Approximate Bayesian Computation (ABC) in Quantile Regression

Antonio Pulcini *

July 12, 2012

Abstract

In this article we propose an approximate Bayesian approach to estimate the joint distribution of the response variable Y and the set of covariates $\mathbf{X}$ based on the notion of quantile distribution. We focus on cases where the quantile regression framework is necessary, but the unknown form of the regression function and the large number of quantiles let be more convenient to estimate directly the conditional distribution. In this cases the use of very flexibly-shaped distributions is of interest. In this context we propose to estimate directly the joint distribution using the multivariate **g**-and-**h** distribution, a member of the quantile distribution family. Such family of distribution provide simple and flexible distribution when quantiles are of principal interest. Due to the lack of the likelihood function, the estimation proceed via an Approximate Bayesian Computation (ABC) algorithm that allows us to easily estimate all the parameters. The performance of the proposed approach is evaluated with simulated data sets.

Key words: Approximate Bayesian Computation; Joint Distribution; Kurtosis; Quantile Distribution; Quantile Regression; Skewness.

*Dipartimento SEFEMEQ. Università di Roma TOR VERGATA. Email: antoniopulcini@hotmail.com.

1 Introduction

The usual assumptions of the standard linear model imply that the conditional distribution of the response variable is, at least approximately, Gaussian. In practice, this assumption is rarely acceptable. In many observational studies, the conditional distribution is far from being symmetric and, even worse, its shape depends on the value of the covariates (Yu et al. (2003)). For these situations, methods based on quantile regression are nowadays valid and common alternatives (Koenker & Bassett (1978), Koenker (2005)). Nevertheless when influence concerns more than one quantile and the relationship between quantiles and covariates is not easy to model, it may be convenient to take a more general perspective and to consider the problem of directly estimating the conditional distribution of the response variable given the explanatory variables (Peracchi (2002), Foresi & Peracchi (1995)). Standard methods for this problem rely upon relatively complex models, like mixture models and non-parametric approaches, which may be difficult to implement and interpret. In this context quantile distributions, due to their flexibility and the small number of parameters, may represent a good alternative to the existing methods. Field & Genton (2006) have proposed a generalization of the univariate g-and-h distribution to the multivariate case, so providing a flexible family of multivariate distributions that are characterized by only four parameters related to location, scale, skewness and kurtosis.

In this paper we exploit the use of the multivariate **g**-and-**h** distribution for the estimation of the joint distribution of the response variable Y given a vector of covariates $\mathbf{X} = (X_1, \ldots, X_K)$.

A drawback of quantile distributions, which has represented an obstacle to their use, is the lack of a closed form expression of the likelihood function. On the other hand, the way in which quantile distributions are defined, makes the problem of generating random values from them, an easy task. The last two issues, namely the lack of explicit likelihood function and the case of random generation, strongly suggest the use of Approximate Bayesian Computation (ABC) approach.

ABC is a Monte Carlo base algorithm for parameter estimation which has been developed in population genetics studies (Allingham et al. (2009), Tavar et al. (1997)). In a few words, ABC allows to produce a sample from an approximate version of the posterior

distribution. No likelihood evaluation is required, only a way to sample from the model distribution.

Section 2 gives an overview of the quantile regression method. Section 3 describes the notion of quantile distribution. Section 4 illustrates the notions of multivariate quantile and recall the definition of multivariate **g**-and-**h** distribution. Section 5 illustrates the ABC approach. In Section 6 we detail a specific ABC algorithm for the estimation of the parameters of the multivariate **g**-and-**h** distribution. Section 7 presents the results for two simulations studies.

2 The Quantile Regression and the Conditional Distribution Function

2.1 The Distribution function and The Quantile Function

The probability distribution function P of a real random variable Z can be described either by its distribution function or by its quantile function. If Z has a continuous distribution with strictly positive density, the distribution function F is given by

$$F(z) = P\{Z \leq z\}, \quad \forall z \in \mathbb{R}$$

whereas the quantile function Q is a continuous and strictly increasing function defined on the interval $[0, 1]$ and values in $\mathbb{R}$ given by

$$Q(\tau) = \{z \in \mathbb{R} : F(z) = \tau\}.$$

Clearly Q and F are the inverse of each other, that is,

$$Q(F(z)) = z \quad \text{and} \quad F(Q(\tau)) = \tau.$$

It can be show that $Q(\tau)$ can be characterized as the solution of the following minimization problem

$$\min_{z \in \mathbb{R}} \mathbb{E}[\rho_\tau(Z - z)] \tag{1}$$

where ρ_τ is the asymmetric absolute loss function

$$\rho_\tau(\nu) = \nu(\tau - I(\nu \leq 0)) \quad \nu \in \mathbb{R}$$

with $I(\cdot)$ is the indicator function. For a motivation and illustration of this particular loss in quantile estimation see Peracchi (2002).

When $\tau = \frac{1}{2}$, $Q(\tau)$ is the median of the random variable Z. The generalization to conditional distributions is straightforward. Given the vector $(\mathbf{X}, Y) \in \mathbb{R}^{K+1}$, where Y is a real random variable with strictly positive density, the conditional distribution $P_{\mathbf{x}}$ of Y given $\mathbf{X} = \mathbf{x}$ can be described either by the conditional distribution function (c.d.f.) or by the conditional quantile function (c.q.f.). The c.d.f. is defined on $\mathbb{R} \times \mathbb{R}^K$ and is given by

$$F(y|\mathbf{x}) = P(Y \leq y|\mathbf{X} = \mathbf{x})$$

whereas the c.q.f. is defined on $(0,1) \times \mathbb{R}^K$ and is given by

$$Q(\tau|\mathbf{x}) = \{y \in \mathbb{R} : F(y|\mathbf{x}) = \tau\}.$$

Obviously, also in the conditional case, one has that

$$Q(\tau|\mathbf{x}) = F^{-1}(\tau|\mathbf{x}) \quad \text{and} \quad F(y|\mathbf{x}) = Q^{-1}(y|\mathbf{x})$$

and, for any fixed $\mathbf{x}$, $F(y|\mathbf{x})$ and $Q(\tau|\mathbf{x})$ must satisfy all the properties of a distribution function and a quantile function, respectively. In particular, one can see that, for example, different quantile functions cannot cross. This amounts to say that for $\tau < \tau'$, $Q(\tau'|\mathbf{x}) < Q(\tau|\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ (Peracchi (2002)).

2.2 The Quantile Regression

Given the sample $(y_i, \mathbf{x}_i')$, for $i = 1, \ldots, n$, where y_i's are independent and continuous random variables distributed according to F_Y, and $\mathbf{x}_i'$'s are K-vector of predictors, the quantile regression (QR) problem consists on the estimation of the c.q.f. $Q(\tau|\mathbf{x})$ using the information provided by the sample. The usual approach considers $Q(\tau|\mathbf{x})$ as a function of $\mathbf{x}$ given τ, then, it can interpreted as the solution of the minimization problem

$$\min_{h \in \Xi} \mathbb{E}[\rho_\tau(Y_i - h(\mathbf{X}_i))] \quad 0 < \tau < 1, \tag{2}$$

where Ξ is the set of real-valued functions in $\mathbb{R}^K$.
In practical applications, one usually select a specific subclass of Ξ, say $\mathcal{H}$, for $h(\cdot)$ and solve the sample analogue of (2)

$$\min_{h\in\mathcal{H}} \sum_{i=1}^{n} \rho_\tau(y_i - h(\mathbf{x}_i)).$$

In the usual approach, introduced by Koenker & Bassett (1978), $\mathcal{H}$ is the family of linear functions of $\mathbf{x}$, that is $Q(\tau|\mathbf{x}') = \mathbf{x}_i'\beta(\tau)$, and the problem (2) reduces to estimate the τ-dependent K-dimensional parameter $\beta(\tau)$ through the solution $\hat{\beta}(\tau)$ of

$$\min_{\beta\in\mathbb{R}^K} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i'\beta). \tag{3}$$

