Estimation of Conditional Quantiles by a New Smoothing Approximation of Asymmetric Loss Functions

G.H. Zhao [1], K.L. Teo [2], and K.S. Chan [3]

November, 2002

Abstract

In this paper, nonparametric estimation of conditional quantiles of a nonlinear time series model is formulated as a nonsmooth optimization problem involving asymmetric loss function. This asymmetric loss function is nonsmooth and is of the same structure as the so-called 'lobsided' absolute valued function. Using an effective smoothing approximation method introduced for this 'lobsided' absolute value function, we obtain a sequence of approximate smooth optimization problems. Some important convergence properties of the approximation are established. Each of these smooth approximate optimization problems is solved by the optimization algorithm based on the sequential quadratic programming approximation with active set strategy. The proposed approach is compared with an approach proposed by Yao and Tong (1996) through some empirical numerical studies using simulated data and the classic lynx pelt series. In particular, the empirical performance of the proposed approach is comparable with that of the Yao-Tong approach in the ideal cases for the latter, but otherwise it outperforms the Yao-Tong approach at the expense of increasing the computation time by 4 to 8 folds for the numerical examples studied here.

Some Key Words: Asymmetric loss function, Nonlinear Time Series, Optimization, Prediction, Smooth approximation, Lynx data

1 Introduction

In forecasting, it is often required and, indeed, is desirable to compute prediction intervals. Prediction intervals may be derived from a model-based approach or a nonparametric approach. Here, we focus on nonparametric prediction intervals. Two general approaches for constructing nonparametric prediction intervals are highest density prediction intervals and equal tail prediction intervals (Box and Tiao, 1973). The latter approach is more popular owing to its ease of construction. An equal tail interval is constructed by excluding equal percent of smallest and largest values. In the case of predicting a continuous random variable, say, Y, given the covariate $X = x$, a $(1-\alpha) \times 100\%$ prediction interval is equal to $(q_{\alpha/2}(x), q_{1-\alpha/2}(x))$, where $q_\alpha(x)$ is the conditional α-quantile defined by the equation

$$\alpha = P\{Y \leq q_\alpha(x) \mid X = x\}, \tag{1}$$

for $0 < \alpha < 1$. We assume that the conditional quantiles are uniquely defined. It is well-known (Koenker and Bassett, 1978) that the conditional α-quantile $q_\alpha(x)$ minimizes the following asymmetric loss function.

$$q_\alpha(x) = \arg\min_a E\{L_{1,\alpha}(Y-a) \mid X = x\}, \tag{2}$$

where E denotes the mathematical expectation, and

$$L_{1,\alpha}(y) = \begin{cases} (1-\alpha)|y| & y \leq 0, \\ \alpha y & y > 0. \end{cases} \tag{3}$$

Estimation of $q_\alpha(x)$ can be achieved by minimizing a sample version of $E\{L_{1,\alpha}(Y - a) \mid X = x\}$. Often, $q_\alpha(x)$ is assumed to belong to a parameterized class of functions, e.g., linear functions or splines. In some such cases, the conditional α-quantiles can be estimated by solving some linear programming problems which admit several efficient numerical algorithms; see Portnoy and Koenker (1997) and the references therein.

We consider the case for which $q_\alpha(x)$ is a smooth function of x. Furthermore, we focus on the case of m-step ahead prediction with data generated from a stationary pth order nonlinear autoregression, i.e., $Y = Y_t$ and $X = (Y_{t-m}, Y_{t-m-1}, \cdots, Y_{t-m-p+1})^T$, where the superscript T denotes the transpose and m often equals 1. Yao and Tong (1996) (see also Chan and Tong, 2001) developed an approximate solution to this problem within the framework of the following multi-step forecasting model:

$$Y_t = h_m(X_t) + g_m(X_t)e_t, \tag{4}$$

where $h_m : R^d \to (0, \infty)$, $g_m : R^d \to (0, \infty)$, e_t is a sequence of independent and identically distributed (iid) random variables of zero mean and finite non-zero variance, and e_t is independent of X_t. Yao and Tong (1996) adopted the local polynomial approach (Fan and Gijbels, 1996) to approximate $q_\alpha(\cdot)$ around x by its first order Taylor approximation, i.e.,

