doi:10.1111/j.1467-9892.2007.00551.x

QUANTILE SELF-EXCITING THRESHOLD AUTOREGRESSIVE TIME SERIES MODELS

By Yuzhi Cai and Julian Stander

School of Mathematics and Statistics, University of Plymouth

First Version received September 2005

Abstract. In this paper we present a Bayesian approach to quantile self-exciting threshold autoregressive time series models. The simulation work shows that the method can deal very well with nonstationary time series with very large, but not necessarily symmetric, variations. The methodology has also been applied to the growth rate of US real GNP data and some interesting results have been obtained.

Keywords. Bayesian methods; MCMC; quantile SETAR model; simulation; US GNP.

1. INTRODUCTION

Generally speaking, there are two different types of quantile approach to statistical modelling. One is parametric proposed by Gilchrist (2000), while the other is semiparametric suggested by Koenker and Bassett (1978).

Quantile regression is an important statistical technique that offers a mechanism for estimating models for the conditional median and the full range of the other conditional quantiles. Compared with the usual estimation of the conditional mean function, quantile regression is capable of providing a more complete statistical analysis of the stochastic relationships among random variables.

Quantile regression techniques have been applied in many areas, such as finance and economics, medicine, survival analysis and environment modelling. A good review of quantile regression with some applications can be found in Yu *et al.* (2003).

Some work on quantile methods for autoregressive time series models can be found in the literature. For example, Weiss (1987) discussed how to estimate nonlinear dynamic models using least absolute error estimation, and Koenker and Xiao (2004) studied statistical inference in quantile autoregression models when the largest autoregressive coefficient may be unity. Koenker and Xiao (2005) considered quantile autoregression models in which the autoregressive coefficients can be expressed as monotone functions of a single, scalar random variable. However, to the authors' knowledge, no paper can be found in the literature on quantile regression for self-exciting threshold autoregressive time series (SETAR) and other smoothed version of nonlinear autoregressive time series such as smooth transition autoregressive time series models. In this paper we first define a quantile SETAR (QSETAR) model, after which we present a Baysian approach for performing inference.

The arrangement of the paper is as follows. In Section 2 we define a QSETAR model. Details about the Bayesian approach are given in Section 3, and simulation studies are presented in Section 4. In Section 5, we discuss an interesting application of our methodology to the growth rate of US real GNP from the first quarter of 1947 to the first quarter of 1991. We finish with some comments and conclusions in Section 6.

2. THE MODEL

Let ξ_{it} be a sequence of independent and identically distributed (iid) Gaussian random variables with mean zero and variance σ_i^2, let $-\infty = r_0 < r_1 < \cdots < r_{m+1} = \infty$ be threshold values, and let $\Omega_i = (r_{i-1}, r_i]$, $i = 1, \ldots, m$, and $\Omega_{m+1} = (r_m, r_{m+1})$. A SETAR model is defined by:

$$x_t = \sum_{i=1}^{m+1} \left(\beta_{i0} + \beta_{i1} x_{t-1} + \cdots + \beta_{ip} x_{t-p} + \xi_{it}\right) I_{[x_{t-d} \in \Omega_i]}, \tag{1}$$

where β_{ij}, $\sigma_i > 0$ and $d > 0$ are the parameters of the model, with d being called the delay parameter. The order of the model is p, and $I_{[x \in \Omega]}$ is an indication function of the event $x \in \Omega$.

This model was first proposed by Tong and Lim (1980). It is interesting to note that the stationarity of x_t in model (1) does not require all of the roots of $1 - \beta_{i1} B - \cdots - \beta_{ip} B^p = 0$ to be outside of the unit circle. For necessary and sufficient conditions for stationarity of simple SETAR models, see Petruccelli and Woolford (1984) and Chen and Tsay (1991).

The SETAR model has been used extensively in the field of financial economics. Under this framework the variable of interest follows different autoregressions according to the thresholds and the delay parameter. For example, in modelling currency exchange rate, the thresholds may represent interventions levels, and the delay parameter d, say $d = 1$, indicates that the most recent change is important for effecting the intervention behaviour of a central bank. It is noted that model (1) is a mean model in the sense that it is used to model the mean of a time series. For a more complete statistical analysis of a time series we consider quantile time series models in this paper.

To define a QSETAR model, we rewrite (1) as

$$x_t = \sum_{i=1}^{m+1} \left\{\beta_{i0} + \beta_{i1} x_{t-1} + \cdots + \beta_{ip} x_{t-p} + S_i(a_t)\right\} I_{[x_{t-d} \in \Omega_i]}, \tag{2}$$

where $\{a_t\}$ is a sequence of iid uniform random variables on (0,1), and S_i are quantile functions (i.e. the inverse distribution functions) without any requirements of symmetry. The definition of the SETAR model (2) is more general than that of model (1). For example, if S_i is the quantile function of ξ_{it}, then (1) and (2) are equivalent.

Since the right-hand side of (2) is monotone increasing in a_t, it follows that the θth $(0 < \theta < 1)$ conditional quantile of x_t given previous values $\mathbf{x}_{t-1} = (x_{t-1}, x_{t-2}, \ldots, x_0)^{\mathrm{T}}$ can be written as

$$q^{\theta}_{x_t|\mathbf{x}_{t-1}} = \sum_{i=1}^{m+1} \left(\beta^{\theta}_{i0} + \beta^{\theta}_{i1} x_{t-1} + \cdots + \beta^{\theta}_{ip} x_{t-p} \right) I_{[x_{t-d^{\theta}} \in \Omega_i]}, \tag{3}$$

where the parameters dependent on the values of θ. We refer to model (3) as a QSETAR model.

The reasons for the parameters in (3) being dependent on θ are as follows. Let us write $\xi_{it} = \sigma_i \xi'_{it}$ in (1), where ξ'_{it} has zero mean and unit variance. Then immediately we have that $S_i(\theta) = \sigma_i S'_i(\theta)$, where $S'_i(\theta)$ is the quantile function of ξ'_{it}. If σ_i is a constant, then

$$\beta^{\theta}_{i0} + \beta^{\theta}_{i1} x_{t-1} + \cdots + \beta^{\theta}_{ip} x_{t-p} = \beta_{i0} + \beta_{i1} x_{t-1} + \cdots + \beta_{ip} x_{t-p} + \sigma_i S'_i(\theta),$$

giving that $\beta^{\theta}_{i0} = \beta_{i0} + \sigma_i S'_i(\theta)$, and $\beta^{\theta}_{ij} = \beta_{ij}$ for $i = 1, \ldots, m+1, j = 1, \ldots, p$. If, on the other hand, σ_i is a function of $\mathbf{x}_{t-1}$, say, a linear function of $x_{t-1}, \ldots, x_{t-p}$, then we have that the constant term and the coefficients of $x_{t-1}, \ldots, x_{t-p}$ on the right side may be the sum of a constant coefficient and a function of θ.