Obviously, the estimate of $Q(\tau|\mathbf{x})$ will be given by $\hat{Q}(\tau|\mathbf{x}) = \mathbf{x}_i'\hat{\beta}(\tau)$. A nice features of this approach is that a fraction τ of y_i's lies on or below the estimated quantile regression hyperplane $\hat{Q}(\tau|\mathbf{x})$. Since $Q(\tau|\mathbf{x})$ can be viewed as a function of both τ and $\mathbf{x}$, it is common to select L different probability values $0 < \tau_1 < \ldots < \tau_l < \ldots < \tau_L < 1$ and then estimate L different c.q.f. $Q_l(\mathbf{x}) = Q(\tau_l|\mathbf{x})$, $l = 1, \ldots, L$. By suitably choosing their number and position, one may get a reasonable estimate of the c.d.f. $F(y|\mathbf{x})$. Hence, given the estimated c.q.f. $\hat{Q}(\tau|\mathbf{x})$ the corresponding estimate of the c.d.f is simply given by

$$\hat{F}(y|\mathbf{x}) = \sup\{\tau \in (0,1) : \hat{Q}(\tau|\mathbf{x}) \leq y\}$$

The estimated c.d.f will be a proper c.d.f. if and only if $\hat{Q}(\tau|\mathbf{x})$ is a proper c.q.f.
The standard Bayesian approach to the estimation of the linear Quantile Regression is based on the notion of the Asymmetric Laplace Distribution (ALD) (Yu & Moyeed (2001),Yu et al. (2005)).
Just like the Normal distribution in standard regression problems, the ALD is the parametric distributional link between the previous minimization problem (3) and the maximum likelihood theory for quantile regression. The ALD (Koenker & Machado (1999), Yu & Moyeed (2001)) is a skewed distribution with three parameters: a skewness parameter $0 < \tau < 1$, a scale parameter $\sigma > 0$ and a location parameter $\mu \in \mathbb{R}$. A random variable

$Y \in \mathbb{R}$, is distributed according to an ALD(μ, σ, τ) if its density is given by:

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\}.$$

Yu & Zhang (2005) gave properties and generalization of this distribution. In particular one can easily see that $F(\mu|\mu, \sigma, \tau) = \tau$ for any suitable combination of the parameter values. Hence setting $\mu_i = \mathbf{x}_i'\beta$, and assuming that $y_i \sim$ ALD(μ_i, σ, τ), the likelihood associated to n independent observations, $\mathbf{y} = (y_1, \ldots, y_n)$, can be written as

$$L(\beta, \sigma; \mathbf{y}, \tau) = \sigma^{-n} \exp\left\{-\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i - \mathbf{x}_i'\beta}{\sigma}\right)\right\} \tau^n (1-\tau)^n. \tag{4}$$

When σ is considered a nuisance parameter, the maximization of the likelihood in (4) with respect to β is equivalent to the minimization of (3). The introduction of a proper likelihood function in the problem, allows to use the likelihood and Bayesian tools for parametric inference (Koenker & Machado (1999), Yu & Moyeed (2001), Yu et al. (2003)). Although the linear quantile regression models are usually adopted in the empirical studies, they present several drawbacks. The most significant issues are, in general:

1. The validity of the linearity assumption
2. The choice of the function $h(\cdot)$
3. The number and the positions of the L different probability values τ_l
4. The defined model needs to be separately fitted for each corresponding value τ_l
5. The non-crossing condition among the c.q.f.'s might be not satisfied

Following Peracchi (2002), the trouble arising from the first issue can be illustrated as follows: let X be a scalar random variable and let the c.d.f. of Y depend on x only through the location parameter $\mu(x)$, that is,

$$F(y|x) = F(y - \mu(x)).$$

By definition,

$$\tau = F(Q(\tau|x) - \mu(x))$$

and

$$Q(\tau|x) = Q(\tau) + \mu(x). \tag{5}$$

Hence

$$Q(\tau'|x) - Q(\tau|x) = Q(\tau') - Q(\tau) \quad \forall x \in \mathcal{X},$$

that is the distance between any pair of conditional quantiles is independent of x and it will depend just on (τ, τ'). In particular, for location models, where $\mu(x) = \alpha + x\beta$ the (5) becomes

$$Q(\tau|x) = Q(\tau) + \alpha + x\beta$$

and the set of conditional quantiles of Y is the class of parallel lines with slope equal to β.

Furthermore, if the conditional distribution of Y depends on x both through the location parameter $\mu(x) = \alpha + x\beta$ and the scale parameter $\sigma(x)$, (5) becomes

$$Q(\tau|x) = \sigma(x)Q(\tau) + \alpha + x\beta$$

and the linear assumption for the conditional quantiles of Y fails when $\sigma(x)$ is not linear in x.

This implies that the linear quantile regression model can be a bad approximation when data present some degree of heteroskedasticity and the conditional standard deviation is not linear in x. When $\sigma(x)$ is linear in x, the conditional quantiles are linear in x but with possibly different slopes.

In the general case of location models where $\mu(x)$ is a general function, (5) becomes

$$Q(\tau|x) = \mu(x) + \sigma(x)Q(\tau)$$

and the nonlinearity of $Q(\tau|x)$ may be due to the nonlinearity of $\mu(x)$, to heteroskedasticity or both.

The presence of heteroskedasticity implies that the estimated linear models may cross each other, violating one of the main properties of quantiles, so invalidating the estimation results.

The linearity assumption can be relaxed using nonparametric estimators which leave $\mathcal{H}$

unrestricted. These estimators are essentially based on kernel or nearest neighbour methods - in the classical literature - and on Dirichlet process in the Bayesian approach. Also these approaches show some potential drawbacks: Bayesian non-parametric methods are computational demanding, and it is challenging to make comparison among them. Also it may be difficult to generalize them to cases in which the number of covariates is large. A possible solution to some of these problems is offered by the use of approximate likelihoods. Dunson & Taylor (2005) proposed a Bayesian approach to quantile regression using the Substitution Likelihood, an approximate likelihood based on the notion of quantile, introduced by Jeffreys (1961) and later developed by Lavine (1995). This likelihood allows to simultaneously estimate all the L different quantiles and to impose the non-crossing condition among them. We have discussed this approach in the first chapter. As already said, mainly because it is difficult to deal with the non crossing condition for large L, it may be convenient to directly estimate the joint distribution of $(Y, \mathbf{X})$.
To avoid to specify the form of the c.d.f $F(Y|\mathbf{x})$, much attention has been devoted in the literature to the non parametric methods. Unfortunately most of these methods tend to do rather poorly when data are sparse, which is the case with more than two covariates, or, as in a Bayesian framework, are quite computationally demanding.
For all these reasons, we explore in this paper a flexible and easy to interpret family of distributions that encompasses a considerable variety of distributional shapes, providing accurate results, and that can be implemented in a Bayesian framework with little computational effort. Summarizing, since the interest concern the conditional quantile of Y given $\mathbf{X}$, we propose the use of the multivariate **g**-and-**h** distribution, a family of multivariate distribution defined in terms of quantiles, introduced by Field & Genton (2006), to estimate the joint distribution $F(Y, \mathbf{X})$.

3 Quantile Distributions

Distributional families tend to fall into two broad classes:

1. Those defined in terms of the probability density function (PDF).

2. Those defined in terms of a transformation of a base distribution.

Most common distributions fall into the first category, whereas a quantile distribution defined by a quantile function fall into the second one. Some distributional families defined by their q.f. are the Johnson system, the Tukey lambda distribution and its generalization, the g-and-k distribution and finally the g-and-h distribution (Rayner & MacGillivray (2002a)). Except for the Johnson system, which can be written also in terms of its density, all the other quantile distributions, defined by their quantile functions, have not a closed form expression for their PDFs and an explicit likelihood function is not available. Quantile distributions admit great flexibility in shape and have a small number of parameters which can very well describe distributional characteristics like skewness and kurtosis.

