parameter space to achieve that goal. This motivates the identifiability analysis of independent Student-t MEM with WNDE from a Bayesian point of view.

2.1 Bayesian identification

Reiersol (1950) and Madansky (1959) established important identifiability results for the structural MEM under normal errors. In our context, we consider the Student-t distribution for the error terms. Although in Bayesian analysis identification issues present no formal difficulties, from a computational point of view the identification problems could imply ill-behaved posterior surfaces and, in such cases, MCMC methods can be difficult to implement; for more details and references related to the identifiability and convergence of the Gibbs sampler, see Hobert et al. (1997).

As mentioned above, in this section, we focus on the Bayesian identifiability of the independent Student-t MEM with WNDE. We conclude that the Student-t MEM with WNDE is identified under specific parameter restrictions. It is important to stress that the Bayesian identifiability is analyzed using the concept of parameter sufficiency, introduced by Barankin (1961).

Definition 2.1 A function $g(\boldsymbol{\theta})$ of parameter $\boldsymbol{\theta}$ is sufficient for the sample $\mathbf{x}$, if the conditional distribution of the sample $\mathbf{x}$ given $\boldsymbol{\theta}$ is the same as the conditional distribution of the sample $\mathbf{x}$ given $g(\boldsymbol{\theta})$, that is,

$$(\mathbf{x} \perp \boldsymbol{\theta}) |\, g(\boldsymbol{\theta}) . \tag{6}$$

The condition established in Definition 2.1 means that the distribution of $\mathbf{x}$ is completely determined by $g(\boldsymbol{\theta})$, i.e., $\boldsymbol{\theta}$ is redundant once $g(\boldsymbol{\theta})$ is known. It is relevant to remark that, by the symmetry of a conditional independence relation, the parameter $g(\boldsymbol{\theta})$ is a sufficient parameter if the conditional distribution of $\boldsymbol{\theta}$ given the sufficient parameter $g(\boldsymbol{\theta})$ is not updated by the sample, i.e., $p(\boldsymbol{\theta}|\mathbf{x}, g(\boldsymbol{\theta})) = p(\boldsymbol{\theta}|g(\boldsymbol{\theta}))$; see Dawid (1979) and Florens et al. (1990).

The conditional independence relation given in Equation (6) establishes that the sample $\mathbf{x}$ does not increase the knowledge about $\boldsymbol{\theta}$ given the sufficient parameter $\boldsymbol{\psi} = g(\boldsymbol{\theta})$. Therefore, the parameterization $\boldsymbol{\theta}$ is not identified by the data $\mathbf{x}$. This situation can be avoided if $\boldsymbol{\theta}$ is a minimal sufficient parameter, that is, if $\boldsymbol{\theta}$ is a sufficient parameter and if this is a function of any other sufficient parameter. Consequently, if the parameterization of a statistical model is based on a minimal sufficient parameter, then the parameterization does not contain redundant information. These considerations motivate the following definition; for more details, see Florens and Rolin (1984).

Definition 2.2 A sufficient parameter $\boldsymbol{\psi} = g(\boldsymbol{\theta})$ is said to be Bayesian identified if $\boldsymbol{\psi}$ is a minimal sufficient parameter.

Definition 2.2 establishes that a Bayesian identified parameter fully characterizes the learning process underlying a Bayesian model. It is important to stress that if a parameter $\boldsymbol{\theta}$ is identified in the classical sense, then $\boldsymbol{\theta}$ is Bayesian identified too. The reciprocal result is not necessarily true and is a matter of prior null sets; see (Florens et al., 1990, Chapter 4). Moreover, a Bayesian identified parameter is always a function of a countable number of sampling expectations; see Florens et al. (1990, Chapter 4). This means that a parameterization of interest $\boldsymbol{\psi} = g(\boldsymbol{\theta})$ for some function g, is identified if there exist measurable functions f and h such that

$$\boldsymbol{\psi} = h\{\mathbb{E}(f(\mathbf{x})|\boldsymbol{\theta})\}.$$

It is relevant to remark that after integrating out the incidental component, namely $\boldsymbol{\xi}$, in the pdf given in Equation (5), the identification is lost even though the joint distribution of $(\mathbf{y}, \mathbf{x}, \boldsymbol{\xi})$ is identified. Specifically, when a model is reduced by marginalization or conditionalization, the identification is typically lost; see, e.g., Florens et al. (1990, Chapter 4). Consequently, the next result proposes an identification restriction for the marginal Student-t model of $(\mathbf{y}, \mathbf{x})$.

Theorem 2.3 For the MEM given by Equations (3) and (5), the marginal pdf of $(y_i, x_i)^\top$ with parameter $\boldsymbol{\psi} = (\tilde{\boldsymbol{\beta}}, \psi_1, \psi_2, \psi_3, \nu) \in \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$, where $\tilde{\boldsymbol{\beta}} = (\beta_0 + \beta_1\mu, \mu)^\top$, $\psi_1 = \beta_1^2\phi_3^{-1} + \phi_1^{-1}$, $\psi_2 = \phi_3^{-1} + \phi_2^{-1}$, $\psi_3 = \beta_1\phi_3^{-1}$ and $\nu > 4$, is Bayesian identified by $(y_i, x_i)^\top$, for $i = 1, \ldots, n$.

Proof It consists of writing the parameters of the marginal model induced by $(y_i, x_i)^\top$ as a function of sampling expectations of the form $\mathbb{E}\left(f\left((y_i, x_i)^\top\right)|\boldsymbol{\theta}\right)$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi_1, \phi_2, \phi_3, \mu, \nu)$ and f is a measurable function, for $i = 1, \ldots, n$. From Equations (3) and (5) and after integrating out the incidental components, we have the marginal model for $(y_i, x_i)^\top$ given by

$$[y_i, x_i|\boldsymbol{\theta}] = t_2\left(\begin{pmatrix} y_i \\ x_i \end{pmatrix} \middle| \begin{pmatrix} \beta_0 + \beta_1\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta_1^2\phi_3^{-1} + \phi_1^{-1} & \beta_1\phi_3^{-1} \\ \beta_1\phi_3^{-1} & \phi_3^{-1} + \phi_2^{-1} \end{pmatrix}, \nu\right), \quad i = 1, \ldots, n.$$

Thus,

$$\mathbb{E}\big((y_i, x_i)^\top|\boldsymbol{\theta}\big) = (\beta_0 + \beta_1\mu, \mu)^\top, \quad \nu > 1,$$

$$\mathrm{Cov}\left((y_i, x_i)^\top|\boldsymbol{\theta}\right) = \frac{\nu}{\nu - 2}\begin{pmatrix} \beta_1^2\phi_3^{-1} + \phi_1^{-1} & \beta_1\phi_3^{-1} \\ \beta_1\phi_3^{-1} & \phi_3^{-1} + \phi_2^{-1} \end{pmatrix}, \quad \nu > 2,$$

and, from Kim and Mallick (2003),

$$\mathbb{E}(x_i^4) = 6\nu(\nu-2)\mu^2(\phi_2^{-1}+\phi_3^{-1}) + \frac{3\,\nu^2}{(\nu-2)(\nu-4)}(\phi_2^{-1}+\phi_3^{-1})^2 + \mu^4, \quad \nu > 4, \quad i = 1, \ldots, n.$$