Under this new framework, the quantiles of the variable of interest follow different autoregressions according to the values of the delay parameter and thresholds. By estimating a sequence of quantiles of the variable of interest, we actually obtain an estimate of the distribution of the variable, not just of its mean.

Let $\beta^{\theta} = (\beta^{\theta}_{10}, \ldots, \beta^{\theta}_{1p}, \ldots, \beta^{\theta}_{m+10}, \ldots, \beta^{\theta}_{m+1p}, d^{\theta})^{\mathrm{T}}$ be a vector of parameters depending on θ. Furthermore, let $k = \max\{p, d_{\max}\}$, where $d_{\max}$ is the maximum value of d, assumed to be known. Then for an observed time series $x_0, x_1, \ldots, x_n$, estimation of the QSETAR model (3) can be performed by solving the minimization problem

$$\min_{\boldsymbol{\beta}^{\theta}} \sum_{t=k+1}^{n} \rho_{\theta}(u_t) \tag{4}$$

as in Koenker (2005), where

$$\rho_{\theta}(u_t) = u_t(\theta - I_{[u_t<0]}),$$

and

$$u_t = x_t - \sum_{i=1}^{m+1} \left(\beta^{\theta}_{i0} + \beta^{\theta}_{i1} x_{t-1} + \cdots + \beta^{\theta}_{ip} x_{t-p} \right) I_{[x_{t-d^{\theta}} \in \Omega_i]}. \tag{5}$$

Different methods have been proposed to solve (4); see, for example, Koenker and Bassett (1978), and Bassett and Koenker (1982). Koenker and D'Orey

(1987, 1993) and Koenker (2005) provided computing algorithms for these procedures.

It is worth mentioning that minimizing (4) does not require us to supply a specific form for S_i. The information about S_i has been absorbed into the parameter vector $\boldsymbol{\beta}^\theta$. It is also worth mentioning that Gilchrist's (2000) method requires a specific formula for S_i, and estimates the parameters of S_i as well as $\boldsymbol{\beta}^\theta$. In this paper we follow Yu and Moyeed (2001) and perform statistical inference about $\boldsymbol{\beta}^\theta$ in a Bayesian framework based on (4).

3. BAYESIAN APPROACH

It can be easily shown that minimizing (4) is equivalent to maximizing

$$\ell(x_{k+1}, \ldots, x_n | \boldsymbol{\beta}^\theta, \mathbf{x}_k) = \theta^{n-k}(1-\theta)^{n-k} e^{-\sum_{t=k+1}^{n} \rho_\theta(u_t)} \tag{6}$$

over $\boldsymbol{\beta}^\theta$. This expression turns out to be the likelihood function of x_{k+1}, $x_{k+2}, \ldots, x_n$ given $\mathbf{x}_k = (x_k, x_{k-1}, \ldots, x_0)^{\mathrm{T}}$ if we assume that x_t follows the model

$$x_t = \sum_{i=1}^{m+1} \left(\beta_{i0}^\theta + \beta_{i1}^\theta x_{t-1} + \cdots + \beta_{ip}^\theta x_{t-p} + \epsilon_t^\theta \right) I_{[x_{t-d^\theta} \in \Omega_i]},$$

where ϵ_t^θ are iid asymmetric Laplace random variables with density function

$$f(\epsilon) = \theta(1-\theta)e^{-\rho_\theta(\epsilon)}.$$

It is worth emphasizing that the likelihood function used here does not, in general, correspond to the actual distribution of x_t. However, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}^\theta$ of $\boldsymbol{\beta}^\theta$ obtained by maximizing the likelihood (6) is exactly the solution of the minimization problem (4). Koenker (2005) presented important asymptotic properties of the estimates of the parameters obtained from the minimization problem (4), and because of the equivalence of the optimization problems these properties also apply to $\hat{\boldsymbol{\beta}}^\theta$. Koenker (2005) has noted that finding the minimum in (4) is computationaly very challenging. This motivated us to develop a Bayesian approach so that an efficient simulation procedure can be used. For a comprehensive treatment of the Bayesian approach, see O'Hagan and Forster (2004). We now discuss our Bayesian formulation for QSETAR time series.

Let $\pi(\boldsymbol{\beta}^\theta)$ be a prior distribution for $\boldsymbol{\beta}^\theta$. Then, using Bayes' theorem, the posterior distribution of $\boldsymbol{\beta}^\theta$ is given by

$$p(\boldsymbol{\beta}^\theta | \mathbf{x}) \propto \ell(x_{k+1}, \ldots, x_n | \boldsymbol{\beta}^\theta, \mathbf{x}_k)\pi(\boldsymbol{\beta}^\theta).$$

If we take a flat prior distribution for $\boldsymbol{\beta}^\theta$, then the posterior distribution is proportional to the likelihood function. In the following we show that the likelihood function itself as a function of $\boldsymbol{\beta}^\theta$ is proportional to a proper probability density function. Hence the posterior is proper and a Markov chain

Monte Carlo (MCMC) algorithm can be designed to sample from it, so allowing us to make inferences about the parameters. Moreover, in the case of a flat prior distribution for $\boldsymbol{\beta}^{\theta}$, the mode of the posterior distribution corresponds to the maximum-likelihood estimate $\hat{\boldsymbol{\beta}}^{\theta}$ or, equivalently, to the solution of (4).

LEMMA 1. *For any* $0 < \theta < 1$, *let* $h(\theta) = \min(\theta, 1 - \theta)$, *then*

$$e^{-\rho_\theta(u_t)} \le e^{-h(\theta)|u_t|} \le 1.$$

PROOF. First note that since $h(\theta) > 0$ we have $e^{-h(\theta)|u_t|} \le 1$.

Now we prove the first inequality. Consider the case of $\theta \le 1 - \theta$. In this case, $h(\theta) = \theta$, and for $u_t < 0$, we have $-(1 - \theta)(-u_t) \le -\theta(-u_t)$. Hence

$$e^{-\rho_\theta(u_t)} = \left\{ \begin{array}{ll} e^{-\theta u_t}, & u_t \ge 0 \\ e^{(1-\theta)u_t} = e^{-(1-\theta)(-u_t)} \le e^{-\theta(-u_t)} = e^{\theta u_t}, & u_t < 0 \end{array} \right\} = e^{-\theta|u_t|} = e^{-h(\theta)|u_t|}.$$

Similarly, it can be shown that the result also holds in the case of $\theta > 1 - \theta$. This completes the proof. □

THEOREM 1. *If the prior* $\pi(\boldsymbol{\beta}^{\theta}) \propto 1$ *and if* d^{θ} *is any fixed integer between* 1 *and* $d_{\max}$, *then for any non-negative integers*, n_{ij} *we have*

$$0 < \int_{R^w} \prod_{i=1}^{m+1} \prod_{j=0}^{p} |\beta_{ij}^{\theta}|^{n_{ij}} p(\boldsymbol{\beta}^{\theta}|\mathbf{x}) \mathrm{d}\beta_{ij}^{\theta} < \infty,$$

where w is the dimension of the coefficient parameter space.

