Biometrika (2010), **97**, 1, *pp.* 199–208
doi: 10.1093/biomet/asp070

Printed in Great Britain

Forecasting for quantile self-exciting threshold autoregressive time series models

BY YUZHI CAI

School of Computing and Mathematics, University of Plymouth, Plymouth PL4 8AA, U.K.
ycai@plymouth.ac.uk

SUMMARY

Self-exciting threshold autoregressive time series models have been used extensively, and the conditional mean obtained from these models can be used to predict the future value of a random variable. In this paper we consider quantile forecasts of a time series based on the quantile self-exciting threshold autoregressive time series models proposed by Cai and Stander (2008) and present a new forecasting method for them. Simulation studies and application to real time series show that the method works very well.

Some key words: Forecasting method; Monte Carlo method; Predictive density function; Quantile forecast.

1. INTRODUCTION

Quantile regression can be used to estimate conditional quantile functions and can provide a complete statistical analysis of stochastic relationships between random variables. Some work on quantile estimation for time series can be found in the literature. For example, Weiss (1987) discussed how to fit nonlinear dynamic models using least absolute error estimation, and Koenker and Xiao (2004) studied statistical inference in quantile autoregressive models when the largest autoregressive coefficient may be unity. Koenker and Xiao (2006) also considered quantile autoregressive models with coefficients that can be expressed as monotone functions of a single, scalar random variable. Most recently, Cai and Stander (2008) proposed a quantile self-exciting threshold autoregressive time series model and developed a Markov chain Monte Carlo method to make Bayesian inferences about its parameters.

Many forecasting methods for time series focus on point forecasts and use the conditional mean to predict the future value of a random variable. Tong (1990) and Granger and Teräsvirta (1993) gave general reviews of this topic. Cai (2003, 2005) presented a forecasting method to obtain an approximate m-step ahead predictive probability density or distribution function, and hence any predictive quantities such as mean and variance, for a range of nonlinear autoregressive time series models including self-exciting threshold autoregressive models (Tong and Lim, 1980). Examples in these papers show that the procedure can yield accurate forecasts for a range of nonlinear autoregressive time series models. Due to the wide applications of self-exciting threshold autoregressive models, Cai (2007) also developed a forecasting method for quantile self-exciting threshold autoregressive time series models that enabled m-step ahead predictive quantiles to be obtained. However, a limitation of this method is that it is difficult to measure the forecasting accuracy. In this paper we develop a new forecasting method for quantile self-exciting threshold autoregressive time series models that overcomes this difficulty.

2. THE MODEL

We first consider a self-exciting threshold autoregressive time series model. Let ξ_{it} be a sequence of independent and identically distributed Gaussian random variables with mean zero and variance σ_i^2, let $-\infty = r_0 < \cdots < r_w = \infty$ be the threshold values, and let $\Omega_i = (r_{i-1}, r_i]$ $(i = 1, \ldots, w-1)$ and $\Omega_w = (r_{w-1}, r_w)$. Then a self-exciting threshold autoregressive time series model is defined by

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

where β_{ij}, σ_i and $d > 0$ are the model parameters and d is called the delay parameter. The order of the model is p, and $I_{(x \in \Omega)}$ is an indicator of the event $x \in \Omega$. This model defines a data-generating process in the sense that we can recursively use expression (1) to generate a time series conditional on some initial values.

Cai and Stander (2008) proposed a quantile self-exciting threshold autoregressive time series model. This new model says that the θth $(0 < \theta < 1)$ quantile of x_t conditional on $x_{t-1}^* = (x_{t-1}, \ldots, x_0)^{\mathrm{T}}$ is given by

$$q^{\theta}_{x_t | x^*_{t-1}} = \sum_{i=1}^{w} (\beta^{\theta}_{i0} + \beta^{\theta}_{i1} x_{t-1} + \cdots + \beta^{\theta}_{ip} x_{t-p}) I_{(x_{t-d^{\theta}} \in \Omega_i)}, \tag{2}$$

where β^{θ}_{ij} and d^{θ} are model parameters depending on θ. Under this new framework, the conditional quantiles of x_t follow different autoregressions according to the values of the delay parameter and thresholds. The conditional distribution of x_t can be obtained by interpolating the conditional quantiles in a proper way. The fundamental difference between models (1) and (2) is that the former is motivated by a state equation approach while the latter is motivated by a conditional distribution approach. As pointed out by a referee, for Markov systems the two approaches are equivalent under quite general conditions (Tong, 1990; Chan and Tong, 2002). Alternative formulations of the conditional distribution approach are based on different motivations. For example, Wong and Li (2000) proposed a mixture autoregressive model that also has the ability to describe the conditional distribution of a time series. Yao and Tong (1996) proposed a nonparametric approach to the estimation of conditional distribution based on an asymmetric L_1 norm minimization. In this paper we will discuss how to obtain predictive quantiles from model (2).