Using the previous moments and after some algebra, we can write the parameter ν in terms of the first, second and fourth moment, say

$$\nu = \frac{2\{2\mathbb{E}(x_i^4|\boldsymbol{\theta}) + 7\mathbb{E}(x_i|\boldsymbol{\theta})^4 - 3\mathbb{E}(x_i^2|\boldsymbol{\theta})\left[\mathbb{E}(x_i^2|\boldsymbol{\theta}) + 2\mathbb{E}(x_i|\boldsymbol{\theta})^2\right]\}}{\mathbb{E}(x_i^4|\boldsymbol{\theta}) - 3\mathbb{E}(x_i^2|\boldsymbol{\theta})^2 + 2\mathbb{E}(x_i|\boldsymbol{\theta})^4}.$$

Since $\tilde{\boldsymbol{\beta}}$, $\psi_1 = \beta_1^2\phi_3^{-1} + \phi_1^{-1}$, $\psi_2 = \phi_3^{-1} + \phi_2^{-1}$, $\psi_3 = \beta_1\phi_3^{-1}$ and ν are functions of countable number of sampling expectations, we conclude that $\boldsymbol{\psi} = (\tilde{\boldsymbol{\beta}}, \psi_1, \psi_2, \psi_3, \nu)$ is the minimal sufficient parameter and, therefore, identified by observations. ∎

Using the result given in Theorem 2.3, we are able to propose some restrictions for the identification of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi_1, \phi_2, \phi_3, \mu, \nu)$. For example, we could use any of the identification restrictions commonly used in the normal structural MEM.

Summarizing, in this work we consider a model given by Equations (1), (2) and the error distribution

$$[\boldsymbol{\epsilon}, \mathbf{u}, \boldsymbol{\xi}\,|\phi_1, \phi_3, \mu, \nu, \lambda] = \prod_{i=1}^{n} t_3\left((0, 0, \mu)^\top, \mathrm{diag}^{-1}(\phi_1, \lambda\phi_1, \phi_3), \nu\right), \tag{7}$$

where $\lambda > 0$ is a known value.

In order to obtain known distributions for the full conditional distributions, we chose normal prior distributions for mean parameters and gamma prior distributions for precision parameters. It is important to stress that normal and gamma distributions give enough flexibility to represent a great variety of prior information. For example, it is simple to represent little information with such distributions. Thus, we chose the following prior distributions:

- $\boldsymbol{\beta}|\,\mathbf{b},\mathbf{B},\phi_1 \sim \mathrm{N}_2\left(\mathbf{b},\phi_1^{-1}\mathbf{B}\right)$, where $\mathbf{b}_{(2\times 1)}$ and $\mathbf{B}_{(2\times 2)}$ are known.
- $\mu|\,m,v,\phi_3 \sim \mathrm{N}\left(m,\phi_3^{-1}v^2\right)$, where m and v are known.
- $\phi_j|\,a_j,b_j \overset{\text{ind.}}{\sim} \mathrm{Ga}\left(a_j/2,b_j/2\right)$, where a_j and b_j are known for $j=1,3$. Here, $\mathrm{Ga}(a,b)$ denotes the gamma distribution with expected value equal to a/b.
- $\nu|\,c \sim \mathrm{Ga}\left(1,c/2\right)$, where c is known.

3. Influential Observations

As mentioned early, inference in measurement error models (MEM) can be strongly modified by the inclusion or deletion influential observations. In this section, we study influence measures whose computation is based on the idea of perturbation functions introduced by Kass et al. (1989) and Weiss (1996).

Perturbation functions are used to deal with the problem of assessing the influence of model assumptions on a posterior distribution $[\boldsymbol{\theta}\,|\mathbf{y},M_0]$ in a general context. Assume that $[\boldsymbol{\theta}\,|\mathbf{y},M_0]$ is the posterior distribution of $\boldsymbol{\theta}$ under the M_0 model. Similarly, $[\boldsymbol{\theta}\,|\mathbf{y},M_1]$ is the posterior distribution of the same parameter, but under the M_1 model. Then, the perturbation function is defined by

$$h\left(\boldsymbol{\theta}\right)=\frac{[\boldsymbol{\theta}\,|\mathbf{y},M_1]}{[\boldsymbol{\theta}\,|\mathbf{y},M_0]}.$$

Perturbation functions also can be used to assess the influence of a subset of data on the posterior distribution of the structural parameters. The following lemma gives the perturbation function for the structural parameters of the MEM given by Equations (1), (2) and (7) when a subset of data is deleted. As usual, I denotes any subset with k elements of the set $\{1,\ldots,n\}$, and when a subset I has been deleted from the data, $(\mathbf{y}_I,\mathbf{x}_I)$ and $\left(\mathbf{y}_{(I)},\mathbf{x}_{(I)}\right)$ are the corresponding eliminated and remaining data. In this case, the perturbation function for deletion cases is

$$h\left(\boldsymbol{\theta}\right)=\frac{\left[\boldsymbol{\theta}\,\middle|\mathbf{y}_{(I)},\mathbf{x}_{(I)}\right]}{[\boldsymbol{\theta}\,|\mathbf{y},\mathbf{x}]}.$$

Lemma 3.1 The perturbation function for the structural parameters $\boldsymbol{\theta}=\left(\boldsymbol{\beta},\phi_1,\phi_3,\mu,\nu\right)^\top$ of the MEM given by Equations (1), (2) and (7) corresponding to the deletion cases is given by

$$h\left(\boldsymbol{\theta}\right)=\frac{\left(\prod_{i\in I}t_2\left(y_i,x_i|\,\boldsymbol{\theta}\right)\right)^{-1}}{\mathbb{E}_{\pi^*}\left(\left(\prod_{i\in I}t_2\left(y_i,x_i|\,\boldsymbol{\theta}\right)\right)^{-1}\middle|\,\mathbf{y},\mathbf{x}\right)}, \tag{8}$$

where π^* is the posterior distribution of $\boldsymbol{\theta}$ and

$$t_2\left(y_i,x_i|\,\boldsymbol{\theta}\right)=t_2\left(\begin{pmatrix}y_i\\x_i\end{pmatrix}\middle|\begin{pmatrix}\beta_0+\beta_1\mu\\ \mu\end{pmatrix},\begin{pmatrix}\beta_1^2\phi_3^{-1}+\phi_1^{-1} & \beta_1\phi_3^{-1}\\ \beta_1\phi_3^{-1} & \phi_3^{-1}+(\lambda\phi_1)^{-1}\end{pmatrix},\nu\right).$$