PROOF. As the integrand is positive, the integral is positive.

If the sample size n is large enough such that $n - k + 1$ is greater than the number of parameters in the model, then it follows from Lemma 1 that

$$\prod_{t=k+1}^{n} e^{-\rho_\theta(u_t)} \le \prod_{l=1}^{w} e^{-h(\theta)|u_{t_l}|}$$

for any $t_l \in [k + 1, n]$, where $l = 1, \ldots, w$. Therefore,

$$\begin{aligned} p(\boldsymbol{\beta}^{\theta}|\mathbf{x}) \propto \ell(x_{k+1}, \ldots, x_n|\boldsymbol{\beta}^{\theta}, \mathbf{x}_k) &= \theta^{n-k}(1-\theta)^{n-k} \prod_{t=k+1}^{n} e^{-\rho_\theta(u_t)} \\ &\le \theta^{n-k}(1-\theta)^{n-k} \prod_{l=1}^{w} e^{-\rho_\theta(u_{t_l})} \le \theta^{n-k}(1-\theta)^{n-k} \prod_{l=1}^{w} e^{-h(\theta)|u_{t_l}|} \end{aligned}$$

and

$$\int_{R^w} \prod_{i=1}^{m+1} \prod_{j=0}^{p} |\beta_{ij}^\theta|^{n_{ij}} p(\boldsymbol{\beta}^\theta|\mathbf{x}) \mathrm{d}\beta_{ij}^\theta \leq \theta^{n-k}(1-\theta)^{n-k} \int_{R^w} \prod_{i=1}^{m+1} \prod_{j=0}^{p} |\beta_{ij}^\theta|^{n_{ij}} \prod_{l=1}^{w} e^{-h(\theta)|u_{t_l}|} \mathrm{d}\beta_{ij}^\theta.$$

Let $v_l = u_{t_l}$, i.e. transform the variables of integration to v_l, $l = 1, \ldots, w$. To do this we need to consider

$$C := \frac{\partial(v_1, \ldots, v_w)}{\partial(\beta_{10}^\theta, \ldots, \beta_{m+1p}^\theta)} = \left|\frac{\partial v_l}{\partial \beta_{ij}^\theta}\right|_{w \times w},$$

where

$$\frac{\partial v_l}{\partial \beta_{ij}^\theta} = \begin{cases} -I_{[x_{t_l-d^\theta} \in \Omega_i]}, & j = 0, \\ -x_{t_l-j} I_{[x_{t_l-d^\theta} \in \Omega_i]}, & j = 1, \ldots, p \end{cases}$$

Provided only that, for the observed time series, the regression defined by the model (3) is full rank (or nonsingular) we can always find time points t_l, $l = 1, \ldots, w$, such that $C \neq 0$. Hence, the Jacobian of the above transformation can be found as $J = 1/|C|$.

On the other hand, it follows from the transformation that β_{ij}^θ can always be expressed as a linear function of v_l, $l = 1, \ldots, w$. By using the triangle inequality, an upper bound of the product $\prod_{i=1}^{m+1} \prod_{j=0}^{p} |\beta_{ij}^\theta|^{n_{ij}}$ can be expressed as a sum of finite terms of the form $A_{i_s} \prod_{l=1}^{w} |v_l|^{i_l}$, where i_l are non-negative integers and A_{i_s} are constants. Therefore, we only need to show that

$$\int_{R^w} \prod_{l=1}^{w} |v_l|^{i_l} \prod_{l=1}^{w} e^{-h(\theta)|v_l|} \mathrm{d}v_1 \cdots \mathrm{d}v_w < \infty.$$

Now

$$\int_{R^w} \prod_{l=1}^{w} |v_l|^{i_l} \prod_{l=1}^{w} e^{-h(\theta)|v_l|} \mathrm{d}v_1 \cdots \mathrm{d}v_w = \int_R |v_1|^{i_1} e^{-h(\theta)|v_1|} \mathrm{d}v_1 \cdots \int_R |v_w|^{i_w} e^{-h(\theta)|v_w|} \mathrm{d}v_w,$$

and for $l = 1, \ldots, w$,

$$\begin{aligned} \int_R |v_l|^{i_l} e^{-h(\theta)|v_l|} \mathrm{d}v_l &= \int_{-\infty}^{0} |v_l|^{i_l} e^{-h(\theta)|v_l|} \mathrm{d}v_l + \int_0^\infty |v_l|^{i_l} e^{-h(\theta)|v_l|} \mathrm{d}v_l \\ &= \int_{-\infty}^{0} (-1)^{i_l} v_i^{i_l} e^{h(\theta)v_l} \mathrm{d}v_l + \int_0^\infty v_l^{i_l} e^{-h(\theta)v_l} \mathrm{d}v_l = 2\int_0^\infty v_l^{i_l} e^{-h(\theta)v_l} \mathrm{d}v_l. \end{aligned}$$

Then it follows using integration by parts that

$$\int_R |v_l|^{i_l} e^{-h(\theta)|v_l|} \mathrm{d}v_l = \frac{2 i_l!}{h^{i_l+1}(\theta)} < \infty$$

as required. $\square$

It is worth mentioning that the above result is also true if we treat d^θ as a random variable taking integer values between 1 and $d_{\max}$. This is because d^θ only takes a finite number of integers.

Theorem 1 tells us that if we take a flat prior distribution on $\boldsymbol{\beta}^\theta$, then the posterior distribution is proportional to the likelihood function, which as a function of $\boldsymbol{\beta}^\theta$ itself is proportional to a proper density function. Moreover, all posterior integer power moments exist. Thus, an MCMC algorithm can be designed to obtain a sample from the posterior distribution and inferences about the parameters can be made.

For our applications a random walk MCMC method has been used. This sampler is very easy to implement and all our simulation results show that it converges very quickly to its equilibrium distribution. In detail, let β_{ij} and d be the current values of the parameters, and let β'_{ij} and d' be proposed new values. Then a random walk sampler can be constructed as follows. First, we draw β'_{ij} from a $N(\beta_{ij}, \sigma^2)$ distribution, where σ is the proposal standard deviation chosen by the user, and d' from a uniform distribution on $[1, d_{\max}]$, where $d_{\max}$ is also chosen by the user. Then, we accept the proposed β'_{ij} and d' with probability $\min\left\{\frac{p(\boldsymbol{\beta}', d'|\mathbf{x})}{p(\beta, d|\mathbf{x})}, 1\right\}$. We will not discuss the optimal choice of the values of σ here, except to say that the optimal choices of the value is problem dependent. Several test runs will help us to decide a proper value of σ.