Due to the lack of a closed expression for the likelihood function, fitting methods used in the literature have been essentially focused on moments, on some form of quantile matching (Rayner & MacGillivray (2002b)), on numerical maximization of the likelihood in the frequentist paradigm (Rayner & MacGillivray (2002b), Su (2007)) or the Bayesian one (Haynes & Mengersen (2005)). More recently the potentially of ABC approach has been extensively explored in Allingham et al. (2009), McVinish (2010) and Drovandi & Pettitt (2011).

3.1 Definition

The q.f. for a continuous random variable Y is defined as the inverse of its distribution function $F(\cdot)$, when F depends on an unknown parameter vector $\boldsymbol{\vartheta}$, then

$$y = Q(\tau|\boldsymbol{\vartheta}) = F^{-1}(\tau|\boldsymbol{\vartheta}).$$

Then to each realization y_i is associated a probability value $\tau_i = F(y_i|\boldsymbol{\vartheta})$ representing the depth corresponding to y_i.

The usual way of constructing quantile distributions is based on a transformation, say $R_{\boldsymbol{\psi}}(\cdot)$, of quantiles of a given distribution, for example the Gaussian one. Location and scale parameters can then be added, and so one gets

$$y = Q(\tau|\boldsymbol{\vartheta}) = A + B \cdot R_{\boldsymbol{\psi}}(z_\tau) \tag{6}$$

where A and B are respectively the location and scale parameters, $\boldsymbol{\psi}$ is a shape parameter, $R_{\boldsymbol{\psi}}(\cdot)$ is a known function depending on $\boldsymbol{\psi}$, and z_τ is the τ-th quantile of a given distribution.
The corresponding density is given by:

$$f_Y(y_i|\boldsymbol{\vartheta}) = \begin{cases} \frac{1}{Q'_Y(Q_Y^{-1}(y_i|\boldsymbol{\vartheta})|\boldsymbol{\vartheta})} & \inf\{y|y \in \mathcal{Y}\} < y_i < \sup\{y|y \in \mathcal{Y}\}; \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathcal{Y}$ is the support of the random variable Y. This density integrate to one whenever the quantile function is a non decreasing function of the base random variable (Rayner & MacGillivray (2002a)).
Depending on the choice of the function $R_{\boldsymbol{\psi}}(\cdot)$ one has different quantile distributions.

1. *The g-and-h distribution*

   The g-and-h distributional family has been introduced by Tukey (1977) by transforming the standard normal distribution using a skewness parameter g and a kurtosis parameter h, through the following function:

$$R_{\boldsymbol{\psi}}(z_\tau) = R_{g,h}(z_\tau) = \left(\frac{\exp(gz_\tau) - 1}{g}\right) \exp\left(\frac{hz_\tau^2}{2}\right), \tag{7}$$

   The quantile representation of a g-and-h distribution is then

$$Q_{g,h}(\tau; A, B, g, h) = A + B \cdot \left(\frac{\exp(gz_\tau) - 1}{g}\right) \exp\left(\frac{hz_\tau^2}{2}\right). \tag{8}$$

   Here, $g \in \mathbb{R}$ is a measure of skewness: $g < 0$ indicates that the distribution is skewed to the left, $g > 0$ indicates skewness to the right. Also, $h \in \mathbb{R}_0^+$ is a measure of kurtosis. The value $h = 0$ corresponds to no extra kurtosis compared with the normal (base) distribution, which shows the minimum amount of kurtosis for this family of quantile distributions.
   This family of distributions is quite useful because it encompasses a considerable variety of distribution shapes and it has only four parameters, that are easy to interpret. In particular:

- for $g = 0$ and $h = 0$ one gets the normal distribution
- for $g = 0$ and $h > 0$ one gets tails heavier than the normal ones.
- $g \neq 0$ and $h = 0$ coincides with the log-normal distribution.

The g-and-h family only produces leptokurtotic distributions.

2. *The generalized g-and-h distribution*

The generalized g-and-h distribution quantile function is given by:

$$Q_{G_{g,h}}(\tau; A, B, g, h) = A + B \cdot \left(1 + c\frac{1 - \exp(-gz_\tau)}{1 + \exp(-gz_\tau)}\right) \exp\left(\frac{hz_\tau^2}{2}\right) z_\tau$$

where:

- c measures the "overall asymmetry" and it is usually fixed approximately at 0.8 (MacGillivray (1986), Rayner & MacGillivray (2002a))

3. *The g-and-k distribution*

The g-and-k distribution is given by the following quantile function:

$$Q_{g,k}(\tau; A, B, g, k) = A + B \cdot \left(1 + c\frac{1 - \exp(-gz_\tau)}{1 + \exp(-gz_\tau)}\right) (1 + z_\tau^2)^k z_\tau$$

Here again, A and B are respectively the location and the scale parameter, $g \in \mathbb{R}$ is a measure of skewness: $g < 0$ indicates the distribution is skewed to the left, $g > 0$ indicates skewness to the right, $k \in (-\frac{1}{2}, +\infty)$ is a measure of kurtosis. The k parameter plays a role similar to that of h in the previous class; the main difference is that, for negative values of k, the g-and-k distributions are platykurtotic. Finally, c measures the "overall asymmetry" and it is usually fixed approximately at 0.8, and z_τ is the τ-th standardized normal quantile.

3.2 Estimation Methods

Given a random sample $\mathbf{y} = (y_1, \ldots, y_n)$ taken from a quantile distribution, the likelihood function is given by:

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} f_Y(y_i|\boldsymbol{\vartheta}) = \prod_{i=1}^{n} \frac{\partial}{\partial y_i} Q_Y^{-1}(y_i|\boldsymbol{\vartheta}) = \left(\prod_{i=1}^{n} Q'_Y(Q_Y^{-1}(y_i|\boldsymbol{\vartheta})|\boldsymbol{\vartheta}) \right)^{-1}$$

and, if the base distribution is the standard normal one, $z_\tau = \Phi^{-1}(\tau)$, and one obtains

$$Q'_Y(\tau|\boldsymbol{\vartheta}) = \frac{\partial Q_Y}{\partial \tau} = \frac{\partial Q_Y}{\partial z_\tau} \frac{dz_\tau}{d\tau},$$

where

$$\frac{dz_\tau}{d\tau} = \sqrt{2\pi} \exp\left(\frac{z_\tau^2}{2} \right),$$

and a closed form expression for $Q_Y^{-1}(y_i|\boldsymbol{\vartheta})$ doesn't exist. For this reason parameter estimation methods have focused on different approaches which do not require the likelihood evaluation. Rayner & MacGillivray (2002a) and Su (2007) have proposed a maximum likelihood estimator (MLE) using numerical approximations to the likelihood function in a frequentist framework, whereas Haynes & Mengersen (2005) proposeed a Bayesian version of it. Rayner & MacGillivray (2002b) proposed also a parameter estimation based on quantile-matching for the g-and-k distribution. Finally Allingham et al. (2009), McVinish (2010), Drovandi & Pettitt (2011) have proposed ABC methods for the parameter estimation of quantile distributions.

Briefly ABC is a parameter estimation scheme which utilizes simulated data to assess proposed parameter values, without requiring likelihoods, but only a way to sample from the distribution chosen to fit the data. We refer to section 5 for a complete description of the algorithm.

4 The Multivariate Quantile and the Multivariate g-and-h Distribution

In this section, following Chaudhuri (1996) and Field & Genton (2006), we extend the notions of quantile and g-and-h distribution to the multivariate framework.

