Statistica Sinica **20** (2010), 481-496

MULTIVARIATE QUANTILE FUNCTION MODELS

Yuzhi Cai

University of Plymouth

Abstract: Multivariate quantiles have been defined by a number of researchers and can be estimated by different methods. However, little work can be found in the literature about Bayesian estimation of joint quantiles of multivariate random variables. In this paper we present a multivariate quantile function model and propose a Bayesian method to estimate the model parameters. The methodology developed here enables us to estimate the multivariate quantile surfaces and the joint probability without direct use of the joint probability distribution or density functions of the random variables of interest. Furthermore, simulation studies and applications of the methodology to bivariate economics data sets show that the method works well both theoretically and practically.

Key words and phrases: Bayesian method, τth quantile surface, τth quantile curve, multivariate quantile function, MCMC, log return.

1. Introduction

Quantiles play an important role in statistical analysis of many areas such as economics, finance, and coastal engineering. The problem is often to estimate the quantiles of a variable conditional on the values of other variables.

There have been several approaches to quantile functions for multivariate distributions. The simplest defines a quantile vector as one that has the marginal classical quantiles as its components, but this does not take account correlations between the components of the vectors of observations (Chakraborty (2001)).

Substantive extensions are based on ordering multivariate observations, for example, Eddy (1985); Brown and Hettmansperger (1987, 1989); Oja (1983); Donoho and Gasko (1992); He and Wang (1997); Liu, Parelius, and Singh (1999); Zuo and Serfling (2000); Serfling (2002). The definition of multivariate quantile proposed by Chaudhuri (1996) is a generalization of what was proposed by Koenker and Bassett (1978) in one dimension. See also Salibian-Barrera and Wei (2008) and Wei (2008). Gilchrist (2000) defined bivariate quantiles differently, and we find these ideas practically appealing. In this paper, we generalize Gilchrist's (2000) bivariate quantile function definition.

In the literature, different methods have been developed to estimate quantiles, for example, Koenker and D'Orey (1987, 1994); Yu and Moyeed (2001);

Cai (2007); Cai and Stander (2008); Kottas and Gelfand (2001); Dunson, Watson and Taylor (2003); Schennach (2005). We develop an MCMC approach to estimating multivariate quantiles. Compared with other approaches to multivariate quantile estimation, the proposed approach can deal more broadly with multivariate data, for example, when the Box-Cox transformations fail to transform the data into multivariate Gaussian. We can construct proper multivariate quantile function models, see Section 2.1, so that the relationship between the random variables can be explored through joint quantiles. Furthermore, our approach enables us to study extremes by using all the available data, not just the maximum/minimum values of large groups or exceedances above/below a proper threshold value as required by the traditional extreme value analysis. This is of importance in practice.

The arrangement of the paper is as follows. In Section 2, we introduce the model and develop the Bayesian methodology for estimating the model parameters. A simulation study is carried out in Section 3. In Section 4, we apply the methodology developed here to the daily closing prices of major European stock indices during 1991-1998, and compare our approach with that proposed by Wei (2008) for this data set. Finally, some comments and further discussions are given in Section 5.

2. The Methodology

2.1. Multivariate models

We extend the bivariate quantile function of Gilchrist (2000) to general cases. A random vector $\mathbf{X} = (X_1, \ldots, X_{m_1})$ has a distribution that can be defined through quantile functions $Q_{x_k}(\tau, r_1, \ldots, r_{m_1-1})$ with

$$x_k = Q_{x_k}(\tau, r_1, \ldots, r_{m_1-1}), \quad k = 1, \ldots, m_1, \tag{2.1}$$

where $(\tau, r_1, \ldots, r_{m_1-1})$ is in the unit space $\Omega_u = \{(\tau, r_1, \ldots, r_{m_1-1}) \mid 0 \le \tau, r_k \le 1, k = 1, \ldots, m_1-1\}$, and the $Q_{x_k}(\tau, r_1, \ldots, r_{m_1-1})$ $(k = 1, \ldots, m_1)$ are such that, for any $(\tau, r_1, \ldots, r_{m_1}) \in \Omega_u$, there exists $(x_1, \ldots, x_{m_1}) \in R_1^{m_1}$. Accordingly, a random sample $(x_1, \ldots, x_{m_1})$ can be easily obtained by taking $(\tau, r_1, \ldots, r_{m_1-1})$ as a random sample of $(U, U_1, \ldots, U_{m_1-1})$, uniformly distributed in the unit space Ω_u. Similar to the bivariate case, for given values τ_0 and r_{k0}, $k = 1, \ldots, m_1 - 1$, we have

$$\begin{aligned} \tau_0 r_{10} \cdots r_{m_1-1\,0} &= P(U \le \tau_0,\ U_1 \le r_{10}, \ldots, U_{m_1-1} \le r_{m_1-1\,0}) \\ &= P((x_1, \ldots, x_{m_1}) \in A_{\tau_0 r_{10} \cdots r_{m_1-1\,0}}), \end{aligned}$$

where the region $A_{\tau_0 r_{10} \cdots r_{m_1-1\,0}}$ contains all possible values of $(x_1, \ldots, x_{m_1})$ corresponding to those τ and r_k such that $\tau \le \tau_0, r_k \le r_{k0}$ and $k = 1, \ldots, m_1 - 1$. Particularly, if $r_{k0} = 1$ for $k = 1, \ldots, m_1-1$, then $\tau_0 = P(U \le \tau_0) = P((x_1, \ldots, x_{m_1}) \in$

A_{τ_0}), where A_{τ_0} does not depend on r_k for $k = 1, \ldots, m_1 - 1$. The boundary of A_{τ_0} is called the τ_0th quantile surface of $\mathbf{X}$. We can define the r_{k0}th quantile surface of $\mathbf{X}$ for $k = 1, \ldots, m_1 - 1$ in the same way. The relationship between different quantile surfaces needs further investigation. In the rest of the paper, we assume that $\mathbf{X}$ is an m_1-dimensional continuous random variable.