Let $\beta^{\theta} = (\beta^{\theta}_{10}, \ldots, \beta^{\theta}_{1p}, \ldots, \beta^{\theta}_{w0}, \ldots, \beta^{\theta}_{wp}, d^{\theta})^{\mathrm{T}}$; let $L = \max(p, d_{\max})$, where $d_{\max}$ is the maximum value of d^{θ} assumed to be known; and let $\hat{\beta}^{\theta}$ be the estimate of β^{θ}. Then for an observed time series $x_0, \ldots, x_n$, $\hat{\beta}^{\theta}$ can be obtained by solving the minimization problem $\min_{\beta^{\theta}} \sum_{t=L+1}^{n} \rho_{\theta}(u_t)$, where $u_t = x_t - \sum_{i=1}^{w} (\beta^{\theta}_{i0} + \sum_{j=1}^{p} \beta^{\theta}_{ij} x_{t-j}) I_{(x_{t-d^{\theta}} \in \Omega_i)}$, and $\rho_{\theta}(u) = u(\theta - I_{(u<0)})$, proposed by Koenker (2005). He showed that under certain conditions the corresponding estimate $\hat{q}^{\theta}_{x_t|x^*_{t-1}}$ of $q^{\theta}_{x_t|x^*_{t-1}}$ is a monotone function of θ. Cai and Stander (2008) also proposed a Bayesian method for estimating the parameters of model (2). From now on, we assume that model (2) has been fitted to a time series and we focus on the forecasting method based on the fitted model.

3. THE FORECASTING METHOD

The forecasting method for obtaining m-step ahead predictive quantiles, i.e., the quantiles of x_{t+m} given x^*_t, consists of the following five steps.

Step 1. Define an auxiliary process. Let $y_t^* = x_t^*$. The m-step auxiliary model is defined by

$$y_{t+m} = \sum_{i=1}^{w} \{\beta_{i0}^\theta + \beta_{i1}^\theta y_{t+m-1} + \cdots + \beta_{ip}^\theta y_{t+m-p} + s_i^\theta(\epsilon_{t+m}^\theta - \mu^\theta)\} I_{(y_{t+m-d^\theta} \in \Omega_i)}, \tag{3}$$

where ϵ_{t+m}^θ are independently and identically distributed random variables with mean μ^θ and standard deviation b^θ, and $s_i^\theta = a_i/b^\theta$, in which a_i is the standard deviation of the x_t process in Ω_i. Furthermore, let $y_{t+j} = x_{t+j}$ $(j = 1, \ldots, m-1)$, in which x_{t+j} is simulated from the predictive distribution of x_{t+j} with density function $g(x_{t+j} \mid x_t^*)$ obtained below in Step 5. The auxiliary model (3) uses the estimates of the parameters of model (2).

Step 2. Obtain predictive means $E(y_{t+m} \mid y_t^*)$ from the auxiliary model (3) using two different methods.

Method 1. *Numerical method*

Let $f_\theta(y_{t+m} \mid y_t^*)$, depending on θ, be the m-step ahead predictive density function of y_{t+m} given $y_t^* = x_t^*$. Then the numerical method consists of the following three steps.

(i) Apply the method proposed by Cai (2003, 2005) to model (3) to obtain a discrete version of $f_\theta(y_{t+m} \mid y_t^*)$ for a wide range of possible y_{t+m} values: $\{y_{t+m}^{(j)},\ f_\theta(y_{t+m}^{(j)} \mid y_t^*)\}$ $(j = 1, \ldots, N,$ $m = 1, 2, \ldots)$, where $y_{t+m}^{(j)} < y_{t+m}^{(j+1)}$ for $j < N$ and N is the total number of points where the values of $f_\theta(y_{t+m} \mid y_t^*)$ are evaluated.

(ii) Use $\{y_{t+m}^{(j)},\ f_\theta(y_{t+m}^{(j)} \mid y_t^*)\}$ $(j = 1, \ldots, N)$ to construct a continuous function $\hat{f}_\theta(y_{t+m} \mid y_t^*)$ to approximate $f_\theta(y_{t+m} \mid y_t^*)$. The accuracy of such an approximation can be improved by increasing the accuracy of each $f_\theta(y_{t+m}^{(j)} \mid y_t^*)$ $(j = 1, \ldots, N)$ and by increasing N. For illustration, we used a piecewise linear function to approximate $f_\theta(y_{t+m} \mid y_t^*)$ in this paper.

(iii) Estimate $E_\theta(y_{t+m} \mid y_t^*)$ by $\int_{-\infty}^{\infty} y_{t+m} \hat{f}_\theta(y_{t+m} \mid y_t^*) dy_{t+m}$.

Method 2. *Monte Carlo method*

Using (3), we can also simulate a large sample of y_{t+m} given $y_t^* = x_t^*$ for each m. Then the m-step ahead predictive density function and hence any other predictive quantities can be estimated from this sample. In particular, we can estimate $E_\theta(y_{t+m} \mid y_t^*)$.

Step 3. Estimate the θth quantile of x_{t+m} given x_t^*. Note that $E_\theta(y_{t+m} \mid x_t^*) = E_\theta(y_{t+m} \mid y_t^*)$. Let $q_{x_{t+m}|x_t^*}^\theta = E_\theta(y_{t+m} \mid x_t^*)$. Then we have a θth quantile forecast of x_{t+m} given x_t^*. In particular, the predictive median $q_{x_{t+m}|x_t^*}^{0\cdot5}$ can be used as a point forecast of x_{t+m} given x_t^*. We will see in § 4 that this point forecast has a high accuracy.

Step 4. Construct a predictive density function. Use a sequence of quantile forecasts $q_{x_{t+m}|x_t^*}^{\theta_i}$, where $\theta_i < \theta_{i+1}$ $(i = 1, \ldots, M-1)$ to construct a predictive density function $g(x_{t+m} \mid x_t^*)$.