In §2.1 we have defined the univariate quantile as the solution of the minimization problem (1). That minimization problem can be reformulated by considering, for each $\tau \in [0,1]$, the quantity $u = 2\tau - 1$. Given the sample $(y_1, \ldots, y_n)$ the τ-th quantile q can be obtained as the

$$\arg\min_{q \in \mathbb{R}} \sum_{i=1}^{n} |y_i - q| - nuq \tag{9}$$

This formulation can be easily extended in the multivariate setting. The multivariate quantile $\mathbf{q}$ in $\mathbb{R}^{K+1}$ for the multivariate sample $\mathbf{w}_1, \ldots, \mathbf{w}_n$ must be specified in terms of the direction we are interested in. For each vector $\mathbf{u} \in \mathbb{R}^{K+1}$, such that $\|\mathbf{u}\|_p < 1$, we define the multivariate quantile $\mathbf{q}$, along the direction $\mathbf{u}$ and depth $\|\mathbf{u}\|_p$, as

$$\arg\min_{\mathbf{q} \in \mathbb{R}^{K+1}} \sum_{i=1}^{n} \|\mathbf{w}_i - \mathbf{q}\|_p - n \sum_{j=1}^{K+1} u_j q_j \tag{10}$$

where $\|.\|_p$ denotes the l_p-norm for $1 \leq p < \infty$.

Depending on the chosen l_p-norm we obtain different multivariate quantiles. By taking $p = 1$, (10) can be written as

$$\arg\min_{\mathbf{q} \in \mathbb{R}^{K+1}} \sum_{j=1}^{K+1} \left(\sum_{i=1}^{n} |w_{ij} - q_j| - nu_j q_j \right),$$

giving the result that the multivariate quantile is computed componentwise, since the minimization can be done separately for each dimension, so that q_j is the $\tau = (u_j + 1)/2$-th quantile for the j-th component.

Unlike the l_1-norm, that gives the usual univariate quantiles for each component, the l_2-norm incorporate the dependence among the variables w_j, $\forall j = 1, \ldots, K+1$, and produces less common quantiles in one dimension. For this reason we essentially focus on l_2-norm.

Given the definition of the multivariate quantile we can define the multivariate **g**-and-**h** distribution. A random vector $\mathbf{W} \in \mathbb{R}^{K+1}$ is said to have a multivariate **g**-and-**h** distribution if it can be represented as

$$\mathbf{W} = \boldsymbol{\Sigma}^{1/2}\mathbf{R}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) + \boldsymbol{\mu} \tag{11}$$

where:
- $\boldsymbol{\mu} \in \mathbb{R}^{K+1}$ is the location parameter vector
- $\boldsymbol{\Sigma}$ is the covariance matrix
- $\mathbf{g} = (g_1, g_2, \ldots, g_{K+1}) \in \mathbb{R}^{K+1}$ controls for the skewness
- $\mathbf{h} = (h_1, h_2, \ldots, h_{K+1}) \in {\mathbb{R}_+}^{K+1}$ controls for the kurtosis
- $\mathbf{Z} \sim N_{K+1}(0, \mathbf{I})$ is the standard normal random variable
- $\mathbf{R}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) = (R_{g_1,h_1}(Z_1), R_{g_2,h_2}(Z_2), \ldots, R_{g_{K+1},h_{K+1}}(Z_{K+1}))^T$ where $R_{g,h}$ is given by (7).

The componentwise application of the function $R_{g,h}$ allows to maintain the property that, for small g, the transformation is approximately linear.
Since the g-and-h distribution is a quantile distribution, it is worth to define the quantile function in the multivariate case corresponding to (8).
Given the direction $\mathbf{u} \in \{\mathbf{u} | \mathbf{u} \in \mathbb{R}^{K+1}, \|\mathbf{u}\|_p < 1\}$ the quantile function for the multivariate **g**-and-**h** distribution, based on the l_p-norm, is given by

$$\mathbf{Q}_{\mathbf{Y}}(\mathbf{u}) = \boldsymbol{\Sigma}^{1/2}\mathbf{R}_{\mathbf{g},\mathbf{h}}(\mathbf{Q}_{\mathbf{Z}}(\tilde{\mathbf{u}})) + \boldsymbol{\mu}$$

where $\tilde{\mathbf{u}} = \boldsymbol{\Sigma}^{-1/2}\mathbf{u}\|\mathbf{u}\|_r / \|\boldsymbol{\Sigma}^{-1/2}\mathbf{u}\|_r$ with $\frac{1}{p} + \frac{1}{r} = 1$, if $p = 1$ ($r = \infty$) we have the sup norm. This definition is consistent with the univariate case given by (8).
To estimate the multivariate **g**-and-**h** distribution, Field & Genton (2006) proposed a quantile matching technique based on a transformation of the original data that give affine-equivariant quantiles (Chakraborty (2001)). In what follows, we illustrate an ABC algorithm to estimate the parameter vector $(\mathbf{g}, \mathbf{h}, \boldsymbol{\mu}, \Sigma)$.

5 Likelihood-free Markov Chain Monte Carlo: the Approximate Bayesian Computation (ABC) method

Many Bayesian estimation methods rely on stochastic simulation approaches for generating observation from the posterior distribution. These approaches, as Markov Chain Monte Carlo methods, require the knowledge of the likelihood function, and become either computational prohibitive for complex probability models or even impossible for those models, such as the quantile distributions, which don't have an explicit likelihood function. In these case ABC methods are recommendable.

ABC methods are parameter estimation schemes that utilize simulated data to assess proposed parameter values; unlike MCMC methods, they do not require likelihoods but only a way to sample from the model chosen to describe the data.

Marjoram et al. (2003) details several Monte Carlo parameters estimation methods and present MCMC Without Likelihoods algorithms establishing its theoretical properties, Sisson & Fan (2010) give an exhaustive description of the algorithm.

Briefly, the ABC algorithm proceeds by proposing a parameter value θ' by a transition kernel based on current parameter value θ. A random sample $\mathcal{D}'$ is then drawn from the model $\mathcal{M}$, chosen to describe the data $\mathcal{D}$, using the proposed parameter θ'. Summary statistics for the observed data $\mathcal{D}$ and for the simulated data $\mathcal{D}'$ are computed and compared through a distance metric ρ. Then the proposed parameter is accepted or rejected assessing the computed distance.

In detail, following Allingham et al. (2009), consider the sample $\mathcal{D}$ with n observations generated from the model $\mathcal{M}$ with parameter θ, let $\rho(\cdot)$ a distance metric and $S(\cdot)$ a summary statistics. The proposed ABC algorithm proceed as follows:

A1. Being at θ, propose a move to θ' according to a transition kernel $q(\theta \to \theta')$

A2. Generate $\mathcal{D}'$ using model $\mathcal{M}$ with parameters θ'

A3. If $\rho(S(\mathcal{D}), S(\mathcal{D}')) < \epsilon$ go to A4 otherwise stay at θ and return to A1

A4. Calculate $\alpha = \min(1, \frac{\pi(\theta')q(\theta' \to \theta)}{\pi(\theta)q(\theta \to \theta')})$

A5. Accept θ' with probability α, otherwise stay at θ; then return to A1,

where $S(\cdot)$ is a summary statistics, $\rho(\cdot)$ is a distance metric and ϵ is an acceptance threshold. The length of $\mathcal{D}'$ could be different from n, although we always consider in the following the same sample size for $\mathcal{D}$ and $\mathcal{D}'$. All the accepted values of θ' are used for estimating the parameter θ.
The initial value of θ is obtained in a similar way using the prior distribution $\pi(\theta)$, thus:

(I) Sample a parameter θ from $\pi(\theta)$

(II) Draw a random sample, $\mathcal{D}'$ from $\mathcal{M}(\theta)$

(III) Accept θ if $\rho(S(\mathcal{D}), S(\mathcal{D}')) < \epsilon_0$ otherwise return to (I)

where ϵ_0 may be different from ϵ.
If $S(\cdot)$ is a sufficient statistics for θ then in the limit as $\epsilon \to 0$, accepted values of θ are samples from the true posterior distribution; otherwise the accepted parameters become a sample from an approximation of the posterior; for this reason the algorithm is named "Approximate".
An equivalent and more efficient algorithm can be obtained by introducing a standard smoothing kernel density, $K(\cdot)$, on the distance metric $\rho(\cdot)$, centered at the point $S(\mathcal{D}') = S(\mathcal{D})$ with scale given by ϵ (Sisson & Fan (2010)). In this manner, introducing the kernel density

$$K_\epsilon\left(\rho(S(\mathcal{D}), S(\mathcal{D}'))\right) \tag{12}$$

we can weigh the proposed parameters θ' with hight values when $S(\mathcal{D}) \approx S(\mathcal{D}')$ and with values close to zero otherwise. In this way one obtains the following ABC algorithm