In practice, for a given data set we need to estimate the τth quantile surface of $\mathbf{X}$, i.e. the boundary of A_τ. One way to achieve this is to find a continuous function, say $h(\mathbf{X})$, and a one-dimensional quantile function, say $Q(\tau)$ for $0 \leq \tau \leq 1$, such that $Q^{-1}(h(\mathbf{X})) = U \sim U(0,1)$, where Q^{-1} is the inverse function of Q. So $\tau = P(U \leq \tau) = P(Q^{-1}(h(\mathbf{X})) \leq \tau) = P(h(\mathbf{X}) \leq Q(\tau))$. Therefore, the boundary of $A_\tau = \{\mathbf{x} \mid h(\mathbf{x}) \leq Q(\tau)\}$ can be determined by $h(\mathbf{x}) = Q(\tau)$. Now the functions h and Q may contain parameters $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ that can be estimated from the given data. Note that $\mathbf{x} = (x_1, \ldots, x_{m_1})$. Therefore,

$$h(x_1, \ldots, x_{m_1}, \boldsymbol{\eta}) = Q(\tau, \boldsymbol{\gamma}) \tag{2.2}$$

defines a statistical model with $\boldsymbol{\beta} = (\boldsymbol{\eta}, \boldsymbol{\gamma})$ as the model parameter vector. We refer to model (2.2) as a multivariate quantile function model.

Some remarks follow.

(a) To simplify the notation, we use small letter x to stand for both a realization of the random variable X and a variable of the h function.

(b) Model (2.2) defines a m_1-dimensional parametric joint distribution function whose τth quantile surface is the boundary of $A_\tau = \{(x_1, \ldots, x_{m_1}) \mid h(x_1, \ldots, x_{m_1}, \boldsymbol{\eta}) \leq Q(\tau, \boldsymbol{\gamma})\}$. Thus (2.2) does not necessarily hold for any m_1-dimensional random variable.

(c) Model (2.2) uses a univariate continuous quantile function $Q(\tau, \boldsymbol{\gamma})$ so that estimation of the τth quantile surface of $\mathbf{X}$ is easier.

(d) If model (2.2) holds and h is a continuous function of $x_1, \ldots, x_{m_1}$, then the resulting quantile regions are unions of compact sets and the resulting quantile surfaces are nested within each other.

(e) Function h describes the basic shape of a quantile surface and one can explore the form of h for given data. For example, in the two dimensional cases, a scatter plot may be suggestive.

(f) Q describes the distribution of the random component of the model. Its construction is very flexible since one-dimensional quantile functions have some good properties. For example, they can be added and, under certain conditions, multiplied to form new quantile functions; proper transformations of quantile functions of simple distributions can result in quantile functions of complex distributions. These properties enable us to obtain a model for a distribution by combining simple component models. The consequence of

this quantile function modelling flexibility is that we can produce models in practice that are appropriate for a whole distribution including the tails. Working in this way with model (2.2) is much more appealing and intuitive than working with distribution functions.

Different h and Q lead to different fitted models. Gilchrist (2000) described several methods that can be used to identify the best fitted model for a given data set. Here we consider a special form of h given by

$$\sum_{k=1}^{m_1}\Big(x_k-\sum_{j=0}^{k-1}a_{kk-j-1}x_{k-j-1}\Big)^2=Q(\tau,\boldsymbol{\gamma}), \tag{2.3}$$

where $x_0=1$ and $Q(\tau,\boldsymbol{\gamma})$ is a non-negative one-dimensional quantile function. The model parameters are $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_{m_2})$ and $\boldsymbol{\eta}=(a_{10},a_{20},a_{21},\ldots,a_{m_1\,0},\ldots,$ $a_{m_1\,m_1-1})$. Note that (2.3) is also restrictive. However, such a model is suitable for clustered data and has the estimated quantile surfaces sharing the same shape but differing in scale. We find the model to be useful in practice.

2.2. MCMC method

Let $f(u,\boldsymbol{\gamma})$ be the probability density function corresponding to the quantile function $Q(\tau,\boldsymbol{\gamma})$, let $\mathbf{x}=\{(x_{1i},\ldots,x_{m_1 i}),i=1,\ldots,n\}$ be a set of independent samples of size n, and let $u_i=\sum_{k=1}^{m_1}\Big(x_{ki}-\sum_{j=0}^{k-1}a_{kk-j-1}x_{k-j-1\,i}\Big)^2$, $i=1,\ldots,n$. Then it follows from (2.3) that the likelihood of $\mathbf{u}=(u_1,\ldots,u_n)$ is given by $L(\mathbf{u}\mid\boldsymbol{\beta})=f(u_1,\boldsymbol{\gamma})\cdots f(u_n,\boldsymbol{\gamma})$. We write this as $L(\mathbf{x}\mid\boldsymbol{\beta})$. It is seen that each u_i depends on $\boldsymbol{\eta}$ and each u_i corresponds to a value τ_i such that

$$u_i=Q(\tau_i,\boldsymbol{\gamma}). \tag{2.4}$$

So τ_i depends on $\boldsymbol{\beta}$ implicitly and the likelihood can be further written as

$$L(\mathbf{x}\mid\boldsymbol{\beta})=f_\tau(\tau_1,\boldsymbol{\beta})f_\tau(\tau_2,\boldsymbol{\beta})\cdots f_\tau(\tau_n,\boldsymbol{\beta}),$$

where $f_\tau(\tau_i,\boldsymbol{\beta})=f(Q(\tau_i,\boldsymbol{\gamma}),\boldsymbol{\gamma})$.

Theorem 1. *For any $0\le\tau\le1$, we have that $f_\tau(\tau,\boldsymbol{\beta})=(\partial Q(\tau,\boldsymbol{\gamma})/\partial\tau)^{-1}$.*

Proof. If $u=Q(\tau,\boldsymbol{\gamma})$, then $\tau=F(u,\boldsymbol{\gamma})$, where F is the inverse function of Q. It follows from $(\partial u/\partial\tau)(\partial\tau/\partial u)=1$ that the result holds.

Therefore, once $Q(\tau,\boldsymbol{\gamma})$ is known, the likelihood can be calculated. However, it is important to remember that τ_i is determined by (2.4). In general cases, a numerical method is required to find τ_i; here we use the Newton-Raphson method as it is good enough for our purposes.

Table 2.1. A General Random walk sampler for multivariate quantile function models

```
Sample
    η′ ∼ g1(η′) such that η′ ∈ Ω1 and γ′ ∼ g2(γ′) such that γ′ ∈ Ω2
Solve
    Σ_{k=1}^{m1} ( x_ki − Σ_{j=0}^{k−1} a′_kj x_{k−j−1 i} )^2 = Q(τ′_i, γ′) for τ′_i, i = 1, ..., n.
Sample p ∼ U(0,1). If p ≤ min{ABC, 1}, accept β′.
```

If $\pi(\boldsymbol{\beta}) = \pi(\boldsymbol{\eta})\pi(\boldsymbol{\gamma})$ is a prior distribution of the parameters, the posterior distribution of $\boldsymbol{\beta}$ is $\pi(\boldsymbol{\beta} \mid \mathbf{x}) \propto L(\mathbf{x} \mid \boldsymbol{\beta})\pi(\boldsymbol{\beta}) = L(\mathbf{x} \mid \boldsymbol{\beta})\pi(\boldsymbol{\eta})\pi(\boldsymbol{\gamma})$.