Step 5. Obtain a random sample x_{t+m} from $g(x_{t+m} \mid x_t^*)$. Finally, let $y_{t+m} = x_{t+m}$, $m = m+1$, and go to Step 1.

A formal proof of the convergence of the method can be constructed, as we will now outline. First let us consider the case $m = 1$. It follows from model (3) that $q_{x_{t+1}|x_t^*}^\theta = E_\theta(y_{t+1} \mid x_t^*)$ exactly for any θ. Therefore, $g(x_{t+1} \mid x_t^*)$ can be estimated accurately provided that M is large enough. A sample x_{t+1} can also be drawn from $g(x_{t+1} \mid x_t^*)$ with high accuracy.

For $m = 2$, it also follows from model (3) that $q_{x_{t+2}|x_{t+1}^*}^\theta = E_\theta(y_{t+2} \mid y_{t+1}^*)$, where $y_{t+1} = x_{t+1}$. This equation holds for any x_{t+1} sampled from $g(x_{t+1} \mid x_t^*)$. Hence, it is equivalent to $q_{x_{t+2}|x_t^*}^\theta = E_\theta(y_{t+2} \mid x_t^*)$. On the other hand, $E_\theta(y_{t+2} \mid x_t^*)$ can be estimated accurately if the numerical method proposed by Cai (2003, 2005) is used. Therefore, $g(x_{t+2} \mid x_t^*)$ and hence a new sample x_{t+2} can also be obtained with high accuracy. Generally, for $m \geqslant 3$, by repeating the above procedure we can obtain $q^\theta_{x_{t+m}|x_t^*} = E_\theta(y_{t+m} \mid x_t^*)$.

It can be shown that the convergence of the new method does not depend on the distribution of ϵ_t^θ provided that this is continuous, because the method of Cai (2003, 2005) does not depend on the distribution of the error terms. For the purposes of illustration and further demonstration of the performance of Cai's (2003, 2005) method in non-Gaussian error term cases, we let ϵ_t^θ follow the asymmetric Laplace distribution, see for example Koenker & Machado (1999), throughout the paper. This distribution has density $f(\epsilon) = \theta(1-\theta)e^{-\rho_\theta(\epsilon)}$ with mean $\mu^\theta = (1-2\theta)/\{\theta(1-\theta)\}$ and standard deviation $b^\theta = [(1-2\theta+2\theta^2)/\{\theta(1-\theta)\}]^{1/2}$.

We remark that we need to draw a sample from $g(x_{t+m} \mid x_t^*)$ in order to obtain the $(m+1)$-step ahead quantile forecast. This is because model (2) is not a data-generating process. On the other hand, although model (3) is a data-generating process, it is not one for the original time series x_t.

We should also highlight the differences between the method developed above and the approach of Cai (2003, 2005). Cai's (2003, 2005) method is based on a nonlinear autoregressive time series model, where the error term of the model follows a distribution with zero mean. This model can be used to estimate the conditional mean of a time series. The new method is based on the quantile model (2) which is used to estimate the conditional quantiles of a time series and which does not contain an error term. Hence model (2) is semiparametric and is suitable for non-Gaussian time series. Because of the semiparametric nature of model (2), the forecasting method proposed here is much more complicated than that of Cai (2003, 2005). It is a combination of defining auxiliary processes, applying the numerical method of Cai (2003, 2005) in Step 2(i), other numerical methods and simulation procedures.

As mentioned above, the quality of $f_\theta(y_{t+m} \mid y_t^*)$ in Step 2(i) plays an important role in the accuracy of the new method. In the following section, we first compare the numerical and the Monte Carlo methods in Step 2(i) with available theoretical results. Then we carry out further simulation studies for more general cases.

4. Examples and simulation studies

4·1. *Some theoretical results*

Although no explicit formula for $f_\theta(y_{t+m} \mid y_t^*)$ is available in general cases, for the special case

$$y_t = \sum_{i=1}^{w} \{\beta_i^\theta + s_i^\theta(\epsilon_t^\theta - \mu^\theta)\} I_{(y_{t-1}\in\Omega_i)} \tag{4}$$

we can work out the theoretical m-step ahead predictive density function, mean and variance of y_{t+m} given y_t^*, where ϵ_t^θ follows the asymmetric Laplace distribution.

Theorem 1 and Corollary 1 show that, for model (4), the m-step ahead predictive density function can be calculated exactly. The m-step ahead predictive mean and predictive variance are also obtained exactly in Theorems 2 and 3 and Corollary 2. The proofs of the results are outlined in the Appendix.

THEOREM 1. *Suppose that y_t follows model* (4). *Then $f_\theta(y_{t+m} \mid y_t^*)$ takes the following form: for $m = 1$, $f_\theta(y_{t+1} \mid y_t^*) = (s_i^\theta)^{-1} f\{(y_{t+1} - \beta_i^\theta + s_i^\theta\mu^\theta)/s_i^\theta\}$, if $y_t \in \Omega_i (i = 1, \ldots, w)$; for $m > 1$, $f_\theta(y_{t+m} \mid y_t^*) = h_m^{\mathrm{T}} H^{m-2} g$, where $H^0 = I$ is an identity matrix, $h_m^{\mathrm{T}} = (h_m^{(i)})_{1\times w}$, $H = (h_{ij})_{w\times w}$,* *and* $g^{\mathrm{T}} = (g_i)_{1\times w}$, *where*

$$h_m^{(i)} = (s_i^\theta)^{-1} f\{(y_{t+m} - \beta_i^\theta + s_i^\theta \mu^\theta)/s_i^\theta\}, \quad g_i = \int_{\Omega_i} f_\theta(y_{t+1} \mid y_t^*) dy_{t+1},$$

$h_{ij} = \int_{\Omega_i} (s_j^\theta)^{-1} f\{(y - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta\} dy$, *and* f *is the density function of* ϵ_t^θ.

