Local Influence in Null Intercept Measurement Error Regression under a Student_t Model

Filidor V. Labra[1], Reiko Aoki[2] and Heleno Bolfarine[3]

(1) Departamento de Estatística, IMECC/UNICAMP, Caixa Postal 6065
13083-970 Campinas, SP, Brazil
(2) Departamento de Ciências da Computação e Estatística,
ICMC/USP-São Carlos, Caixa Postal 668,
13560-970, São Carlos, SP, Brazil
(3) Departamento de Estatística, IME/USP, Caixa Postal 66281,
05311-970, São Paulo, SP, Brazil

February 12, 2003

Summary

In this paper we discuss the application of local influence in measurement error regression model with null intercepts under a Student_t model with dependent populations. The Student_t distribution is a robust alternative to modeling data sets involving errors with longer than Normal tails. We derive the appropriate matrices for assessing the local influence for different perturbation schemes and use a real data as an illustration of the usefulness of the application.

KEY WORDS: Influence diagnostic; Student_t model; Likelihood displacement; Lack of fit; pretest/posttest data; Measurement error models.

1 Introduction

In this paper, we discuss an application of the local influence method (Cook, 1986) in the measurement error regression models with null intercept. The motivation comes from the need of such a model for dependent populations involving errors with longer than normal tails. This approach is applied to the data from a pretest/posttest study presented in Singer and Andrade (1997). In that study, designed to compare two types of toothbrushes with respect to the efficacy in removing dental plaque, 26 preschoolers were evaluated with respect to a dental plaque index before and after toothbrushing either with a conventional or with an experimental (hugger) toothbrush. The reason for considering null intercepts is that null pretest dental plaque indices imply null expected posttest values. As the same individuals were evaluated under two different experimental conditions (toothbrushes), we need a model which takes into account the possible within subjects correlation structure. The analysis of such a model considering the Normal Distribution was studied in Aoki (2001). See also Aoki et al. (2001). The extension of the model considering Student_t distribution is discussed in detail in the next section.

Influence diagnostic is an important step in the analysis of a data set, as it provides us indication of bad model fitting or of influential observations. This analysis has received a great deal of attention since the paper by Cook (1977). Usually the analysis is based on the case-weight perturbation scheme where the case (observation) is either deleted or retained, so that the individual impact of cases is assessed in the estimation process (see, for example, Cook (1986)), however deletion can be viewed as one of the many ways of perturbing a problem formulation. Cook (1986) proposed a method of assessing the local influence of minor perturbations of a statistical model. Since then several papers have been written with respect to the local influence, but little work has been found in the literature for the measurement error regression models. Lee and Zhao (1996) employed local influence approach in generalized linear measurement error models, while Abdullah (1995) compared several methods for detecting influential observations in functional measurement error models. Recently, Kim (2000) applied the local influence method in structural measurement error models. Section 2 presents the model. Section 3 reviews the concept of the local influence, as well as, the application to the model defined in Section 2 and the appropriate matrices necessary to construct the influence graphs are given in closed form expressions. Finally, in Section 4 we present the illustrative application using a real data from pretest/posttest study described earlier in this section.

2 Null Intercept Measurement Error Regression under a Student_t Model

The basic model is given by

$$Y_{ij} = \beta_i x_{ij} + e_{ij}, \tag{2.1}$$

$$X_{ij} = x_{ij} + u_{ij}, \tag{2.2}$$

where Y_{ij} and X_{ij}, respectively, denote the observed values of the response and explanatory variables for population i and subject j, ($i = 1, \cdots, p$, $j = 1,\cdots,n$), x_{ij}, correspond to the true values of the latter, β_i, $i = 1, \cdots, p$ stand for the (unknown) slopes. Let us denote by $\boldsymbol{Z}_j$, the vector of observations, i.e., $\boldsymbol{Z}_j = (\boldsymbol{X}_j^\top, \boldsymbol{Y}_j^\top)^\top$, with $\boldsymbol{X}_j{=}(X_{1j}, \cdots, X_{pj})^\top$, $\boldsymbol{Y}_j{=}(Y_{1j}, \cdots, Y_{pj})^\top$ and assume that $\boldsymbol{Z}_j \sim t_{2p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, where $t_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ denotes a k-variate Student_t distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom and $\boldsymbol{\mu} = \mu\mathbf{b}$, $\boldsymbol{\Sigma} = \sigma_x^2\mathbf{b}\mathbf{b}^\top + \sigma^2\boldsymbol{D}(\mathbf{1}_p, \boldsymbol{\lambda})$, with $\mathbf{b} = (\mathbf{1}_p^\top, \boldsymbol{\beta}^\top)^\top$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)^\top$, $\boldsymbol{D}$ denotes the diagonal matrix, $\mathbf{1}_p$ the vector composed by p ones, so that $\boldsymbol{D}(\mathbf{1}_p, \boldsymbol{\lambda})$ denotes the diagonal matrix with diagonal elements $1, \cdots 1, \lambda_1, \cdots, \lambda_p$. The log-likelihood function of model (2.1)-(2.2) is given by

$$L(\boldsymbol{\theta}) = \sum_{j=1}^{n} l_j(\boldsymbol{\theta}), \tag{2.3}$$

where

$$l_j(\boldsymbol{\theta}) = const - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\nu + 2p)\log(\nu + d_j(\boldsymbol{\theta})), \tag{2.4}$$

with

$$d_j(\boldsymbol{\theta}) = d_j = (\boldsymbol{Z}_j - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_j - \boldsymbol{\mu}), \tag{2.5}$$

$j = 1, ..., n$ and $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^\top, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}^\top)^\top$.

Maximum Likelihood estimates for the vector of parameters $\boldsymbol{\theta}$ may be obtained by using iterative procedures, based on EM algorithm (which are described in the Appendix A), for example.