Suppose the posterior distribution of $\boldsymbol{\beta}$ is well defined on $\Omega = \Omega_1 \times \Omega_2$, where $\boldsymbol{\eta} \in \Omega_1$ and $\boldsymbol{\gamma} \in \Omega_2$. It would be difficult to design a Gibbs sampler to estimate $\boldsymbol{\beta}$ in this case because τ_i depends on $\boldsymbol{\beta}$ only implicitly, hence the full conditional distributions are not available. We propose a random walk sampler for the parameter estimation. Specifically, let $\boldsymbol{\beta}$ be the current parameter value and τ_i, $i = 1, \ldots, n$, the corresponding probabilities associated with $\boldsymbol{\beta}$. Let $\boldsymbol{\beta}'$ be the proposed value and τ_i' the associated probabilities. Then a general random walk MCMC sampler for multivariate quantile function models can be as given in Table 2.1 with

$$A = \frac{\pi(\boldsymbol{\beta}' \mid \mathbf{x})}{\pi(\boldsymbol{\beta} \mid \mathbf{x})} = \frac{L(\mathbf{x} \mid \boldsymbol{\beta}')\pi(\boldsymbol{\eta}')\pi(\boldsymbol{\gamma}')}{L(\mathbf{x} \mid \boldsymbol{\beta})\pi(\boldsymbol{\eta})\pi(\boldsymbol{\gamma})}, \quad B = \frac{q(\boldsymbol{\eta}' \to \boldsymbol{\eta})}{q(\boldsymbol{\eta} \to \boldsymbol{\eta}')}, \quad C = \frac{q(\boldsymbol{\gamma}' \to \boldsymbol{\gamma})}{q(\boldsymbol{\gamma} \to \boldsymbol{\gamma}')},$$

where $q(a' \to a)$ is the transition probability density function of a given a', and g_1 and g_2 are the probability density functions from which the proposal $\boldsymbol{\beta}'$ is obtained.

Theorem 2. *In Table 2.1, let* $g_1(\boldsymbol{\eta}') = \prod_{k=1}^{m_1}\prod_{j=0}^{k-1}(2\pi\sigma_{a_{kj}}^2)^{-1/2}\exp\{-(a_{kj}' - a_{kj})^2/(2\sigma_{a_{kj}}^2)\}$, *and* $g_2(\boldsymbol{\gamma}') = \prod_{\ell=1}^{m_2}(2\pi\sigma_{\gamma_\ell}^2)^{-1/2}\exp\{-(\gamma_\ell' - \gamma_\ell)^2/(2\sigma_{\gamma_\ell}^2)\}$, *where* $\boldsymbol{\eta}' \in \Omega_1$, $\boldsymbol{\gamma}' \in \Omega_2$ *and* $\sigma_{a_{kj}}$, σ_{γ_ℓ} *are given values. Then*

$$B = \frac{\int_{\Omega_1}\prod_{k=1}^{m_1}\prod_{j=0}^{k-1}(\sqrt{2\pi}\sigma_{a_{kj}})^{-1}\exp\{-(a_{kj}' - a_{kj})^2/(2\sigma_{a_{kj}}^2)\}da_{kj}'}{\int_{\Omega_1}\prod_{k=1}^{m_1}\prod_{j=0}^{k-1}(\sqrt{2\pi}\sigma_{a_{kj}})^{-1}\exp\{-(a_{kj} - a_{kj}')^2/(2\sigma_{a_{kj}}^2)\}da_{kj}},$$

$$C = \frac{\int_{\Omega_2}\prod_{\ell=1}^{m_2}(\sqrt{2\pi}\sigma_{\gamma_\ell})^{-1}\exp\{-(\gamma_\ell' - \gamma_\ell)^2/(2\sigma_{\gamma_\ell}^2)\}d\gamma_\ell'}{\int_{\Omega_2}\prod_{\ell=1}^{m_2}(\sqrt{2\pi}\sigma_{\gamma_\ell})^{-1}\exp\{-(\gamma_\ell - \gamma_\ell')^2(2\sigma_{\gamma_\ell}^2)\}d\gamma_\ell}.$$

Proof. The above choice of $g_1(\boldsymbol{\eta}')$ means that the proposals a_{kj}', $k = 1, \ldots, m_1$, $j = 0, \ldots, k-1$, are obtained from $N(a_{kj}, \sigma_{a_{kj}}^2)$ independently such that $\boldsymbol{\eta}' \in \Omega_1$. Therefore it follows from

$$q(\boldsymbol{\eta}' \to \boldsymbol{\eta}) = \frac{\prod_{k=1}^{m_1}\prod_{j=0}^{k-1}(2\pi\sigma_{a_{kj}}^2)^{-1/2}\exp\{-(a_{kj} - a_{kj}')^2/(2\sigma_{a_{kj}}^2)\}}{\int_{\Omega_1}\prod_{k=1}^{m_1}\prod_{j=0}^{k-1}(2\pi\sigma_{a_{kj}}^2)^{-1/2}\exp\{-(a_{kj} - a_{kj}')^2/(2\sigma_{a_{kj}}^2)\}da_{kj}}$$

that B holds. Finally, C can be derived in the same way.

Note that, generally, the values of B and C can be difficult to calculate because the shapes of Ω_1 and Ω_2 can be very complicated in multivariate cases. We propose to use a simulation method to estimate the values of the integrals involved in B and C if necessary. For example, to estimate

$$\int_{\Omega_1} \prod_{k=1}^{m_1} \prod_{j=0}^{k-1} (2\pi\sigma_{a_{kj}}^2)^{-1/2} \exp\left\{-\frac{(a'_{kj} - a_{kj})^2}{2\sigma_{a_{kj}}^2}\right\} da'_{kj},$$

we simulate $a_{kj}^v \sim N(a_{kj}, \sigma_{a_{kj}}^2)$, $k = 1, \ldots, m_1$, $j = 0, \ldots, k-1$, for $v = 1, \ldots, M_1$. Let N_1 be the number of the simulated samples such that $\boldsymbol{\eta}^v \in \Omega_1$. Then the integral is estimated by N_1/M_1.