Theorem 1 holds for any continuous distributions that ϵ_t^θ may follow, while other results depend on the asymmetric Laplace distribution.

COROLLARY 1. *Under the conditions of Theorem* 1, *if* ϵ_t^θ *follows the asymmetric Laplace distribution and if* $y_t \in \Omega_j$, *then* $g_i = h_{ij}$, *where*

$$h_{ij} = \begin{cases} \theta\left\{e^{(1-\theta)(r_i - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta} - e^{(1-\theta)(r_{i-1} - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta}\right\} & (r_i - \beta_j^\theta + s_j^\theta \mu^\theta < 0), \\ 1 - \theta e^{(1-\theta)(r_{i-1} - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta} - (1-\theta) e^{-\theta(r_i - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta} & \\ \qquad (r_i - \beta_j^\theta + s_j^\theta \mu^\theta > 0, \ r_{i-1} - \beta_j^\theta + s_j^\theta \mu^\theta < 0), & \\ (1-\theta)\left\{e^{-\theta(r_{i-1} - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta} - e^{-\theta(r_i - \beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta}\right\} & (r_{i-1} - \beta_j^\theta + s_j^\theta \mu^\theta > 0). \end{cases}$$

THEOREM 2. *Suppose that* y_t *follows model* (4) *and* ϵ_t^θ *follows the asymmetric Laplace distribution. Let* $e^{\mathrm{T}} = (e_1, \ldots, e_w)$. *Then* $E_\theta(y_{t+m} \mid y_t^*)$ *takes the following form*:

$$\text{for } m = 1, \ e_i = E_\theta(y_{t+1} \mid y_t^*) = s_i^\theta(1-2\theta)/\{\theta(1-\theta)\} + \beta_i^\theta - s_i^\theta \mu^\theta, \ \text{if } y_t \in \Omega_i;$$

$$\text{for } m > 1, \ E_\theta(y_{t+m} \mid y_t^*) = e^{\mathrm{T}} H^{m-2} g.$$

THEOREM 3. *Suppose that* y_t *follows model* (4) *and* ϵ_t^θ *follows the asymmetric Laplace distribution. Then* $E_\theta(y_{t+m}^2 \mid y_t^*)$ *takes the following form: for* $m = 1$, $E_\theta(y_{t+1}^2 \mid y_t^*) = v_i$ *if* $y_t \in \Omega_i$; *for* $m > 1$, $E_\theta(y_{t+m}^2 \mid y_t^*) = v^{\mathrm{T}} H^{m-2} g$, *where* $v^{\mathrm{T}} = (v_1, \ldots, v_w)$, *and*

$$v_i = (s_i^\theta)^2\{2\theta/(1-\theta)^2 + 2(1-\theta)/\theta^2\} + 2s_i^\theta(\beta_i^\theta - s_i^\theta\mu^\theta)(1-2\theta)/\{\theta(1-\theta)\} + (\beta_i^\theta - s_i^\theta \mu^\theta)^2.$$

COROLLARY 2. *Suppose that* y_t *follows model* (4) *and* ϵ_t^θ *follows the asymmetric Laplace distribution. Then* $var(y_{t+m} \mid y_t^*)$ *takes the following form: for* $m = 1$, $var(y_{t+1} \mid y_t^*) = v_i - e_i^2$ *if* $y_t \in \Omega_i$; *for* $m > 1$, $var(y_{t+m} \mid y_t^*) = v^{\mathrm{T}} H^{m-2} g - (e^{\mathrm{T}} H^{m-2} g)^2$.

4·2. *Example*

We have considered several examples, but to save space we only present one. Suppose that y_t follows the model

$$y_{t+1} = \begin{cases} -3{\cdot}5 + 0{\cdot}5(\epsilon_{t+1}^\theta - \mu^\theta) & (y_t \leqslant -3), \\ 1{\cdot}5 + 1{\cdot}0(\epsilon_{t+1}^\theta - \mu^\theta) & (-3 < y_t \leqslant 1), \\ 6{\cdot}5 + 1{\cdot}5(\epsilon_{t+1}^\theta - \mu^\theta) & (y_t > 1), \end{cases}$$

where ε_t^θ follows the asymmetric Laplace distribution. We take $y_0 = 1$. For other initial values that we tried, we got very similar results to those that we now present.