Several authors have considered the Student_t distribution as an alternative to the normal distribution as it can naturally accommodate outliers present in the data. Lange et. al. (1989) discussed the use of the Student_t distribution in regression models, as well as in problems related to multivariate analysis; Bolfarine and Arellano-Valle (1994) introduced Student_t functional and structural measurement error models, Bolfarine and Galea (1996) considered the Student_t distribution in comparative calibration models and Aoki et al. (2003) studied the null intercept structural measurement error model defined in (2.1) and (2.2) considering a bayesian approach.

The Student_t distribution incorporates an additional parameter, ν, namely the degrees of freedom, which allows adjusting for the kurthosis of the distribution. This parameter can be fixed previously. In Lange et al. (1989) and Berkane et al. (1994) it was recommended to take $\nu = 4$ or, otherwise, to get information about it from the data set. For some difficulty in the estimation of ν, see Fernández and Steel (1999).

3 Local influence diagnostics

Case deletion is a popular way to asses the individual impact of cases on the estimation process. This approach can be regarded as a global measure of influence. An alternative methodology for the identification of groups of cases which may require some concern is local influence wich is based on differential geometry instead of complete deletion. It employs a differential comparison of parameter estimates before and after perturbation to data values or model assumptions. As considered in Cook (1986), the likelihood displacement is used as the metric to assess the local influence.

Let $L(\boldsymbol{\theta})$ denote the log-likelihood function given in (2.3), $\boldsymbol{\omega}$, $q \times 1$, the perturbation introduced in the model, where $\boldsymbol{\omega} \in \boldsymbol{\Omega} \subseteq \mathbb{R}^q$, $\boldsymbol{\Omega}$ an open subset and $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ the log-likelihood function corresponding to the perturbed data or model. Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_\omega$ denote the maximum likelihood estimates under the model defined by $L(\boldsymbol{\theta})$ and $L(\boldsymbol{\theta}|\boldsymbol{\omega})$, respectively, and assume that there is an $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ representing no perturbation, such that $L(\boldsymbol{\theta})$=$L(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$ for all $\boldsymbol{\theta}$. The influence of $\boldsymbol{\omega}$ can be assessed by the log-likelihood displacement

$$LD(\boldsymbol{\omega}) = 2[L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_\omega)], \tag{3.1}$$

where $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\omega_0}$. Because evaluation of $LD(\boldsymbol{\omega})$ for all $\boldsymbol{\omega}$ is practically unfeasible, Cook (1986) proposed to study the local behaviour of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$, which can be performed by evaluating the normal curvature C_l of $LD(\boldsymbol{\omega})$ at $\boldsymbol{\omega}_0$ in the direction of some unit vector $\boldsymbol{l}$.

Cook (1986) showed that the normal curvature in the direction $\boldsymbol{l}$ takes the form

$$C_l = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top \ \boldsymbol{I}^{-1} \boldsymbol{\Delta} \boldsymbol{l}|, \tag{3.2}$$

where $\|\boldsymbol{l}\| = 1$, $\boldsymbol{I} = -\dfrac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$ is a $(2p+3) \times (2p+3)$ observed information matrix, and

$$\boldsymbol{\Delta} = \frac{\partial^2 L(\boldsymbol{\theta}/\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top} \tag{3.3}$$

are both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega_0}$.

There are many ways of studying the influence of minor perturbations considering C_l. Let $\boldsymbol{l}_{\max}$ be the direction of the maximum normal curvature (C_{max}). Then, it is the perturbation that produces the greatest local change in $\widehat{\boldsymbol{\theta}}$. The most influential elements of the data may be identified by looking at the components of the vector $\boldsymbol{l}_{\max}$, which are relatively large. Furthermore, $\boldsymbol{l}_{\max}$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Delta}^\top \ \boldsymbol{I}^{-1} \boldsymbol{\Delta}$, which is C_{max}. Other important direction is $\boldsymbol{l} = \boldsymbol{e}_j$, denoting that the element of the jth position is one. In that case, the normal curvature, called the total local influence of individual j, is given by $C_j = 2\boldsymbol{\Delta}_j^\top \boldsymbol{I}^{-1} \boldsymbol{\Delta}_j$, where $\boldsymbol{\Delta}_j$ is the jth column of $\boldsymbol{\Delta}$, $j = 1, ..., n$. We use $\boldsymbol{l}_{\max}$ and C_{max} as diagnostics for local influence. From (2.3), it follows that $\boldsymbol{I}$ takes the form

$$\boldsymbol{I} = -\left[\left(\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} \right)\right],$$

where, $\boldsymbol{\gamma}, \boldsymbol{\tau} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}$. The elements of the matrix $\boldsymbol{I}$ are presented in the Appendix B.

When a subset $\boldsymbol{\theta}_1$ from the partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is of interest, influence diagnostics can be based on (Cook, 1986)

$$\boldsymbol{\Delta}^\top(\boldsymbol{I}^{-1} - \boldsymbol{B}_{22})\boldsymbol{\Delta},$$

with

$$\boldsymbol{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{22}^{-1} \end{pmatrix}$$

and $\boldsymbol{I}_{22}$ is determined by the partition of $\boldsymbol{I}$ accordingly with the partition of $\boldsymbol{\theta}$. We consider several perturbation schemes for the model defined in (2.1) and (2.2), which is given in the next subsections.

3.1 Perturbation of case weights

Consider the vector $\boldsymbol{w} = (w_1, ..., w_n)^\top$ of case-weights, so that the perturbed log-likelihood function is given by

$$L(\boldsymbol{\theta}/\boldsymbol{w}) = \sum_{j=1}^{n} w_j l_j(\boldsymbol{\theta}),$$

where $l_j(\boldsymbol{\theta})$ is as in (2.4). The vector of no perturbations is denoted by $\boldsymbol{w}_0 = \mathbf{1}_n$. Under this perturbation scheme the matrix $\boldsymbol{\Delta}$ defined in (3.3) is a $(2p+3) \times n$ matrix and given by $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1(\boldsymbol{\theta}), ..., \boldsymbol{\Delta}_n(\boldsymbol{\theta}))$, where $\boldsymbol{\Delta}_j(\boldsymbol{\theta}) = \dfrac{\partial l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ with individual elements given by

$$\frac{\partial l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2}\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\gamma}} - \frac{1}{2}\frac{\nu + 2p}{(\nu + d_j)} d_j\boldsymbol{\gamma},$$

with $d_j\boldsymbol{\gamma} = \dfrac{\partial d_j}{\partial \boldsymbol{\gamma}}$, $\boldsymbol{\gamma} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}$ and d_j as given in (2.5), $j = 1, ..., n$. The components of matrix $\boldsymbol{\Delta}$ is presented in Appendix B.