$$q_\alpha(u) \approx a + b^T(u - x), \tag{5}$$

where the superscript T denotes the transpose, so that $a = q_\alpha(x)$. Now, a can be estimated by minimizing the loss function

$$\sum_s L_{1,\alpha}(Y_s - a - b^T(X_s - x))k(X_s - x; h), \tag{6}$$

where the summation is over all data cases for which Y_s and X_s are defined, and $k(x; h)$ is kernel function often taken as the multivariate normal probability density function (pdf) with zero mean and covariance matrix $h^2 I$, where I is the identity matrix, and h is the bandwidth parameter. In practice, the optimal bandwidth parameter has to be estimated from the data. Here, we focus on the problem of developing an efficient and accurate method for minimizing the loss function (6), assuming that the bandwidth parameter is given. Minimizing the loss function (6) is a nonsmooth optimization problem, as $L_{1,\alpha}(y)$ is not a smooth function. Thus, all the available efficient smooth optimization techniques are not applicable. To tackle this problem, Yao and Tong (1996) introduced the α-conditional expectile $\tau_{\alpha,m}$ of Y given $X = x$ which is defined by

$$\tau_{\alpha,m} = \arg\min_a E\{L_{2,\alpha}(Y - a) \mid X = x\} \tag{7}$$

where

$$L_{2,\alpha}(y) = \begin{cases} (1-\alpha)y^2 & y \le 0, \\ \alpha y^2 & y > 0. \end{cases}$$

Similar to (6), the conditional expectile can be estimated by minimizing the following loss function

$$\sum_s L_{2,\alpha}(Y_s - a - b^T(X_s - x))k(X_s - x; h). \tag{8}$$

Yao and Tong (1996) have shown that the conditional α-quantile of Y given $X = x$, denoted as $q_{\alpha,m}(x)$, is equal to a corresponding conditional expectile $\tau_{\beta,m}$, where

$$\beta = \beta(\alpha, x) = \frac{\alpha q_\alpha^e - E\{e_t I_{\{e_t \le q_\alpha^e\}}\}}{2E\{e_t I_{\{e_t - q_\alpha^e\}}\} - (1-2\alpha)q_\alpha^e}, \tag{9}$$

with q_α^e being the α-quantile of the noise distribution.

The significance of this result is that by solving a smooth optimization problem (8) with α taken as β (here, β is defined by (9)), one gets the optimal solution to the original optimization problem (6). Although the approach of Yao and Tong (1996) is numerically convenient and mathematically elegant, it is only applicable to cases under the following two conditions.

Condition 1. The model defined by (4) holds, which assumes that all m-step conditional distributions are obtained from the common distribution of e_t by some location-scale change.

Condition 2. The common distribution of e_t has to be specified in order to compute (9).

For convenience, the common distribution of e_t is usually taken as the normal distribution. Although the second condition can be relaxed by further modeling the shape of the noise distribution, this will greatly complicate the Yao-Tong method. Here, we introduce an alternative approach for solving (4) by approximating $L_{1,\alpha}(y)$ in (6) by a new smoothing approximation introduced by Jennings, Wong and Teo (1996), leading to an approximate smooth optimization problem for the estimation of $q_{\alpha,m}(x)$, the conditional α-quantile of Y given $X = x$. (Below, we shall focus on the case $m = 1$, for simplicity.) This approach relaxes the preceding two conditions. In particular, our approximate smooth optimization approach can be readily extended to provide an alternative approach to computing parametric nonlinear quantile regression estimates (Koenker and Park, 1996), which is an interesting future research problem.

For fixed x, the conditional quantile $q_\alpha(x)$ is an increasing function of α. This monotonicity property guarantees that the (true) equal tail prediction intervals enjoy a desirable property, namely, they are nested so that any prediction interval is contained in another prediction interval that has a higher prediction probability. As an illustration, a 90% prediction interval is always a sub-interval of its 95% counterpart for predicting Y given $X = x$. An estimation scheme that preserves the monotonicity property of the conditional quantiles is said to be monotone. The Yao-Tong estimator and the proposed estimator of the conditional quantiles need not be monotone, as shown by an example in section 3. However, an estimation scheme, say S, can be easily modified to be monotone by the following device. Let $\{0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{2m+1} < 1\}$ be an equally-spaced fine partition of the unit interval $[0, 1]$. We first get the estimate $\hat{q}_{\alpha_{m+1}}(x)$ using the estimation scheme S. Then we estimate $q_{\alpha_k}(x)$ for $m \geq k \geq 1$ one by one in a downward manner.