2.3. Bivariate models

To apply our methodology, we consider the case where $m_1 = 2$ and $Q(\tau, \boldsymbol{\gamma}) = \tau^{\gamma_1}/(1-\tau)^{\gamma_2}$, $\gamma_1 > 0$, $\gamma_2 > 0$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$. This quantile function is the product of a power quantile function and a Pareto quantile function. This special choice is made since (a) it has few applications in the literature due to the fact that its inverse function is not available, though it is a natural generalization of the Pareto distribution that is widely used in extreme value analysis, and (b) it is able to describe both the main body and the tails of a distribution. Here (2.3) can be rewritten as

$$(x_2 - a_{21}x_1 - a_{20})^2 + (x_1 - a_{10})^2 = \tfrac{\tau^{\gamma_1}}{(1-\tau)^{\gamma_2}}, \tag{2.5}$$

where $0 < \epsilon \le \tau < 1$, and ϵ is a fixed small positive number. This last requirement does not affect applications because such an ϵ value always exists for a finite data set. Theoretically, it guarantees that the posterior distribution of the parameters is well-defined.

Theorem 3. *Given* (2.5)*, the likelihood of an independent sample of size n is*

$$L(\mathbf{x} \mid \boldsymbol{\beta}) = \prod_{i=1}^{n} f_\tau(\tau_i, \boldsymbol{\gamma}) = \prod_{i=1}^{n} \frac{\tau_i^{1-\gamma_1}(1-\tau_i)^{1+\gamma_2}}{\gamma_1(1-\tau_i) + \gamma_2\tau_i}, \tag{2.6}$$

where the τ_i satisfy

$$(x_{2i} - a_{21}x_{1i} - a_{20})^2 + (x_{1i} - a_{10})^2 = \tau_i^{\gamma_1}(1-\tau_i)^{-\gamma_2}, \qquad i = 1, \ldots, n. \tag{2.7}$$

Furthermore, if

$$\begin{aligned} \pi(\boldsymbol{\eta}) &= \prod_{k=1}^{m_1} \prod_{j=0}^{k-1} \pi(a_{kj}) = \prod_{k=1}^{m_1} \prod_{j=0}^{k-1} (2\pi\sigma_{kj}^2)^{-1/2}\, e^{-a_{kj}^2/(2\sigma_{kj}^2)}, \\ \pi(\boldsymbol{\gamma}) &= \prod_{\ell=1}^{2} \pi(\gamma_\ell) = \prod_{\ell=1}^{2} \lambda_\ell \gamma_\ell^{-2} e^{-\lambda_\ell/\gamma_\ell}, \end{aligned} \tag{2.8}$$

then the posterior distribution of the parameters is given by

$$\pi(\boldsymbol{\beta} \mid \mathbf{x}) \propto \prod_{i=1}^{n} \frac{\tau_i^{1-\gamma_1}(1-\tau_i)^{1+\gamma_2}}{\gamma_1(1-\tau_i)+\gamma_2\tau_i} \prod_{\ell=1}^{2} \lambda_\ell \gamma_\ell^{-2} e^{-\lambda_\ell/\gamma_\ell} \prod_{k=1}^{2}\prod_{j=0}^{k-1} \frac{1}{\sqrt{2\pi}\sigma_{kj}}\, e^{-a_{kj}^2/2\sigma_{kj}^2}, \tag{2.9}$$

and is well-defined on $\Omega_1 \times \Omega_2$*, where* $\Omega_1 = (-\infty,\infty)^3$*,* $\Omega_2 = (0, M] \times (0,\infty)$*, and* M *is any fixed positive real number.*

Proof. It follows from $\partial Q(\tau,\boldsymbol{\gamma})/\partial\tau = \tau^{\gamma_1-1}(1-\tau)^{-\gamma_2-1}[\gamma_1(1-\tau)+\gamma_2\tau]$, and $f_\tau(\tau_i,\boldsymbol{\gamma}) = \tau_i^{1-\gamma_1}(1-\tau_i)^{1+\gamma_2}/(\gamma_1(1-\tau_i)+\gamma_2\tau_i)$ that (2.6) holds, where τ_i satisfy (2.7). Therefore, it follows from $\pi(\boldsymbol{\beta} \mid \mathbf{x}) \propto L(\mathbf{x} \mid \boldsymbol{\beta})\pi(\boldsymbol{\eta})\pi(\boldsymbol{\gamma})$ that (2.9) holds.

Note that (2.5) is well-defined for any $\boldsymbol{\eta}$. Hence $\Omega_1 = (-\infty,\infty)^3$. Let $\Omega_2 = (0, M] \times (0,\infty)$. Since τ_i satisfies (2.7), we have $\tau_i \geq \epsilon$, so

$$\begin{aligned}
\int_{\Omega_1\times\Omega_2} \pi(\boldsymbol{\beta} \mid \mathbf{x})d\boldsymbol{\beta} &= \int_{\Omega_1\times\Omega_2} L(\mathbf{x},\mathbf{y} \mid \boldsymbol{\beta})\pi(\gamma_1)\pi(\gamma_2) \prod_{k=1}^{2}\prod_{j=0}^{k-1} \pi(a_{kj})da_{kj}d\gamma_1 d\gamma_2 \\
&\leq \int_{\Omega_1\times\Omega_2} \frac{(\tau_1\cdots\tau_n)^{1-\gamma_1}\pi(\gamma_1)\pi(\gamma_2)}{\gamma_2^n\tau_1\cdots\tau_n} \prod_{k=1}^{2}\prod_{j=0}^{k-1} \pi(a_{kj})da_{kj}d\gamma_1 d\gamma_2 \\
&\leq \frac{1}{\epsilon^{nM}} \int_{\Omega_1\times\Omega_2} \frac{\pi(\gamma_2)}{\gamma_2^n}\pi(\gamma_1) \prod_{k=1}^{2}\prod_{j=0}^{k-1} \pi(a_{kj})da_{kj}d\gamma_2 d\gamma_1 \\
&\leq \frac{1}{\epsilon^{nM}} \int_0^\infty \frac{\lambda_2}{\gamma_2^{n+2}} e^{-\lambda_2/\gamma_2} d\gamma_2 < \infty,
\end{aligned}$$

as required. Therefore, the posterior density function is well-defined on $\Omega_1 \times \Omega_2$.