3.2 Perturbation of the response variables

One way of perturbing the response variable, when our interest is to detect the sensitivity of the model when this kind of perturbation happens, we can consider for example, a sequence of scale factors $S_1, ..., S_n$, where

$$\boldsymbol{Y}_{wj} = \boldsymbol{Y}_j + \boldsymbol{S} * \boldsymbol{w}_j,$$

with $\boldsymbol{S} = (S_1, ..., S_p)^\top$, $\boldsymbol{w}_j = (w_{1j}, ..., w_{pj})^\top$ and $*$ denotes Rademaker product. The scale factor S_i can be taken as $S_i = S_{Y_i}$, where S_{Y_i} denotes for example, the sample standard deviation of $Y_{i1}, ..., Y_{in}$, $i = 1, ..., p$. The perturbed log-likelihood function is given by $L(\boldsymbol{\theta}/\boldsymbol{w}) = \sum_{j=1}^{n} l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)$, where $l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)$ is as given in (2.4), switching $\boldsymbol{Y}_{wj}$ with $\boldsymbol{Y}_j$ and $\boldsymbol{w} = (\boldsymbol{w}_1^\top, ..., \boldsymbol{w}_n^\top)^\top$. Under this perturbation scheme the vector $\boldsymbol{w}_0$, representing no perturbation is given by $\boldsymbol{w}_0 = \boldsymbol{0}$ and the $(2p+3) \times np$ matrix $\boldsymbol{\Delta}$, which is given in (3.3) can be expressed as $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1(\boldsymbol{\theta}; \boldsymbol{w}_1), ..., \boldsymbol{\Delta}_n(\boldsymbol{\theta}; \boldsymbol{w}_n))$, where $\boldsymbol{\Delta}_j(\boldsymbol{\theta}, \boldsymbol{w}_j)$ is given by

$$\boldsymbol{\Delta}_j(\boldsymbol{\theta}, \boldsymbol{w}_j) = \left(\frac{\partial^2 l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{w}_j^\top} \right), \quad \boldsymbol{\gamma} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda},$$

with

$$\frac{\partial^2 l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{w}_j^\top} = \frac{1}{2} \frac{\nu + 2p}{(\nu + d_j(w))^2} d_j\boldsymbol{\gamma}(w) d_j^\top \boldsymbol{w}_j(w) - \frac{1}{2} \frac{\nu + 2p}{\nu + d_j(w)} d_j \boldsymbol{\gamma} \boldsymbol{w}_j(w),$$

and

$$d_j\boldsymbol{\gamma}(w) = \frac{\partial d_j(w)}{\partial \boldsymbol{\gamma}}, \quad d_j\boldsymbol{w}_j(w) = \frac{\partial d_j(w)}{\partial \boldsymbol{w}_j}, d_j\boldsymbol{\gamma}\boldsymbol{w}_j(w) = \frac{\partial^2 d_j(w)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{w}_j^\top}, \tag{3.4}$$

and $d_j(w)$ as defined in (2.5), switching $\boldsymbol{Y}_{w_j}$ with $\boldsymbol{Y}_j$, $j = 1, ..., n$. The elements that compose the matrix $\boldsymbol{\Delta}$, can be found in Appendix B1.

3.3 Perturbation of the explanatory variables

If we are interested in investigating the sensitivity of minor perturbation in the explanatory variable, we can define the following perturbation scheme for the explanatory variable in the same way that was defined in the last subsection for the response variable. Let

$$\boldsymbol{X}_{wj} = \boldsymbol{X}_j + \boldsymbol{S} * \boldsymbol{w}_j,$$

where $\boldsymbol{S} = (S_1, ..., S_p)^\top$, $\boldsymbol{w}_j = (w_{1j}, ..., w_{pj})^\top$ and $*$ denotes the Rademaker product. The scale factor S_i can be defined as $S_i = S_{X_i}$, with S_{X_i} denoting the sample standard deviation of $X_{i1}, ..., X_{in}$, $i = 1, ..., p$. The log-likelihood function for the perturbed model is denoted by $L(\boldsymbol{\theta}/\boldsymbol{w}) = \sum_{j=1}^{n} l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)$, where $l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)$ is as defined in (2.4), switching $\boldsymbol{X}_{wj}$ with $\boldsymbol{X}_j$. The vector $\boldsymbol{w}_0$ representing no perturbation is given by $\boldsymbol{w}_0 = \mathbf{0}$ and the $(2p+3) \times np$ matrix $\boldsymbol{\Delta}$ defined in (3.3) is given by

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1(\boldsymbol{\theta}; \boldsymbol{w}_1), ..., \boldsymbol{\Delta}_n(\boldsymbol{\theta}; \boldsymbol{w}_n)), \quad \text{with}$$

$$\boldsymbol{\Delta}_j(\boldsymbol{\theta}, \boldsymbol{w}_j) = \left(\frac{\partial^2 l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{w}_j^\top} \right), \quad \boldsymbol{\gamma} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}, \quad j = 1, \ldots, n, \quad \text{and}$$

$$\frac{\partial^2 l_j(\boldsymbol{\theta}/\boldsymbol{w}_j)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{w}_j{}^\top} = \frac{1}{2} \frac{\nu + 2p}{(\nu + d_j(w))^2} d_{j\boldsymbol{\gamma}}(w) d_{j\boldsymbol{w}_j}^\top(w) - \frac{1}{2} \frac{\nu + 2p}{\nu + d_j(w)} d_{j\boldsymbol{\gamma}\boldsymbol{w}_j}(w),$$

where $d_{j\boldsymbol{\gamma}}(w)$, $d_{j\boldsymbol{w}_j}(w)$ and $d_{j\boldsymbol{\gamma}\boldsymbol{w}_j}(w)$ are as defined in (3.4) and $d_j(w)$ as given in (2.5), switching $\boldsymbol{X}_{wj}$ with $\boldsymbol{X}_j$, $j = 1, ..., n$. The components of the matrix $\boldsymbol{\Delta}$ are given in the Appendix B2.