Specifically, given the estimate $\hat{q}_{\alpha_k}(x)$, we estimate $q_{\alpha_{k-1}}(x)$ using the method S with the constraint that $q_{\alpha_{k-1}}(x) \leq \hat{q}_{\alpha_k}(x)$. Similarly, we estimate $q_{\alpha_k}(x)$ for $m+2 \leq k \leq 2m+1$ one by one in an upward manner so that $q_{\alpha_k}(x)$ is estimated by method S subject to the constraint that $q_{\alpha_k}(x) \geq \hat{q}_{\alpha_{k-1}}(x)$. For any arbitrary $0 < \alpha < 1$, $q_\alpha(x)$ can be estimated by linear interpolation or extrapolation on the logistic scale of α. The last step is not needed if the grid has already contained all the αs of interest. In section 3, we illustrate by an example that using this device may improve the performance of both the Yao-Tong method and the proposed method.

The rest of the paper is organized as follows. Section 2 introduces this new smoothing approximation of $L_{1,\alpha}(y)$ that is continuously differentiable but the second derivative is discontinuous at a finite number of points. We derive two important convergence properties of the global minimizer of the approximate smooth objective function to that of the original objective function, as the amount of smoothing decreases to zero; all proofs are deferred to Appendices 1 and 2. Section 3 reports some empirical numerical studies comparing the performance of the new approach with that of the Yao-Tong approach, using simulated data and the classic lynx pelt series.

2 Smoothing Approximation

As discussed in Section 1, the estimation of the conditional α-quantile $q_\alpha(\cdot)$ defined by (2) can be achieved via minimizing a sample version of $E\{L_{1,\alpha}(Y-a) \mid X = x\}$. Following Yao and Tong (1996), we consider the approximation of $q_\alpha(x)$ around x by its Taylor approximation, which gives (6) such that $a = q_\alpha(x)$. Now, (a, b) can be estimated by minimizing the loss function (6), which is recalled in the following.

$$f(a,b) = \sum_s L_{1,\alpha}(Y_s - a - b^T(X_s - x))k(X_s - x; h) \tag{10}$$

where $\alpha \in [0,1]$ and $L_{1,\alpha}$ is defined by (3). Clearly, minimizing the cost function (10) is a nonsmooth optimization problem, as it contains a nonsmooth function $L_{1,\alpha}$. Currently there is no efficient optimization algorithm for minimizing the nonsmooth loss function f.

On the other hand, there are many efficient numerical methods for smooth optimization problems, such as the sequential quadratic programming approximation with active set strategy; see, for example, Teo, Goh and Wong (1991). To use these smooth optimization techniques, we need to smooth the corner of the asymmetric loss function $L_{1,\alpha}$. Of the many ways of doing this, we choose the following smooth approximation due to Jenning, Wong and Teo (1996).

$$L_{1,\alpha}^{\delta}(x) = \begin{cases} \alpha x & \text{if } \alpha x > \delta \\ (\delta^2 + (\alpha x)^2)/2\delta & \text{if } 0 \le \alpha x \le \delta \\ (\delta^2 + ((1-\alpha)x)^2)/2\delta & \text{if } -\delta \le (1-\alpha)x \le \delta \\ (\alpha - 1)x & \text{if } (\alpha-1)x < -\delta. \end{cases} \tag{11}$$

The main reasons for such a choice are given as follows:

(i) $L_{1,\alpha}^{\delta}(x)$ is continuously differentiable.

(ii) It has the minimum at the same place as the original function $L_{1,\alpha}(x)$, and satisfies the following properties.

$$0 \le L_{1,\alpha}^{\delta}(x) - L_{1,\alpha}(x) \le \delta/2 \tag{12}$$

(iii) The smooth approximation possesses some desired convergence properties to be presented in the following two theorems.