B1. If we are at θ_t propose a move to θ' according to a transition kernel $q(\theta_t \to \theta')$

B2. Generate $\mathcal{D}'$ using model $\mathcal{M}$ with parameters θ'

B3. Calculate $\alpha = \min\left(1, \frac{K_\epsilon(\rho(S(\mathcal{D}),S(\mathcal{D}')))\pi(\theta')q(\theta'\to\theta_t)}{K_\epsilon(\rho(S(\mathcal{D}),S(\mathcal{D}'_t)))\pi(\theta_t)q(\theta_t\to\theta')}\right)$

B4. Accept θ' with probability α, otherwise stay at θ_t; then return to B1

by choosing a uniform kernel density, K_U, where $S(\mathcal{D})$ is uniformly distributed on the sphere centered at $S(\mathcal{D}')$ with radius ϵ, one reobtains the previous algorithm A. Alternative kernel density of interest are the Epanechnikov kernel (Beaumont et al. (2002)) and the Gaussian kernel density (Peters et al. (2008)).
Furthermore, since the interest concerns on targeting the posterior distribution of θ, the algorithm B can be improved through a Monte Carlo integration by replacing the steps B2 and B3 with

C2. Generate M samples, $\mathcal{D}'_1, \ldots, \mathcal{D}'_M$, using model $\mathcal{M}$ with parameters θ'

C3. Calculate $\alpha = \min\left(1, \frac{\frac{1}{M}\sum_{m=1}^{M} K_\epsilon(\rho(S(\mathcal{D}),S(\mathcal{D}'_m)))\pi(\theta')q(\theta'\to\theta_t)}{\frac{1}{M}\sum_{m=1} MK_\epsilon(\rho(S(\mathcal{D}),S(\mathcal{D}'_{m;t})))\pi(\theta_t)q(\theta_t\to\theta')}\right)$

In all the above described ABC algorithms, the performance of the algorithm depends on:

- The easiness of simulating from $\mathcal{M}$.
- The choice of the summary statistics, $S(\cdot)$, the distance metric, $\rho(\cdot)$, and the acceptance threshold ϵ.

For the algorithm C, Sisson & Fan (2010) showed that the choice of the kernel function, $K_\epsilon(\cdot)$, has a larger impact on the sampler performance than the number of generated datasets M, while the Gaussian kernel outperforms the Uniform kernel.
All the other components of the algorithm are identical to the most common MCMC algorithm.
Priors distributions are chosen in such a way that they reflect either specific prior knowledge of the problem or the degree at which the evidence should inform the conclusion. Adopting quantile distributions prior knowledge about the mode, the expected shape, in terms of skewness or asymmetry, can be easily incorporated into the prior.
With respect to the choice of the transition kernel, one has a trade-off. With the transition kernel, the algorithm should explore the entire parameter space, suggesting broad kernel, and, at the same time, the acceptance rate of proposed parameters should not be very small, suggesting narrow kernels. Looking at the algorithm's structure, it is expected that "valid" values (referred to the threshold ϵ) of θ' have other "valid" parameters in the neighbourhood, and this suggests narrow kernels.

The choice of the summary statistics $S(\cdot)$ and of the distance metric ρ depends on the empirical study of interest. In this context it is worth to remind that the ability of ABC to estimate the parameter θ relies on the summary statistics chosen to describe the data features, because the algorithm accepts only those parameters that reflect those features. The use of sufficient statistics is a requirement to ensure that the accepted parameters belongs to the true posterior distribution. If sufficient statistics are not available, the choice must be driven by the characteristics of the model used to fit the data. Finally the euclidean norm is usually chosen for ρ.
The choice of ϵ is made through a preliminary execution of the algorithm with large values of ϵ. This preliminary execution give a posterior distribution close to the prior but give the possibility of study the histogram of the $\rho(\mathcal{D}, \mathcal{D}')$'s obtained and set the acceptance threshold close to the left-hand end of this histogram.

6 The Model

Given the data,$(\mathcal{D})$, $\mathbf{w}_i = (y_i, \mathbf{x}_i')' \in \mathbb{R}^{K+1}$ for $i = 1, \ldots, n$, suppose that the interest concerns the estimation of a large number L of conditional quantiles

$$Q_Y(\tau_l|\mathbf{x}) = h(\mathbf{x}) \quad \text{for} \quad l = 1, \ldots, L$$

but the lack of prior informations about the function $h(\cdot)$ and about the conditional distribution $F(Y|\mathbf{X})$ make the standard Bayesian quantile regression methods and the approximate ones, based on approximate likelihood (Dunson & Taylor (2005)), unfeasible.
Furthermore the large number of quantiles to be estimated make necessary to estimate directly the conditional distribution $F(Y|\mathbf{X})$. In this context, due to the lack of any information about its form, the only possible way to estimation is to apply non-parametric methods. But Bayesian non-parametric methods are computational demanding and are difficult to implement.
Our solution rely to the fact that some "rough" prior informations are instead available. These informations concern the principal characteristics of any distribution like the location, the dispersion, the asymmetry and the kurtosis.

Since the interest concern the conditional quantiles of Y given $\mathbf{X}$, we propose to use the Multivariate **g**-and-**h** distribution in order to estimate the joint distribution $F_{(Y,\mathbf{X})}$. As we saw in the section §4 this family of distribution is flexible, easy to interpret, encompasses a considerable variety of distribution shapes providing accurate results, and finally can be implemented in a Bayesian framework with little computational efforts.
Thus given the estimate $\hat{F}_{(Y,\mathbf{X})}$ is straightforward to estimate the conditional distribution function $F_{(Y|\mathbf{X})}$ and its quantiles

$$\hat{F}_{(Y|\mathbf{X})} = \frac{\hat{F}_{(Y,\mathbf{X})}}{\hat{F}_{(\mathbf{X})}}$$

and

$$\hat{Q}_Y(\tau_l|\mathbf{X}) = \hat{F}^{-1}_{(Y|\mathbf{X})}(\tau_l) \quad l = 1, \ldots, L.$$

where the distribution $\hat{F}_{(\mathbf{X})}$ is estimated either in the same way of $\hat{F}_{(Y,\mathbf{X})}$ or with the standard Bayesian methods if prior information is available.
Given the data, $(\mathcal{D})$, $\mathbf{w}_i = (y_i, \mathbf{x}_i')' \in \mathbb{R}^{K+1}$ for $i = 1, \ldots, n$, we assume that the joint distribution of the response variable $Y \in \mathbb{R}$, and the set of covariates $\mathbf{X} \in \mathbb{R}^K$ is given by the multivariate **g**-and-**h** distribution:

$$\mathbf{W} = (Y, \mathbf{X}')' = \boldsymbol{\Sigma}^{1/2}\mathbf{R}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) + \boldsymbol{\mu} \tag{13}$$

with
- $\mathbf{g} \in \mathbb{R}^{K+1}$ is the skewness parameter
- $\mathbf{h} \in \mathbb{R}^{K+1}_+$ is the kurtosis parameter
- $\boldsymbol{\Sigma} \in \left(\mathbb{R}^{K+1} \times \mathbb{R}^{K+1}\right)$ is the variance covariance matrix
- $\boldsymbol{\mu} \in \mathbb{R}^{K+1}$ is the location parameter

and where $\mathbf{Z}$ is the standard multivariate normal distribution.