4. SIMULATION RESULTS

A large number of simulation studies have been performed to check our methodology. In each simulation study, we generated a time series from a given model, and estimated the parameters in the corresponding QSETAR model. Our simulation studies show that very good results can be obtained. We present three of them below.

Consider the following model:

$$x_t = \begin{cases} 0.5 + 0.4x_{t-1} + 0.1x_{t-2} + S_1(u_t), & \text{if } x_{t-d} \le r_1, \\ -0.5 + 0.3x_{t-1} + 0.2x_{t-2} + S_2(u_t), & \text{if } r_1 < x_{t-d} \le r_2, \\ -0.7 + 0.4x_{t-1} + 0.6x_{t-2} + S_3(u_t), & \text{if } r_2 < x_{t-d}, \end{cases} \tag{7}$$

where $S_i(u_t)$, $i = 1, 2, 3$ are independent quantile functions of the error terms in each region.

The theoretical θth conditional quantile of x_t given $\mathbf{x}_{t-1}$ is given by

$$q^\theta_{x_t|\mathbf{x}_{t-1}} = \begin{cases} 0.5 + 0.4x_{t-1} + 0.1x_{t-2} + S_1(\theta), & \text{if } x_{t-d} \le r_1, \\ -0.5 + 0.3x_{t-1} + 0.2x_{t-2} + S_2(\theta), & \text{if } r_1 < x_{t-d} \le r_2, \\ -0.7 + 0.4x_{t-1} + 0.6x_{t-2} + S_3(\theta), & \text{if } r_2 < x_{t-d}. \end{cases} \tag{8}$$

By comparison with (3), we have

$$\beta_{10}^{\theta} = 0.5 + S_1(\theta), \quad \beta_{11}^{\theta} = 0.4, \quad \beta_{12}^{\theta} = 0.1,$$
$$\beta_{20}^{\theta} = -0.5 + S_2(\theta), \quad \beta_{21}^{\theta} = 0.3, \quad \beta_{22}^{\theta} = 0.2,$$
$$\beta_{30}^{\theta} = -0.7 + S_3(\theta), \quad \beta_{31}^{\theta} = 0.4, \quad \beta_{32}^{\theta} = 0.6.$$

and $d^{\theta} = d$.

Simulation Study 1

In this simulation study, we consider a time series with Gaussian errors. Let $S_i(u_t)$ be the quantile function of standard normal distribution, i.e. $S_i(u_t) = \Phi^{-1}(u_t)$, for $i = 1, 2, 3$, and let $r_0 = -\infty$, $r_1 = -1$, $r_2 = 0$, $r_3 = \infty$, so that $\Omega_1 = (-\infty, -1]$, $\Omega_2 = (-1, 0]$ and $\Omega_3 = (0, \infty)$. Finally, let the delay $d = 5$. The first part of Table I gives the true values of β_{i0}^{θ} for typical θ values in this simulation study.

Simulation Study 2

In this simulation we consider time series with heavy but symmetric tails. So we would expect that large outliers appear in the time series.

Let $S_1(u_t)$, $S_2(u_t)$ and $S_3(u_t)$ be independent quantile functions of t-distributions with 2, 3 and 3 degrees of freedom, respectively, and let $r_0 = -\infty$, $r_1 = -1$, $r_2 = 1$, $r_3 = \infty$ so that $\Omega_1 = (-\infty, -1]$, $\Omega_2 = (-1, 1]$ and $\Omega_3 = (1, \infty)$. Finally, let the delay $d = 1$. The second part of Table I gives the true values of β_{i0}^{θ} for typical θ values in this simulation study.

Simulation Study 3

In this simulation, we consider a time series with asymmetric errors. Let $S_1(u_t)$, $S_2(u_t)$ and $S_3(u_t)$ be independent quantile functions of exponential distributions with rates 1.5, 2.5 and 1.5, respectively, and let $r_0 = -\infty$, $r_1 = 0$, $r_2 = 1.5$ and $r_3 = \infty$ so that $\Omega_1 = (-\infty, 0]$, $\Omega_2 = (0, 1.5]$ and $\Omega_3 = (1.5, \infty)$. Finally, let the delay $d = 3$.

The θth quantile of an exponential distribution with rate λ is given by $-\frac{1}{\lambda}\ln(1 - \theta)$. Hence, the true values of β_{i0}^{θ} can be calculated and are shown in the third part of Table I.

TABLE I

True Values of β_{i0}^{θ} in the Three Simulation Studies

θ	0.05	0.25	0.5	0.75	0.95
Simulation Study 1					
β_{10}^{θ}	−1.145	−0.175	0.500	1.174	2.145
β_{20}^{θ}	−2.145	−1.175	−0.500	0.174	1.145
β_{30}^{θ}	−2.345	−1.374	−0.700	−0.026	0.945
Simulation Study 2					
β_{10}^{θ}	−2.420	−0.316	0.500	1.316	3.420
β_{20}^{θ}	−2.853	−1.265	−0.500	0.265	1.853
β_{30}^{θ}	−3.053	−1.465	−0.700	0.065	1.653
Simulation Study 3					
β_{10}^{θ}	0.534	0.692	0.962	1.424	2.497
β_{20}^{θ}	−0.479	−0.385	−0.223	−0.055	0.698
β_{30}^{θ}	−0.666	−0.508	−0.238	0.224	1.297

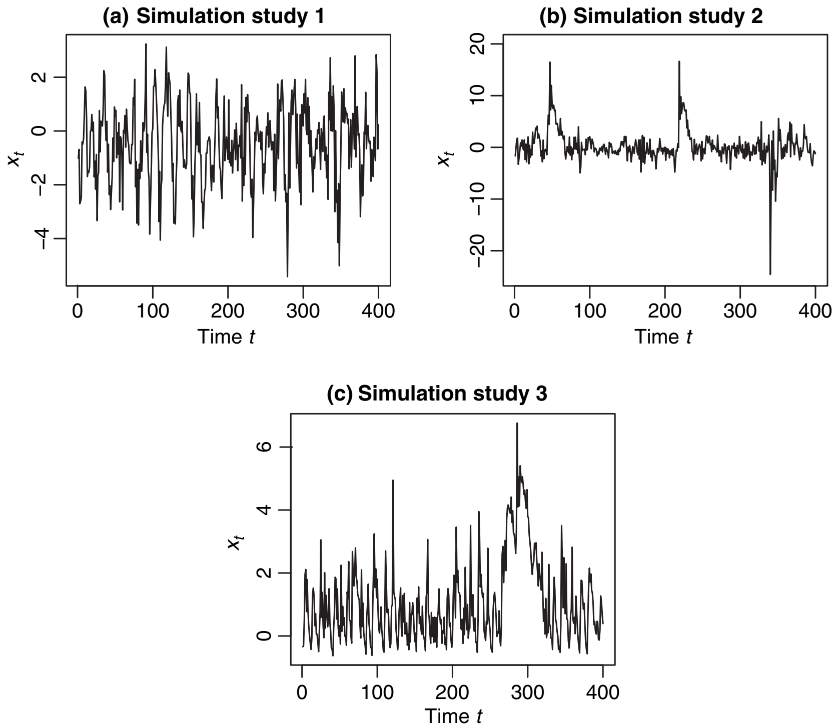

FIGURE 1. Plots of the randomly generated time series used in our three simulation studies.