By approximating $L_{1,\alpha}(x)$ by $L_{1,\alpha}^{\delta}(x)$, the objective function (10) becomes:

$$f^{\delta}(a,b) = \sum_s L_{1,\alpha}^{\delta}(Y_s - a - b^T(X_s - x))k(X_s - x; h). \tag{13}$$

It is easy to see that

$$0 \le f^{\delta}(a,b) - f(a,b) \le \frac{\delta}{2} \sum_s k(X_s - x; h) \tag{14}$$

Theorem 2.1 *Let $(a^{\delta,*}, b^{\delta,*})$ and (a^*, b^*) be optimal solutions of the optimization problems (12) and (10), respectively. Then,*

$$0 \le f(a^{\delta,*}, b^{\delta,*}) - f(a^*, b^*) \le \frac{\delta}{2} \sum_s k(X_s - x; h).$$

Theorem 2.1 shows that, by minimizing the approximate smooth loss function (13), the solution $(a^{\delta,*}, b^{\delta,*})$ obtained will give rise to the cost for which its deviation from the optimal cost can be made as tight as we please by reducing the parameter δ. The following theorem shows that the minimizer of the approximate objective function converges to that of the original objective function as the smoothing parameter $\delta \to 0$.

Theorem 2.2 *Let* $(a^*, b^*) = \arg\min f(a,b)$ *and* $(a^{\delta,*}, b^{\delta,*}) = \arg\min f^\delta(a,b)$. *If there exists a unique minimizer of* $f(a,b)$, *then* $\lim_{\delta\to 0}(a^{\delta,*}, b^{\delta,*}) = (a^*, b^*)$.

This approximation scheme requires the specification of the smoothing parameter δ and a stopping rule. A simple stopping rule is to keep halving δ until, for example, the difference between the successive minima is less than a pre-specified tolerance level, say, 0.0001; see Appendix 3 for the more complex stopping rule used in the studies reported below.

3 Numerical Experiment

The proposed approach will be applied to two examples. The first one uses some simulated data, while the second one uses the classic lynx pelt series. The optimization algorithm detailed in the previous section was implemented in C using the optimization software package CFSQP by E.R. Panier, A.L. Tits, J.L. Zhou and C.T. Lawrence, see Lawrence, Zhou and Tits (1997). Our algorithm codes are used to smooth two examples and are available upon request to the authors; see Appendix 3.

3.1 Simulated Data

We evaluate our new approach using Monte Carlo. The simulation model is the first order Auto-Regressive (AR(1)) Model:

$$Y_t = \phi Y_{t-1} + e_t, \tag{15}$$

where $X_t = Y_{t-1}$. This linear model implies that the conditional α-quantile is linear in x so that $q_\alpha(x)$ can be estimated by minimizing the loss function (6) with the kernel function there dropped. We deliberately choose this simple model so that the new approach can be compared with the Yao-Tong approach without the confounding issue of having to choose the bandwidth. For this linear case, a simplex-based modified Barrodale-Roberts algorithm (see Koenker and d'Orey, 1987) is a more efficient algorithm. However, as our main focus is on nonlinear time series, we do not include the simplex-based algorithm in the comparison. Furthermore, we assess the performance of each of these two approaches as follows. We first estimate the α-conditional quantile of Y_ℓ given $X_\ell = Y_{\ell-1}$ using all past observations up to and including $s \leq \ell - 1$ in the summation in (6), with the kernel function there dropped. Then, we compute the empirical relative frequencies that Y_ℓ is less than the conditional α-quantile for $M \leq \ell \leq 200$. Each experiment is replicated 1000 times, where M is set to be 60.

Table 1 reports the simulation results where ϕ is varied over $\{-0.9, -0.5, 0.0, 0.5, 0.9\}$. For each fixed ϕ, the errors are taken as t-distributed with d.f. equal to $\infty, 30, 20, 15, 10, 5, 3$ (recall that t with infinite d.f. is the standard normal distribution). We computed the conditional α-quantiles with $\alpha = 0.05$ to 0.95 with an increment of 0.05. When computing the Yao-Tong estimates of the conditional quantiles, (9) is computed as if the errors were normally distributed. This was done to examine the effect of misspecification in the Yao-Tong approach. Hence, for t-distributed errors with small d.f., the Yao-Tong estimate will suffer from systematic bias. We compare the two approaches in terms of the average (respectively, maximum) absolute deviation between the empirical relative frequencies and the nominal probabilities that the observations are less than the conditional quantiles, and also in terms of computation time (in seconds). All computations were done with HP PC with Intel(R) Pentium(R) 4 CPU 1800MHz AT/AT compatible and 523760 KB RAM on Microsoft Windows 2000 Service Pack 2.