6.1 The estimation method

In order to estimate the statistical model given by (13) we propose the ABC algorithm C as described in §5.
Prior distributions and transition kernels are chosen as in the standard Normal Symmetric

Random Walk Metropolis (N-SRWM) methods. Thus, given the following independent priors

$$\pi(\mathbf{g}), \quad \pi(\mathbf{h}), \quad \pi(\boldsymbol{\Sigma}), \quad \pi(\boldsymbol{\mu})$$

and the following transition kernels

$$q(\mathbf{g}^{t+1}|\mathbf{g}^t) = \mathbf{g}^t + \boldsymbol{\xi}^{t+1}, \quad q(\mathbf{h}^{t+1}|\mathbf{h}^t) = \mathbf{h}^t + \boldsymbol{\xi}^{t+1},$$
$$q(\boldsymbol{\Sigma}^{t+1}|\boldsymbol{\Sigma}^t) = \boldsymbol{\Sigma}^t + \boldsymbol{\xi}^{t+1}, \quad q(\boldsymbol{\mu}^{t+1}|\boldsymbol{\mu}^t) = \boldsymbol{\mu}^t + \boldsymbol{\xi}^{t+1}$$

where $\boldsymbol{\xi}^{t+1}$ is a multivariate normal distribution with zero mean, for the other components of the algorithm, we propose

(i) A multivariate quantile-based summary statistic, $S(\cdot)$

(ii) The Euclidean norm $\rho(\cdot) = \|.\|_2$

(iii) A data driven choice for the multivariate threshold $\boldsymbol{\epsilon}$

(iv) A multivariate Gaussian kernel, $K_{\boldsymbol{\epsilon}}(\cdot)$

The choice of a multivariate quantile-based summary statistics $S(\cdot)$ has been made by considering that, since the multivariate **g**-and-**h** distribution models the quantiles, the distance of the multivariate quantiles of $\mathcal{D}$ from those of $\mathcal{D}'$ seems a good method to evaluate the proposed parameters.

Thus, at the beginning of the algorithm we define L different directions $\mathbf{u}_l$, $l = 1, \ldots, L$, and we calculate the corresponding multivariate quantiles $\mathbf{q}_{\mathbf{u}_l}(\mathcal{D})$ as a summary statistics of our sample $\mathbf{w}_i = (y_i, \mathbf{x}_i')' \in \mathbb{R}^{K+1}$

$$S_l(\mathcal{D}) = \mathbf{q}_{\mathbf{u}_l}(\mathcal{D}) \quad \text{for} \quad l = 1, \ldots, L$$

using the l_2-norm to incorporate in $S(\cdot)$ the dependence of the response variable Y on the covariates $\mathbf{X}$.

After that a parameter has been proposed and a corresponding sample $\mathcal{D}'$ has been generated, to evaluate the distance between $\mathcal{D}$ and the proposed $\mathcal{D}'$, we propose the Euclidean norm between the quantiles $S_l(\mathcal{D})$ and $S_l(\mathcal{D}')$, for $l = 1, \ldots, L$

$$\rho_l(S_l(\mathcal{D}) - S_l(\mathcal{D}')) = \|S_l(\mathcal{D}) - S_l(\mathcal{D}')\|_2 = \|\mathbf{q}_{\mathbf{u}_l}(\mathcal{D}) - \mathbf{q}_{\mathbf{u}_l}(\mathcal{D}')\|_2 \quad \forall l$$

obtaining the vector of distances $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_L)' \in \mathbb{R}^L$.
Then the proposed parameter is weighed by a multivariate Gaussian kernel on $\boldsymbol{\rho}$,

$$K_\epsilon(\boldsymbol{\rho}) \sim N\left(\mathbf{0}, \frac{1}{12} \cdot \mathrm{diag}((\epsilon_1^2, \ldots, \epsilon_L^2))\right)$$

evaluated for M different draws of $\mathcal{D}'_m$, for $m = 1, \ldots, M$. The covariance matrix has been chosen in order to have marginally the same variance of a uniform distribution on $(0, \epsilon_l)$.
The multivariate acceptance threshold $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_L)' \in \mathbb{R}^L$, is selected as follows

(a) The ABC algorithm C is executed with very large values ϵ_l, $\forall l$, getting posterior distributions close to the priors.

(b) The histograms of the values ρ_l, $\forall l = 1, \ldots, L$ are constructed

(c) Values of ρ_l, $\forall l$ for the best performing parameters can be observed from the left-hand tail of these histograms

(d) Good choices of ϵ_l, $\forall l$ are taken to be values lying close to the left-hand end of these histograms.

Finally, since step C3 integrates out the randomness due to sampling datasets $\mathcal{D}'_m$, in order to leave the trade off describe in §5 for the variance of the normal transition kernels $q(\cdot)$, and getting an acceptance probability of approximately 0.3 for each parameter (Roberts et al. (1997)), we propose - inside the ABC algorithm C - an Adaptive MCMC (AMCMC) scheme. AMCMC are special MCMC methods that allows to get the required acceptance probability without specify apriori the variance of the proposal density, since the chain, being self-learning, specifies this value at each iteration. Specifically, we use the Algorithm 4 in Andrieu & Thoms (2008). In this context the adaptive structure must be not confused with the adaptive ABC of Beaumont et al. (2009).
At the end, the inference is based on the accepted parameters that are samples from an approximation of the posterior distribution. Modes and averages, as a point parameter estimation, and the 95% credible intervals, as a interval estimation, are proposed in order to estimate the multivariate **g**-and-**h** distribution's parameters.

7 Simulation studies

Two simulations studies are conducted in order to evaluate the performance of the proposed method.
In the first simulation data have been generated from the linear model

$$y_i = 2x_i + \varepsilon_i \quad i = 1, ..., 1000$$

with the x_i's and ε_i's generated from the following distributions

$$X \sim N(5, 4)$$
$$\varepsilon \sim N(0, 9)$$

obtaining the following distribution for the generated dataset

$$\mathbf{w} = [\mathbf{y}, \mathbf{x}] \sim N_2 \left(\begin{pmatrix} 10 \\ 5 \end{pmatrix}; \begin{bmatrix} 25 & 8 \\ 8 & 4 \end{bmatrix} \right).$$

To apply the proposed method we model the obtained dataset as it come from a multivariate **g**-and-**h** distribution

$$\mathbf{W} = \mathbf{\Sigma}^{1/2} \mathbf{R}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) + \boldsymbol{\mu}.$$

The following non-informative priors are chosen for each parameter:

$$\mathbf{g} \sim N_2(\mathbf{0}; \mathbf{I}_2 \cdot 10^2) \qquad \mathbf{h} \sim N_2(\mathbf{0}; \mathbf{I}_2 \cdot 10^2)$$
$$\boldsymbol{\mu} \sim N_2(\mathbf{0}; \mathbf{I}_2 \cdot 10^2) \qquad \mathbf{vec}(\mathbf{C}) \sim N_3(\mathbf{0}; \mathbf{I}_3 \cdot 10^2)$$

where $\mathbf{C}$ is the lower triangular matrix of the Cholesky decomposition of $\mathbf{\Sigma}$.
The empirical l_2-quantiles have been computed in the directions $2\pi/j$, $j = 1, \ldots, 16$ for the levels 15%, 25%, 50%, 75%, 90%, 95%, and $\boldsymbol{\epsilon}$ has been set equal to the median of the distances obtained in a previous run of the algorithm using informative priors centered on the true values , large $\boldsymbol{\epsilon}$ and a uniform kernel on the support of $\boldsymbol{\rho}$.
Chains of 100.000 iterations have been run with optimal acceptance rate fixed to 0.3 taking one draw every 40, considering $M = 2$ generations of $\mathcal{D}'_m$ for the proposed parameters in each iteration. The obtained chains and the related approximate posterior densities

	Mean	Mode	95% HPD	True value
g_1	-0.03	-0.02	$[-0.41, 0.33]$	0
g_2	-0.50	-0.54	$[-1.59, 0.56]$	0
h_1	0.15	0.10	$[0, 0.35]$	0
h_2	0.38	0.28	$[0, 0.84]$	0
μ_1	10.37	10.39	$[9.89, 10.97]$	10
μ_2	5.51	5.22	$[4.85, 5.56]$	5
σ_1^2	20.14	20.27	$[13.66, 26.80]$	25
σ_{12}^2	5.47	5.34	$[1.98, 9.00]$	8
σ_2^2	2.95	2.38	$[0.39, 5.90]$	4

Table 1: Estimated parameters (*First simulation*)

are shown in Figure 1 and in Figure 2, the Figure 3 depict the mean of the acceptance probabilities across the iterations for each parameter.

