Likelihood Based Inference for Quantile Regression Using the Asymmetric Laplace Distribution

Luis B. Sánchez Victor H. Lachos *

Filidor V. Labra

Departamento de Estatística, Universidade Estadual de Campinas, Brazil

Abstract

To make inferences about the shape of a population distribution, the widely popular mean regression model, for example, is inadequate if the distribution is not approximately Gaussian (or symmetric). Compared to conventional mean regression (MR), quantile regression (QR) can characterize the entire conditional distribution of the outcome variable, and is more robust to outliers and misspecification of the error distribution. We present a likelihood-based approach to the estimation of the regression quantiles based on the asymmetric Laplace distribution (ALD), a choice that turns out to be natural in this context. The ALD has a nice hierarchical representation which facilitates the implementation of the EM algorithm for maximum-likelihood estimation of the parameters at the pth level with the observed information matrix as a byproduct. Inspired by the EM algorithm, we develop case-deletion diagnostics analysis for QR models, following the approach of Zhu et al. (2001). This is because the observed data log–likelihood function associated with the proposed model is somewhat complex (e.g., not differentiable at zero) and by using Cook's well-known approach it can be very difficult to obtain case-deletion measures. The techniques are illustrated with both simulated and real data. In particular, in an empirical comparison, our approach out-performed other common classic estimators under a wide array of simulated data models and is flexible enough to easily accommodate changes in their assumed distribution. The proposed algorithm and methods are implemented in the R package `ALDqr()`.

Keywords Quantile regression model; EM algorithm; Case-deletion model; asymmetric Laplace distribution.

1 Introduction

QR models have become increasingly popular since the seminal work of Koenker & Gilbert (1978). In contrast to the mean regression model, QR belongs to a robust model family, which can give an overall assessment of the covariate effects at different quantiles of the outcome (Koenker, 2005). In particular, we can model the lower or higher quantiles of the outcome to provide a natural assessment of covariate effects specific for those regression quantiles. Unlike conventional models,

*Corresponding author. Address for correspondence: Departamento de Estatística, Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz, Campinas, São Paulo, Brazil. CEP 13083-859. E-mail: `hlachos@ime.unicamp.br`

which only address the conditional mean or the central effects of the covariates, QR models quantify the entire conditional distribution of the outcome variable. In addition, QR does not impose any distributional assumption on the error, except requiring the error to have a zero conditional quantile. The foundations of the methods for independent data are now consolidated, and some statistical methods for estimating and drawing inferences about conditional quantiles are provided by most of the available statistical programs (e.g., R, SAS, Matlab and Stata). For instance, just to name a few, in the well-known R package `quantreg()` is implemented a variant of the Barrodale & Roberts (1977) simplex (BR) for linear programming problems described in Koenker & d'Orey (1987), where the standard errors are computed by the rank inversion method (Koenker, 2005). Another method implemented in this popular package is Lasso Penalized Quantile Regression (LPQR), introduced by Tibshirani (1996), where a penalty parameter is specified to determine how much shrinkage occurs in the estimation process. QR can be implemented in a range of different ways. Koenker (2005) provided an overview of some commonly used quantile regression techniques from a "classical" framework.

From a Bayesian point of view, Kottas & Gelfand (2001) considered median regression, which is a special case of quantile regression, and discussed non-parametric modeling for the error distribution based on either Pólya tree or Dirichlet process priors. Regarding general quantile regression, Yu & Moyeed (2001) proposed a Bayesian modeling approach by using the ALD, Kottas & Krnjajić (2009) developed Bayesian semi-parametric models for quantile regression using Dirichlet process mixtures for the error distribution. More recently, Kozumi & Kobayashi (2011) developed a simple and efficient Gibbs sampling algorithm for fitting the quantile regression model based on a location-scale mixture representation of the ALD. However, to the best of our knowledge, there are no studies on QR from a likelihood based perspective. Thus, the main contribution of this paper is to propose an alternative method for drawing inferences about conditional quantiles in linear regression problems via maximum-likelihood (ML) estimation. The hierarchical representation of the ALD enables the implementation of an efficient (and easy) EM algorithm for ML estimation of the parameters at the pth level with the standard error as a byproduct. As will be seen in a simulation study, our EM algorithm outperformed the competing BR and LPQR algorithms.

Since the traditional normal model is very sensitive to outlying observations, the assessment of robustness aspects of the parameter estimates is an important concern. The deletion method, which consists of studying the impact on the parameter estimates after dropping individual observations, is probably the most employed technique to detect influential observations, see Cook & Weisberg (1982) and the references therein. In the context of the QR model the marginal log-likelihood function is complex (e.g., not differentiable at zero), and with direct application of Cook & Weisberg's well-known approach it can be very difficult to obtain case-deletion measures. The work of Zhu *et al.* (2001) presents an approach to perform diagnostic analysis for general statistical models with missing data by working with a Q-displacement function closely related to the conditional expectation of the complete-data log-likelihood at the E-step of the EM algorithm. This method or modifications of it have been applied successfully to perform influence analysis in several regression models. See, for example, Zeller *et al.* (2010) and Matos *et al.* (2013), among others. Using this general method and taking advantage of the likelihood structure imposed by the ALD, in this paper we develop a case-deletion diagnostic approach for the QR model and show that it leads to simple influence measures. Since QR methods complement and improve established means regression models, we feel that the assessment of robustness aspects of the parameter estimates in QR is also an important concern at a given quantile level $p \in (0,1)$.

The rest of the paper is organized as follows. Section 2 introduces the connection between QR and ALD as well as outlining the main results related to ALD. In addition, we develop an EM-type algorithm to proceed with ML estimation for the parameters at the *p*th level and analytically derive the observed information matrix. In section 4 we develop influence diagnostic techniques, based on case deletion. Sections 5 and 6 are dedicated to the analysis of real and simulated data sets, respectively. Section 6 concludes with a short discussion of issues raised by our study and some possible directions for the future research.

2 The quantile regression model

As discussed in Yu & Moyeed (2001), we say that a random variable Y is distributed as an ALD with location parameter μ, scale parameter $\sigma > 0$ and skewness parameter $p \in (0,1)$, if its probability density function (pdf) is given by

$$f(y|\mu,\sigma,p) = \frac{p(1-p)}{\sigma}\exp\Big\{-\rho_p\big(\frac{y-\mu}{\sigma}\big)\Big\}, \tag{1}$$

where $\rho_p(.)$ is the so called check (or loss) function defined by $\rho_p(u) = u(p-\mathbb{I}\{u<0\})$, with $\mathbb{I}\{.\}$ denoting the usual indicator function. This distribution is denoted by $ALD(\mu,\sigma,p)$. It is easy to see that $W = \rho_p\big(\frac{Y-\mu}{\sigma}\big)$ follows an exponential distribution $\text{Exp}(\sigma)$ with mean σ.

The following result is useful to obtain some properties of this distribution, as for example, the moments, moment generating function (mgf), and estimation algorithm. The result that involves a stochastic representation can be found in Kotz *et al.* (2001).

Lemma 1. *Let $U \sim \text{Exp}(\sigma)$ and $Z \sim N(0,1)$ be two independent random variables. Then $Y \sim ALD(\mu,\sigma,p)$ can be represented by*

$$Y \overset{d}{=} \mu + \vartheta_p U + \tau_p\sqrt{\sigma U}Z,$$

where $\vartheta_p = \frac{1-2p}{p(1-p)}$ and $\tau_p^2 = \frac{2}{p(1-p)}$, and $\overset{d}{=}$ denotes equality in distribution.