Proof Marginalizing on Equation (7) and considering Equations (1) and (2), we have $[\mathbf{y},\mathbf{x}\,|\boldsymbol{\beta},\phi_1,\phi_3,\mu,\nu]=\prod_{i=1}^n t_2\left(y_i,x_i|\,\boldsymbol{\theta}\right)$. Thus, the posterior pdf of $\boldsymbol{\theta}$ is

$$[\boldsymbol{\theta}\,|\mathbf{y},\mathbf{x}]=\frac{[\boldsymbol{\theta}]\prod_{i=1}^n t_2\left(y_i,x_i|\,\boldsymbol{\theta}\right)}{\int[\boldsymbol{\theta}]\prod_{i=1}^n t_2\left(y_i,x_i|\,\boldsymbol{\theta}\right)d\boldsymbol{\theta}}.$$

Therefore, the perturbation function for deletion cases is

$$
\begin{aligned}
h(\boldsymbol{\theta}) &= \frac{\left[\boldsymbol{\theta}\left|\mathbf{y}_{(I)}, \mathbf{x}_{(I)}\right.\right]}{[\boldsymbol{\theta}\,|\mathbf{y}, \mathbf{x}]} \\
&= \frac{[\boldsymbol{\theta}] \prod_{i \notin I} t_2\left(y_i, x_i \right| \boldsymbol{\theta})}{\int [\boldsymbol{\theta}] \prod_{i \notin I} t_2\left(y_i, x_i \right| \boldsymbol{\theta})\, d\boldsymbol{\theta}} \times \frac{\int [\boldsymbol{\theta}] \prod_{i=1}^{n} t_2\left(y_i, x_i \right| \boldsymbol{\theta})\, d\boldsymbol{\theta}}{[\boldsymbol{\theta}] \prod_{i=1}^{n} t_2\left(y_i, x_i \right| \boldsymbol{\theta})} \\
&= \frac{1}{\prod_{i \in I} t_2\left(y_i, x_i \right| \boldsymbol{\theta})} \times \left(\int \frac{1}{\prod_{i \in I} t_2\left(y_i, x_i \right| \boldsymbol{\theta})} \frac{[\boldsymbol{\theta}] \prod_{i=1}^{n} t_2\left(y_i, x_i \right| \boldsymbol{\theta})}{\int [\boldsymbol{\theta}] \prod_{i=1}^{n} t_2\left(y_i, x_i \right| \boldsymbol{\theta})\, d\boldsymbol{\theta}} d\boldsymbol{\theta} \right)^{-1} \\
&= \frac{\left(\prod_{i \in I} t_2\left(y_i, x_i \right| \boldsymbol{\theta})\right)^{-1}}{\int \left(\prod_{i \in I} t_2\left(y_i, x_i \right| \boldsymbol{\theta})\right)^{-1} [\boldsymbol{\theta}\,|\mathbf{y}, \mathbf{x}]\, d\boldsymbol{\theta}}.
\end{aligned}
$$

■

From Lemma 3.1, we can see that the perturbation function for the structural parameters in deletion cases can be computed through MCMC techniques by sampling from the posterior distribution.

3.1 Influence measures based on the posterior Bayes risk

Kempthorne (1986) defined different influence measures in a Bayesian decision theory framework. In this context, the influence of a subset I of observations on the decision problem is defined by its impact on the posterior Bayes risk. Consequently, if we choose the action $\mathbf{a}$ and $\boldsymbol{\theta}$ is the true state of the world, then preferences among actions are determined by their posterior Bayes risk given by

$$r(\pi^*, \mathbf{a}) = \mathbb{E}_{\pi^*}\left(L(\boldsymbol{\theta}, \mathbf{a})\right),$$

where $L(\boldsymbol{\theta}, \mathbf{a})$ is a loss function. In the context of parametric inference, objectives might include estimation, prediction, hypothesis testing, model selection, etc., such as have been established by Berger (1985), O'Hagan (1994) and Bernardo and Smith (1994). In this section, we consider the estimation problem following Kempthorne (1986).

Specifically, we consider the quadratic loss function given by

$$L(\boldsymbol{\theta}, \mathbf{a}) = (\boldsymbol{\theta} - \mathbf{a})^\top \mathbf{W} (\boldsymbol{\theta} - \mathbf{a}) = \|\boldsymbol{\theta} - \mathbf{a}\|_{\mathbf{W}}^2, \tag{9}$$

with $\mathbf{W}$ being a known symmetric positive semi-definite matrix. In this case, the optimal action is the Bayes action $\mathbf{a}^* = \mathbb{E}_{\pi^*}(\boldsymbol{\theta})$, which gets the smallest posterior Bayes risk.

Two ways of measuring the influence of a subset I of observations on $\boldsymbol{\theta}$ are

$$M_{1,\boldsymbol{\theta}}(I) = r\left(\pi^*, \mathbf{a}^*_{(I)}\right) - r(\pi^*, \mathbf{a}^*)$$

and

$$M_{2,\boldsymbol{\theta}}(I) = r\left(\pi^*_{(I)}, \mathbf{a}^*\right) - r\left(\pi^*_{(I)}, \mathbf{a}^*_{(I)}\right),$$

where $\pi^*_{(I)}$ denotes the posterior distribution on $\boldsymbol{\theta}$ when the subset I of observations is excluded from the analysis and $\mathbf{a}^*_{(I)}$ is the corresponding Bayes action. On one hand, if we consider that all data follow the same model, then the influence measure M_1 is the cost of excluding the subset I of observations from the analysis in terms of the posterior Bayes risk. On the other hand, if we consider that all data follow the same model except the subset I, then M_2 measures the increment of the posterior Bayes risk when the subset I is incorrectly included in the data. For more details about this and other influence measures, see Kempthorne (1986).

The next lemma gives a general expressions for these two measures.

LEMMA 3.2 Under the quadratic loss function given by Equation (9),

$$M_{1,\boldsymbol{\theta}}(I) = M_{2,\boldsymbol{\theta}}(I) = \left\|\mathbf{a}^* - \mathbf{a}^*_{(I)}\right\|^2_{\mathbf{W}}.$$

PROOF It is an immediate consequence of the well-known expression

$$\mathbb{E}_{\pi^*}\left((\boldsymbol{\theta} - \mathbf{a})^\top \mathbf{W}(\boldsymbol{\theta} - \mathbf{a})\right) = \operatorname{tr}(\mathbf{W}\mathbf{V}^*) + \|\mathbf{a}^* - \mathbf{a}\|^2_{\mathbf{W}},$$

where $\mathbf{V}^*$ is the posterior variance of $\boldsymbol{\theta}$. ■

The next proposition provides an expression for M_1 and M_2 that only involves the expected value of the unperturbed posterior distribution and the perturbation function h.

PROPOSITION 3.3 Under the quadratic loss function given by Equation (9),

$$M_{1,\boldsymbol{\theta}}(I) = M_{2,\boldsymbol{\theta}}(I) = \|\mathbb{E}_{\pi^*}([1 - h(\boldsymbol{\theta})]\,\boldsymbol{\theta})\|^2_{\mathbf{W}}, \tag{10}$$

where $h(\cdot)$ is the perturbation function of π^* to $\pi^*_{(I)}$.