As can be seen from Table 1, Figure 1 and Figure 2, our proposed method estimate quite accurately all the parameters without any prior informations, but just setting the multivariate tolerance $\boldsymbol{\epsilon}$, reaching quite quickly the fixed optimal acceptance probability in the M-H step (Figure 3).

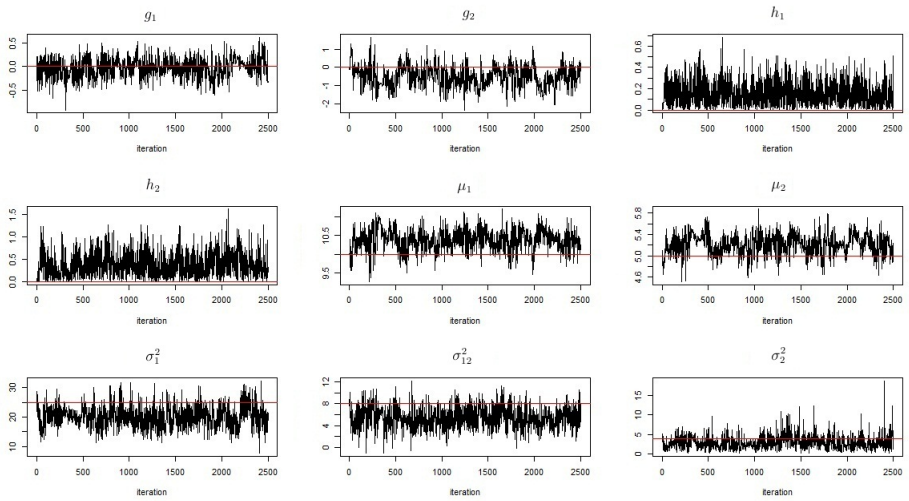

Figure 1: Chains of the parameters and the true parameters (red lines) (*First simulation*)

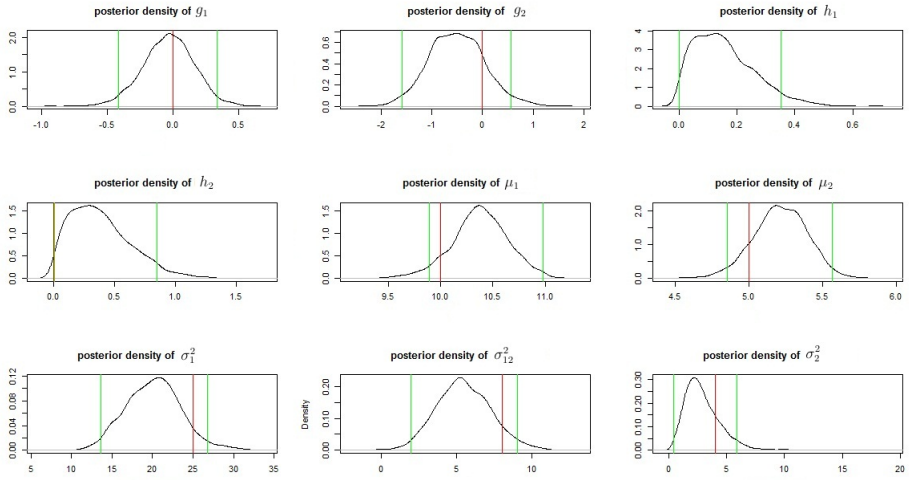

Figure 2: Approximate posterior densities with 95% HPD (green lines) and the true parameters (red lines) (*First simulation*)

In the second simulation data have been generated from a multivariate **g**-and-**h** distribution

$$\mathbf{W} = \boldsymbol{\Sigma}^{1/2}\mathbf{R}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) + \boldsymbol{\mu}.$$

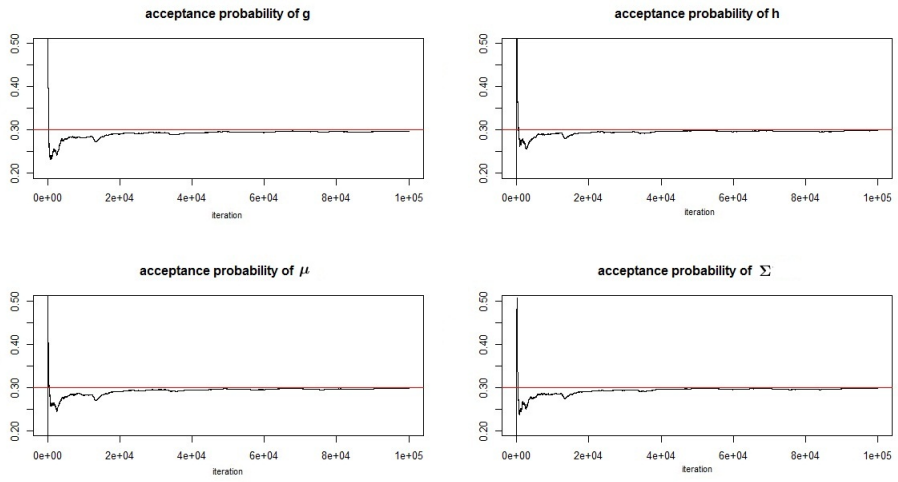

Figure 3: Acceptance probabilities and the optimal ones (red lines)(*First simulation*)

with parameters:

$$\mathbf{g} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \mathbf{h} = \begin{pmatrix} 0 \\ 0.3 \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} -5 \\ 2 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$$

The following informative priors are chosen for each parameter:

$$\mathbf{g} \sim N\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}; \begin{bmatrix} 0.0625 & 0 \\ 0 & 0.0625 \end{bmatrix}\right)$$

$$\mathbf{h} \sim N\left(\begin{pmatrix} 0 \\ 0.3 \end{pmatrix}; \begin{bmatrix} 0.005 & 0 \\ 0 & 0.005 \end{bmatrix}\right)$$

$$\boldsymbol{\mu} \sim N\left(\begin{pmatrix} -5 \\ 2 \end{pmatrix}; \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\right)$$

$$\mathbf{vec}(\mathbf{C}) \sim N\left(\begin{pmatrix} 2 \\ 0.5 \\ 1.936 \end{pmatrix}; \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}\right)$$

where $\mathbf{C}$ is the lower triangular matrix of the Cholesky decomposition of $\boldsymbol{\Sigma}$.
The empirical l_2-quantiles have been computed in the directions $2\pi/j$, $j = 1, \ldots, 8$ for the levels 15%, 25%, 50%, 75%, 90% and $\boldsymbol{\epsilon}$ has been set equal to the median of the distances

obtained in a previous run of the algorithm with a large value of $\boldsymbol{\epsilon}$ and a uniform kernel on the support of $\boldsymbol{\rho}$.

Chains of 100.000 iterations have been run with optimal acceptance rate fixed to 0.3 taking one draw every 40, considering $M = 2$ generations of $\mathcal{D}'_m$ for the proposed parameters in each iteration. The obtained chains and the related approximate posterior densities are shown in Figure 4 and in Figure 5, the Figure 6 depict the mean of the acceptance probabilities across the iterations for each parameter.

As can be seen from Figure 5, our proposed method estimate accurately all the parameters improving the priors and reaching quite quickly the fixed optimal acceptance probability in the M-H step (Figure 6).