Table 1 about here

Results from Table 1 suggest the following. For the conditional quantile estimator based on the proposed method, the empirical relative frequencies are generally close to the nominal probabilities, both in terms of mean absolute deviation and mean maximum absolute deviation. There is a very slight tendency of larger discrepancy between the empirical relative frequencies and the nominal probabilities with larger magnitude of the AR(1) coefficient, more so for a positive coefficient than the negative coefficient of equal magnitude. Overall speaking, the mean absolute deviations between the empirical relative frequencies and the nominal probabilities are about same for different degrees of freedom of the t-distributed noise. On the other hand, for the conditional quantile estimator based on the Yao-Tong method, the mean absolute deviations between the empirical relative frequencies and the nominal probabilities increase with the extent of misspecification of the error distribution, i.e. the smallness of the degrees of freedom of the t-distribution. When the error distribution is almost correctly specified, the proposed method and the Yao-Tong method are of comparable accuracy but the proposed method outperforms the Yao-Tong method when the specified error distribution is further from the true distribution. Note that the proposed method requires longer computational time than the Yao-Tong method, about seven to eight times longer.

3.2 Real Data Analysis

For real data analysis, we take up the classic set of (log-transformed) lynx pelt data collected from the McKenzie river region of Canada from 1821 to 1934; see Tong (1990) and Chan and Tong (2001). Consider the following general nonlinear time series model for the lynx data.

$$Y_t = f(X_t) + g(X_t)e_t, \tag{16}$$

where Y_t is the log-transformed lynx data at year t, $X_t = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})^T$, and $\{e_t\}$ is a sequence of iid errors that are independent of past values of Y. Clearly, $E(Y_t|X_t) = f(X_t)$. The order p is generally taken as between 2 and 4. We have repeated the analysis of the previous sub-section to the (standardized) log lynx data with the order from $p = 1$ to 4; the first order is included for curiosity. Specifically, we first estimated the α-conditional quantile of Y_ℓ given $X_\ell = (Y_{\ell-1}, Y_{\ell-2}, \cdots, Y_{\ell-p})^T$ using all past observations up to and including $s \leq \ell - 1$ in the summation in (6); α takes values from 0.05 to 0.95 with increment 0.05. Then, we computed the empirical relative frequencies that Y_ℓ is less than the conditional α-quantile for $M \leq \ell \leq 114$. The Yao-Tong estimators were computed by assuming that the error distribution is normal. Table 2 reports the average absolute deviation and the maximum absolute deviation between the empirical relative frequencies of the conditional quantiles and their nominal counterparts with $M = 30$ (60) and the bandwidth $h = 0.57$ that was adopted by Yao and Tong (1996). The proposed method results in more accurate estimates than the Yao-Tong estimates with ten to eighty percent of reduction in average (maximum) absolute deviation.

Table 2 about here

Recall that an estimation scheme is monotone if $\hat{q}_\alpha(x) < \hat{q}_\beta(x)$ whenever $\alpha < \beta$. Both the Yao-Tong method and the proposed method turned out to be non-monotone for the lynx example. Hence, we adopted the device described in the last but one paragraph in section 1 to render the two methods monotone. The results are reported in Table 3.

Compared with Table 2, it can be seen that both methods have been improved by imposing the monotonicity constraints, with the proposed method gaining more improvement owing to the fact that in its unconstrained form, the proposed method has a higher tendency of reversing the order of the conditional quantiles than the Yao-Tong method.