In each simulation study, we first simulated a time series of size 10400 from model (7), starting from $x_0 = x_1 = 0$, but we only saved the last 400 values to remove the effect of the initial values. Figure 1 gives the time series plots of the simulated time series.

As expected, many extreme values are shown in Figure 1(b) for the data from Simulation Study 2 because the time series was generated from model (7) with heavy tails. Quite abnormal behaviours (asymmetric and a big group of data deviating from the general pattern) are observed in Figure 1(c) for the data from Simulation Study 3. As our method does not depend on the distribution of the error terms of the model, we should expect that the final results should be robust to these unusual behaviors.

When running a MCMC sampler, many methods may be used to collect samples from the posterior distribution of the parameters. For example, one is to run a very long chain and then subsample by saving the values after every s steps after the burn-in period. Another is to run many chains with different initial values and to save the final values. All the saved values from each chain can be used as a sample from the posterior distribution of the parameters. In our simulation studies we use a combination of the above two methods, i.e. we run many very long chains from each of which we subsample. We believe that this is more helpful in investigating the parameter space.

One-hundred Markov chains of length 100,000 were run for each simulation study, starting from randomly chosen values of β_{ij} in $(-1, 1)$ and a randomly chosen value of d in $[1, d_{\max}]$. In the simulations that we present we let $d_{\max} = 10$. For other values of $d_{\max}$ the results are very similar. For each chain we set $s = 100$ and adopted a burning of 90,000 values. Therefore, by the end of the simulation, we had 10,000 samples from the posterior distribution.

Time series plots (not shown to save space) of the simulated samples indicated that the Markov chains converge to their equilibrium distribution no matter what initial values are chosen. Histograms of the collected samples for each parameter in the simulation studies are also examined for $\theta =$ 0.05, 0.25, 0.5, 0.75 and 0.95. Generally speaking, the results are very similar for different values of θ although those for nonextreme θ values are slightly better. To save space we only show the results correspond to the lowest value $\theta = 0.05$ in Figure 2. The vertical lines on the histograms indicate the true values of the parameters and show that these values lie well within the posterior marginals.

To check the fitted conditional quantiles, we used (8) to calculate the true θth conditional quantile, $q^{\theta}_{x_t|\mathbf{x}_{t-1}}$. We then used the posterior means of the parameters to calculate the fitted θth conditional quantile, $\hat{q}^{\theta}_{x_t|\mathbf{x}_{t-1}}$, from (3). The overall difference between the true and fitted conditional θth quantiles is measured by the mean and the standard error of $\hat{q}^{\theta}_{x_t|\mathbf{x}_{t-1}} - q^{\theta}_{x_t|\mathbf{x}_{t-1}}$ for $t = k + 1, \ldots, n$, where $k = \max\{p, d_{\max}{=}10\}$. The results are given in Table II, which shows that, on average, the fitted and true conditional quantiles of x_t are in good agreement with each other.

Figure 3 shows the fitted (Solid curves) and true (dashed curves) θth conditional quantiles of x_t for $\theta = 0.05$ and 0.5 in each simulation study. The plots for other values of θ are very similar. It is seen that even for Simulation Study 2 and Simulation Study 3, where the original time series have many very large observations and odd behaviors, the true and fitted conditional quantiles are also in very good agreement.

In summary, very good results were obtained for all the simulation studies although the results for $\theta = 0.25$, 0.5 and 0.75 are slightly better than those for $\theta = 0.05$ and 0.095. This is because, when θ is very small or large, the distributions of asymmetric Laplace random variables are very skewed with very large variations. However, even when θ is small or large, our results still show that the method is very robust which, in fact, is one of the advantages of quantile methods reported in the literature; see, for example, Koenker (2005).

5. AN EMPIRICAL APPLICATION TO US REAL GNP DATA

In this section we present some empirical results obtained by applying our methodology to the US real GNP data. The time series is the quarterly US real