3.4 Perturbation of the degrees of freedom

When we assume a fixed known value of the degree of freedom, it is of interest to study the effect of the minor perturbation in the degree of freedom in the estimation process. In that way, we are going to consider a known value of the degree of freedom parameter, namely ν_0 and the vector of observed data $\boldsymbol{Z}_j$, $j = 1, \cdots, n$ as defined in Section 2. The perturbation is introduced in the model by considering

$$\boldsymbol{Z}_j \overset{\text{ind}}{\sim} t_{2p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu_0 g(\omega_j)), \tag{3.5}$$

where g is a differentiable positive function and we assume the existence of ω_{0_j}, such that $g(\omega_{0_j})$=1 and $g^{'}(\omega_{0_j}) \neq 0$, $j = 1, ..., n$. Under the perturbed model the log-likelihood function is as given in (2.3), switching ν with $\nu_j = \nu_0 g(\omega_j)$, $j = 1, ..., n$. The matrix $\boldsymbol{\Delta}$ defined in (3.3) is given by

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1(\boldsymbol{\theta}; g(w_1)), ..., \boldsymbol{\Delta}_n(\boldsymbol{\theta}; g(w_n))),$$

where $\boldsymbol{\Delta}_j(\boldsymbol{\theta}, g(w_j))$ is given by

$$\boldsymbol{\Delta}_j(\boldsymbol{\theta}, g(w_j)) = \frac{\nu_0}{2} g'(\omega_{0_j})(d_j - 2p)(\nu_0 + d_j)^{-2} \frac{\partial d_j}{\partial \boldsymbol{\theta}},$$

$j = 1, ..., n$, evaluated at $\widehat{\boldsymbol{\theta}}$. The function g can be chosen, for example, as in Escobar and Meeker (1992), where $g(\omega_j) = a^{\omega_j}$, with $a > 0$ and $\omega_j \in [-1, 1]$, $j = 1, ..., n$. In that case, $\nu_j = \nu_0 g(\omega_j) \in [\nu_0/a, a\nu_0]$. For instance, if we assume that $a = 2$, $g(\omega_j) = 2^{\omega_j}$ and $g'(\omega_{0_j}) = \log 2$, for $j = 1, ..., n$.

4 Application

Considering the real data described in the Introduction and the model defined by (2.1) and (2.2), it follows that the observed vectors $\boldsymbol{X}_j = (X_{1j}, X_{2j})^\top$ and $\boldsymbol{Y}_j = (Y_{1j}, Y_{2j})^\top$, $j = 1, \cdots, n$, corresponds respectiverly, to the dental plaque index before and after toothbrushing with the hugger toothbrush ($i = 1$) and the conventional toothbrush ($i = 2$), for the jth preschooler. First, we are going to apply the perturbation of case weights, where each case is represented by the vector $\boldsymbol{Z}_j = (\boldsymbol{X}_j^\top, \boldsymbol{Y}_j^\top)^\top$. Figure 1 corresponds to the index plot of $\boldsymbol{l}_{max}$ to assess the influence of the perturbation $\boldsymbol{\omega}$ on the maximum likelihood estimator of the full parameter vector $\boldsymbol{\theta}$, considering the degree of freedom parameter ν=1, 4 and 50.

If we refer to the influence graph in the model using 50 degrees of freedom, we note that the observations 4 and 13 stand out. The same has happened for 250, 500 and 10000 degrees of freedom and as expected for the normal distribution. On the other hand, if we consider the model using low degrees of freedom there are no influent observations, which means that the Student_t model with low degrees of freedom can accomodate these observations. In Aoki et al. (2003) the model defined by the equations (2.1) and (2.2) was analysed, considering the Bayesian approach and Student_t distribution, as well as the normal distribution. It was concluded that the Student_t distribution with low degrees of freedom, more specifically 4 degrees of freedom

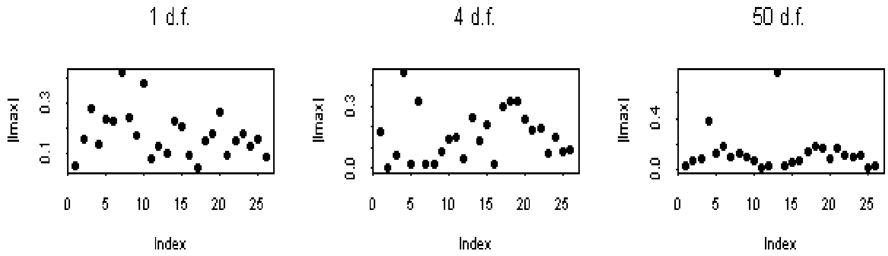

Figure 1: Perturbation of case weights for $\nu = 1, 4$ and 50.

are more appropriate for this data set. Considering the normal distribution, the most influential observation in the data set is the observation 13, which is not the case if we consider the Student_t distribution. In that way we estimated the parameter values considering the normal and Student_t distribution with 4 degrees of freedom with the complete data set and excluding the observation 13 from the data set, wich is given in Table 1. As expected, considering the normal distribution, the observation 13 influences the parameters estimation, while if we consider the student_t distribution with 4 degrees of freedom the same observation has litte influence in the estimation process.

Next, we illustrate the perturbation of the degrees of freedom considering $g(\omega_j) = 2^{\omega_j}$, $j = 1, \ldots, n$. In this case, we obtained the following influence graphs (Figure 2) for ν= 1, 2 and 4 degrees of freedom.