PROOF From

$$\begin{aligned}
\mathbf{a}^* - \mathbf{a}^*_{(I)} &= \mathbb{E}_{\pi^*}(\boldsymbol{\theta}) - \mathbb{E}_{\pi^*_{(I)}}(\boldsymbol{\theta}) \\
&= \mathbb{E}_{\pi^*}(\boldsymbol{\theta}) - \mathbb{E}_{h\pi^*}(\boldsymbol{\theta}) \\
&= \mathbb{E}_{\pi^*}(\boldsymbol{\theta} - \boldsymbol{\theta} h(\boldsymbol{\theta}))
\end{aligned}$$

and Lemma 3.2, the result is obtained. ■

In our case, i.e., under the MEM given by Equations (1), (2) and (7), the function $h(\boldsymbol{\theta})$ is given by Equation (8), and π^* is the posterior pdf of $(\boldsymbol{\beta}, \phi_1, \phi_3, \mu, \nu)$ considering all data. In the next section, we explain how to compute these influence measures for the MEM given by Equations (1), (2) and (7).

3.2 INFLUENCE MEASURES BASED ON q-DIVERGENCE

Another way of quantifying influence is by computing divergence measures between posterior distributions computed with and without a given subset of the data. Csiszár (1967) defined the q-divergence measure between two densities π_1 and π_2 on $\boldsymbol{\theta}$ by

$$d_q(\pi_1, \pi_2) = \int q\left(\frac{\pi_1(\boldsymbol{\theta})}{\pi_2(\boldsymbol{\theta})}\right)\pi_2(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \tag{11}$$

where q is a convex function such that $q(1) = 0$.

A wide class of different divergence measures is obtained from Equation (11). For example:

- When $q(z) = -\log(z)$, the Kullback-Leibler divergence arises;
- When $q(z) = (z-1)\log(z)$, the J-distance (or the symmetric version of Kullback-Leibler divergence) is reached;
- When $q(z) = 1/2\,|z-1|$, the L_1-distance arises; and
- When $q(z) = (z-1)^2$, the χ^2-divergence is obtained.

Thus, taking $\pi_1(\boldsymbol{\theta}) = \pi^*_{(I)}\left(\boldsymbol{\theta}|\,\mathbf{y}_{(I)}\right)$ and $\pi_2(\boldsymbol{\theta}) = \pi^*(\boldsymbol{\theta}|\,\mathbf{y})$ in Equation (11), we have that $d_q(I) = d_q(\pi_1, \pi_2)$ can be interpreted as the q-influence of the data $\mathbf{y}_I$ on the posterior distribution of $\boldsymbol{\theta}$, which can be written as

$$d_q(I) = \mathbb{E}_{\pi^*}\left(q\left(h(\boldsymbol{\theta})\right)\right), \tag{12}$$

where the expected value is taken with respect to the unperturbed posterior distribution. These influence measures have been already used by Weiss and Cook (1992), Peng and Dey (1995), Weiss (1996), Arellano-Valle et al. (2000) and Vidal et al. (2007).

The influence measures $M_{1,\boldsymbol{\theta}}(I)$, $M_{2,\boldsymbol{\theta}}(I)$ and $d_q(I)$ do not determine when an observation is influential. We need to define a cutoff point in order to determine whether a small subset of observations is influential or not. We use the proposal given by Peng and Dey (1995) to determine when an observation is influential by using $d_q(I)$. The idea of this proposal is explained next.

The pdf of a biased coin is

$$\pi_1(x\,|p) = p^x(1-p)^{1-x}, \quad x = 0, 1,\ p \in [0,1],$$

while those of an unbiased coin is

$$\pi_2(x\,|p = 1/2) = 1/2 \quad x = 0, 1,\ p \in [0,1].$$

From Equation (11), it is easy to obtain the q-divergence between a biased and an unbiased coin by

$$d_q^B(p) = \frac{q(2p) + q(2(1-p))}{2}.$$

It is not difficult to see that $d_q^B(p)$ increases as p moves away from 0.5. In addition, $d_q^B(p)$ is symmetric about $p = 0.5$ and $d_q^B(p)$ achieves its minimum at $p = 0.5$. In this point, $d_q^B(0.5) = 0$ and $\pi_1 = \pi_2$. Therefore, if we consider $p \geq 0.75$ (or $p \leq 0.25$) as a strong bias in a coin, then, since

$$d_{L_1}^B(0.75) = d_{\chi^2}^B(0.75) = 0.25,$$

we can indicate an influential observation when $d_{L_1}(i) \geq 0.25$ or $d_{\chi^2}(i) \geq 0.25$. Similarly, for the Kullback-Leibler divergence $d_{\mathrm{KL}}^B(0.75) \approx 0.143841$, and for the J-distance $d_J^B(0.75) \approx 0.274653$. Thus, if we use the Kullback-Leibler divergence, we can consider an influential observation when $d_{\mathrm{KL}}^B(i) > 0.14$. Similarly, using the J-distance, an observation which $d_J(i) > 0.27$ can be considered as influential.

4. Implementation

The influence measures given by Equations (10) and (12) are expected values with respect to the unperturbed posterior distribution of $\boldsymbol{\theta}$. Therefore, for any subset I, we can estimate Equations (10) and (12) using a Monte Carlo estimator. Thus, we could obtain only one sample $\{\boldsymbol{\theta}^{(j)},\ j=1,\ldots,m\}$ from the unperturbed posterior distribution and estimate the measures given by Equations (10) and (12) for any subset I.

Given the analytic intractability of the posterior distribution of the structural parameters, the model inference is usually performed via MCMC methods. Thus, the influence measures mentioned in Section 3 are estimated using Monte Carlo estimators, where the generated data from the unperturbed posterior distribution of $\boldsymbol{\theta}=(\boldsymbol{\beta},\phi_1,\phi_3,\mu,\nu)$ are drawn by using a Metropolis-Hastings (MH) algorithm within a grouped Gibbs sampler algorithm. Details about the implementation of these algorithms can be found in Hastings (1970), Smith and Roberts (1993), Robert and Casella (1999) and Chen et al. (2000).

Expressing Equation (7) as

$$(\epsilon_i,u_i,\xi_i)^\top\,|\phi_1,\phi_3,\mu,\lambda,\omega_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}_3\left((0,0,\mu)^\top,\omega_i\mathrm{diag}^{-1}\left(\phi_1,\lambda\phi_1,\phi_3\right)\right),$$

where $\omega_i^{-1}\big|\,\nu \overset{\text{i.i.d.}}{\sim} \mathrm{Ga}\left(\nu/2,\nu/2\right)$, the full conditional distributions are the following:

- The conditional distribution for $\boldsymbol{\beta}$ depends on $\mathbf{y}$, ϕ_1, $\boldsymbol{\xi}$ and $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_n)^\top$ and is given by

$$[\boldsymbol{\beta}\,|\mathbf{y},\phi_1,\boldsymbol{\xi},\boldsymbol{\omega}]=\mathrm{N}_2\left(\mathbf{b}^*,\phi_1^{-1}\mathbf{B}^*\right),$$

where

$$\mathbf{b}^*=\mathbf{B}^*\left(\mathbf{B}^{-1}\mathbf{b}+\boldsymbol{\Sigma}^\top\mathrm{diag}^{-1}\left(\boldsymbol{\omega}\right)\mathbf{y}\right)$$

and

$$\mathbf{B}^*=\left(\mathbf{B}^{-1}+\boldsymbol{\Sigma}^\top\mathrm{diag}^{-1}\left(\boldsymbol{\omega}\right)\boldsymbol{\Sigma}\right)^{-1}.$$