Figure 1 shows how the skewness of the ALD changes with altering values for *p*. For example, where $p = 0.1$ almost all the mass of the ALD is situated in the right tail. In the case where $p = 0.5$, both tails of the ALD have equal mass and the distribution then equals the more common double exponential distribution. In contrast to the normal distribution with a quadratic term in the exponent, the ALD is linear in the exponent. This results in a more peaked mode for the ALD together with thicker tails. On the other hand, the normal distribution has heavier shoulders compared to the ALD. Lemma 1 yields a further hierarchical representation of Y in the following:

$$Y|U = u \sim N(\mu + \vartheta_p u, \tau_p^2\sigma u), \tag{2}$$

$$U \sim exp(\sigma). \tag{3}$$

It follows that the conditional distribution of U, given Y, is $U|(Y=y) \sim GIG(\frac{1}{2},\delta,\gamma)$, where $\delta = \frac{|y-\mu|}{\tau_p\sqrt{\sigma}}$ and $\gamma = \sqrt{\frac{1}{\sigma}\big(2+\frac{\vartheta_p^2}{\tau_p^2}\big)} = \frac{\tau_p}{2\sqrt{\sigma}}$. The notation $GIG(\nu,a,b)$ denotes the generalized inverse Gaussian (GIG) distribution with its pdf given by

$$f(u|\nu,a,b) = \frac{(b/a)^\nu}{2K_\nu(ab)}u^{\nu-1}\exp\Big\{-\frac{1}{2}\big(a^2u^{-1}+b^2u\big)\Big\},\ u>0,\ \ \nu\in\mathbb{R},\ a,b>0,$$

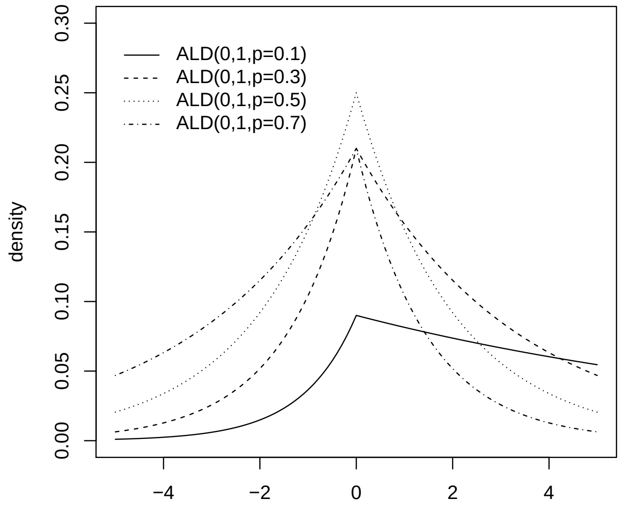

Figure 1: Standard asymmetric Laplace density

where $K_\nu(.)$ is a modified Bessel function of the third kind; see Barndorff-Nielsen & Shephard (2001) for more details. The moments of U are given by

$$E[U^k] = \left(\frac{a}{b}\right)^k \frac{K_{\nu+k}(ab)}{K_\nu(ab)}, \;\; k \in \mathbb{R}.$$

Some properties of the Bessel function of the third kind $K_\lambda(u)$ that will be useful for the developments here are: (i) $K_\nu(u) = K_{-\nu}(u)$; (ii) $K_{\nu+1}(u) = \frac{2\nu}{u}K_\nu(u) + K_{\nu-1}(u)$; (iii) for non-negative integer r, $K_{r+1/2}(u) = \sqrt{\frac{\pi}{2u}}\exp(-u)\sum_{k=0}^{r}\frac{(r+k)!(2u)^{-k}}{(r-k)!k!}$, and in particular, $K_{1/2}(u) = \sqrt{\frac{\pi}{2u}}\exp(-u)$.

2.1 Quantile regression using the asymmetric Laplace distribution

Let $y_i,\; i = 1,\ldots,n$, be a response variable and $\mathbf{x}_i$ a $k\times 1$ vector of covariates for the ith observation, and let $Q_{y_i}(p|\mathbf{x}_i)$ be the pth $(0 < p < 1)$ quantile regression function of y_i given $\mathbf{x}_i$. Suppose that the relationship between $Q_p(\mathbf{x}_i)$ and $\mathbf{x}_i$ can be modeled as $Q_{y_i}(p|\mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}_p$, where $\boldsymbol{\beta}_p$ is a vector $(k\times 1)$ of unknown parameters of interest. Then, we consider the quantile regression model given by

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_p + \varepsilon_i, \;\; i = 1,\ldots,n, \tag{4}$$

where ε_i is the error term whose distribution (with density, say, $f_p(.)$) is restricted to have the pth quantile equal to zero, that is, $\int_{-\infty}^{0} f_p(\varepsilon_i)d\varepsilon_i = p$. The error density $f_p(.)$ is often left unspecified in the classical literature. Thus, quantile regression estimation for $\boldsymbol{\beta}_p$ proceeds by minimizing

$$\widehat{\boldsymbol{\beta}}_p = arg\; min_{\beta\in R^k} \sum_{i=1}^{n} \rho_p\big(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p\big), \tag{5}$$

where $\rho_p(.)$ is as in (1) and $\widehat{\boldsymbol{\beta}}_p$ is the quantile regression estimate for $\boldsymbol{\beta}_p$ at the pth quantile. The special case $p = 0.5$ corresponds to median regression. As the check function is not differentiable at zero, we cannot derive explicit solutions to the minimization problem. Therefore, linear programming methods are commonly applied to obtain quantile regression estimates for $\boldsymbol{\beta}_p$. A connection between the minimization of the sum in (5) and the maximum-likelihood theory is provided by the ALD.

Suppose that $y_i \sim ALD(\mathbf{x}_i^\top \boldsymbol{\beta}_p, \sigma, p)$, $i = 1, \ldots, n$ are independent. Then from (1), the likelihood function for n observations is

$$L(\boldsymbol{\beta}, \sigma|\mathbf{y}) = \frac{p^n(1-p)^n}{\sigma^n} \exp\Big\{ -\sum_{i=1}^{n} \rho_p\big(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p}{\sigma}\big)\Big\}. \tag{6}$$

Note that if we consider σ as a nuisance parameter, then the maximization of the likelihood in (6) with respect to the parameter $\boldsymbol{\beta}_p$ is equivalent to the minimization of the objective function in (5). and hence the relationship between the check function and ALD can be used to reformulate the QR method in the likelihood framework. It is also true that under model in (6), we have

$$d_i = \frac{1}{\sigma}\rho_p\big(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p\big) \sim \mathrm{Exp}(1). \tag{7}$$

The above result is useful to check the model in practice, as will be seen in the Application Section.

We impose the assumption $y \sim ALD(\mu, \sigma, p)$, which implies that the different quantiles of y conditional on x have the same slope. However, we only compute the p-quantile of y if $y \sim ALD(\mu, \sigma, p)$ and for different p, we actually use a different model. Thus as long as $Q_{y_i}(p|\mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}_p$, the likelihood is consistent in the sense that the maximum likelihood estimator (MLE) will converge to the true $\boldsymbol{\beta}_p$ in (5). Thus, when using ALD, we still can get consistent estimation of the quantile function and the slope coefficients might be different for different p as will be seen in the simulation study.

3 Parameter estimation

In this section, we discuss an estimation method for QR based on the EM algorithm to obtain ML estimates. Also, we consider the method of moments (MM) estimators,which can be effectively used as starting values in the EM algorithm.

3.1 The EM algorithm

Here, we show how to employ the EM type algorithms for ML estimation in QR model under the ALD. From the hierarchical representation (2)-(3), the QR model defined in (4) can be expressed as

$$\begin{aligned} Y_i|U_i = u_i &\sim N(\mathbf{x}_i^\top \boldsymbol{\beta}_p + \vartheta_p u_i, \tau_p^2 \sigma u_i), & (8)\\ U_i &\sim Exp(\sigma),\ \ i = 1, \ldots, n, & (9) \end{aligned}$$

where ϑ_p and τ_p^2 are as in Lemma 1. This relation is a convenient hierarchical representation of the QR model, and will be useful in the path E of the algorithm.