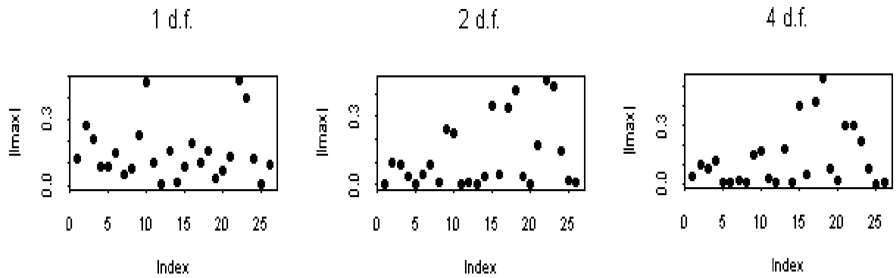

Figure 2: Perturbation of degree of freedom for $\nu = 1, 2$, and 4.

Table 1: Maximum likelihood estimates

Normal Distribution	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}$	$\hat{\sigma}_x^2$	$\hat{\sigma}^2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
complete data set	0,147	0,454	1,759	0,540	0,481	0,102	0,267
without obs. 13	0,135	0,464	1,760	0,594	0,367	0,091	0,310

Student_t Distribution with 4 degrees of freedom	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}$	$\hat{\sigma}_x^2$	$\hat{\sigma}^2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$
complete data set	0,130	0,441	1,654	0,594	0,384	0,083	0,263
without obs. 13	0,125	0,447	1,652	0,608	0,390	0,075	0,278

Considering these graphis, we conclude that there are no influent observations and as we assumed a fixed known value of the degree of freedom it is important to know the effect of a minor perturbation in the degree of freedom in the estimation process.

Appendix A: EM Algorithm

Considering the model defined in Section 2, we are going to present an iterative procedure to obtain the maximum likelihood estimates of the parameter $\boldsymbol{\theta}$, as the log likelihood function given by (2.3) has no explicit solutions for the likelihood equations. In that way, we are going to implement the EM algorithm. Let us define by $\mathbf{T}_j = (x_j, \mathbf{Z}_j^\top)^\top$, with $\mathbf{Z}_j = (\boldsymbol{X}_j^\top, \boldsymbol{Y}_j^\top)^\top$. As defined in Section 2, $\mathbf{Z}_j \sim t_{2p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ and $\mathbf{T}_j \sim t_{2p+1}(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T; \nu)$ where $\boldsymbol{\mu}_T = (\mu, \boldsymbol{\mu}^\top)^\top$, $\boldsymbol{\Sigma}_T = \begin{pmatrix} \sigma_x^2 & \sigma_x^2\mathbf{b}^\top \\ \sigma_x^2\mathbf{b} & \boldsymbol{\Sigma} \end{pmatrix}$. Let us define by $Q_j \sim \frac{\chi^2_{(\nu)}}{\nu}$, $\nu > 0$, $j = 1, \cdots, n$ and $\mathbf{T}_j | (Q_j = q_j) \sim N_{2p+1}(\boldsymbol{\mu}_T, q_j^{-1}\boldsymbol{\Sigma}_T)$, so that $\mathbf{T}_j \sim$

$t_{2p+1}(\boldsymbol{\mu}_T, \boldsymbol{\Sigma}_T; \nu)$. Note that if $f(t_j, q_j)$ denotes the joint density of $(\mathbf{T}_j, Q_j)$, $j = 1, \cdots, n$, then $f(t_j, q_j) = f_1(t_j/q_j)f_2(q_j)$, so that the complete log likelihood function is given by

$$\begin{aligned} L_c(\boldsymbol{\theta}) = \quad & const - \frac{n}{2}\log[(\sigma^2)^{2p}\prod_{i=1}^{p}\lambda_i\sigma_x^2] - \frac{1}{2}\sum_{j=1}^{n} q_j\{\frac{(x_j-\mu)^2}{\sigma_x^2} \\ & + \frac{1}{\sigma^2}(\boldsymbol{Z}_j - \boldsymbol{b}x_j)^\top D^{-1}(\mathbf{1}_p, \boldsymbol{\lambda})(\boldsymbol{Z}_j - \boldsymbol{b}x_j)\} + \sum_{j=1}^{n}\log f_2(q_j), \quad (3.6) \end{aligned}$$

where $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^\top, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}^\top)^\top$. Each cycle of the EM algorithm has two steps, namely the E and M steps.

E Step

The E step is defined by the equations

$$\begin{aligned} 1)\ \hat{q}_j \quad &= \quad E(q_j|\mathbf{Z}, \boldsymbol{\theta}) = \frac{\nu + 2p}{\nu + d_j}, \\ 2)\ \hat{x}_j \quad &= \quad E(x_j|\mathbf{Z}, \boldsymbol{\theta}) = \mu + \frac{\sigma_x^2}{c\sigma^2}\mathbf{b}^\top D^{-1}(\mathbf{1}_p, \boldsymbol{\lambda})(\boldsymbol{Z}_j - \mu\boldsymbol{b}) \end{aligned}$$

and

$$3)\ \hat{x}_j^2 \quad = \quad E(x_j^2|\mathbf{Z}, \boldsymbol{\theta}) = \widehat{x}_j^2 + \frac{\sigma_x^2}{c}\frac{\nu + d_j}{\nu + 2p - 2},$$

with d_j as defined in (2.5), $j = 1, \cdots, n$.