Since $\phi_1|\,a_1,b_1 \overset{\text{ind.}}{\sim} \mathrm{Ga}\left(a_1/2,b_1/2\right)$, then the conditional distribution for $\boldsymbol{\beta}$ can also be expressed as

$$[\boldsymbol{\beta}\,|\mathbf{y},\boldsymbol{\xi},\boldsymbol{\omega}]=t_2\left(\mathbf{b}^*,\frac{a_1}{b_1}\mathbf{B}^*,a_1\right).$$

- The conditional distribution for ϕ_1 depends on $\mathbf{y}$, $\boldsymbol{\beta}$, $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ and is given by

$$[\phi_1\,|\mathbf{y},\boldsymbol{\beta},\boldsymbol{\xi},\boldsymbol{\omega}]=\mathrm{Ga}\left(1+\frac{n+a_1}{2},\frac{1}{2}\left(b_1+\|\mathbf{y}-\boldsymbol{\Sigma}\boldsymbol{\beta}\|^2_{\mathrm{diag}^{-1}(\boldsymbol{\omega})}+\|\boldsymbol{\beta}-\mathbf{b}\|^2_{\mathbf{B}^{-1}}\right)\right).$$

- The conditional distribution for ϕ_3 only depends on $\boldsymbol{\xi}$, μ and $\boldsymbol{\omega}$ and is given by

$$[\phi_3\,|\boldsymbol{\xi},\mu,\boldsymbol{\omega}]=\mathrm{Ga}\left(\frac{n+a_3+1}{2},\frac{1}{2}\left(b_3+\|\boldsymbol{\xi}-\mu\mathbb{1}_n\|^2_{\mathrm{diag}^{-1}(\boldsymbol{\omega})}+\frac{(\mu-m)^2}{v^2}\right)\right).$$

- The conditional distribution for $\boldsymbol{\xi}$ is given by

$$[\boldsymbol{\xi}\,|\mathbf{y},\mathbf{x},\boldsymbol{\beta},\phi_1,\phi_3,\mu,\boldsymbol{\omega}] = \mathrm{N}_n\left(\mathbf{m}^*,\mathbf{V}^*\right),$$

where

$$\mathbf{m}^* = \left(\phi_1\beta_1^2+\lambda\phi_1+\phi_3\right)^{-1}\left[\phi_1\beta_1\left(\mathbf{y}-\beta_0\mathbb{1}_n\right)+\lambda\phi_1\mathbf{x}+\phi_3\mu\mathbb{1}_n\right]$$

and

$$\mathbf{V}^* = \left(\phi_1\beta_1^2+\lambda\phi_1+\phi_3\right)^{-1}\mathrm{diag}\left(\boldsymbol{\omega}\right).$$

- The conditional distribution for μ only depends on ϕ_3, $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ and is given by

$$[\mu\,|\phi_3,\boldsymbol{\xi},\boldsymbol{\omega}] = \mathrm{N}\left(\frac{m+v^2\boldsymbol{\xi}^\top\mathrm{diag}^{-1}\left(\boldsymbol{\omega}\right)\mathbb{1}_n}{1+v^2\mathbb{1}_n^\top\mathrm{diag}^{-1}\left(\boldsymbol{\omega}\right)\mathbb{1}_n},\frac{v^2}{\phi_3\left(1+v^2\mathbb{1}_n^\top\mathrm{diag}^{-1}\left(\boldsymbol{\omega}\right)\mathbb{1}_n\right)}\right).$$

- The conditional distribution for $\boldsymbol{\omega}$ is given by

$$[\boldsymbol{\omega}\,|\mathbf{y},\mathbf{x},\boldsymbol{\beta},\phi_1,\phi_3,\boldsymbol{\xi},\mu,\nu] = \prod_{i=1}^{n}\mathrm{IGa}\left(\frac{\nu+1}{2},\frac{\nu+\phi_1\epsilon_i^2+\lambda\phi_1\left(x_i-\xi_i\right)^2+\phi_3\left(\xi_i-\mu\right)^2}{2}\right),$$

where $\epsilon_i = y_i-\beta_0-\beta_1\xi_i$ and $\mathrm{IGa}(a,b)$ denotes an inverted gamma distribution such that if $X\sim\mathrm{IGa}\left(a,b\right)$, then $X^{-1}\sim\mathrm{Ga}\left(a,b\right)$.
- Finally, the conditional pdf for ν only depends on $\boldsymbol{\omega}$ and is given by

$$[\nu\,|\boldsymbol{\omega}] \propto \exp\left\{-\frac{\nu}{2}\left[c+\mathbb{1}_n^\top\mathrm{diag}^{-1}\left(\boldsymbol{\omega}\right)\mathbb{1}_n+\sum_{i=1}^{n}\ln\left(\omega_i\right)\right]+\frac{n\nu}{2}\ln\left(\frac{\nu}{2}\right)-n\ln\Gamma\left(\frac{\nu}{2}\right)\right\},$$

where $\Gamma\left(z\right)=\int_0^\infty t^{z-1}\exp(-t)dt$.

To obtain a sample of $[\nu\,|\boldsymbol{\omega}]$, we make a step of the MH algorithm with instrumental distribution $\mathrm{Ga}(a^*,b^*)$, where $a^*=b^*\nu^{(t-1)}+1$,

$$b^* = \frac{n+c+\sum_{i=1}^{n}\left(\omega_i^{-1}+\ln\left(\omega_i\right)\right)}{2}$$

and $\nu^{(t-1)}$ is the sample from the previous step. More details about the Metropolis-within-Gibbs (MwG) algorithm can be found in Robert and Casella (1999, Section 7.3).

5. Applications

In this section, we illustrate the estimation of the influence measures given in Section 3 using two real data sets. For both applications, the samples from the unperturbed posterior distributions were obtained using the MwG algorithm explained in the previous section. We obtained 1,001,000 samples from the posterior distribution discarding the first 1,000 iterations. A lag of 10 observations was selected to avoid autocorrelation. This means that a net sample size of 100,000 was used in each sampling process.

5.1 Concrete compressive strengths data

These data were taken from Wellman and Gunst (1991) and consist of 41 pairs of observed (y_i, x_i) values, where the y_i and x_i represent the measured compressive strengths of concrete taken 28 days and 2 days after pouring, respectively; see Table 1. This data set is well-known in the literature.

Table 1. Concrete compressive strength measurements in pounds per square inch (psi).