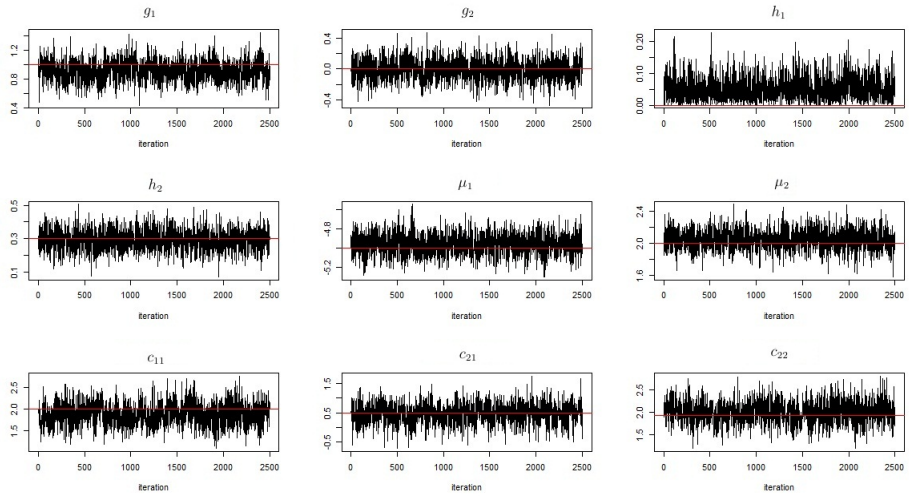

Figure 4: Chains of the parameters and the true parameters (red lines) (*Second simulation*)

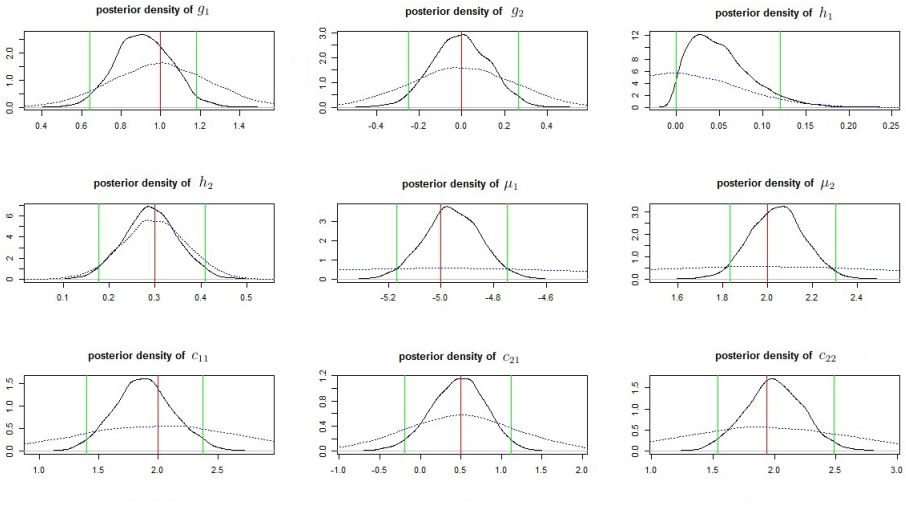

Figure 5: Approximate posterior densities with 95% HPD (green lines), the true parameters (red lines) and the prior distributions (dotted blue lines) (*Second simulation*)

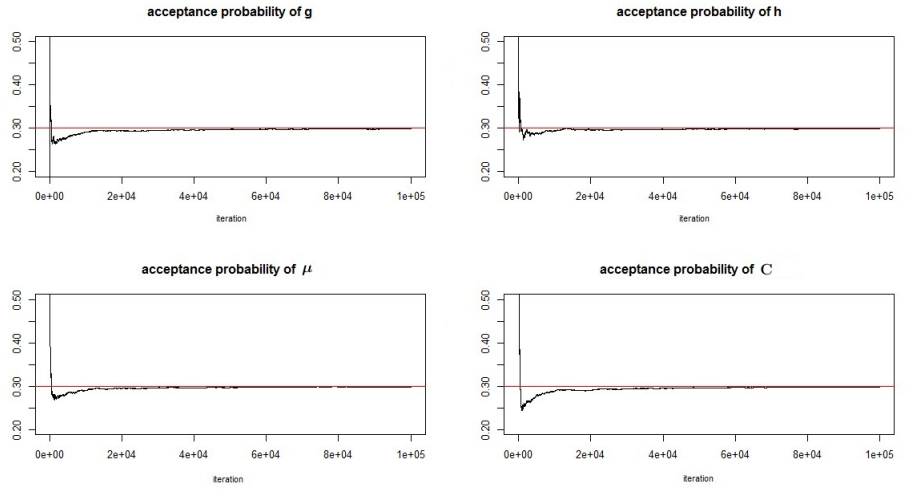

Figure 6: Acceptance probabilities and the optimal ones (red lines) (*Second simulation*)

References

Allingham, D., King, R. A. R. & Mengersen, K. L. (2009). Bayesian estimation of quantile distributions. *Stat. Comput.* 19 189–201.

Andrieu, C. & Thoms, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* 18 343–373.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. & Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika* 96 983–990.

Beaumont, M. A., Zhang, W. & Balding, D. J. (2002). Approximate Bayesian Computation in population genetics. *Genetics* 162 2025–2035.

Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Ann. Inst. Statist. Math.* 53 380–403.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* 91 862–872.

Drovandi, C. C. & Pettitt, A. N. (2011). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Comput. Stat. Data Anal.* 55 2541–2556.

Dunson, D. B. & Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametr. Stat.* 17 385–400.

Field, C. & Genton, M. G. (2006). The multivariate **g**-and-**h** distribution. *Technometrics* 48 104–111.

Foresi, S. & Peracchi, F. (1995). The Conditional Distribution of Excess Returns: An Empirical Analysis. *J. Amer. Statist. Assoc.* 90 451–466.

Haynes, M. & Mengersen, K. (2005). Bayesian estimation of g-and-k distributions using MCMC. *Comput. Statist.* 20 7–30.

Jeffreys, H. (1961). *Theory of probability.* Third edition. Clarendon Press, Oxford.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. & Bassett, G. J. (1978). Regression Quantiles. *Econometrica* 46 33–50.

Koenker, R. & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.* 94 1296–1310.

Lavine, M. (1995). On an approximate likelihood for quantiles. *Biometrika* 82 220–222.

MacGillivray, H. L. (1986). Skewness and asymmetry: Measures and orderings. *The Annals of Statistics* 14 994–1011.

Marjoram, P., Molitor, J., Plagnol, V. & Tavar, S. (2003). Markov chain Monte Carlo without likelihood. *PNAS* 100 15324–15328.

McVinish, R. (2010). Improving ABC for quantile dstributions. *Stat. Comput.* .

Peracchi, F. (2002). On estimating conditional quantiles and distribution functions. *Comput. Statist. Data Anal.* 38 433–447. Nonlinear methods and data mining (Rome, 2000).

Peters, G. W., Fan, Y. & Sisson, S. A. (2008). On sequential Monte Carlo, partial rejection control and approximate Bayesian computation. *ArXiv e-prints* `0808.3466`.

Rayner, G. D. & MacGillivray, H. L. (2002a). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Stat. Comput.* 12 57–75.

Rayner, G. D. & MacGillivray, H. L. (2002b). Weigthed quantile-based estimation for a class of transformation distributions. *Comput. Stat. Data Anal.* 39 401–433.

Roberts, G. O., Gelman, A. & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* 7 110–120.

Sisson, S. & Fan, Y. (2010). Likelihood-free Markov chain Monte Carlo. *ArXiv e-prints* `1001.2058`.

Su, S. (2007). Numerical maximum log likelihood estimation for generalized lambda distributions. *Comput. Stat. Data Anal.* 51 3983–3998.

Tavar, S., Balding, D., Griffiths, R. c. & Donnelly, P. (1997). Inferring Coalescence Times From DNA Sequence Data. *Genetics* 145 505–518.

Tukey, J. W. (1977). Modern techniques in data analysis. NSF - Sponsored Regional Research Conference, Southeastern Massachusetts University, North Dartmouth, Massachusetts .

Yu, K., Lu, Z. & Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician* 52 331–350.

Yu, K. & Moyeed, R. A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* 54 437–447.

Yu, K., Van Kerm, P. & Zhang, J. (2005). Bayesian quantile regression: an application to the wage distribution in 1990s Britain. *Sankhyā* 67 359–377.

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