We used the Monte Carlo method and the numerical method to obtain $f_\theta(y_{t+m} \mid y_t^*)$ for $\theta = 0{\cdot}05$, $0{\cdot}25$, $0{\cdot}5$, $0{\cdot}75$ and $0{\cdot}95$, $m = 1, 2, 3, 4$. We also used the theoretical results developed above to calculate the corresponding theoretical predictive density functions. The Monte Carlo results are based on 10 000 samples. To save space, we only show the results for $f_{0{\cdot}75}(y_{t+3} \mid y_t^*)$

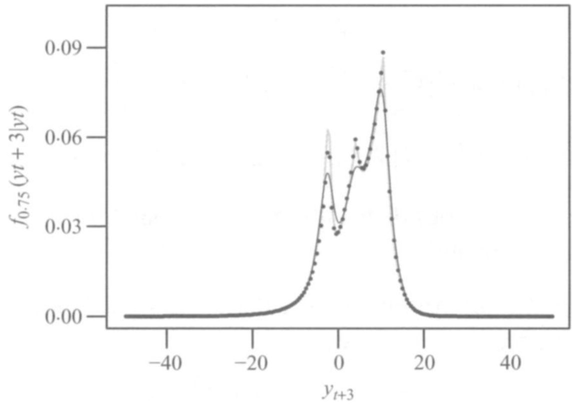

Fig. 1. Predictive density function $f_{0\cdot75}(y_{t+m} \mid y_t^*)$. The lighter curve corresponds to the theoretical results, the dots are the values calculated by using the numerical method, and the darker curve corresponds to the Monte Carlo method.

Table 1. *Mean square errors between theoretical and estimated predictive means, and standard deviations in brackets, of y_{t+m} given y_t^* ($m = 1, 2, 3, 4$)*

θ	0·05	0·25	0·5	0·75	0·95
Theoretical and numerical	0·5 (0·4)	10·1 (0·3)	2·3 (0·3)	1·1 (0·2)	0·1 (4·9)
Theoretical and Monte Carlo	6·2 (11·1)	0·4 (3·5)	0·2 (4·5)	0·9 (3·2)	19·7 (32·6)

All entries are multiplied by 100.

in Fig. 1. Clearly, the results obtained from the numerical method are almost always very close to the theoretical curve.

Once a discrete version of $f_\theta(y_{t+m} \mid y_t^*)$ is available, we can obtain any predictive quantities of interest. Table 1 shows the mean square errors between the theoretical and the predicted means and standard deviations in brackets of y_{t+m} ($m = 1, 2, 3, 4$) given y_t^*. Clearly, all the mean square errors are acceptable and the overall results are very good.

All our examples show that $f_\theta(y_{t+m} \mid y_t^*)$ usually has multiple modes for $m > 1$, and the numerical results are always very close to the theoretical results, while the Monte Carlo results are much less accurate near the multiple modes of the theoretical density functions even for much larger sample sizes. The mean square errors for the numerical method do not depend on the value of θ, while for the Monte Carlo method they usually increase as θ goes to extremes. However, both methods perform similarly in estimating $E_\theta(y_{t+m} \mid y_t^*)$. On the other hand, the Monte Carlo method is faster than the numerical method for large m, and it needs less computer storage space. Therefore, we suggest using the numerical method if possible, otherwise the simulation method could be used, as for the large-scale simulation studies given in the next subsection. Our experience is that a poor accuracy in estimating $f_\theta(y_{t+m} \mid y_t^*)$ may cause the estimated $q^\theta_{x_{t+m}|x_t^*}$ to violate the monotonic property of a quantile function, leading to a failure of the method. Therefore, high accuracy in estimating $f_\theta(y_{t+m} \mid y_t^*)$ plays an important role in the forecasting method.

4·3. *Simulation studies*

We have shown that the methods for obtaining $f_\theta(y_{t+m} \mid y_t^*)$ and $E_\theta(y_{t+m} \mid y_t^*)$ often work very well. Now we study the performance of the method for obtaining $q^\theta_{x_{t+m} \mid x_t^*}$.

Table 2. *True coefficient values of model* (3) *corresponding to model* (5)

	θ	0·005	0·05	0·10	0·25	0·5	0·75	0·90	0·95	0·995
Case A	β_{10}^{θ}	−3·08	−2·15	−1·78	−1·17	−0·50	0·17	0·78	1·15	2·08
	β_{20}^{θ}	−2·08	−1·15	−0·78	−0·17	0·50	1·17	1·78	2·15	3·08
	β_{30}^{θ}	−1·88	−0·95	−0·58	0·03	0·70	1·37	1·98	2·35	3·28
Case B	β_{10}^{θ}	−0·63	−3·42	−2·39	−1·32	−0·50	0·32	1·39	2·42	5·34
	β_{20}^{θ}	−4·10	−1·85	−1·14	−0·27	0·50	1·27	2·14	2·85	5·10
	β_{30}^{θ}	−3·33	−1·43	−0·83	−0·04	0·70	1·44	2·23	2·83	4·73
Case C	β_{10}^{θ}	−0·50	−0·47	−0·43	−0·31	−0·04	0·42	1·04	1·50	3·03
	β_{20}^{θ}	0·50	0·52	0·54	0·62	0·78	1·06	1·42	1·70	2·62
	β_{30}^{θ}	0·70	0·73	0·77	0·89	1·16	1·62	2·24	2·70	4·23

Simulation study 1. Consider a time series model given by

$$x_t = \begin{cases} -0{\cdot}5 - 0{\cdot}4x_{t-1} - 0{\cdot}1x_{t-2} + \xi_{1t} & (x_{t-1} \leqslant -15), \\ 0{\cdot}5 + 0{\cdot}3x_{t-1} - 0{\cdot}2x_{t-2} + \xi_{2t} & (-15 < x_{t-1} \leqslant 0), \\ 0{\cdot}7 + 0{\cdot}4x_{t-1} - 0{\cdot}6x_{t-2} + \xi_{3t} & (0 < x_{t-1}), \end{cases} \tag{5}$$

where ξ_{it} follow the distributions given below.