Sample	Day 28	Day 2	Sample	Day 28	Day 2	Sample	Day 28	Day 2
1	4470	2830	15	4690	2985	29	4650	3335
2	4740	3295	16	4880	3135	30	4680	3800
3	5115	2710	17	3425	2750	31	5165	2680
4	4880	2855	18	4265	3205	32	5075	3760
5	4445	2980	19	4485	3000	33	4710	3605
6	4080	3065	20	5220	3035	34	4200	2005
7	5390	3765	21	7695	4245	35	4645	2495
8	4045	3265	22	3330	1635	36	4725	3205
9	4370	3170	23	4065	2270	37	4695	2060
10	4955	2895	24	4715	2895	38	5470	3425
11	3835	2630	25	4735	2845	39	4330	3315
12	4290	2830	26	3605	2205	40	4950	3825
13	4600	2935	27	4670	3590	41	4460	3160
14	4605	3115	28	4720	3080			

The measured strengths of concrete differ from their respective true underlying values due to various sources of measurement errors. Thus an appropriate model for the data is given by Equations (1) and (2). Figure 1 shows a scatter plot of the data set and four lines fitted from different methods. Wellman and Gunst (1991), Abdullah (1995) and Vidal et al. (2007) used these data to evaluate the performance of various diagnostic techniques in normal MEM. Similar to Galea et al. (2002a), but from a Bayesian perspective, we use these data to detect influential observations in MEM with Student-t errors. Wellman and Gunst (1991), Abdullah (1995), Galea et al. (2002a) and Vidal et al. (2007) concluded that observation 21 exhibits a strong influence on parameters estimates.

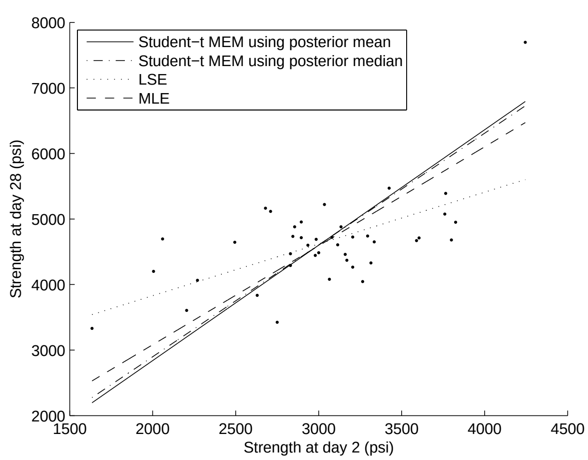

Figure 1. Concrete compressive strengths, in pounds per square inch, at 2 and 28 days.

For the sake of illustration, we assume that expert knowledge leads to the following informative priors:

- $\boldsymbol{\beta}|\,\phi_1 \sim \mathrm{N}_2\left((1500,1)^\top,\, \phi_1^{-1}\left(500^2,4\right)\mathbf{I}_2\right)$;
- $\phi_1 \sim \mathrm{Ga}\left(1,10^6\right)$;
- $\phi_3 \sim \mathrm{Ga}\left(1,10^5\right)$,
- $\mu|\,\phi_3 \sim \mathrm{N}\left(3000, \phi_3^{-1}10^3\right)$; and
- $\nu \sim \mathrm{Ga}\left(1,1\right)$.

The influence measures based on the posterior Bayes risk and the J-distance assign greater influence to observations 21, 22, 37 and 34 (in that order); see Figure 2. However, the L_1-distance and the Kullback-Leibler divergence assign greater influence to observations 21, 22 and 37 (in that order) and, the χ^2-divergence assigns greater influence to observations 21, 22, 37, 34 and 30 (in that order); see Figure 2. From this analysis, the most influential observations are 21, 22 and 37 coinciding with the observations analyzed by Wellman and Gunst (1991), Abdullah (1995), Galea et al. (2002a) and Vidal et al. (2007).

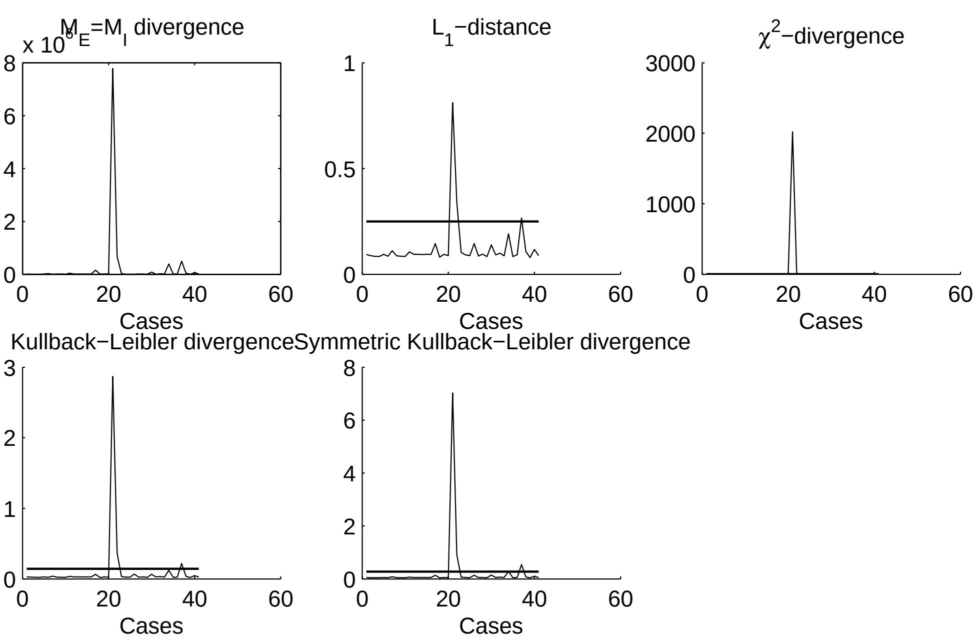

Figure 2. Influence measures for each observation for concrete compressive strengths data.

Figure 3 shows the influences for all subsets of two observations. We can see that there are many pairs of observations with great influence on the posterior distribution. But, the pairs with the biggest influence have the form $(21,\cdot)$ according to M_1, M_2 and q-influences. That suggests observation 21 is an influential case followed by the observation 22.

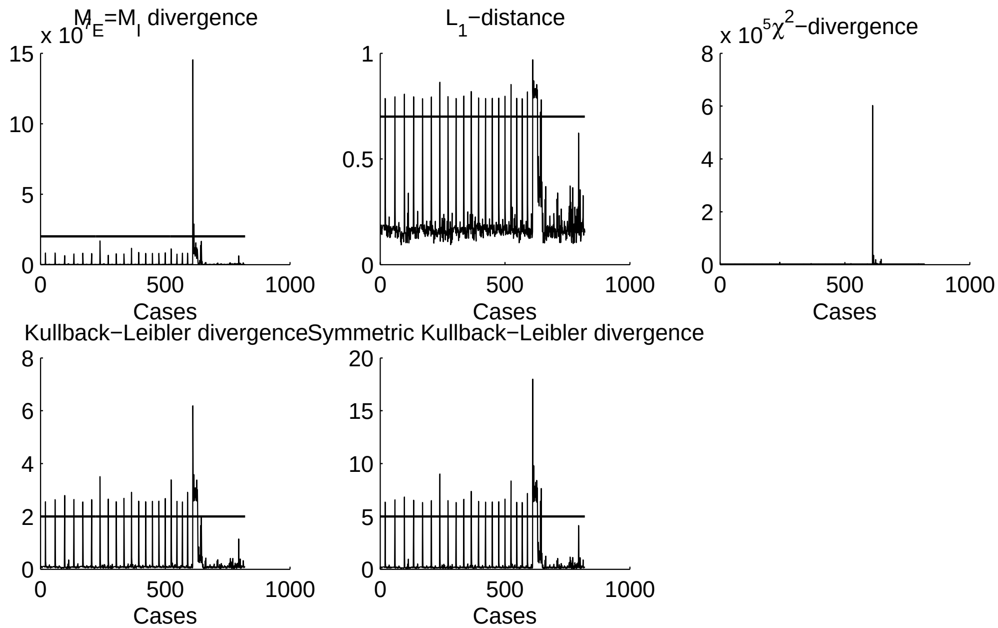

Figure 3. Influence measures for all subsets of two observations for concrete compressive strengths data.