Table 3 about here

From a prediction perspective, a prediction scheme should be judged in terms of two criteria, namely, (i) the closeness of the empirical coverage rate to the nominal prediction probability of the interval and (ii) the shortness of the prediction interval. So far, our discussion is based on the first criterion. Indeed, this is reasonable for cases with a fixed set of covariate. However, with different order of the process, the covariate X_t contains different amount of information for predicting Y_t. The determination of p, the order of the process, should then depend on both criteria. In Table 3, we have also reported the average length of the 90% equal tail prediction intervals for Y_t given X_t and the corresponding empirical coverage rate, enclosed by parentheses, of these prediction intervals for both estimation schemes and under different order. It is interesting to note that the average length of the 90% prediction interval decreases with the order but there is a sharp drop from order 1 to order 2. Also, the Yao-Tong method tends to produce shorter intervals than the proposed method on average. The empirical coverage rates are all lower than the nominal 90%, with the discrepancy being larger with a higher order. Also, in terms of the empirical coverage rate of the 90% prediction interval, the Yao-Tong method and the proposed method are comparable. Thus, the better performance of the proposed method in terms of average absolute deviation and the maximum absolute deviation between the empirical relative frequencies of the conditional quantiles and their nominal probabilities do not carry over to the extreme quantiles; indeed, the situation is clearly presented in Figure 1 which plots the empirical relative frequencies of the conditional quantiles against their nominal counterparts for the proposed method and the Yao-Tong method. The graphs suggests that when the order is equal to 1, the Yao-Tong method performs poorly compared to the proposed method especially for intermediate α values, (ii) both methods do not do well for estimating the extreme quantiles for $p \geq 2$ and (iii) both methods perform very poorly for $p = 3$ and 4.

Figure 1 about here

We then repeat the above exercise by using larger bandwidth and with $M = 60$. The results are reported in Table 4 and plotted in Figure 2. The proposed method clearly outperforms the Yao-Tong method for larger bandwidths. The plots suggest that (i) for intermediate values of α, the conditional α-quantiles, from both methods, tend to have lower empirical frequencies than the nominal probabilities, and (ii) for α close to 0 and 1, the estimators are more accurate when $p = 2$. Note that with larger bandwidth, the 90% prediction intervals have coverage rates closer to 90%, at the expense of lengthening the intervals.

Table 4 and Figure 2 about here

Through the simulations in the previous subsection and the real data analysis here, we have demonstrated that the proposed method outperforms the Yao-Tong method for estimating conditional quantiles of nonlinear time series models, with fixed bandwidth. An interesting future problem is to devise methods for improving the coverage rates of the prediction intervals.

4 Acknowledgement

We thank Howell Tong for encouragement and very helpful comments on an earlier draft of the paper. This research is partially supported by the Research Grant Committee of Hong Kong.

Appendix 1

Proof of Theorem 2.1

Since $(a^{\delta,*}, b^{\delta,*})$ is an optimal solution of the optimization problem (12), we have, by (13),

$$f^{\delta}(a^{\delta,*}, b^{\delta,*}) \leq f^{\delta}(a^*, b^*) \leq f(a^*, b^*) + \frac{\delta}{2}\sum_s k(X_s - x; h) \tag{A1}$$

Using (13) again and (A1), it follows that

$$f(a^{\delta,*}, b^{\delta,*}) \leq f^{\delta}(a^*, b^*) \leq f(a^*, b) + \frac{\delta}{2}\sum_s k(X_s - x; h). \tag{A2}$$

Since (a^*, b^*) is an optimal solution of the optimization problem (10), we have, by (A2),

$$0 \leq f(a^{\delta,*}, b^{\delta,*}) - f(a^*, b^*) \leq \frac{\delta}{2}\sum_s k(X_s - x; h).$$

This completes the proof.