Let $\mathbf{y}=(y_1,\ldots,y_n)$ and $\mathbf{u}=(u_1,\ldots,u_n)$ be the observed data and the missing data, respectively. Then, the complete data log-likelihood function of $\boldsymbol{\theta}=(\boldsymbol{\beta}_p^\top,\sigma)^\top$ given $(\mathbf{y},\mathbf{u})$, ignoring additive constant terms, is given by $\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})=\sum_{i=1}^n \ell_c(\boldsymbol{\theta}|y_i,u_i)$, where

$$\ell_c(\boldsymbol{\theta}|y_i,u_i)=-\frac{1}{2}\log(2\pi\tau_p^2)-\frac{3}{2}\log(\sigma)-\frac{1}{2}\log(u_i)-\frac{1}{2\sigma\tau_p^2}u_i^{-1}(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_p-\vartheta_p u_i)^2-\frac{1}{\sigma}u_i,$$

for $i=1,\ldots,n$. In what follows the superscript (k) indicates the estimate of the related parameter at the stage k of the algorithm. The E-step of the EM algorithm requires evaluation of the so-called Q-function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})=\mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y},\mathbf{u})|\mathbf{y},\boldsymbol{\theta}^{(k)}]$, where $\mathrm{E}_{\boldsymbol{\theta}^{(k)}}[.]$ means that the expectation is being effected using $\boldsymbol{\theta}^{(k)}$ for $\boldsymbol{\theta}$. Observe that the expression of the Q-function is completely determined by the knowledge of the expectations

$$\mathscr{E}_s(\boldsymbol{\theta}^{(k)})=\mathrm{E}[U_i^s|y_i,\boldsymbol{\theta}^{(k)}],\ s=-1,1. \tag{10}$$

Let us consider $\boldsymbol{\xi}_s^{(k)}=\left(\mathscr{E}_{s1}(\boldsymbol{\theta}^{(k)}),\ldots,\mathscr{E}_{sn}(\boldsymbol{\theta}^{(k)})\right)^\top$ the vector that contains all quantities defined in (10). Thus, dropping unimportant constants, the Q-function can be written in a synthetic form as $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})=\sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$, where

$$Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})=-\frac{3}{2}\log\sigma-\frac{1}{2\sigma\tau_p^2}\left[\mathscr{E}_{-1i}(\boldsymbol{\theta}^{(k)})(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_p)^2-2(y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_p)\vartheta_p+\frac{1}{4}\mathscr{E}_{1i}(\boldsymbol{\theta}^{(k)})\tau_p^4\right]. \tag{11}$$

This is, undoubtedly, a computationally attractive and quite useful expression to implement the M-step, which consists of maximizing it over $\boldsymbol{\theta}$. So the EM algorithm that we propose can be summarized in the following steps:

E-step: Given $\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}$, compute $\mathscr{E}_{si}(\boldsymbol{\theta}^{(k)})$, for $s=-1,1$, given by

$$E[U_i^s|y_i,\boldsymbol{\theta}^{(k)}]=\left(\frac{\delta_i^{(k)}}{\gamma^{(k)}}\right)^s\frac{K_{1/2+s}(\lambda_i^{(k)})}{K_{1/2}(\lambda_i^{(k)})}, \tag{12}$$

where $\delta_i^{(k)}=\frac{|y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_p^{(k)}|}{\tau_p\sqrt{\sigma^{(k)}}}$, $\gamma^{(k)}=\frac{\tau_p}{2\sqrt{\sigma^{(k)}}}$ and $\lambda_i^{(k)}=\delta_i^{(k)}\gamma^{(k)}$;

M-step: Update $\boldsymbol{\theta}^{(k)}$ by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ over $\boldsymbol{\theta}$, which leads to the following expressions

$$\begin{aligned}\boldsymbol{\beta}_p^{(k+1)} &= \left(\mathbf{X}^\top D(\boldsymbol{\xi}_{-1}^{(k)})\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(D(\boldsymbol{\xi}_{-1}^{(k)})\mathbf{Y}-\vartheta_p\mathbf{1}_n\right),\\ \sigma^{(k+1)} &= \frac{1}{3n\tau_p^2}\left[Q(\boldsymbol{\beta}^{(k+1)},\boldsymbol{\xi}_{-1}^{(k)})-2\mathbf{1}_n^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}^{(k+1)})\vartheta_p+\frac{\tau_p^4}{4}\mathbf{1}_n^\top\boldsymbol{\xi}_1^{(k)}\right],\end{aligned}$$

where $D(\mathbf{a})$ denotes the diagonal matrix, with the diagonal elements given by $\mathbf{a}=(a_1,\ldots,a_p)^\top$ and $Q(\boldsymbol{\beta},\boldsymbol{\xi}_1)=(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^\top D(\boldsymbol{\xi}_{-1})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$. This process is iterated until some distance involving two successive evaluations of the actual log-likelihood $\ell(\boldsymbol{\theta})$, like $||\ell(\boldsymbol{\theta}^{(k+1)})-\ell(\boldsymbol{\theta}^{(k)})||$ or $||\ell(\boldsymbol{\theta}^{(k+1)})/\ell(\boldsymbol{\theta}^{(k)})-1||$, is small enough. This algorithm is implemented as part of the R package `ALDqr()`, which can be downloaded at not cost from the repository CRAN. Furthermore, following the results given in Yu & Zhang (2005), the MM estimators for $\boldsymbol{\beta}_p$ and σ are solutions of the following equations:

$$\widehat{\boldsymbol{\beta}}_{pM}=\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\mathbf{Y}-\widehat{\sigma}_M\vartheta_p\mathbf{1}_n\right)\ \text{and}\ \widehat{\sigma}_M=\frac{1}{n}\sum_{i=1}^n\rho_p\left(y_i-\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}_{pM}\right), \tag{13}$$

where ϑ_p as defined in Lemma 1. Note that the MM estimators do not have explicit closed form and numerical procedures are needed to solve these non-linear equations. However they can be used as initial values in the iterative procedure for computing the ML estimates based on the EM-algorithm.

In the estimation procedure described in the EM algorithm, one can see that $\mathscr{E}_{-1i}(\boldsymbol{\theta}^{(k)})$ is inversely proportional to $d_i = |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p^{(k)}|/\sigma$. Hence, $u_i(\boldsymbol{\theta}^{(k)}) = \mathscr{E}_{-1i}(\boldsymbol{\theta}^{(k)})$ can be interpreted as a type of weight for the ith case in the estimates of $\boldsymbol{\beta}_p^{(k)}$, which tends to be small for outlying observations. In fact, these weights can be used as tools for identifying outlying observations as will be seen in Section 5.

3.2 The observed information matrix

From (8)-(9) and after some algebraic manipulations we find that the log–likelihood function can be written as:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{14}$$

where $\ell_i(\boldsymbol{\theta}) = c - \frac{3}{2}\log\sigma + \frac{\vartheta_p}{\tau_p^2\sigma}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p) + \log(A_i)$, with c is a constant that is independent of $\boldsymbol{\theta}$ and $A_i = 2\left(\frac{\delta_i}{\gamma}\right)^{1/2} K_{1/2}(\lambda_i) = \frac{\sqrt{2\pi}}{\gamma}\exp(-\lambda_i)$, where δ_i, γ and λ_i are as in (12). Thus, the matrix of second derivatives with respect to $\boldsymbol{\theta}$ is given by $\mathbf{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^\top}\right)$, $\boldsymbol{\gamma},\boldsymbol{\tau} = \boldsymbol{\beta}_p, \sigma$. The derivatives of $\ell_i(\boldsymbol{\theta})$ are presented in Appendix A. Asymptotic confidence intervals and tests of the ML estimator at the pth level can be obtained assuming that the ML estimates $\widehat{\boldsymbol{\theta}}$ has approximately a $N_{k+1}(\boldsymbol{\theta}, \mathbf{L}(\boldsymbol{\theta})^{-1})$ distribution. In practice, $\mathbf{L}(\boldsymbol{\theta})$ is usually unknown and needs to be replaced by its ML estimates $\mathbf{L}(\widehat{\boldsymbol{\theta}})$.

4 Case-deletion measures

Case-deletion is a classical approach to study the effects of dropping the ith case from the data set. In what follows, $\mathbf{y}_c = (\mathbf{y}, \mathbf{u})$ denotes the augmented data set and a quantity with a subscript "$[i]$" denotes the original one with the ith case deleted. Thus, The complete-data log-likelihood function based on the data with the ith case deleted will be denoted by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_{c[i]})$. Let $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{p[i]}^\top, \widehat{\sigma^2}_{[i]})^\top$ be the maximizer of the function $Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left[\ell_c(\boldsymbol{\theta}|\mathbf{Y}_{c[i]})|\mathbf{y}\right]$, where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma^2})^\top$ is the EM estimate of $\boldsymbol{\theta}$. To assess the influence of the ith case on $\widehat{\boldsymbol{\theta}}$, we compare the difference between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$. If the deletion of a case seriously influences the estimates, more attention needs to be paid to that case. Hence, if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential. As $\widehat{\boldsymbol{\theta}}_{[i]}$ is needed for every case, the required computational effort can be quite heavy, especially when the sample size is large. Hence, the following one-step pseudo approximation $\widehat{\boldsymbol{\theta}}_{[i]}^1$ is used to reduce the burden (see Zhu *et al.*, 2001):

$$\widehat{\boldsymbol{\theta}}_{[i]}^1 = \widehat{\boldsymbol{\theta}} + \left\{-\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\right\}^{-1}\dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \tag{15}$$

where

$$\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \text{ and } \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \tag{16}$$