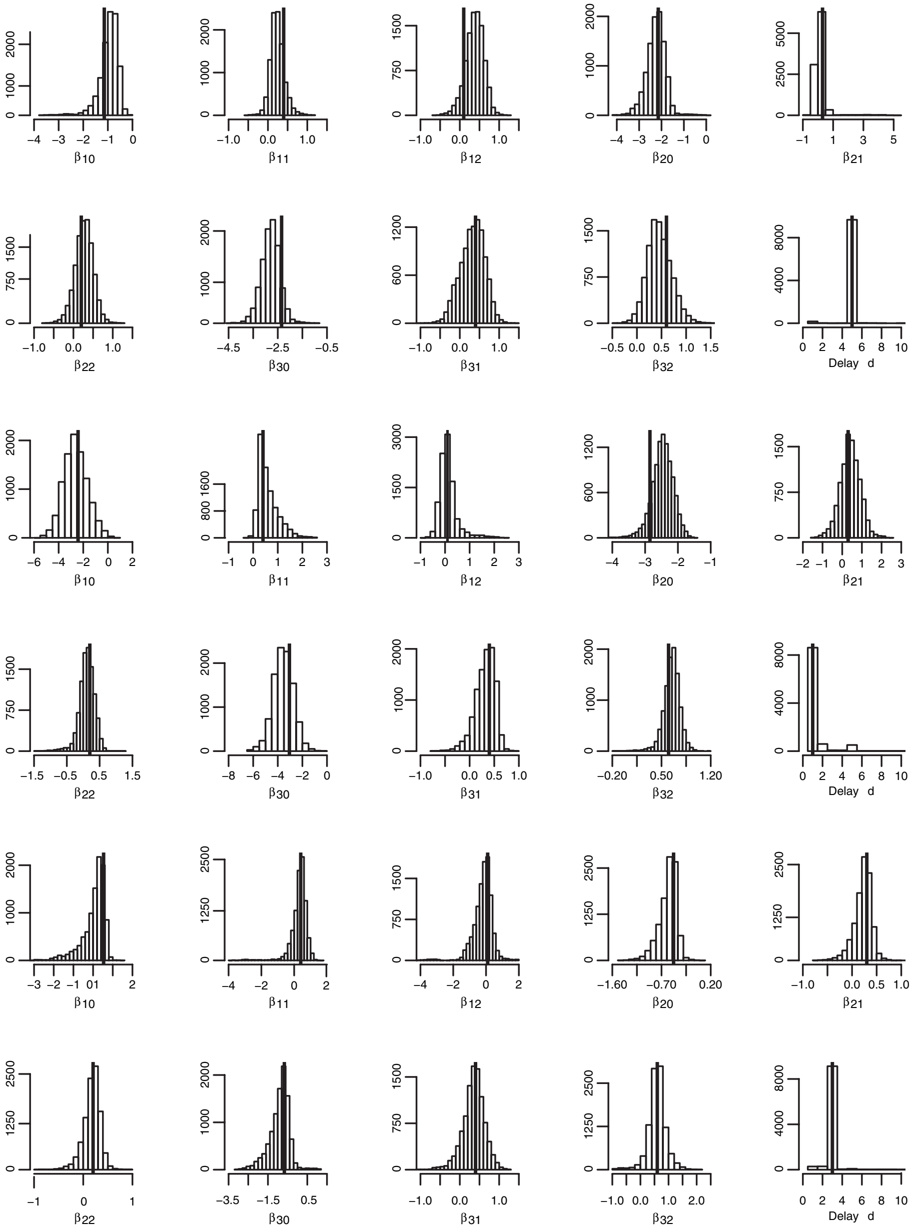

FIGURE 2. Histograms of the simulated parameters when $\theta = 0.05$. The first two rows are for Simulation Study 1, the second two rows are for Simulation Study 2, and the last two rows are for Simulation Study 3 The true values of the parameters are shown by the vertical lines on each histogram. The vertical axis represents frequency.

TABLE II

THE MEANS AND STANDARD ERRORS OF THE DIFFERENCES BETWEEN THE FITTED AND OBSERVED CONDITIONAL QUANTILES $\hat{q}^{\theta}_{x_t|x_{t-1}} - q^{\theta}_{x_t|X_{t-1}}$ FOR TYPICAL VALUES OF θ, WHERE $t = 11, \ldots, 400$

θ	0.05	0.25	0.5	0.75	0.95
Simulation Study 1					
Mean	0.1406	0.0331	0.0136	0.0522	0.1190
Standard error	0.3496	0.1319	0.1094	0.1442	0.1974
Simulation Study 2					
Mean	−0.2166	−0.0018	0.0220	0.1388	0.5639
Standard error	0.6456	0.2841	0.3048	0.4273	0.7317
Simulation Study 3					
Mean	−0.3159	−0.0092	0.0043	0.0416	0.3774
Standard error	0.2209	0.0446	0.0471	0.0696	0.3242

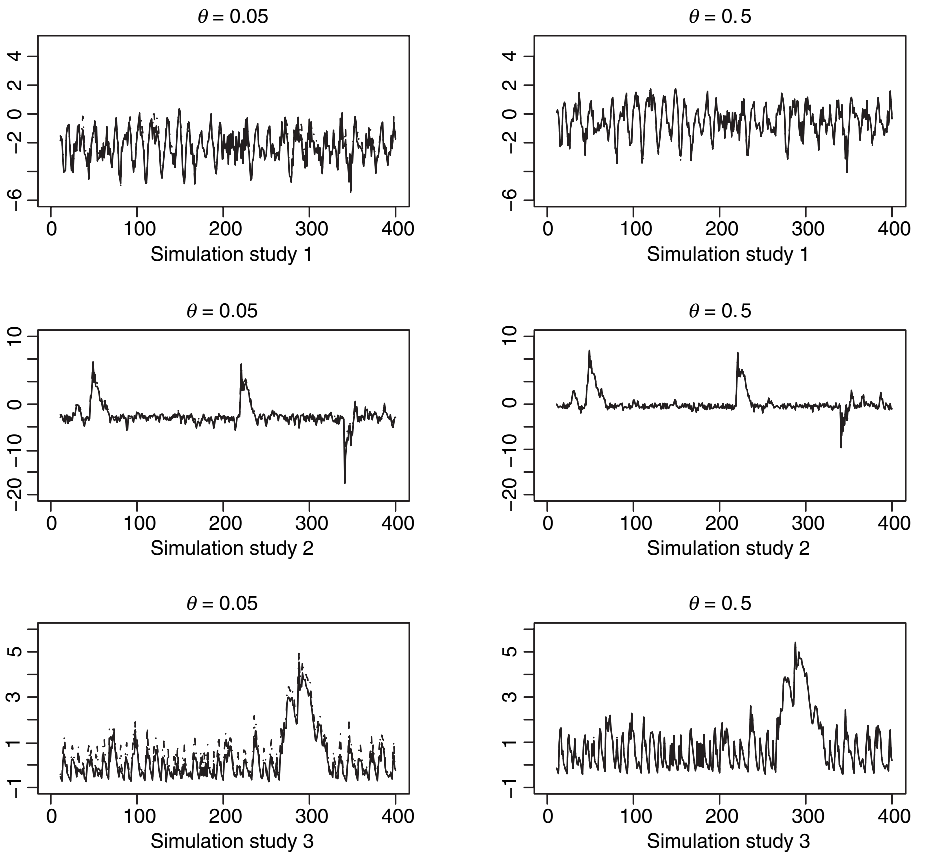

FIGURE 3. Fitted (solid curves) and true (dashed curves) conditional quantiles of x_t.

GNP (in 1982 dollars) from the first quarter of 1947 to the first quarter of 1991. The length of the time series is 177. The data are seasonally adjusted. In this application we consider the growth rate, that is

$$y_{t-1} = \log(x_t) - \log(x_{t-1}), \quad t = 2, \ldots, 177,$$

where x_t is the US GNP at time point t. The time series plot of the growth rate y_t is shown in Figure 4.