Because M can be arbitrarily large and because the estimate of γ_1 is always finite, the MCMC sampler can be applied to this model with $B = 1$, and

$$\begin{aligned}
A &= \frac{\prod_{i=1}^{n} \frac{(\tau_i')^{1-\gamma_1'}(1-\tau_i')^{1+\gamma_2'}}{\gamma_1'(1-\tau_i')+\gamma_2'\tau_i'}}{\prod_{i=1}^{n} \frac{\tau_i^{1-\gamma_1}(1-\tau_i)^{1+\gamma_2}}{\gamma_1(1-\tau_i)+\gamma_2\tau_i}} \frac{\prod_{\ell=1}^{2} \frac{\lambda_\ell}{(\gamma_\ell')^2} e^{-\lambda_\ell/\gamma_\ell'}}{\prod_{\ell=1}^{2} \frac{\lambda_\ell}{\gamma_\ell^2} e^{-\lambda_\ell/\gamma_\ell}} \frac{\prod_{k=1}^{2}\prod_{j=0}^{k-1} \frac{1}{\sqrt{2\pi}\sigma_{kj}} e^{-{a_{kj}'}^2/2\sigma_{kj}^2}}{\prod_{k=1}^{2}\prod_{j=0}^{k-1} \frac{1}{\sqrt{2\pi}\sigma_{kj}} e^{-{a_{kj}}^2/2\sigma_{kj}^2}}, \\
C &= \frac{\Phi((M-\gamma_1)/\sigma_{\gamma_1}) - \Phi(-\gamma_1/\sigma_{\gamma_1})}{\Phi((M-\gamma_1')/\sigma_{\gamma_1}) - \Phi(-\gamma_1'/\sigma_{\gamma_1})} \frac{1-\Phi(-\gamma_2/\sigma_{\gamma_2})}{1-\Phi(-\gamma_2'/\sigma_{\gamma_2})},
\end{aligned}$$

where the Φ terms containing M can be approximated by 1 if M is large enough.

Note that any proper distributions can be used for $\pi(\boldsymbol{\eta})$ and $\pi(\gamma_1)$ because they do not affect the appropriateness of the posterior distribution, but $\pi(\gamma_2)$ needs to be chosen carefully such that the resulting posterior distribution is proper on Ω. Also note that large (small) values of σ_{kj} and λ_ℓ indicate weak (strong) prior information on the parameters. Our experience with the sampler

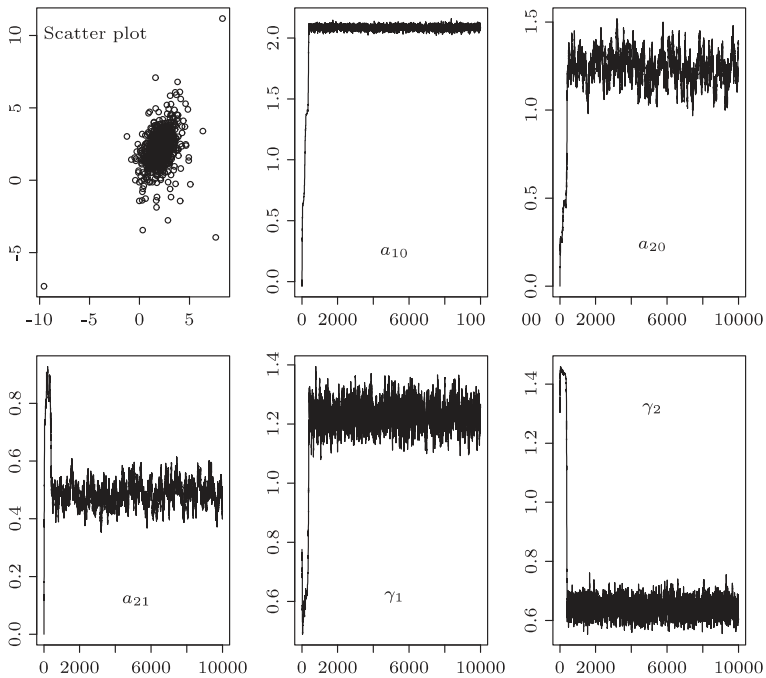

Figure 3.1. Scatter plot of the simulated data and the time series plots of the simulated parameter values.

shows that those values do not play an important role in the convergence of the method. In our study, we chose them to be large.

3. Simulation Study

The model we considered for simulation is

$$(x_2 - 0.5x_1 - 1.2)^2 + (x_1 - 2.1)^2 = \frac{\tau^{1.3}}{(1-\tau)^{0.6}}. \tag{3.1}$$

Figure 3.1 shows the scatter plot of a sample of size 1,000 simulated from (3.1); the simulated data is clustered and with some extreme values. These features are found in many data sets for which a standard regression model may not be appropriate.

We applied our methodology to the simulated data in the hope that we would recover model (3.1). Assuming not much information on the true parameter values, we randomly chose $\sigma_{kj} = 5$ for $k = 1, 2$ and $j = 0, 1$, and $\lambda_\ell = 1$ for $\ell = 1, 2$. The initial values should be in the support of the posterior distribution, hence we also rather arbitrarily assigned $(a_{10}, a_{20}, a_{21}, \gamma_1, \gamma_2) = (0, 0, 0, 0.776, 1.330)$ as the initial values, where γ_1 and γ_2 were sampled from their prior distributions,

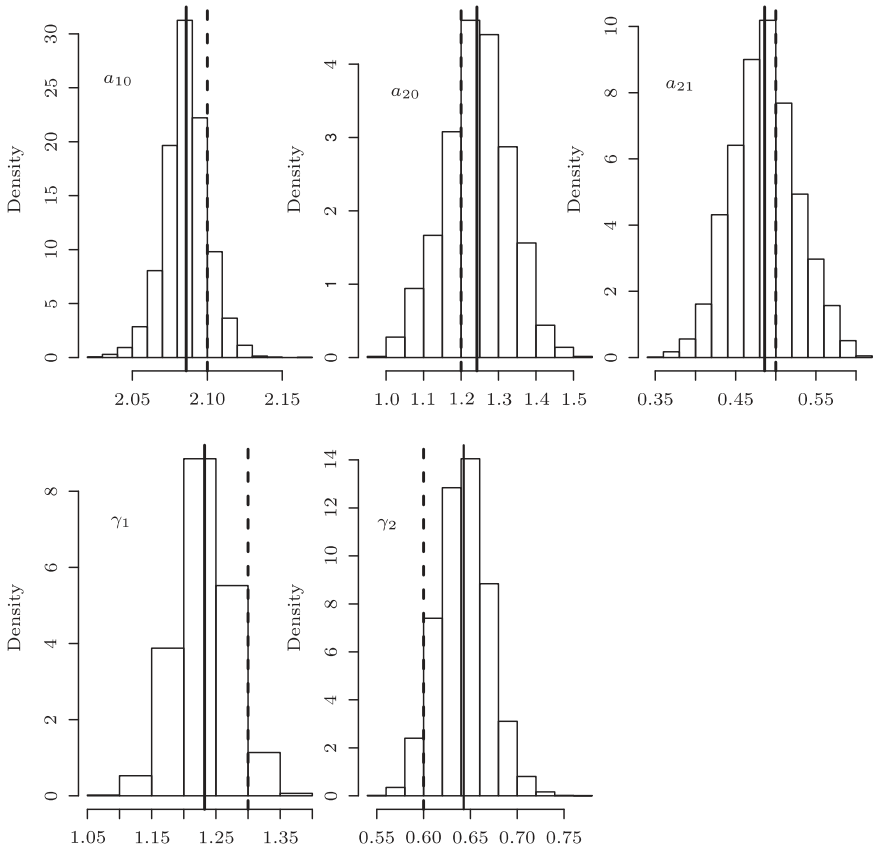

Figure 3.2. The histograms of the samples collected from the MCMC sampler.

respectively. For other initial values and prior information we obtained similar results.