are the Hessian matrix and the gradient vector evaluated at $\widehat{\boldsymbol{\theta}}$, respectively. In particular, the Hessian matrix is an essential element in the method developed by Zhu & Lee (2001) and Zhu *et al.* (2001) to obtain the measures for case-deletion diagnosis and for local influence of a specified perturbation scheme. These formulas can be obtained quite easily from (11). The latter has the following coordinates:

$$\dot{Q}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma}} E_{1[i]} \text{ and } \dot{Q}_{[i]\sigma}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) = \frac{\partial Q_{[i]}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \sigma}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = -\frac{1}{2\widehat{\sigma^2}} E_{2[i]},$$

where

$$E_{1[i]} = \frac{1}{\tau_p^2} \sum_{j \neq i} \left[\mathscr{E}_{-1j}(\widehat{\boldsymbol{\theta}}^{(k)})(y_j - \mathbf{x}_j^\top \widehat{\boldsymbol{\beta}})\mathbf{x}_j - \mathbf{x}_j \vartheta_p \right] \quad \text{and} \tag{17}$$

$$E_{2[i]} = \sum_{j \neq i} \left[3\widehat{\sigma} - \frac{1}{\tau_p^2} \mathscr{E}_{-1j}(\widehat{\boldsymbol{\theta}}^{(k)})(y_j - \mathbf{x}_j^\top \widehat{\boldsymbol{\beta}_p})^2 - 2(y_j - \mathbf{x}_j^\top \widehat{\boldsymbol{\beta}_p})\vartheta_p + \frac{1}{4} \mathscr{E}_{1j}(\widehat{\boldsymbol{\theta}}^{(k)})\tau_p^4 \right]. \tag{18}$$

The elements of the second order partial derivatives of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ evaluated at $\widehat{\boldsymbol{\theta}}$ are

$$\begin{aligned} \ddot{Q}_\beta(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) &= -\frac{1}{\widehat{\sigma}\tau_p^2} \mathbf{X}^\top D(\boldsymbol{\xi}_{-1}^{(k)})\mathbf{X}, \\ \ddot{Q}_\sigma(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\} &= \frac{3}{4\widehat{\sigma^2}} - \frac{1}{2\widehat{\sigma^3}\tau_p^2} \left[Q(\boldsymbol{\beta}, \boldsymbol{\xi}_{-1}^{(k)}) - 2\mathbf{1}_n^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\vartheta_p + \frac{\tau_p^4}{4} \mathbf{1}_n^\top \boldsymbol{\xi}_1^{(k)} \right] \end{aligned}$$

and $\ddot{Q}_{\beta\sigma}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})\} = \mathbf{0}$. In the following result, we will obtain the one-step approximation of $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{p[i]}^\top, \widehat{\sigma}_{[i]})^\top$, $i = 1, \ldots, n$ based on (15), viz., the relationships between the parameter estimates for the full data set and the data with the ith case deleted.

Theorem 4.1. *For the QR model defined in (8)-(9) , the relationships between the parameter estimates for full data set and the data with the ith case deleted are as follows:*

$$\widehat{\boldsymbol{\beta}}_{p[i]}^1 = \widehat{\boldsymbol{\beta}}_p + \tau_p^2 (\mathbf{X}^\top D(\widehat{\boldsymbol{\xi}}_{-1})\mathbf{X})^{-1} \boldsymbol{E}_{1[i]} \text{ and } \widehat{\sigma^2}_{[i]}^1 = \widehat{\sigma^2} - \frac{1}{2\widehat{\sigma^2}} \Big(\ddot{Q}_\sigma(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) \Big)^{-1} E_{2[i]},$$

where $\boldsymbol{E}_{1[i]}$ and $E_{2[i]}$ are as in (17) and (18), respectively.

To assess the influence of the ith case on the EM estimate $\widehat{\boldsymbol{\theta}}$, we compare $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$, and if $\widehat{\boldsymbol{\theta}}_{[i]}$ is far from $\widehat{\boldsymbol{\theta}}$ in some sense, then the ith case is regarded as influential. Based on the metric for measuring the distance between $\widehat{\boldsymbol{\theta}}_{[i]}$ and $\widehat{\boldsymbol{\theta}}$ proposed by Zhu *et al.* (2001), we consider here the following *generalized Cook distance*:

$$GD_i = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top \{ -\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) \} (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}), \quad i = 1, \ldots, n. \tag{19}$$

Upon substituting (15) into (19), we obtain the approximation

$$GD_i^1 = \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}})^\top \big\{ -\ddot{Q}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) \big\}^{-1} \dot{Q}_{[i]}(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}), \quad i = 1, \dots, n.$$

Another measure of the influence of the ith case is the following Q-distance function, similar to the likelihood distance LD_i (Cook & Weisberg, 1982), defined as

$$QD_i = 2\big\{ Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}_{[i]}|\widehat{\boldsymbol{\theta}}) \big\}. \tag{20}$$

We can calculate an approximation of the likelihood displacement QD_i by substituting (15) into (20), resulting in the following approximation QD_i^1 of QD_i:

$$QD_i^1 = 2\big\{ Q(\widehat{\boldsymbol{\theta}}|\widehat{\boldsymbol{\theta}}) - Q(\widehat{\boldsymbol{\theta}}^1_{[i]}|\widehat{\boldsymbol{\theta}}) \big\}.$$

Table 1: AIS data. Results of the parameter estimation via EM, Barrodale and Roberts (BR) and Lasso Penalized Quantile Regression (LPQR) algorithms for three selected quantiles.

		EM		BR		LPQR
p	Parameter	MLE	SE	Estimative	SE	Estimative
0.1	β_0	9.3913	0.6911	9.3915	1.2631	9.8573
	β_1	0.1705	0.0089	0.1705	0.0160	0.1647
	β_2	0.8312	0.2588	0.8209	0.4432	0.6684
	σ	0.2617	0.0064	1.0991	——	1.0959
0.5	β_0	7.6480	0.8713	7.6480	1.1120	7.6480
	β_1	0.2160	0.0115	0.2160	0.0159	0.2160
	β_2	2.2499	0.3007	2.2226	0.4032	2.2226
	σ	0.6894	0.1948	0.6894	——	0.6894
0.9	β_0	5.8000	0.5620	5.8000	1.6461	6.0292
	β_1	0.2700	0.0077	0.2700	0.0256	0.2678
	β_2	3.9596	0.1939	3.9658	0.6203	3.8271
	σ	0.3391	0.0110	1.2677	——	1.2767

5 Application

We illustrate the proposed methods by applying them to the Australian Institute of Sport (AIS) data, analyzed by Cook and Weisberg (1994) in a normal regression setting. The data set consists of several variables measured in $n = 202$ athletes (102 males and 100 females). Here, we focus on body mass index (BMI), which is assumed to be explained by lean body mass (LBM) and gender (SEX). Thus, we consider the following QR model:

$$BMI_i = \beta_0 + \beta_1 LBM_i + \beta_2 SEX_i + \varepsilon_i, \quad i = 1, \dots, 202,$$

where ε_i is a zero p quantile. This model can be fitted in the R software by using the R package `quantreg()`, where one can arbitrarily use the BR or the LPQR algorithms. In order to compare with our proposed EM algorithm, we carry out quantile regression at three different quantiles,