M Step

In this step the complete data log likelihood function given in (3.6) is maximized. Equating the likelihood equations to zero, we obtain after algebraic manipulations

$$\hat{\mu} = \frac{\sum_{j=1}^{n}\hat{q}_j\hat{x}_j}{\sum_{j=1}^{n}\hat{q}_j},\ \hat{\sigma}^2 = \frac{1}{np}\sum_{j=1}^{n}\hat{q}_j\sum_{i=1}^{p}(X_{ij}^2 - 2\hat{x}_jX_{ij} + \hat{x}_j^2),\ \hat{\sigma}_x^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{q}_j(\hat{x}_j^2 - \hat{\mu}^2),$$

$$\hat{\beta}_i = \frac{\sum_{j=1}^{n} \hat{q}_j \hat{x}_j Y_{ij}}{\sum_{j=1}^{n} \hat{q}_j \hat{x}_j^2} \quad \text{and} \quad \hat{\lambda}_i = \frac{1}{np\hat{\sigma}^2} \sum_{j=1}^{n} \hat{q}_j (Y_{ij}^2 - \hat{x}_j^2 \hat{\beta}_i^2), \quad i = 1, \cdots, p.$$

The EM algorithm cycles between equations given in the E step and the equations given in M step until convergence (Dempster et.al., 1977). Note that as no additional iterative procedure is requered to solve the M step within each cycle of the algorithm, this procedure is extremely simple to implement and computationally inexpensive. Considering the model defined in Section 2, the following maximum likelihood estimates were obtained for fixed values of the degrees of freedom.

Table 3: Maximum likelihood estimates (MLE) of the parameters under the model defined in Section 2, via EM algorithmith for the data presented in Singer and Andrade (1997).

degrees of	Parameter						
freedom	β_1	β_2	μ	σ_x^2	σ^2	λ_1	λ_2
1	0.123	0.431	1.614	0.900	0.588	0.066	0.245
2	0.126	0.436	1.628	0.676	0.434	0.074	0.254
3	0.128	0.439	1.642	0.617	0.397	0.080	0.259
4	0.130	0.441	1.654	0.594	0.384	0.084	0.263
20	0.140	0.450	1.724	0.559	0.414	0.098	0.274
50	0.143	0.453	1.744	0.551	0.446	0.101	0.272
10000	0.147	0.454	1.759	0.539	0.481	0.102	0.267
Normal	0.147	0.454	1.758	0.539	0.482	0.102	0.267

Appendix B: Computing the observed information matrix in the Student_t structural model

In this appendix we present the elements of the observed information matrix. From (2.3), it follows that

$$\frac{\partial l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\frac{1}{2}\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\gamma}} - \frac{1}{2}\frac{\nu + 2p}{(\nu + d_j)} d_j\boldsymbol{\gamma}, \tag{B.1}$$

with $d_j\boldsymbol{\gamma} = \dfrac{\partial d_j(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}}$, $\boldsymbol{\gamma} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}$ and d_j as given in (2.5) $j = 1, ..., n$. After some algebraic manipulations it follows that

$$\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \mu} = 0, \quad \frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta}} = 2c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}, \quad \frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \sigma_x^2} = c^{-1}\frac{c-1}{\sigma_x^2},$$

$$\frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \sigma^2} = -c^{-1}\frac{c-1}{\sigma^2} + \frac{2p}{\sigma^2}, \quad \frac{\partial \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\lambda}} = -c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta} + \boldsymbol{D}^{-1}(\boldsymbol{\lambda})\mathbf{1}_p,$$

$$\begin{aligned}
d_{j\mu} =& \ -2\frac{c^{-1}}{\sigma^2}A_j \\
d_{j\boldsymbol{\beta}} =& \ -2\frac{\mu}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu) \\
& +2c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j^2\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta} - 2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu), \\
d_{j\sigma_x^2} =& \ -\frac{c^{-2}}{\sigma^4}A_j^2 \\
d_{j\sigma^2} =& \ -\frac{1}{\sigma^2}d_j + c^{-2}\frac{\sigma_x^2}{\sigma^6}A_j^2 \\
d_{j\boldsymbol{\lambda}} =& \ -\frac{1}{\sigma^2}\boldsymbol{D}(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu)\boldsymbol{D}^{-2}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu) \\
& -c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j^2\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta} + 2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu),
\end{aligned}$$

where $c = 1 + \frac{\sigma_x^2}{\sigma^2}(p + \boldsymbol{\beta}^\top\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta})$ and $A_j = (\boldsymbol{X}_j - \mathbf{1}_p\mu)^\top\mathbf{1}_p + \boldsymbol{\beta}^\top\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu)$.

From (A.1) it follows that the per element observed information matrix is given by

$$I_j = I_j(\boldsymbol{\theta}/\boldsymbol{Z}_j) = -\left(\frac{\partial^2 l_j(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top}\right),$$

where

$$\frac{\partial^2 l_j(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = -\frac{1}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} + \frac{1}{2}\frac{\nu + 2p}{(\nu + d_j)^2} d_{j\boldsymbol{\gamma}} d_{j\boldsymbol{\tau}}^\top - \frac{1}{2}\frac{\nu + 2p}{\nu + d_j} d_{j\boldsymbol{\gamma}\boldsymbol{\tau}},$$

with $d_{j\boldsymbol{\gamma}\boldsymbol{\tau}} = \dfrac{\partial^2 d_j}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top}$ and $\boldsymbol{\gamma}, \boldsymbol{\tau} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}$. The component of $\mathbf{I}_j$ can be expressed as

$$\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \mu \partial \boldsymbol{\gamma}^\top} = 0, \quad \boldsymbol{\gamma} = \mu, \boldsymbol{\beta}, \sigma_x^2, \sigma^2, \boldsymbol{\lambda}$$

$$\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda}) - 4c^{-2}\frac{\sigma_x^4}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{D}^{-1}(\boldsymbol{\lambda})$$

$$\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \sigma_x^2} = 2\frac{c^{-2}}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta},$$

$$\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \sigma^2} = -2c^{-2}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta},$$