5.2 Serum kanamycin data

These data were taken from Kelly (1984) and consist of simultaneous pairs of measurements of serum kanamycin levels in blood samples drawn from twenty premature babies; see Table 2.

Table 2. Serum kanamycin levels in blood samples.

Baby	Heelstick	Catheter
1	23.0	25.2
2	33.2	26.0
3	16.6	16.3
4	26.3	27.2
5	20.0	23.2
6	20.0	18.1
7	20.6	22.2
8	18.9	17.2
9	17.8	18.8
10	20.0	16.4
11	26.4	24.8
12	21.8	26.8
13	14.9	15.4
14	17.4	14.9
15	20.0	18.1
16	13.2	16.3
17	28.4	31.3
18	25.9	31.2
19	18.9	18.0
20	13.8	15.6

One of the measurements was obtained by a heelstick method (x) and the other by using an umbilical catheter (y). Since there is a measurement error in both methods, the model given by Equations (1) and (2) seems to be appropriated for escribing these data. Figure 4 shows a scatter plot of the data set and four lines fitted from different methods.

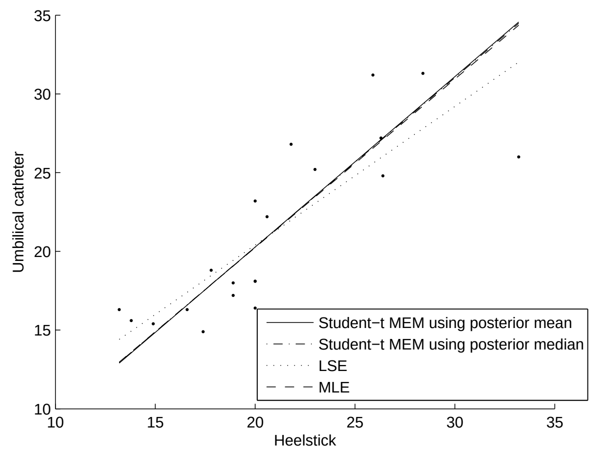

Figure 4. Serum Kanamycin levels in blood samples data.

As in the previous case, and for the sake of illustration, we assume that expert knowledge leads to the following informative priors:

- $\boldsymbol{\beta}|\,\phi_1 \sim \mathrm{N}_2\left((1500,1)^\top, \phi_1^{-1}\left(500^2,4\right)\mathbf{I}_2\right)$;
- $\phi_1 \sim \mathrm{Ga}\left(1,10^6\right)$;
- $\phi_3 \sim \mathrm{Ga}\left(1,10^5\right)$;
- $\mu|\,\phi_3 \sim \mathrm{N}\left(3000, \phi_3^{-1}10^3\right)$; and
- $\nu \sim \mathrm{Ga}\,(1,1)$.

The influence measures based on the posterior Bayes risk assigns greater influence to observations 2, 18, 16 and 17; see Figure 5. However, the L_1 -distance, the Kullback-Leibler divergence and the J-distance assign greater influence to observations 2, 18 and 17 (in that order) and, the χ^2-divergence assigns greater influence to observations 2, 18, 16 and 17 (in that order); see Figure 5. Therefore, the most influential observations are 2, 18 and 17. Kim (2000), Galea et al. (2002a) and Quintana et al. (2005) detected these three observations as the most influential. Kelly (1984) and Abdullah (1995) mentioned observation 16 too. We also determinate the influence of every pair of data. Again, many pairs of data showed great influence. But, the pairs with the biggest influence are the subset $(17,18)$ and those which form is $(2,\cdot)$. That suggests observation 2 is an influential case followed by the observations 18 and 17.

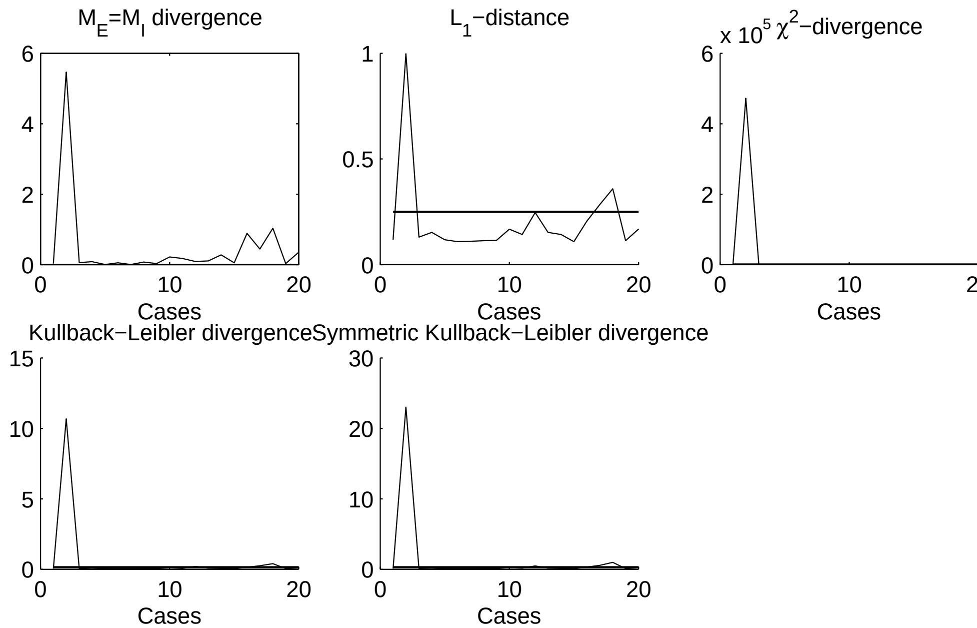

Figure 5. Influence measures for each observation for serum kanamycin levels in blood samples data.

6. Final Remarks

The problem of detecting influential observations is an important step in data analysis. There are several forms of assessing the influence of observation perturbation on parameters estimation. In this work, we have detected influential observations in the independent Student-t measurement error model with weak nondifferential error from a Bayesian point of view. The influence measures presented in this work quantify the impact of any subset of data on the posterior distributions of the structural parameters of an specific measurement error model. These measures are easy to compute and allow to evaluate measure the influence of any subset of data by using only one sample drawn from the posterior distribution. These influence measures were applied to two well-known data sets. The obtained results were similar to those found by other authors.