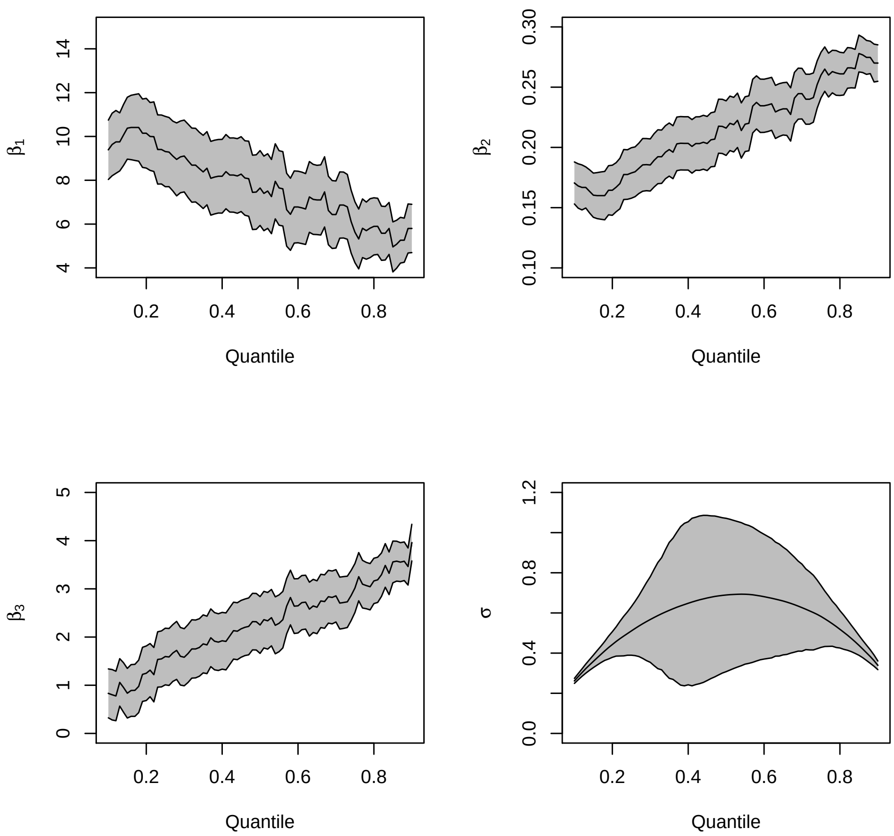

Figure 2: AIS data: ML estimates and 95% confidence intervals for various values of p.

namely $p = \{0.1, 0.5, 0.9\}$ by using the ALD distribution as described in Section 2. The ML estimates and associated standard errors were obtained by using the EM algorithm and the observed information matrix described in subsections 3.1 and 3.3, respectively. Table 1 compares the results of our EM, BR and the LPQR estimates under the three selected quantiles. The standard error of the LPQR estimates are not provided in the R package `quantreg()` and are not shown in Table 1. From this table we can see that estimates under the three methods only exhibit slight differences, as expected. However, the standard errors of our EM estimates are smaller than those via the BR algorithm. This suggests that the EM algorithm seems to produce more accurate estimates of the regression parameters at the pth level.

To obtain a more complete picture of the effects, a series of QR models over the grid $p = \{0.1, 0.15, \ldots, 0.95\}$ is estimated. Figure 2 gives a graphical summary of this analysis. The shaded area depicts the 95% confidence interval from all the parameters. From Figure 2 we can observe some interesting evidences which cannot be detected by mean regression. For example, the effect of the two variables (LBM and gender) become stronger for the higher conditional quantiles, indicating that the BMI are positively correlated with the quantiles. The robustness of the median regression $(p = 0.5)$ can be assessed by considering the influence of a single outlying observation on the EM estimate of $\boldsymbol{\theta}$. In particular, we can assess how much the EM estimate of $\boldsymbol{\theta}$ is influ-

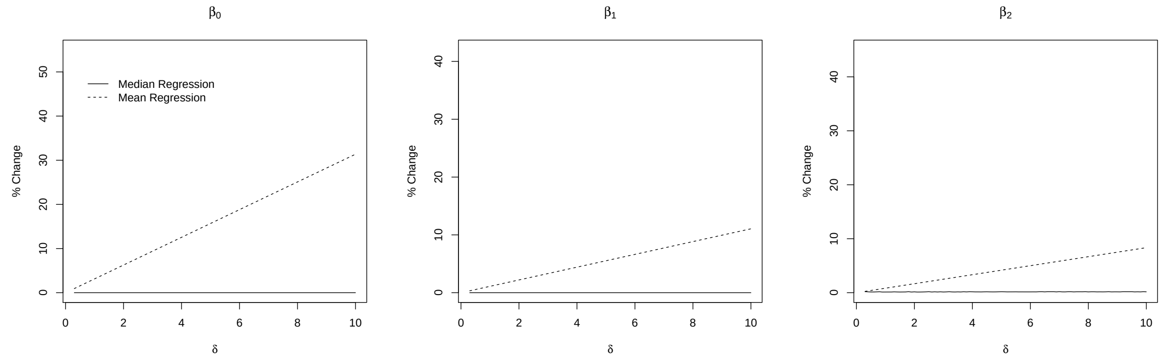

Figure 3: Percentage of change in the estimation of β_0, β_1 and β_2 in comparison with the true value, for median $(p = 0.5)$ and mean regression, for different contaminations δ.

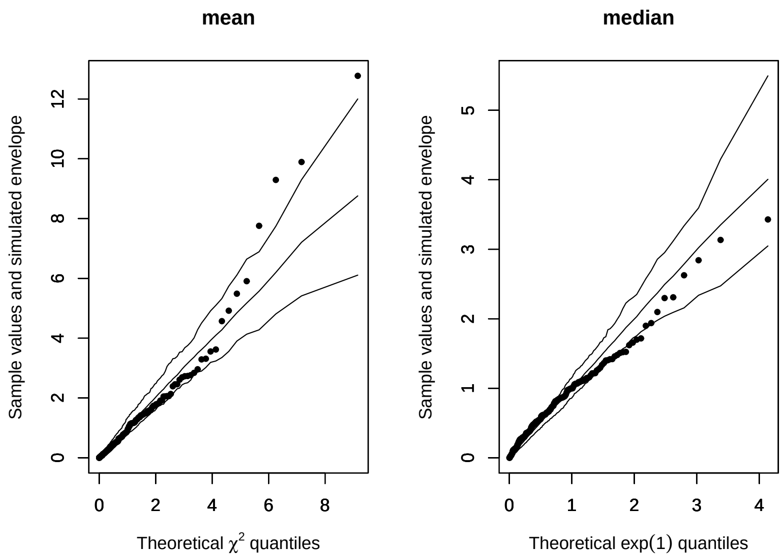

Figure 4: AIS data: Q–Q plots and simulated envelopes for mean and median regression.

enced by a change of δ units in a single observation y_i. Replacing y_i by $y_i(\delta) = y_i + \delta sd(\mathbf{y})$, where $sd(.)$ denotes the standard deviation. Let $\widehat{\beta}_j(\delta)$ be the EM estimates of β_j after contamination, $j = 1,2,3$. We are particularly interested in the relative changes $|(\widehat{\beta}_j(\delta) - \widehat{\beta}_j)/\widehat{\beta}_j|$. In this study we contaminated the observation corresponding to individual $\{\#146\}$ and for δ between 0 and 10. Figure 3 displays the results of the relative changes of the estimates for different values of δ. As expected, the estimates from the median regression model are less affected by variations on δ than those of the mean regression. Moreover, The Q-Q plots and envelopes shown in Figure 4 are based on the distribution of $d_i = |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p|/\sigma$, given in (7), which is $Exp(1)$. The lines in these figures represent the 5th percentile, the mean and the 95th percentile of 100 simulated points for each observation. These figures clearly show that the median regression distribution provides a better-fit than the standard mean regression to the AIS data set.

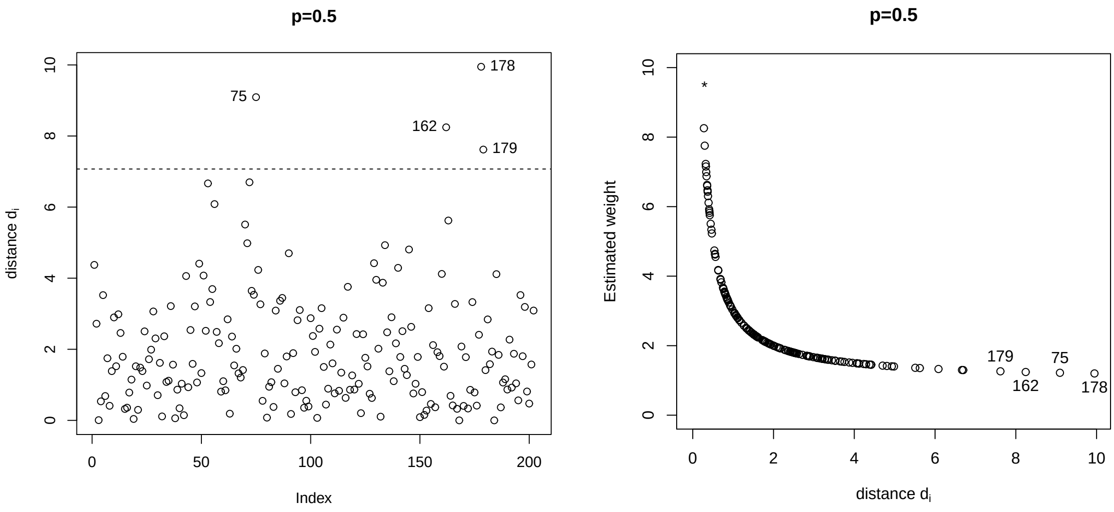

Figure 5: AIS data: Index plot of the distance d_i and the estimated weights u_i.