$$\begin{aligned}
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}^\top} &= 2c^{-1}\frac{\sigma_x^2}{\sigma^2}[c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta}) - \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})], \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \sigma_x^2 \partial \sigma_x^2} &= -c^{-2}(c-1)^2/\sigma_x^4, \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \sigma_x^2 \partial \sigma^2} &= -c^{-2}(c-1)/(\sigma^2\sigma_x^2), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \sigma_x^2 \partial \boldsymbol{\lambda}^\top} &= -\frac{c^{-2}}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \sigma^2 \partial \sigma^2} &= \frac{c^{-2}}{\sigma^4}(c^2(1-2p) - 1), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \sigma^2 \partial \boldsymbol{\lambda}^\top} &= c^{-2}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda}), \\
\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} &= -c^{-1}\frac{\sigma_x^2}{\sigma^2}[c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta}) - 2\boldsymbol{D}^{-3}(\boldsymbol{\lambda})\boldsymbol{D}^2(\boldsymbol{\beta})] \\
&\quad -\boldsymbol{D}^{-2}(\boldsymbol{\lambda}), \\
d_{j\mu\mu} &= 2\frac{c^{-1}(c-1)}{\sigma_x^2},
\end{aligned}$$

$$
\begin{aligned}
d_{j\mu\boldsymbol{\beta}} &= -2\frac{c^{-1}}{\sigma^2}(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)^\top \boldsymbol{D}^{-1}(\boldsymbol{\lambda}) + 4c^{-2}\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{\beta}^\top \boldsymbol{D}^{-1}(\boldsymbol{\lambda}),\\
d_{j\mu\sigma_x^2} &= 2\frac{c^{-2}(c-1)}{\sigma_x^2\sigma^2}A_j,\\
d_{j\mu\sigma^2} &= 2\frac{c^{-2}}{\sigma^4}A_j,\\
d_{j\mu\boldsymbol{\lambda}} &= -2\frac{c^{-1}}{\sigma^2}[c^{-1}\frac{\sigma_x^2}{\sigma^2}A_j\boldsymbol{\beta}^\top \boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda}) - (\boldsymbol{Y}_j - \boldsymbol{\beta}\mu)^\top \boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})],\\
d_{j\boldsymbol{\beta}\boldsymbol{\beta}} &= 2\frac{\mu^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda}) + 2c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j^2(-4c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{D}^{-1}(\boldsymbol{\lambda}) + \boldsymbol{D}^{-1}(\boldsymbol{\lambda}))\\
&\quad +4c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j(\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)\boldsymbol{\beta}^\top \boldsymbol{D}^{-1}(\boldsymbol{\lambda})\\
&\quad +\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)^\top \boldsymbol{D}^{-1}(\boldsymbol{\lambda}))\\
&\quad -2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)^\top \boldsymbol{D}^{-1}(\boldsymbol{\lambda}) + 4c^{-1}\mu\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda}),\\
d_{j\boldsymbol{\beta}\sigma_x^2} &= 4c^{-3}\frac{\sigma_x^2}{\sigma^6}A_j^2\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta} - 2\frac{c^{-2}}{\sigma^4}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu),\\
d_{j\boldsymbol{\beta}\sigma^2} &= 2\frac{\mu}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu) - 2c^{-3}(c+2)\frac{\sigma_x^4}{\sigma^8}A_j^2\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\\
&\quad +2c^{-2}(c+1)\frac{\sigma_x^2}{\sigma^6}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu),\\
d_{j\boldsymbol{\beta}\boldsymbol{\lambda}} &= 2c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j^2[2c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta}) - \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})]\\
&\quad -2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j[2c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu)^\top \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})\\
&\quad +c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)\boldsymbol{\beta}^\top \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta}) - \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)],\\
&\quad +2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j - 2\boldsymbol{\beta}\mu)(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu)^\top \boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})\\
&\quad +2\frac{\mu}{\sigma^2}\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{Y}_j - \boldsymbol{\beta}\mu)\\
d_{j\sigma_x^2\sigma_x^2} &= 2\frac{c^{-3}(c-1)}{\sigma_x^2\sigma^4}A_j^2,\\
d_{j\sigma_x^2\sigma^2} &= 2\frac{c^{-3}}{\sigma^6}A_j^2,
\end{aligned}
$$

$$
\begin{aligned}
d_{j\sigma_x^2\boldsymbol{\lambda}} &= -2c^{-3}\frac{\sigma_x^2}{\sigma^6}A_j^2\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda}) + 2\frac{c^{-2}}{\sigma^4}A_j(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)^\top\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda}),\\
d_{j\sigma^2\sigma^2} &= \frac{2}{\sigma^4}d_j - 2c^{-3}(c+1)\frac{\sigma_x^2}{\sigma^8}A_j^2,\\
d_{j\sigma^2\boldsymbol{\lambda}} &= c^{-3}(c+2)\frac{\sigma_x^4}{\sigma^8}A_j^2\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\\
&\quad -2c^{-2}(c+1)\frac{\sigma_x^2}{\sigma^6}A_j(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)^\top\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\\
&\quad +\frac{1}{\sigma^4}(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)^\top\boldsymbol{D}(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{D}^{-2}(\boldsymbol{\lambda}),\\
d_{j\boldsymbol{\lambda}\boldsymbol{\lambda}} &= -2c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j^2[c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta}) - \boldsymbol{D}^{-3}(\boldsymbol{\lambda})\boldsymbol{D}^2(\boldsymbol{\beta})]\\
&\quad +2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j[c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{\beta}^\top\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})\\
&\quad +c^{-1}\frac{\sigma_x^2}{\sigma^2}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta}(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)^\top\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})\\
&\quad -2\boldsymbol{D}^{-3}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{D}(\boldsymbol{\beta})]\\
&\quad -2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)^\top\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{\beta})\\
&\quad +\frac{2}{\sigma^2}\boldsymbol{D}^2(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{D}^{-3}(\boldsymbol{\lambda}),
\end{aligned}
$$

$j = 1, ..., n$. The complete observed information matrix is $\mathbf{I} = \sum_{j=1}^{n}\mathbf{I}_j(\boldsymbol{\theta}/\boldsymbol{Z}_j)$.