Case A: ξ_{it} $(i = 1, 2, 3)$ follow a $N(0, 1)$ distribution. Case B: ξ_{it} $(i = 1, 2, 3)$ follow a t-distribution with degrees of freedom 3, 4, 5, respectively. Case C: ξ_{it} $(i = 1, 2, 3)$ follow an exponential distribution with rate 1·5, 2·5, 1·5, respectively.

In each case, 1000 samples each of size 2000 were generated from x_{t+m} given x_t^* $(m = 1, \ldots, 20)$ by using model (5). From each sample we estimated the θth quantile $Q_i^{\theta}(x_{t+m} \mid x_t^*)$ $(i = 1, \ldots, 1000)$. Let $Q^{\theta}(x_{t+m} \mid x_t^*) = \sum_{i=1}^{1000} Q_i^{\theta}(x_{t+m} \mid x_t^*)/1000$ be the final estimate of the θth quantile of x_{t+m} given x_t^*, where $\theta = 0{\cdot}005, 0{\cdot}05, 0{\cdot}10, 0{\cdot}25, 0{\cdot}5, 0{\cdot}75, 0{\cdot}90, 0{\cdot}95$ and $0{\cdot}995$. We also used our forecasting method to estimate $q^{\theta}_{x_{t+m}|x_t^*}$ from model (3), where the coefficients β_{i0}^{θ} $(i = 1, 2, 3)$ are given in Table 2 and all other parameter values are the same as those in model (5). We should expect that $q^{\theta}_{x_{t+m}|x_t^*} \approx Q^{\theta}(x_{t+m} \mid x_t^*)$ if our forecasting method works well.

The three rows of Fig. 2 show the two sets of quantile forecasts obtained from model (3) and model (5) corresponding to cases A, B and C, respectively. Again, to save space, we only show plots for $m = 1$, 10 and 20 in each case. These show that the overall differences between the two sets of quantile forecasts are very small except for extreme quantiles, and the performance of the method does not depend on the distribution of the error terms in (5). We also calculated the mean square errors between the two sets of quantile forecasts over $m = 1, \ldots, 20$ for each θ value. We found that all the mean square errors are less than unity except for two of them corresponding to extreme quantiles. Lack of information may be the reason for having large mean square errors in these cases.

Simulation study 2. The effects of the standard deviation of ξ_{it} in (5) on the performance of the forecasting method were investigated. We found that predictive medians always have a much higher accuracy than other quantiles, independently of the standard deviation of ξ_{it}. Therefore, the predictive medians can be taken as point estimates of the future values.

5. Forecasting for United States Gross National Product data

Consider a time series z_t comprising the quarterly United States Gross National Product in 1982 dollars from the first quarter of 1947 to the first quarter of 1991. The data have been seasonally

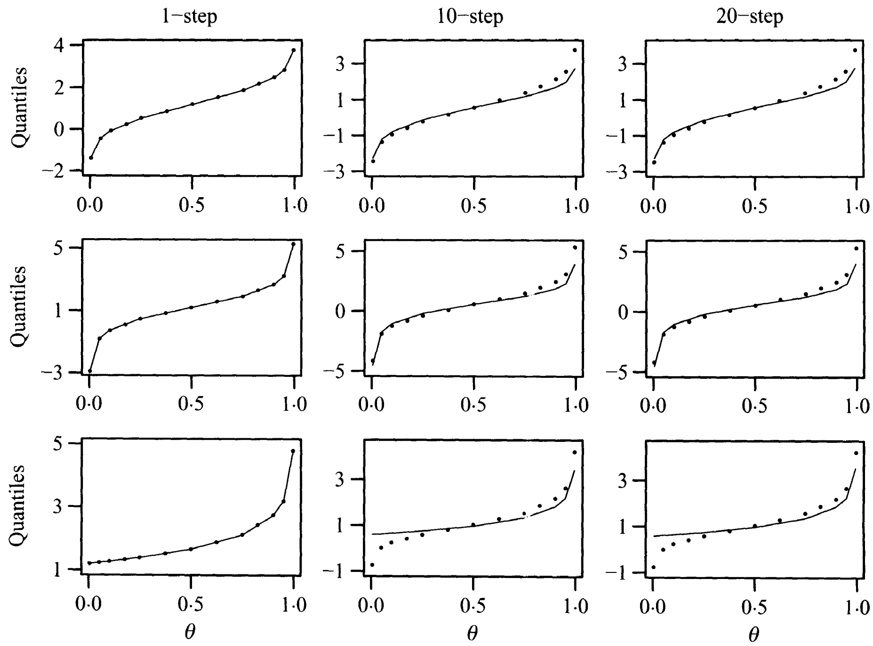

Fig. 2. Predictive quantiles of x_{t+m} given x_t^* in Simulation Study 1. First row, Case A; second row, Case B; third row, Case C. Dotted curves were obtained from model (5), continuous curves from model (3).