We ran the MCMC algorithm 10,000 steps and save the parameter values every 10 steps. The time series plots of the simulated parameter values are given in Figure 3.1, they show that the simulated Markov chain converged very quickly to its equilibrium (posterior) distribution. After burn-in period (we took the first 1,000 steps), we used the sample means as the estimate of the model parameters. The fitted model was given by $(x_2 - 0.486x_1 - 1.242)^2 + (x_1 - 2.086)^2 = \tau^{1.232}/\ (1-\tau)^{0.643}$. Figure 3.2 shows the histograms of the collected samples of the parameters after burn-in, where the solid vertical lines represent the locations of the sample estimates of the parameters and the dashed vertical lines the true parameter values; the true parameter values are well within the posterior marginals.

Figure 3.3 compares the fitted quantile curves (dashed) and the true quantile

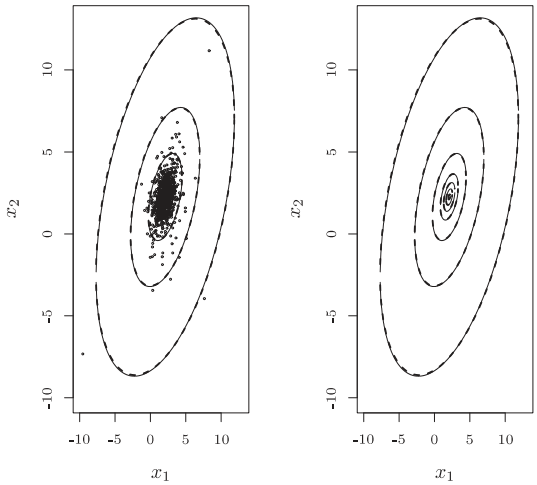

Figure 3.3. The fitted (dashed) and the true (continuous) quantile curves for the simulation study.

curves (continuous). From inside to outside, the quantile curves correspond to $\tau = 0.05$, 0.25, 0.5, 0.75, 0.95, 0.995, and 0.9995 respectively. The left panel of Figure 3.3 also shows the simulated data. All the results show that the fitted and the true quantile curves are almost the same.

Note that the parameters in this study were estimated under the correct model. In practice, an initial data exploration may suggest several possible models for the same data set, and hence we may end up with several competitive models corresponding to different h and Q functions. Gilchrist (2000) discussed several model diagnostic methods which can be used to identify the best fitted model for the same data set. We do not show the diagnostic plots for this simulation study to save space.

4. Applications to Data

We consider the daily closing prices of major European stock indices in the period of 1991-1998, specifically, the Germany DAX and the Switzerland SMI. The Germany DAX is the most commonly cited benchmark for measuring the returns posted by stocks on the Frankfurt Stock Exchange. It is a performance-based index, which means that any dividends and other events are rolled into the index's final calculation. The Switzerland SMI is Switzerland's key equity index. It represents about 85% of the free-float capitalization of the Swiss equity market. The time series plots of the daily closing prices are given in Figure 4.4 (a) and (b), which show that the time series are not stationary. Let $\tilde{x}_t$ be the SMI, and $\tilde{y}_t$ the DAX. Consider the log returns: $x_{1t} = 100 * (\log(\tilde{x}_{t+1}) - \log(\tilde{x}_t))$, and

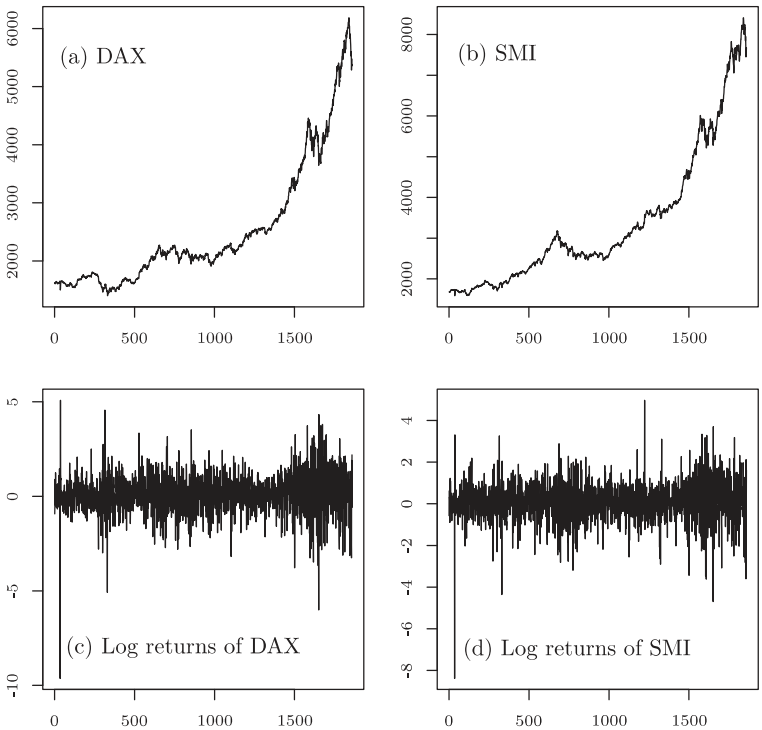

Figure 4.4. The time series plots of the daily closing prices of major European stock indices and the corresponding log returns.

$x_{2t} = 100 * (\log(\tilde{y}_{t+1}) - \log(\tilde{y}_t))$, where the factor 100 is for convenience. Figure 4.4 (c) and (d) show the time series plots of the log returns. It appears that the log return series is stationary but with some extreme values. We further checked the autocorrelation and partial autocorrelation function plots of the log return series and found that there is almost no autocorrelation structure remaining in the return series. In the following we ignore the autocorrelation structure and concentrate on the relation between the two log returns. The scatter plot of the log return data is given in Figure 4.5 and shows an obvious positive relation between the two log returns.

As the data are clustered with many points away from the center of the cluster, it is not appropriate to fit a linear regression model to the data. The joint distribution of the two returns would be more appropriate, so we fit a quantile function model to the data in order to estimate the probability of (x_1, x_2) falling in a certain region of interest.