Appendix B1: Perturbation of the response variables

Elements of the $\boldsymbol{\Delta}$ matrix evaluated at $\boldsymbol{w} = 0$.

$$
\begin{aligned}
d_{j\boldsymbol{w}} &= \frac{2}{\sigma^2}\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu) - 2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta},\\
d_{j\mu\boldsymbol{w}} &= -2\frac{c^{-1}}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda}),\\
d_{j\boldsymbol{\beta}\boldsymbol{w}} &= -2\frac{\mu}{\sigma^2}\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda}) + 4c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{S}_Y)\\
&\quad -2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-2\boldsymbol{\beta}\mu)\boldsymbol{\beta}^\top\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{S}_Y) - 2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{D}(\boldsymbol{S}_Y),
\end{aligned}
$$

$$
\begin{aligned}
d_{j\sigma_x^2\boldsymbol{w}} &= -2\frac{c^{-2}}{\sigma^4}A_j\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda}),\\
d_{j\sigma^2\boldsymbol{w}} &= -\frac{1}{\sigma^2}d_{j\boldsymbol{w}}^\top + 2c^{-2}\frac{\sigma_x^2}{\sigma^6}A_j\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda}),\\
d_{j\boldsymbol{\lambda}\boldsymbol{w}} &= -\frac{2}{\sigma^2}\boldsymbol{D}(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-2}(\boldsymbol{\lambda}) - 2c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\\
&\quad +2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{\beta}^\top\boldsymbol{D}(\boldsymbol{S}_Y)\boldsymbol{D}^{-1}(\boldsymbol{\lambda}),
\end{aligned}
$$

Appendix B2: Perturbation of the explanatory variables

Elements of the $\boldsymbol{\Delta}$ matrix evaluated at $\boldsymbol{w}=0$.

$$
\begin{aligned}
d_{j\boldsymbol{w}} &= \frac{2}{\sigma^2}\boldsymbol{D}(\boldsymbol{S}_X)(\boldsymbol{X}_j-\boldsymbol{1}_p\mu) - 2c^{-1}\frac{\sigma_x^2}{\sigma^4}A_j\boldsymbol{D}(\boldsymbol{S}_X)\boldsymbol{1}_p,\\
d_{j\mu\boldsymbol{w}} &= -2\frac{c^{-1}}{\sigma^2}\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X),\\
d_{j\boldsymbol{\beta}\boldsymbol{w}} &= 4c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j\boldsymbol{D}^{-1}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X) - 2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}^{-1}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-2\boldsymbol{\beta}\mu)\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X),\\
d_{j\sigma_x^2\boldsymbol{w}} &= -2\frac{c^{-2}}{\sigma^4}A_j\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X),\\
d_{j\sigma^2\boldsymbol{w}} &= -\frac{1}{\sigma^2}d_{j\boldsymbol{w}}^\top + 2c^{-2}\frac{\sigma_x^2}{\sigma^6}A_j\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X),\\
d_{j\boldsymbol{\lambda}\boldsymbol{w}} &= -2c^{-2}\frac{\sigma_x^4}{\sigma^6}A_j\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})\boldsymbol{\beta}\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X)\\
&\quad +2c^{-1}\frac{\sigma_x^2}{\sigma^4}\boldsymbol{D}(\boldsymbol{\beta})\boldsymbol{D}^{-2}(\boldsymbol{\lambda})(\boldsymbol{Y}_j-\boldsymbol{\beta}\mu)\boldsymbol{1}_p^\top\boldsymbol{D}(\boldsymbol{S}_X),
\end{aligned}
$$

References

[1] Abdullah, M. B. (1995), Detection of Influential Observations in Functional Errors-in-Variables Model. Communications in Statistics - Theory Meth., vol. 24 (6), 1585-1595.

[2] Aoki, R. (2001), Modelos de Regressão com Erros nas Variáveis com Intercepto Nulo. Doctoral Thesis, IME-USP (in Portuguese).

[3] Aoki, R., Bolfarine, H. and Singer, J. M. (2001), Null Intercept Measurement Error Regression Models, Test, 10 (2), 441-457.

[4] Aoki, R., Achcar, J.A., Bolfarine, H. and Singer, J.M. (2003), Bayesian Analysis of Null Intercept Error-in-Variables Regression for Pretest/Posttest Data. Journal of Applied Statistics, 30, 5-14.

[5] Bolfarine, H. and Arellano-Valle, R. B.(1994), Robust modeling in measurement error models using the Student-t distributions. Brazilian Journal of Probability and Statistics, 8, 67-84.

[6] Bolfarine, H. and Galea-Rojas, M.(1996), One structural comparative calibration under a t-models. Computational Statistics, 11, 63-85.

[7] Berkane, M., Kano, Y. and Bentler, P.M.(1994), Pseudo maximun likelihood estimation in elliptical theory: Effects of misspecification. Computational Statistics & Data Analysis, 18, 255-267.

[8] Cook, D. (1977), Detection of influential observations in linear regression, Technometrics, vol. 19, 15-18.

[9] Cook, D. (1986), Assessment of Local Influence, Journal of Royal Statistical Society, vol. 48, B (2), 133-169.

[10] Dempster, A. , Laird, L. & D. B. Rubin (1977), Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion), Journal of the Royal Statistical Society, Series B, 39, 1-38.

[11] Escobar, E. and Meeker, W. (1992), Assessing influence in regression analysis with censored data. Biometrics, 48, 507-528.

[12] Fernández, C. and Steel, M. (1999), Multivariate Student regression Models : Pitfaalls and Inference. Biometrika, 86, 156-167.

[13] Kim, M. G. (2000), Outliers and influential observations in the structural errors-in-variables model, Journal of Applied Statistics, vol 24, 461-473.

[14] Lange, K.L., Little, R.J. and Taylor, J. (1989), Robust statistical modelling using the t-distribution. Journal of the American Statistical Association, 84, 881-896.

[15] Lee, A. H. and Zhao, Y. (1996), Assessing local influence in measurement error models, Biometrical Journal, vol. 38 (7), 829-841.

[16] Singer, J. M. and Andrade, J. M. (1997), Regression Models for the Analysis of Pretest, Posttest Data, Biometrics, vol 3 (3), 729-7359.