Table 3. *Estimated parameter values and corresponding standard errors in brackets for the United States Gross National Product*

θ	0·05	0·25	0·5	0·75	0·95
β_{10}^{θ}	−0·51 (0·54)	−0·06 (0·11)	0·01 (0·08)	0·09 (0·12)	0·52 (0·51)
β_{20}^{θ}	−0·14 (0·15)	−0·02 (0·03)	0·01 (0·02)	0·03 (0·03)	0·16 (0·15)

adjusted. In this application, we consider the growth rate $x_t = \log(z_t) - \log(z_{t-1})$ yielding a time series of length $n = 176$ shown in Fig. 3.

Cai and Stander (2008) fitted model (2) to the data up to time $t = 172$. The fitted model is $q^{\theta}_{x_t|x^*_{t-1}} = \beta^{\theta}_{i0}$, where $i = 1$ if $x_{t-d^{\theta}} \leqslant 0$ and $i = 2$ otherwise, with $d^{\theta} = 3$ and β^{θ}_{i0} being shown in Table 3. Consider the auxiliary model $y_t = \beta^{\theta}_{i0} + s^{\theta}_i(\epsilon^{\theta}_t - \mu^{\theta})$, where $i = 1$ if $y_{t-3} \leqslant 0$ and $i = 2$ otherwise. Let $y_t = x_t$ $(t = 1, \ldots, 172)$. Note that $s^{\theta}_i = \sigma_i / b^{\theta}$ and σ_1 and σ_2 are the sample standard deviations of the observed process in each subregion. In this case, we have $\sigma_1 = 0{\cdot}014$, $\sigma_2 = 0{\cdot}0099$.

For $\theta = 0{\cdot}05, 0{\cdot}25, 0{\cdot}5, 0{\cdot}75, 0{\cdot}95$ and $m = 1, 2, 3, 4$, we calculated $E_{\theta}(y_{t+m} \mid x^*_t)$ first by the Monte Carlo method and then by the numerical method. By letting $q^{\theta}_{x_{t+m}|x^*_t} \approx E_{\theta}(y_{t+m} \mid x^*_t)$, we obtained two sets of quantile forecasts that are also shown in Fig. 3. The intercepts between the curves and the vertical lines give the positions of the predicted medians. It can be seen that the two sets of predicted quantiles are very similar, with the predicted medians being almost identical. This is in good agreement with what we have found in the simulation studies. We also compared the observed values with the predicted medians obtained by the method developed in this paper and by that of Cai (2007). We found that the largest mean square error is 0·14 and the largest standardized mean square error is 0·75, indicating that for this dataset all the forecasting methods provide reasonable point forecasts.

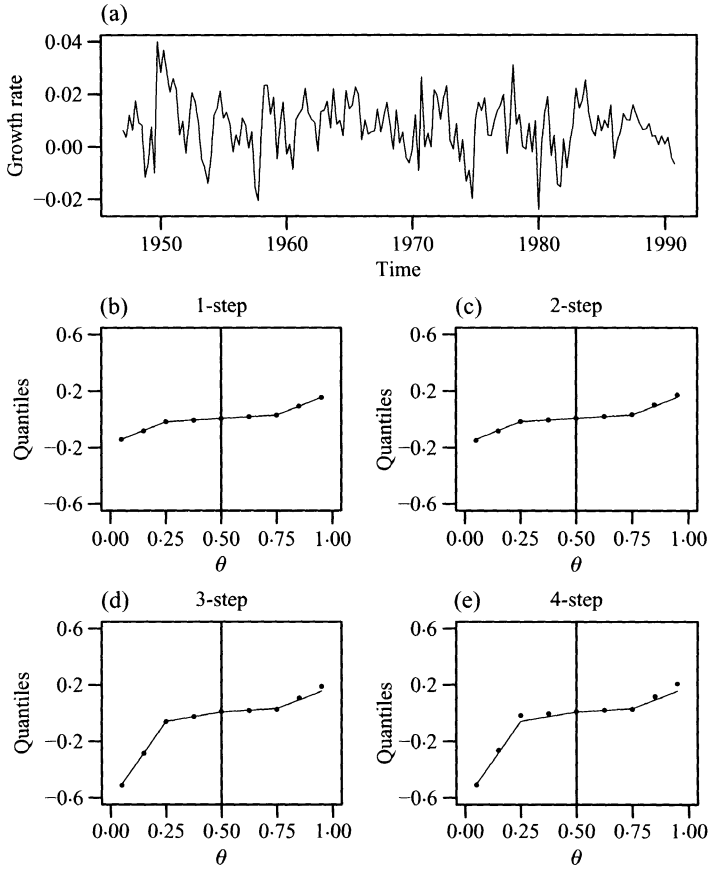

Fig. 3. (a) Time series plot of the growth rate of the United States Gross National Product. (b–e) Quantile forecasts for the growth rates of the United States Gross National Product data. Continuous curves correspond to the Monte Carlo method, dotted curves correspond to the numerical method and the vertical lines correspond to $\theta = 0{\cdot}5$.

Acknowledgement

We thank the editor, three referees and Dr. J. Stander for their very constructive comments and suggestions, which have significantly enhanced the quality and presentation of this paper.