As discussed in Subsection 3.1 the estimated distance $\widehat{d_i}$ can be used as a goodness-of-fit measure and also to identify possible outlying observations. Figure 5(left panel) displays the index plot of the distance d_i for the median regression model $(p = 0.5)$. We see from this figure that observations #75, #162, #178 and #179 appear as possible outliers. From the EM-algorithm, the estimated weights $u_i(\widehat{\boldsymbol{\theta}})$ for these observations are the smallest ones (see right panel in Figure 5), confirming the robustness aspects of the maximum likelihood estimates against outlying observations of the QR models. Thus, larger d_i implies a smaller $u_i(\widehat{\boldsymbol{\theta}})$, and the estimation of $\boldsymbol{\theta}$ tends to give smaller weight to outlying observations in the sense of the distance d_i.

Figure 6 shows the estimated quartiles of two levels of gender at each LBM point from our EM algorithm along with the estimates obtained via mean regression. From this figure we can see clear attenuation in β_1 due to the use of the median regression related to the mean regression. It is possible to observe in this figure some atypical individuals that could have an influence on the ML estimates for different values of quantiles. In this figure, the individuals $\#75,\ \#130,\ \#140\ \#162,\ \#160$ and #178 were marked since they were detected as potentially influential.

In order to identify influential observations at different quantiles when some observation is eliminated, we can generate graphs of the generalized Cook distance GD_i^l, as explained in Section 4. A high value for GD_i^l indicates that the ith observation has a high impact on the maximum likelihood estimate of the parameters. Following Barros *et al.* (2010), we can use $2(p+1)/n$ as benchmark for the GD_i^l at different quantiles. Figure 7 (first row) presents the index plots of GD_i^l. We note from this figure that, only observation #140 appears as influential in the ML estimates at $p = 0.1$ and observations $\#75, \#178$ as influential at $p = 0.5$, whereas observations $\#75, \#162, \#178$ and #179 appear as influential in the ML estimates at $p = 0.9$. Figure 7 (second row) presents the index plots of QD_i^1. From this figure, it can be noted that observations $\#76, \#130, \#140$ appear to be influential at $p = 0.1$, whereas observations $\#75, \#162$ and #178 seem to be influential in the ML estimates at $p = 0.1$, and in addition observation #179 appears to be influential at $p = 0.9$.

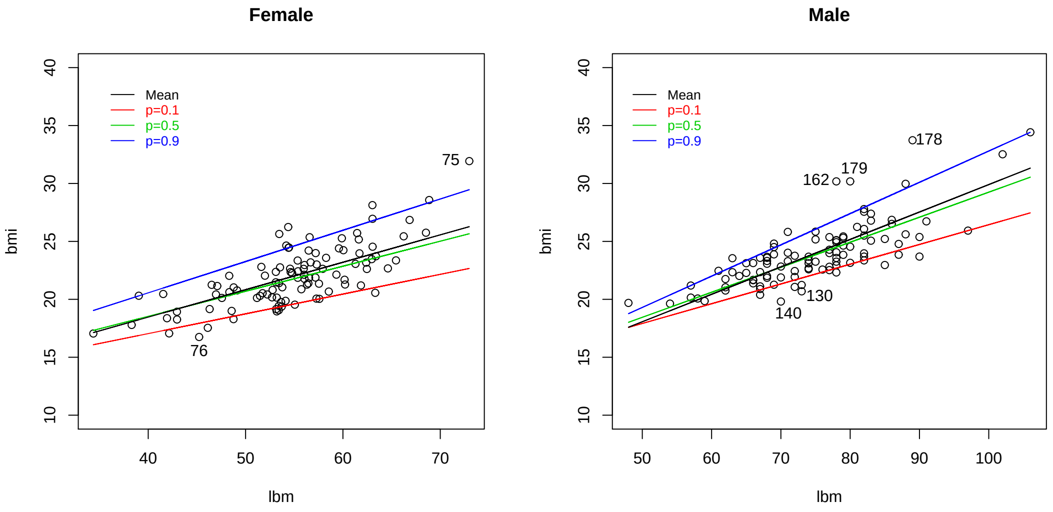

Figure 6: AIS data: Fitted regression lines for the three selected quantiles along with the mean regression line. The influential observations are numbered.

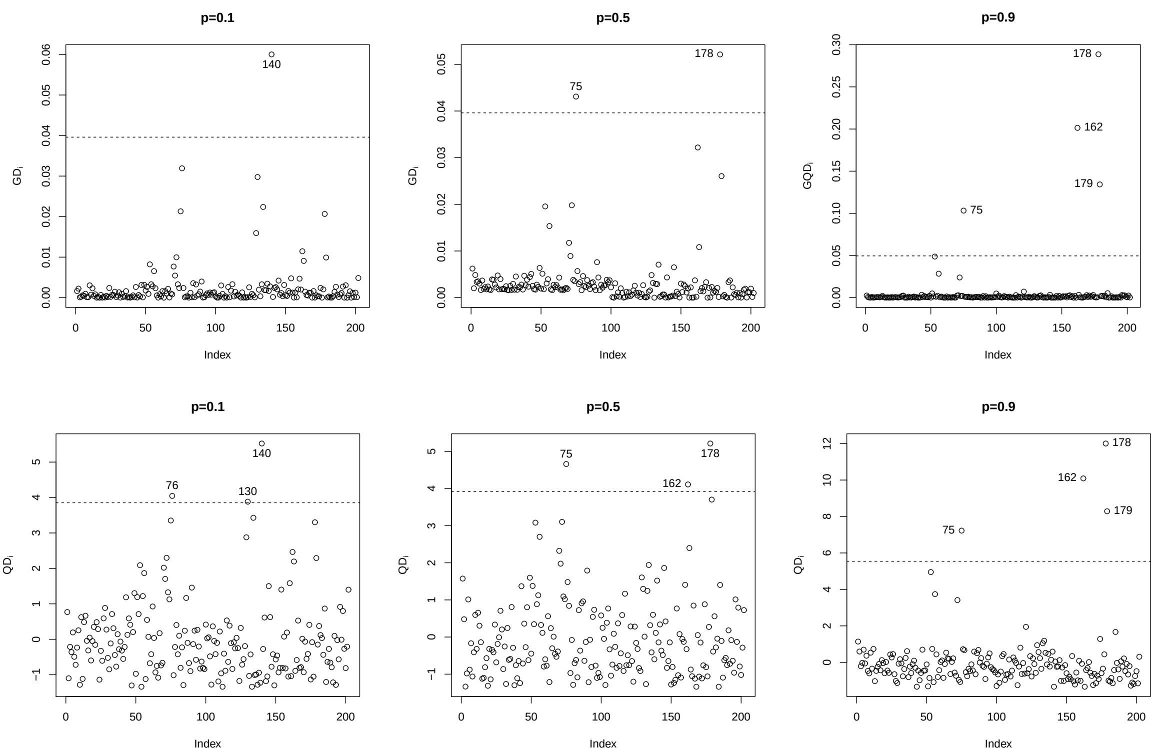

Figure 7: Index plot of (first row) approximate likelihood distance GD_i^1. (second row). Index plot of approximate likelihood displacement QD_i^1. The influential observations are numbered.

6 Simulation studies

In this section, the results from two simulation studies are presented to illustrate the performance of the proposed method.

6.1 Robustness of the EM estimates (Simulation Study 1)

We conducted a simulation study to assess the performance of the proposed EM algorithm, by mimicking the setting of the AIS data by taking the sample size $n = 202$. We simulated data from the model

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \quad i = 1, \ldots, 202, \tag{21}$$

where the $x'_{ij}s$ are simulated from a uniform distribution (U(0,1)) and the errors ε_{ij} are simulated from four different distributions: (i) the standard normal distribution $N(0,1)$, (ii) a Student-t distribution with three degrees of freedom, $t_3(0,1)$, (iii) a heteroscedastic normal distribution, $(1+x_{i2})N(0,1)$ and, (iv) a bimodal mixture distribution $0.6t_3(-20,1)+0.4t_3(15,1)$. The true values of the regression parameters were taken as $\beta_1 = \beta_2 = \beta_3 = 1$. In this way, we had four settings and for each setting we generated 10000 data sets.