Acknowledgements

The work of I. Vidal was partially supported by a FONDECYT 1090109 grant from Chile. The work of L.M. Castro was partially supported by DIUC 209.014.017-1.0 grant from Universidad de Concepción, Chile.

References

Abdullah, M.B., 1995. Detection of influential observations in functional errors-in-variables model. Communications in Statistics — Theory and Methods, 24, 1585-1595.

Aoki, R., Achcar, J.A., Bolfarine, H., Singer, J.M., 2003. Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, 30, 3-12.

Arellano-Valle, R.B., Galea-Rojas, M., Iglesias, P., 2000. Bayesian sensitivity analysis in elliptical linear regression models. Journal of Statistical Planning and Inference, 86, 175-199.

Barankin, E.W., 1961. Sufficient parameters: solution to the minimal dimensionality problem. Annals of the Institute of Statistical Mathematics, 12, 91-118.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York.

Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. Wiley, Chichester.

Bolfarine, H., Arellano-Valle, R.B., 1998. Weak nondifferential measurement error models. Statistics and Probability Letters, 40, 279-287.

Bolfarine, H., Arellano-Valle, R.B., 2005. Elliptical measurement error models - a Bayesian approach. In Dey, D.K., Rao, C.R., (eds.). Handbook of Statistics: Volume 25, Bayesian Thinking - Modeling and Computation, Elsevier B.V., North-Holland, pp. 669-688.

Bolfarine, H., Cordani, L.K., 1993. Estimation of a structural regression model with know reliability ratio. Annals of the Institute of Statistical Mathematics, 45, 531-540.

Bolfarine, H., Lachos, V.H., 2006. Skew binary regression with measurement errors. Journal of Theoretical and Applied Statistics, 40, 485-494.

Carrol, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective. Chapman & Hall/CRC, Boca Raton.

Chen, M.H., Shao, Q.M., Ibrahim, J.G., 2000. Monte Carlo Methods in Bayesian Computation. Springer-Verlag, New York.

Cheng, C.-L., Van Ness, J.W., 1999. Statistical Regression with Measurement Error. Arnold, London.

Csiszár, I., 1967. Information-type measures of difference of probability distributions and indirect observations. Studia Scientiarum Mathematicarum Hungarica, 2, 299-318.

Dawid, A.P., 1979. Conditional independence in statistical theory (with discussion). Journal of The Royal Statistical Society Series B—Statistical Methodology, 41, 1-31.

Dellaportas, P., Stephens, D.A., 1995. Bayesian analysis of errors in variable regression models. Biometrics, 51, 1085-1095.

Florens, J.P., Rolin, J.M., 1984. Asymptotic sufficiency and exact estimability. In Florens, J.-P., Mouchart, M., Raoult, J.-P., Simar, L., (eds.). Alternative Approaches to Time Series Analysis. Publications des Facultés Universitaires Saint-Louis, Bruxelles, pp. 121-142.

Florens, J.P., Mouchart, M., Richard, J.F., 1974. Bayesian inference in error-in-variables models. Journal of Multivariate Analysis, 4, 419-452.

Florens, J.P., Mouchart, M., Rolin, J.M., 1990. Elements of Bayesian Statistics. Marcel Dekker, New York.

Fuller, W.A., 1987. Measurement Error Models – Volume 1. Wiley, New York.

Galea, M., Bolfarine, H., Vilca-Labra, F., 2002a. Influence diagnostics for the structural errors-in-variables model under the student-t distribution. Journal of Applied Statistics, 29, 1191-1204.

Galea, M., Bolfarine, H., de Castro, M., 2002b. Local influence in comparative calibration models. Biometrical Journal, 41, 59-81.

Gustafson, P., 2004. Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. Chapman & Hall, Boca Raton.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57, 97-109.

Hobert, J.P., Robert, C.P., Goutis, C., 1997. Connectedness conditions for the convergence of the Gibbs sampler. Statistics and Probability Letters, 33, 235-240.

Kass, R.E., Tierney, L., Kadane, J.B., 1989. Approximate methods for assessing influence and sensitivity in Bayesian analysis. Biometrika, 76, 663-674.

Kelly, G.E., 1984. The influence function in the errors in variables problem. The Annals of Statistics, 12, 87-100.

Kempthorne, P.J., 1986. Decision-theoretic measures of influence in regression. Journal of The Royal Statistical Society Series B—Statistical Methodology, 48, 370-378.

Kim, H., Mallick, B., 2003. Moments of random vectors with skew t distribution and their quadratics forms. Statistics and Probability Letters, 63, 417-423.

Kim, M.G., 2000. Outliers and influential observations in the structural errors-in-variables model. Journal of Applied Statistics, 27, 451-460.

Lee, A.H., Zhao, Y., 1996. Assessing local influence in measurement error models. Biometrical Journal, 38, 829-841.

Lindley, D.V., El-Sayyad, G.M., 1968. The Bayesian estimation of a linear functional relationship. Journal of The Royal Statistical Society Series B—Statistical Methodology, 30, 190-202.

Madansky, A., 1959. The fitting of straight lines when both variables are subject to error. Journal of the American Statistical Association, 54, 173-205.

Montenegro, L.C., Bolfarine, H., Lachos, V.H., 2009. Influence diagnostics for a skew extension of the Grubbs measurement error model. Communications in Statistics — Simulation and Computation, 38, 667-681.

O'Hagan, A., 1994. Bayesian Inference. Kendall's Advanced Theory of Statistics – Volume 2B. Arnold, London.

Peng, F., Dey, D.K., 1995. Bayesian analysis of outlier problems using divergences measures. The Canadian Journal of Statistics, 23, 199-213.

Quintana, F.A., Iglesias, P., Bolfarine, H., 2005. Bayesian identification of outliers and change points in measurement error models. Advances in Complex Systems, 8, 433-449.

Reiersol, O., 1950. Identifiability of a linear relation between variables which are subject to errors. Econometrica, 18, 375-389.

Richardson, S., 1996. Measurement error. In Gilks, W.R., Richardson, S., Spiegelhalter, D.J., (eds.). Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 401-414.

Richardson, S., Gilks, W.R., 1993. Conditional independence models for epidemiological studies with covariate measurement error. Statistics in Medicine, 12, 1703-1722.

Robert, C.P., Casella, G., 1999. Monte Carlo Statistical Methods. Springer-Verlag, New York.

Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. Journal of The Royal Statistical Society Series B—Statistical Methodology, 55, 3-23.

Vidal, I., Iglesias, P., Galea, M., 2007. Influential observations in the functional measurement error model. Journal of Applied Statistics, 34, 1165-1183.

Villegas, C., 1972. Bayesian inference in linear relations. The Annals of Mathematical Statistics, 43, 1767-1791.

Weiss, R.E., 1996. An approach to Bayesian sensitivity analysis. Journal of The Royal Statistical Society Series B—Statistical Methodology, 58, 739-750.

Weiss, R.E., Cook, R.D., 1992. A graphical case statistic for assessing posterior influence. Biometrika, 79, 51-55.

Wellman, M.J., Gunst, R.F., 1991. Influence diagnostics for linear measurement error models. Biometrika, 78, 373-380.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.