Appendix
Technical details

Proof of Theorem 1. For $m = 1$, the result holds. For $m = 2$, we have

$$\begin{aligned} f_\theta(y_{t+2} \mid y_t^*) &= \int_{-\infty}^{\infty} f_\theta(y_{t+2} \mid y_{t+1}^*) f_\theta(y_{t+1} \mid y_t^*) dy_{t+1} \\ &= \sum_{j=1}^{w} \left(s_j^\theta\right)^{-1} f\left\{\left(y_{t+2} - \beta_j^\theta + s_j^\theta \mu^\theta\right)/s_j^\theta\right\} \int_{\Omega_j} f_\theta(y_{t+1} \mid y_t^*) dy_{t+1} = h_2^{\mathrm{T}} H^0 g \end{aligned}$$

as required. Then the result can be proved by using induction. □

Proof of Corollary 1. If $y_t \in \Omega_j$, then

$$g_i = \int_{\Omega_i} f_\theta(y_{t+1} \mid y_t^*) dy_{t+1} = \int_{\Omega_i} \left(s_j^\theta\right)^{-1} f\left\{\left(y_{t+1} - \beta_j^\theta + s_j^\theta \mu^\theta\right)/s_j^\theta\right\} dy_{t+1} = h_{ij}$$

as required. Now it follows from the definition that we have the required expression

$$\begin{aligned} h_{ij} &= \int_{r_{i-1}}^{r_i} \left(s_j^\theta\right)^{-1} \theta(1-\theta) e^{-\left\{\left(y-\beta_j^\theta + s_j^\theta \mu^\theta\right)/s_j^\theta\right\}\left\{\theta - I_{[(y-\beta_j^\theta + s_j^\theta \mu^\theta)/s_j^\theta < 0]}\right\}} dy \\ &= \int_{\left(r_{i-1}-\beta_j^\theta + s_j^\theta \mu^\theta\right)/s_j^\theta}^{\left(r_i-\beta_j^\theta + s_j^\theta \mu^\theta\right)/s_j^\theta} \theta(1-\theta) e^{-u\left(\theta - I_{[u<0]}\right)} du. \end{aligned}$$

□

Proof of Theorem 2. For $m = 1$ the result is true. For $m \geqslant 2$, it follows from Theorem 1 that we have $E_\theta(y_{t+m} \mid y_t^*) = \int_{-\infty}^{\infty} y_{t+m} h_m^{\mathrm{T}} H^{m-2} g dy_{t+m} = (e_1, \ldots, e_w)^{\mathrm{T}} H^{m-2} g$, where

$$\begin{aligned} e_i &= \int_{-\infty}^{\infty} \left(s_i^\theta\right)^{-1} y_{t+m} f\left\{\left(y_{t+m} - \beta_i^\theta + s_i^\theta \mu^\theta\right)/s_i^\theta\right\} dy_{t+m} \\ &= \int_{-\infty}^{\infty} \theta(1-\theta)\left(s_i^\theta u - s_i^\theta \mu^\theta + \beta_i^\theta\right) e^{-u\left(\theta - I_{[u<0]}\right)} du. \end{aligned}$$

□

Proof of Theorem 3. We only need to find the values of v_i. But

$$v_i = \left(s_i^\theta\right)^{-1} \int_{-\infty}^{\infty} y^2 \theta(1-\theta) e^{-\left\{\left(y-\beta_i^\theta + s_i^\theta \mu^\theta\right)/s_i^\theta\right\}\left\{\theta - I_{[(y-\beta_i^\theta + s_i^\theta \mu^\theta)/s_i^\theta < 0]}\right\}} dy.$$

□

Proof of Corollary 2. The result follows from Theorems 2 and 3. □

REFERENCES

CAI, Y. (2003). Convergence theory of a numerical method for solving Chapman–Kolmogorov equation. *SIAM J. Numer. Anal.* **40**, 2337–51.

CAI, Y. (2005). A forecasting procedure for nonlinear autoregressive time series models. *J. Forecasting* **24**, 335–51.

CAI, Y. (2007). A quantile approach to US GNP. *Econ. Mod.* **24**, 969–79.

CAI, Y. & STANDER, J. (2008). Quantile self-exciting autoregressive time series models. *J. Time Ser. Anal.* **29**, 186–202.

CHAN, K. S. & TONG, H. (2002). A note on the equivalence of two approaches for specifying a Markov process. *Bernoulli* **8**, 117–22.

GRANGER, C. W. J. & TERÄSVIRTA, T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford: Oxford University Press.

KOENKER, R. (2005). *Quantiles Regression.* Cambridge: Cambridge University Press.

KOENKER, R. & MACHADO, J. (1999). Goodness of fit and related inference processes for quantile regression. *J. Am. Statist. Assoc.* **94**, 1296–1309.

KOENKER, R. & XIAO, Z. (2004). Unit root quantile autoregression inference. *J. Am. Statist. Assoc.* **99**, 775–87.

KOENKER, R. & XIAO, Z. (2006). Quantile autoregression. *J. Am. Statist. Assoc.* **101**, 980–90.

TONG, H. (1990). *Non-linear Time Series, A Dynamical System Approach.* Oxford: Oxford University Press.

TONG, H. & LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. R. Statist. Soc.* B **42**, 245–92.

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

WONG, C. S. & LI, W. K. (2000). On a mixture autoregressive model. *J. R. Statist. Soc.* B **62**, 95–115.

YAO, Q. & TONG, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Nonparam. Statist.* **6**, 273–92.

[*Received July* 2008. *Revised June* 2009]