The US GNP is one of the most examined univariate time series in modern macroeconomics. Potter (1995) presented an extensive discussion of a graphical and testing approach to estimating the delay parameter d and threshold values r_i for SETAR models. These testing and graphical techniques produced estimates of $d = 2$ and one threshold $r = 0$. Tiao and Tsay (1994) used a similar set of techniques and produced identical results. It is interesting to note that $y_t \leq 0$ corresponds to a negative growth rate (or contraction), and that $y_t > 0$ corresponds to a positive growth rate (or expansion), so the economy behaves differently after contraction and expansion. Tiao and Tsay's (1994) fitted mean model is

$$y_t = \begin{cases} -0.0039 + 0.44y_{t-1} - 0.79y_{t-2} + \epsilon_{1t}, & \text{if } y_{t-2} \leq 0, \\ 0.0038 + 0.31y_{t-1} + 0.20y_{t-2} + \epsilon_{2t}, & \text{if } y_{t-2} > 0, \end{cases} \tag{9}$$

where $\epsilon_{1t} \sim N(0, 0.012^2)$, $\epsilon_{2t} \sim N(0, 0.0087^2)$.

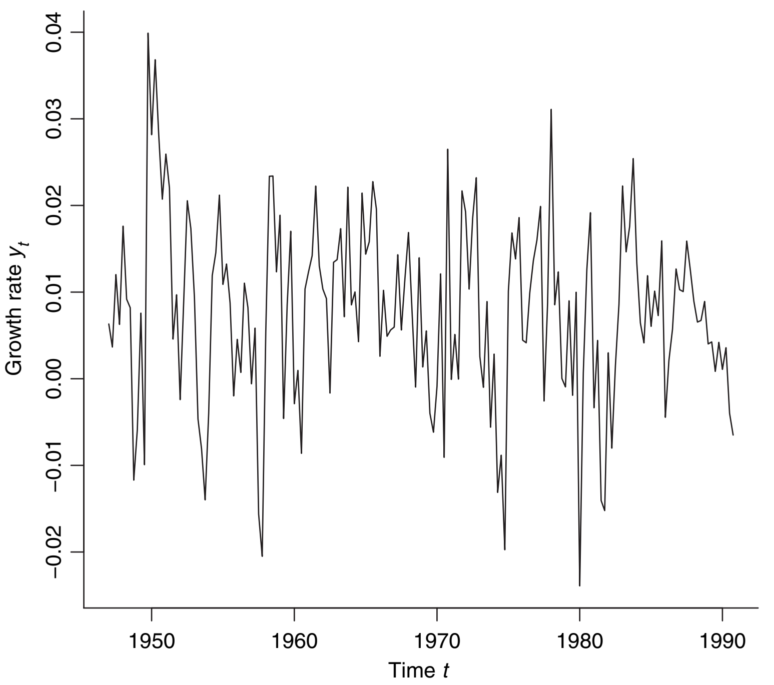

FIGURE 4. Time series plot of the growth rate y_t of US quarterly real GNP from the first quarter of 1947 to the first quarter of 1991.

To fit a quantile model we need to select the order of the model. As pointed out by Koenker (2005), a modified Akaike Information Criterion (AIC) can be used in quantile regression cases. For the Mth model, this takes the form

$$\mathrm{AIC}(M) = \log(\hat{\sigma}_M) + p_M,$$

where

$$\hat{\sigma}_M = \frac{1}{n-k} \sum_{t=k+1}^{n} \rho_\theta\left(u_t^M\right),$$

and u_t^M is calculated according to (5) but corresponds to the Mth model using the estimate $\hat{\boldsymbol{\beta}}^\theta$ of the parameters which minimizing (4), and p_M is the number of parameters in the Mth model. The chosen model should correspond to the minimum value of the AIC.

For reasons mentioned in Section 3, in our Bayesian approach when we have a flat prior it is appropriate to use this AIC, as $\hat{\boldsymbol{\beta}}^\theta$ corresponds to the posterior mode. It is worth mentioning that the above AIC criterion is not appropriate if the prior distribution is not flat, and further investigations on this are needed.

By using the same threshold value $r = 0$ as Tiao and Tsay (1994), a sequence of models with order $p = 0, 1, 2, 3$ and 4, and $\theta = 0.05, 0.25, 0.5, 0.75$ and 0.95 have been fitted to the growth rate time series. For each θ, the AIC value was calculated for each model and is shown in Figure 5. It is interesting to note that in every case the minimum value of AIC corresponds to a QSETAR model of order $p = 0$, which has the form of

$$q^\theta_{y_t|\mathbf{y}_{t-1}} = \begin{cases} \beta^\theta_{10}, & \text{if } y_{t-d^\theta} \le 0, \\ \beta^\theta_{20}, & \text{if } y_{t-d^\theta} > 0, \end{cases} \tag{10}$$

where the estimates of the parameters are shown in Table III.

The estimated delay parameter d^θ is 3 for all the fitted zero-order models. Table IV gives the estimated delay parameter for all the fitted models, and shows that, except for the fitted model with $\theta = 0.75$ and order $p = 4$, all the other models have estimated delays 2 or 3. It is interesting to see that all the fitted QSETAR models with order $p = 2$ have the same estimated delays as found by Tiao and Tsay's (1994). More precisely, the p-values of their F-tests for deciding the value of d are 0.026, 0.058 and 0.051 for $d = 2, 3$ and 4, respectively, and the p-values for other values of delay are all greater than 0.15. The advantage of our approach is that we provide a more complete picture of the distribution of the growth rate, not just the mean.

Figure 6 gives the fitted quantiles and means at different time points. We see that the distribution of the growth rate is heavy tailed, but the mean and the median are very close to each other.

Based on all the empirical results obtained we conclude that a zero-order QSETAR model can be used to describe the θth conditional quantile of the growth rate of US GNP data and, hence, the distribution of the growth rate of US GNP data at any time.