Once the simulated data were generated, we fit a QR model, with $p = 0.1$, 0.5 and 0.9, under Barrodale and Roberts (BR), Lasso (Lasso) and EM algorithms by using the "quantreg()" package and our `ALDqr()` package, from the R language, respectively. For the four scenarios, we computed the bias and the square root of the mean square error (RMSE), for each parameter over the $M = 10,000$ replicas. They are defined as:

$$Bias(\gamma) \;=\; \bar{\hat{\gamma}} - \gamma \text{ and } RMSE(\gamma) = \sqrt{SE(\gamma)^2 + Bias(\gamma)^2} \tag{22}$$

where $\bar{\hat{\gamma}} = \frac{1}{M}\sum_{i=1}^{M}\hat{\gamma}_i$ and $SE(\gamma)^2 = \frac{1}{M-1}\sum_{i=1}^{M}\left(\hat{\gamma}_i - \bar{\hat{\gamma}}\right)^2$, with $\gamma = \beta_1, \beta_2, \beta_3$ or σ, $\hat{\gamma}_i$ is the estimate of γ obtained in replica i and γ is the true value. Table 2 reports the simulation results for $p = 0.1$, 0.5 and 0.9. We observe that the EM yields lower biases and RMSE than the other two estimation methods under all the distributional scenarios. This finding suggests that the EM would produce better results than other alternative methods typically used in the literature of QR models.

6.2 Asymptotic properties (Simulation study 2)

We also conducted a simulation study to evaluate the finite-sample performance of the parameter estimates. We generated artificial samples from the regression model (21) with $\beta_1 = \beta_2 = \beta_3 = 1$ and $x_{ij} \sim U(0,1)$. We chose several distributions for the random term ε_i a little different than the simulation study 1, say, (i) normal distribution $N(0,2)$ (N1), (ii) a Student-t distribution $t_3(0,2)$ (T1), (iii) a heteroscedastic normal distribution, $(1+x_{i2})N(0,2)$ (N2) and, (iv) a bimodal mixture distribution $0.6t_3(-20,2)+0.4t_3(15,2)$ (T2). Finally, the sample sizes were fixed at $n = 50, 100, 150, 200, 300, 400, 500, 700$ and 800.

For each combination of parameters and sample sizes, 10000 samples were generated under the four different situations of error distributions (N1, T1, N2, T2). Therefore, 36 different simulation runs are performed. Once all the data were simulated, we fit the QR model with $p = 0.5$ and the bias (22) and the square root of the mean square error (22) were recorded. The results are shown in Figure 8. We can see a pattern of convergence to zero of the bias and MSE when n increases. As a general rule, we can say that bias and MSE tend to approach to zero when the sample size increases, indicating that the estimates based on the proposed EM-type algorithm do provide good asymptotic properties. This same pattern of convergence to zero is repeated considering different levels of the quantile p.

Table 2: Simulation study. Bias and root mean-squared error (RMSE) of $\boldsymbol{\beta}$ under different error distributions. The estimates under Barrodale and Roberts (BR) and Lasso (Lasso) algorithms were obtained by the "quantreg()" package from the R language.

		β_1		β_2		β_3	
Method	p	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\varepsilon \sim N(0,1)$							
BR	0.1	-1.2639	1.3444	0.0076	0.5961	-0.0030	0.5934
	0.5	0.0064	0.3376	-0.0048	0.4390	-0.0051	0.4453
	0.9	1.2640	1.3460	0.0030	0.6051	0.0069	0.6039
LPQR	0.1	-0.9664	1.0464	-0.3072	0.6165	-0.3110	0.6187
	0.5	0.1474	0.3628	-0.1463	0.4534	-0.1462	0.4576
	0.9	1.5901	1.6460	-0.3164	0.6173	-0.3076	0.6179
EM	0.1	-1.2551	1.3362	-0.0055	0.5964	-0.0090	0.6020
	0.5	0.0040	0.3286	-0.0050	0.4332	-0.0031	0.4363
	0.9	1.2694	1.3484	-0.0071	0.6019	-0.0120	0.5955
$\varepsilon \sim t_3(0,1)$							
BR	0.1	-1.2446	1.3364	-0.0290	0.6274	-0.0313	0.6259
	0.5	0.1049	0.4870	0.1213	0.6714	0.1123	0.6708
	0.9	2.3618	2.8408	1.0056	2.4928	0.9459	2.4332
LPQR	0.1	-0.9315	1.0219	-0.3478	0.6422	-0.3412	0.6354
	0.5	0.3007	0.5410	-0.0928	0.6310	-0.0831	0.6237
	0.9	3.0443	3.2880	0.1911	1.6375	0.2231	1.6601
EM	0.1	-1.2287	1.3213	-0.0402	0.6209	-0.0374	0.6265
	0.5	0.0965	0.4866	0.1352	0.6789	0.1304	0.6758
	0.9	2.3781	2.8459	0.9464	2.4082	0.9264	2.4167
$\varepsilon \sim (1+x_2)N(0,1)$							
BR	0.1	-1.2869	1.4256	0.0130	0.8706	-1.2554	1.5381
	0.5	-0.0051	0.4468	0.0049	0.6336	0.0061	0.6509
	0.9	1.2868	1.4259	0.0018	0.8686	1.2307	1.5256
LPQR	0.1	-1.1393	1.2272	-0.3694	0.7773	-1.1450	1.2756
	0.5	0.1834	0.4520	-0.1906	0.6193	-0.1963	0.6304
	0.9	1.6972	1.7933	-0.3621	0.7925	0.7494	1.1587
EM	0.1	-1.2772	1.4140	0.0051	0.8646	-1.2341	1.5195
	0.5	0.0954	0.4892	0.1289	0.6724	0.1316	0.6694
	0.9	1.2599	1.3987	0.0076	0.8723	1.2488	1.5315
$\varepsilon \sim 0.6t_3(-20,1)+0.4t_3(15,1)$							
BR	0.1	-1.2350	1.3268	-0.0395	0.6160	-0.0396	0.6192
	0.5	0.1029	0.4896	0.1214	0.6780	0.1212	0.6741
	0.9	2.3857	2.8737	0.9657	2.4574	0.9558	2.4585
LPQR	0.1	-0.9664	1.0464	-0.3072	0.6165	-0.3110	0.6187
	0.5	0.1474	0.3628	-0.1463	0.4534	-0.1462	0.4576
	0.9	1.5901	1.6460	-0.3164	0.6173	-0.3076	0.6179
EM	0.1	-0.9327	1.0201	-0.3491	0.6433	-0.3355	0.6372
	0.5	0.2880	0.5343	-0.0745	0.6216	-0.0717	0.6159
	0.9	3.0624	3.3102	0.1702	1.6627	0.2221	1.6575