The prior distributions of the parameters are given by (2.8) with $\sigma_{kj} = 13$, $k = 1, 2$, $j = 0, \ldots, k-1$, and $\lambda_\ell = 1$, $\ell = 1, 2$. The chain was run 50,000 steps and samples were collected every 30 steps. Time series plots of the simulated

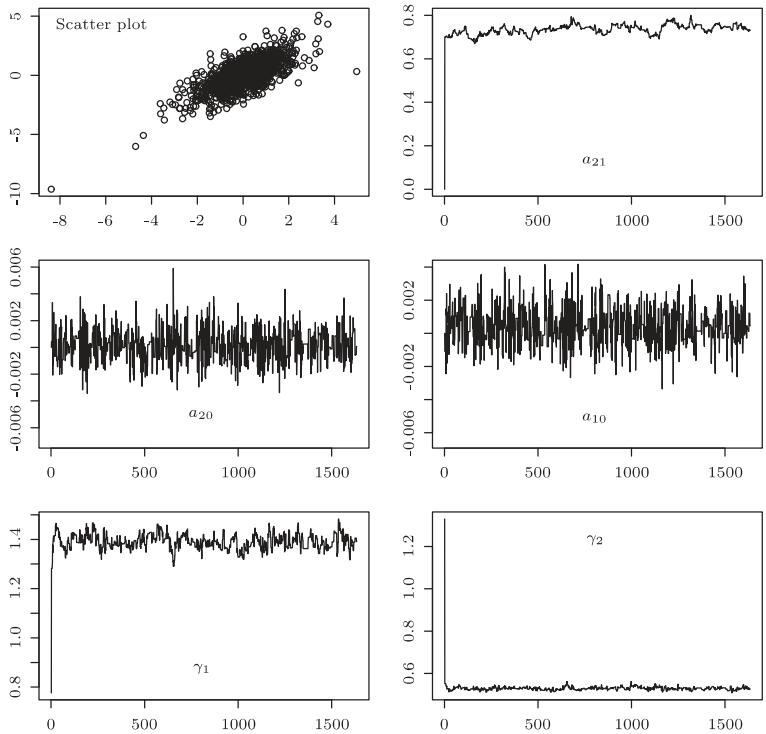

Figure 4.5. Scatter plot of the two log return data and the time series plots of the simulated parameter values of model (2.5).

parameter values (also see Figure 4.5) indicate that the simulated Markov chain converges to its equilibrium distribution after 3,000 steps. The fitted model is

$$(x_2 - 0.7370x_1 - 0.0001)^2 + (x_1 - 0.0004)^2 = \frac{\tau^{1.3908}}{(1-\tau)^{0.5290}}. \tag{4.1}$$

If the fitted model is good, then we would expect that the points on the plot of the quantiles of the residuals $(x_{2i} - 0.7370x_{1i} - 0.0001)^2 + (x_{1i} - 0.0004)^2$, $i = 1, \ldots, n$, against the quantiles of the power-Pareto distribution $Q(\tau, \boldsymbol{\gamma}) = \tau^{1.3908}/(1-\tau)^{0.5290}$ should be roughly along a straight line (Gilchrist (2000)). This plot is given in Figure 4.6(a), confirming that the fitted model is very good.

The fitted τth quantile curve together with the observed data are shown in Figure 4.6(b), where $\tau = 0.995, 0.95, 0.90, 0.75, 0.5, 0.25$, and 0.05. To make it clearer, we also reproduced quantile curves separately in Figure 4.6(c), where darker curves are the quantile curves, and the lighter curves are the contours obtained by fitting a bivariate normal distribution to the data using R. The fitted normal distribution has mean vector $(0.0818, 0.0652)^\top$ and variance-covariance matrix $S = (s_{ij})_{2\times 2}$, where $s_{11} = 0.8552$, $s_{12} = s_{21} = 0.6696$, $s_{22} = 1.0605$. Although the plot of the contour curves of the bivariate normal density function

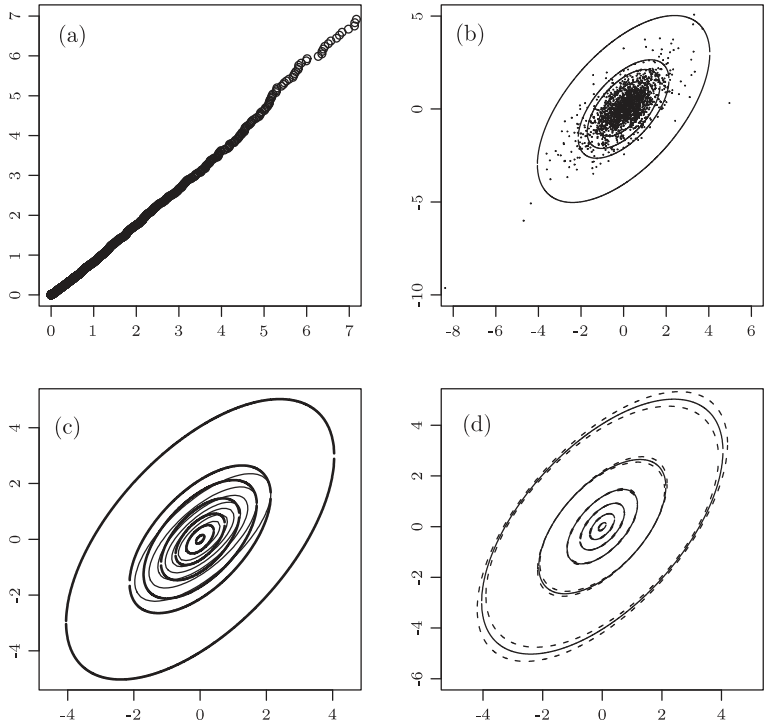

Figure 4.6. QQ-plot, fitted quantile curves and corresponding 95% credible intervals.

Table 4.2. Percentage of points which are within each fitted quantile curves

Fitted quantiles	5%	25%	50%	75%	95%	99.5%
Number of points	90	376	828	1306	1737	1853
Percentage	4.84%	20.2%	55.5%	70.3%	93.4%	99.7%

looks similar to that of the quantile curves, they have completely different interpretations. With a τth quantile curve we mean that the probability of the two log returns falling inside the region defined by the τth quantile curve is τ, while with a level c contour curve we mean that for any point in the region defined by the level c contour curve the value of the joint probability density function is c. Figure 4.6(d) shows a 95% credible interval (dashed curves) for each fitted quantile curve (solid curve). It is seen that reasonable variations of the fitted quantile curves are achieved.

Table 4.2 further shows the percentage of the observed points which fall inside each quantile curve. Clearly the coverage of the fitted quantile curves is in good agreement with the observed data.