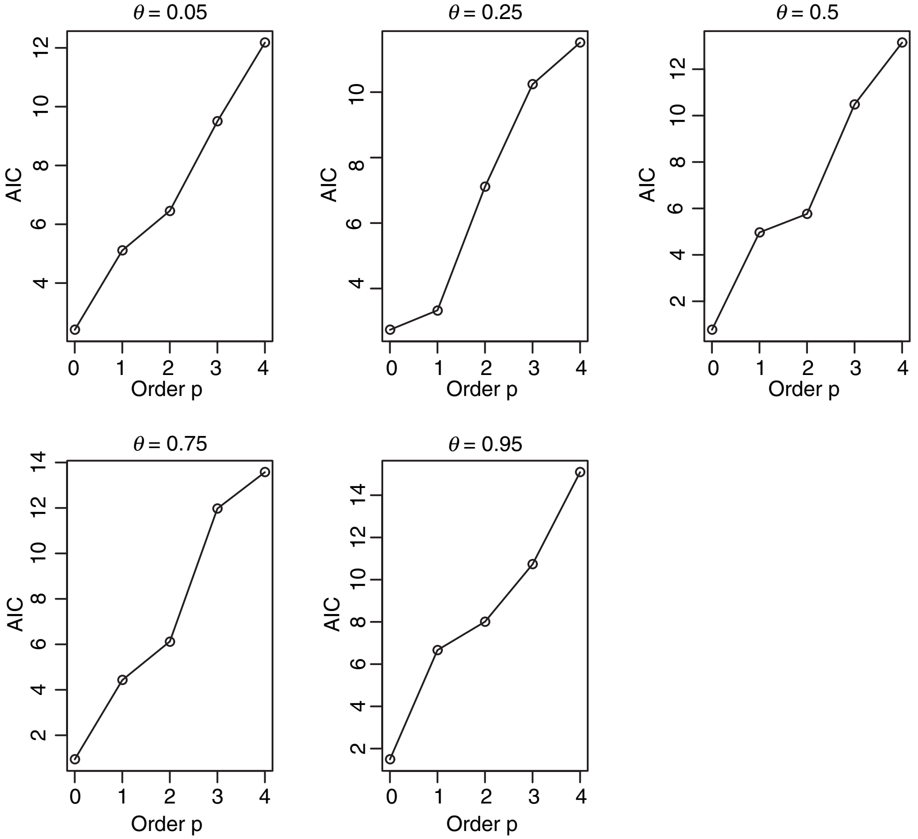

FIGURE 5. AIC values for a sequence of QSETAR models for US GNP data.

TABLE III
ESTIMATED PARAMETER VALUES FOR MODEL (10)

θ	0.05	0.25	0.5	0.75	0.95
β_{10}^{θ}	−0.505	−0.062	0.010	0.085	0.518
β_{20}^{θ}	−0.141	−0.015	0.008	0.031	0.156
d^{θ}	3	3	3	3	3

6. COMMENTS AND CONCLUSION

We have defined a QSETAR time series model and presented a Bayesian approach to QSETAR time series modelling. The simulation results show that the method is robust to outliers for a large variety of time series. It is possible to generalize the method to estimate the number and the positions of the thresholds. Areas for further research include using our QSETAR methodology for

TABLE IV.
ESTIMATED DELAY VALUE FOR ALL FITTED MODELS

Order of the model	$d^{0.05}$	$d^{0.25}$	$d^{0.5}$	$d^{0.75}$	$d^{0.95}$
0	3	3	3	3	3
1	3	3	2	3	2
2	2	2	2	2	2
3	2	2	3	2	2
4	2	3	2	4	3

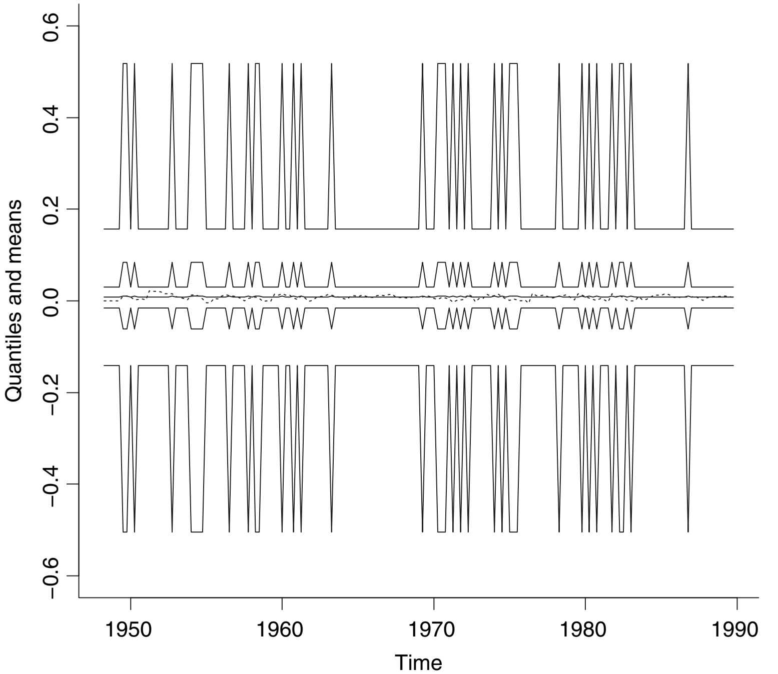

FIGURE 6. Fitted means (dotted curve) and the fitted θth quantiles of y_t given $\mathbf{y}_{t-1}$ (solid curves) for the growth rate of US quarterly real GNP from the second quarter of 1947 to the first quarter of 1991 based on model (10). Curves (starting from the top) correspond to $\theta = 0.95, 0.75, 0.5, 0.25, 0.05$.

forecasting and generalizing our approach to other types of nonlinear autoregressive time series models. We will also investigate further some of the theoretical properties of our Bayesian approach in the case of the prior distributions of the parameters not being flat.

ACKNOWLEDGEMENT

We are grateful thank to Dr K Yu and Professor W Gilchrist for their helpful comments, and to Dr P Hewson for assistance with one of the figures. We would like to express our sincere thanks to a referee for critical comments, which have greatly enhanced the quality and presentation of the paper.

Corresponding author: Dr Yuzhi Cai, School of Mathematics and Statistics, University of Plymouth, Plymouth PL4 8AA, United Kingdom. Tel.: 752 232728; Fax: 752 232780; E-mail: ycai@plymouth.ac.uk

REFERENCES

Bassett, G. and Koenker, R. (1982) An empirical quantile function for linear models with i.i.d. errors. *Journal of American Statistical Association* 77, 407–15.

Chen, R. and Tsay, R. S. (1991) On the ergodicity of TAR(1) processes. *Annals of Applied Probability* 1, 613–34.

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

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, 383–93.

Koenker, R. and D'Orey, V. (1993) Computing regression quantiles. *Applied Statistics* 43, 410–14.

Koenker, R. and Xiao, Z. (2004) Unit root quantile autoregression inference. *Journal of the American Statistical Association* 99, 775–87.

Koenker, R. and Xiao, Z. (2005) Quantile autoregression. Preprint. Department of Economics University of Illinois, USA.

O'Hagan, A. and Forster, J. J. (2004) *Bayesian Inference*. London: Arnold.

Petruccelli, J. and Woolford, S. W. (1984) A threshold AR(1) model. *Journal of Applied Probability* 21, 270–86.

Potter, S. M. (1995) A nonlinear approach to US GNP. *Journal of Applied Econometrics* 10, 109–25.

Tiao, G. C. and Tsay, R. S. (1994) Some advances in non-linear and adaptive modelling in time series. *Journal of Forecasting* 13, 109–31.

Tong, H. and Lim, K. S. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Staistical Society, Series* B 42, 245–92.

Weiss, A. (1987) Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* 7, 46–68.

Yu, K. and Moyeed, R. A. (2001) Bayesian quantile regression. *Statistics and Probability Letters* 54, 437–47.

Yu, K., Lu, Z. and Stander, J. (2003) Quantile regression: applications and current research area. *The Statistician* 52, 331–50.