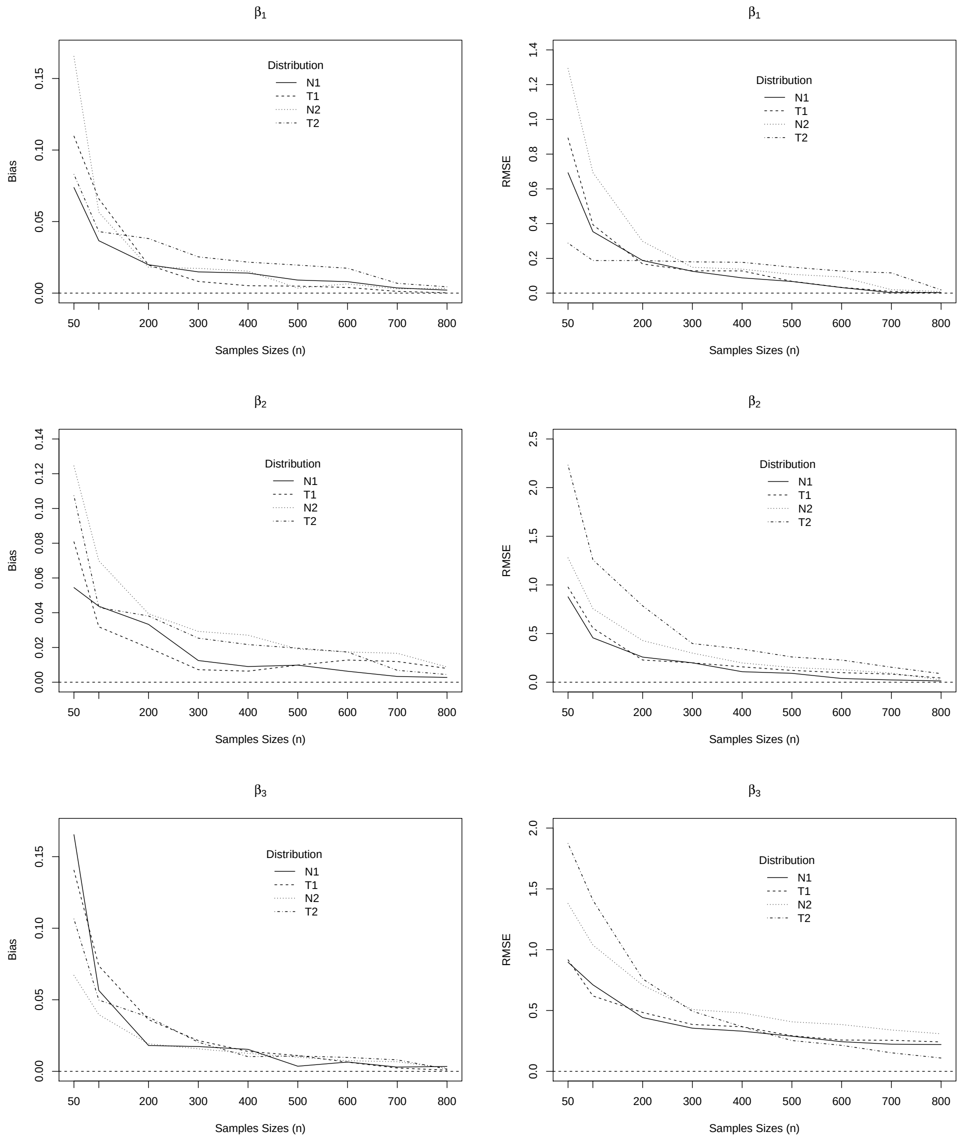

Figure 8: Simulation study 2. Average bias (first column) and average MSE (second column) of the estimates of β_1,β_2, β_3 with $p = 0.5$ (median regression), where $N1 = N(0,2)$, $T1 = t_3(0,2)$, $N2 = (1+x_2)N(0,2)$ and $T2 = 0.6t_3(-20,2)+0.4t_3(15,2)$.

7 Conclusion

We have studied a likelihood-based approach to the estimation of the QR based on the asymmetric Laplace distribution (ALD). By utilizing the relationship between the QR check function and the ALD, we cast the QR problem into the usual likelihood framework. The mixture representation of the ALD allows us to express a QR model as a normal regression model, facilitating the implementation of an EM algorithm, which naturally provides the ML estimates of the model parameters with the observed information matrix as a by product. The EM algorithm was implemented as part of the R package `ALDqr()`. We hope that by making the code of our method available, we will lower the barrier for other researchers to use the EM algorithm in their studies of quantile regression. Further, we presented diagnostic analysis in QR models, which was based on the case-deletion technique suggested by Zhu *et al.* (2001) and Zhu & Lee (2001), which are the counterparts for missing data models of the well-known ones proposed by Cook (1977) and Cook (1986). The simulation studies demonstrated the superiority of the proposed methods to the existing methods, implemented in the package `quantreg()`. We applied our methods to a real data set (freely downloadable from R) in order to illustrate how the procedures can be used to identify outliers and to obtain robust ML parameter estimates. From these results, it is encouraging that the use of ALD offers a better alternative in the analysis of QR models.

Finally, the proposed methods can be extended to more general framework, such as, censored (Tobit) regression models, measurement error models, nonlinear regression models, stochastic volatility models, etc and should yield satisfactory results at the expense of additional complexity in implementation. An in-depth investigation of such extensions is beyond the scope of the present paper, but these are interesting topics for further research.

Acknowledgements

The research of V. H. Lachos was supported by Grant 305054/2011-2 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil) and by Grant 2011/17400-6 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP-Brazil).

Appendix A: Elements of the observed information matrix

$$
\begin{aligned}
\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_p \partial \boldsymbol{\beta}_p^\top} &= -\frac{1}{A_i^2}\frac{\partial A_i}{\partial \boldsymbol{\beta}}\frac{\partial A_i}{\partial \boldsymbol{\beta}^\top} + \frac{1}{A_i}\frac{\partial^2 A_i}{\partial \boldsymbol{\beta}_p \partial \boldsymbol{\beta}_p^\top}, \quad \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_p \partial \sigma} = \frac{\vartheta_p}{\tau_p^2\sigma^2}\mathbf{x}_i - \frac{1}{A_i^2}\frac{\partial A_i}{\partial \boldsymbol{\beta}_p}\frac{\partial A_i}{\partial \sigma} + \frac{1}{A_i}\frac{\partial^2 A_i}{\partial \boldsymbol{\beta}\partial \sigma}, \\
\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \sigma \partial \sigma} &= \frac{3}{2\sigma^2} + \frac{2\vartheta_p}{\tau_p^2\sigma^3}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) - \frac{1}{A_i^2}\left(\frac{\partial A_i}{\partial \sigma}\right)^2 + \frac{1}{A_i}\frac{\partial^2 A_i}{\partial \sigma \partial \sigma},
\end{aligned}
$$

where

$$
\frac{\partial A_i}{\partial \boldsymbol{\beta}_p} = \frac{\sqrt{2\pi}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p)\mathbf{x}_i exp(-\lambda_i)}{\tau_p^2 \sigma \delta_i}, \quad \frac{\partial A_i}{\partial \sigma} = \frac{1}{\sigma}\frac{\sqrt{2\pi}}{2} exp(-\lambda_i)\left[\delta_i - \frac{\gamma}{\delta_i^3}(1+\lambda_i)\right],
$$

and

$$
\begin{aligned}
\frac{\partial^2 A_i}{\partial \boldsymbol{\beta}_p \partial \boldsymbol{\beta}_p^\top} &= \frac{\sqrt{2\pi} exp(-\lambda_i)}{\delta_i \tau_p^2 \sigma}\left[-I + \frac{(1+\lambda_i)(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p)^2}{\tau_p^2 \sigma \delta_i^2}\right]\mathbf{x}_i \mathbf{x}_i^\top \\
\frac{\partial^2 A_i}{\partial \boldsymbol{\beta}_p \partial \sigma} &= -\frac{\sqrt{2\pi}}{2}\frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_p)\mathbf{x}_i exp(-\lambda_i)}{\tau_p^2 \sigma^2 \delta_i}(\delta_i^2 + \lambda_i - 1), \\
\frac{\partial^2 A_i}{\partial \sigma \partial \sigma} &= -\frac{\sqrt{2\pi}}{\sigma^2} exp(-\lambda_i)\left[2\delta_i + \frac{1}{\gamma} - 2\lambda_i \gamma - \frac{1}{4}\frac{\delta_i^{1/2}}{\gamma^{1/2}}(\lambda_i^2 + 3\lambda_i + 1)\right].
\end{aligned}
$$

References

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, **63**, 167–241.

Barrodale, I. & Roberts, F. (1977). Algorithms for restricted least absolute value estimation. *Communications in Statistics-Simulation and Computation*, **6**, 353–363.

Barros, M., Galea, M., González, M. & Leiva, V. (2010). Influence diagnostics in the tobit censored response model. *Statistical Methods & Applications*, **19**, 716–723.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC.

Koenker, R. (2005). *Quantile regression*, volume 38. Cambridge University Press.

Koenker, R. & Gilbert, B. J. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33–50.

Koenker, R. W. & d'Orey, V. (1987). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **36**, 383–393.

Kottas, A. & Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **96**, 1458–1468.

Kottas, A. & Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, **36**, 297–319.

Kotz, S., Kozubowski, T. & Podgorski, K. (2001). *The laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*. Number 183. Birkhauser.

Kozumi, H. & Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, **81**, 1565–1578.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**, 450–464.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, pages 267–288.

Yu, K. & Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54**, 437–447.

Yu, K. & Zhang, J. (2005). A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics-Theory and Methods*, **34**, 1867–1879.

Zeller, C. B., Labra, F. V., Lachos, V. H. & Balakrishnan, N. (2010). Influence analyses of skew-normal/independent linear mixed models. *Computational Statistics & Data Analysis*, **54**, 1266–1280.

Zhu, H. & Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society, Series B*, **63**, 111–126.

Zhu, H., Lee, S.-Y., Wei, B.-C. & Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika*, **88**, 727–737.