As our approach is different from the quantile regression approach that is

commonly used, we compared our method with that of Wei (2008). Again, R was used to obtain reference quantile contours (Wei (2008)) for this data set. Specially, the R library COBS was used to obtain the constrained smoothing function. COBS is based on the method developed by He and Ng (1999). In using it, the main tuning parameter values we used were as follows: the number of internal knots is 20 chosen by the default method of COBS; the degree of the splines is 2; the penalty parameter is chosen by a Schwarz-type information criterion. Figure 4.7 shows the fitted quantile curves by using different methods for $\tau = 0.995, 0.95, 0.5, 0.05$. It is seen that there is no restriction on the shape of the fitted reference quantile contours estimated by Wei's method due to the non-parametric nature of the method. When τ is not too large or too small, both methods provide similar results, but when τ goes to extremes, we encountered problems with Wei's method. In fact, when $\tau > 0.995$, we failed to obtain the τth reference quantile contour by using R. On the other hand, unlike Wei's method, the performance of the proposed method does depend on the choice of Q. We have seen that model (2.5) performs very well for this data set, but it may not be appropriate for normally distributed data or some other data due to the specific function Q used in the model. Because of these reasons, we suggest using both methods in practice. Further comparisons between the two approaches are certainly required in the future.

5. Comments and Conclusions

The proposed multivariate quantile function model provides a flexible way to study multivariate quantiles, and the Bayesian approach developed here make it straightforward to estimate the model parameters and to deal with real data sets.

Different types of quantile function models can be developed by constructing different functions h and Q according to different features in a data set. For example, if the data consist of several clusters, then a mixture quantile function model might be considered. The main complexities involved in such a generalization are in the appropriateness of the posterior distribution function of the parameters, but dealing with mixture models via quantile functions is much easier than that via distribution or density functions. Further research on this issue will be reported elsewhere.

Finally, Mu and He (2007) studied the effect of power transformations on conditional quantile modelling under the framework of Koenker (2005). We believe that power transformations on the variables may also help in constructing appropriate multivariate quantile function models. Further investigations are planned.

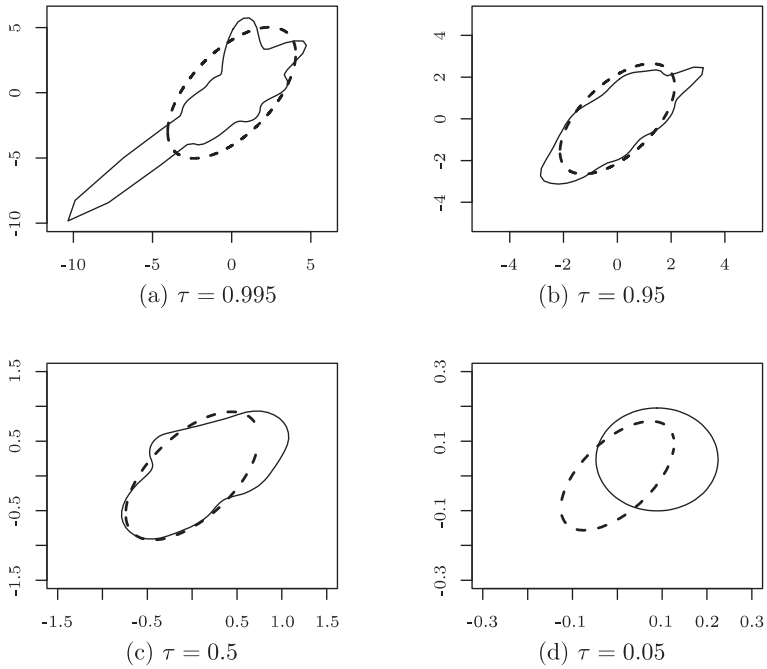

Figure 4.7. Fitted quantile curves by using our method (dashed curves) and fitted reference quantile contours by using Wei's (2008) method (continuous curves).

Acknowledgement

We would like to express our sincere thanks to an associate editor, a referee, and Professor W.G. Gilchrist for their constructive comments and suggestions which have greatly enhanced the quality and presentation of the paper.

References

Brown, B. M. and Hettmansperger, T. P. (1987). Affine invariant rank methods in the bivariate location model. *J. Roy. Statist. Soc. Ser. B* **49**, 301-310.

Brown, B. M. and Hettmansperger, T. P. (1989). An affine invariant bivariate versions of the sign test. *J. Roy. Statist. Soc. Ser. B* **51**, 117-125.

Cai, Y. (2007). A quantile approach to US GNP. *Economic Modelling* **24**, 969-979.

Cai, Y. and Stander, J. (2008). Quantile self-exciting threshold autoregressive time series models. *J. Time Ser. Anal.* **29**, 186-202.

Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *The Institute of Statistical Mathematics* **53**, 380-403.

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

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20**, 1803-1827.

Dunson, D. B., Watson, M. and Taylor, J. A. (2003). Bayesian latent variable models for median regression on multiple outcomes. *Biometrics* **56**, 296-304.

Eddy, W. F. (1985). Ordering of multivariate data. *Computer Science and Statistics: The Interface.* (Edit by L. Billard), North-Holland, Amsterdam, 25-30.

Gilchrist, W. G. (2000). *Statistical Modelling with Quantile Functions.* Chapman & Hall/CRC.

He, X. and Ng, P. (1999). COBS: Constrained smoothing via linear programming. *Computat. Statist.* **14**, 315-337.

He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25**, 495-504.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Koenker, R. and D'Orey, V. (1987). Computing regression quantiles. *Applied Statistics* **36**, 383-393.

Koenker, R. and D'Orey, V. (1994). A remark on Algorithm AS229: Computing dual regression quantiles and regression rank scores. *Applied Statistics* **43**, 410-414.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *J. Amer. Statist. Assoc.* **96**, 1458-1468.

Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* **27**, 783-858.

Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *J. Amer. Statist. Assoc.* **102**, 269-279.

Oja, H. (1983). Descriptive statistics for multivariate trimming. *Statist. Probab. Lett.* **1**, 327-332.

Salibian-Barrera, M. and Wei, Y. (2008). Weighted quantile regression with non-elliptically structured covariates. *Canad. J. Statist.* 36, 595-611.

Schennach, S. M. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika* **92**, 31-46.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica* **56**, 214-232.

Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *J. Amer. Statist. Assoc.* **103**, 397-409.

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

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28**, 461-482.

School of Computing and Mathematics, University of Plymouth, Plymouth, PL4 8AA, United Kingdom.

E-mail: ycai@plymouth.ac.uk

(Received May 2008; accepted February 2009)