Appendix 2

Proof of Theorem 2.2

Observe that $f(a, b)$ and $f^{\delta}(a, b)$ are convex functions. Furthermore, it is clear that $f^{\delta}(a, b)$ decreases and converges to $f(a, b)$ uniformly as δ approaches to zero. If

$$\lim_{\delta \to 0}(a^{\delta,*}, b^{\delta,*}) \neq (a^*, b^*)$$

then there exists a monotonic sequence $\{\delta_i > 0\}$ such that $\lim_{i\to\infty} \delta_i = 0$ and $\lim_{i\to\infty}(a^{\delta_i,*}, b^{\delta_i,*}) = (\tilde{a}, \tilde{b}) \neq (a^*, b^*)$. However, in this case, we will have, by Theorem 2.1, (13) and the continuity of the function f,

$$f(a^*, b^*) = \lim_{i\to\infty} f^{\delta_i}(a^{\delta_i,*}, b^{\delta_i,*}) \geq \lim_{i\to\infty} f^{\delta_i}(a^{\delta_i}, b^{\delta_i}) \geq \lim_{i\to\infty} f(a^{\delta_i}, b^{\delta_i}) = f(\tilde{a}, \tilde{b})$$

This is a contradiction, and hence the proof is complete.

Appendix 3

Pseudo-codes for the Numerical Experiment and Data Analysis

Let N be the sample size. Set $\epsilon = 0.0001$, say, or some number that is small on the scale of the data spread. Below, the true objective function in the inner loop is $f_{t,\alpha}(a,b,x) = \sum_{s=1}^{t-1} L_{1,\alpha}(Y_s - a - b^T(X_s - x))k(X_s - x; h)$ while the corresponding approximate objective function is $f_{t,\alpha}^{\delta}(a,b,x) = \sum_{s=1}^{t-1} L_{1,\alpha}^{\delta}(Y_s - a - b^T(X_s - x))k(X_s - x; h)$; X_s and Y_s are data. For the simulation experiment, the kernel function $k(X_s - x; h) \equiv 1$. For the real analysis, the kernel function was set as the normal probability density function with zero mean and covariance matrix equal to hI where I is the identity matrix.

Set $index = [10, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19]$

Let quantile be an $N - M + 1$ by 19 array cleared.

Loop for $i = 1 : 19$

$\alpha = index[i]/20.0;$

$j = 0;$

Loop for $t = M : N$

$oldf = f = 10000$, $olda = a = 0$, $oldb = b = 0$,

$\delta = 0.1$,

$x = X_t$.

while $\delta > \epsilon$

if($index[i] < 10$)

$$(a_\delta, b_\delta) = \arg \min_{a<quantile[t-M+1][index[i+1]]} f_{t,\alpha}^{\delta}(a,b,x)$$

else if($index[i] > 10$)

$$(a_\delta, b_\delta) = \arg \min_{a>quantile[t-M+1][index[i-1]]} f_{t,\alpha}^{\delta}(a,b,x)$$

else

$$(a_\delta, b_\delta) = \arg\min f_{t,\alpha}^{\delta}(a,b,x)$$

endif

$quantile[t-M+1][index[i]]{=}a_\delta$

$f = f_{t,\alpha}(a_\delta, b_\delta, x)$

$g = f'_{t,\alpha}(a_\delta, b_\delta, x)$ which is the gradient

if($|f| < \epsilon$ and $||g|| < \epsilon$)

if($|f - oldf|/|oldf| < \epsilon$ or $|olda - a_\delta| + |oldb - b_\delta| < \epsilon$) break;

else if($|f - oldf| < \epsilon$ and $|olda - a_\delta| + |oldb - b_\delta| < \epsilon$) break;

endif

$oldf = f$, $olda = a = a_\delta$, $oldb = b = b_\delta$,

$\delta = \delta/2$

End while

if($Y_t \leq a$) $j = j + 1$;

End loop for t

$\alpha_i = j/(N - M + 1)$

End loop for i

As remarked earlier, we implemented the optimization required in the inner loop via the optimization software package CFSQP. At the end of the inner loop, a is the conditional α-quantile of Y_t given X_t. The outputs at the end of the outer loop are α_j which are the relative frequencies of Y_t that are less than or equal to the conditional $j/20$-quantiles of Y_t given X_t for $M \leq t \leq N$ where the conditional quantiles are estimated based on data collected up to and including $t - 1$.

Affiliation of Authors:

[1] Department of Applied Mathematics, Dalian University of Technology, Dalian, 116024, Liaoning, China, and Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. email: maghzhao@polyu.edu.hk

[2] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. email: mateokl@polyu.edu.hk

[3] Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, Iowa 52242, USA. email: kchan@stat.uiowa.edu

References

[1] Box, G. and Tiao, G. 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading.

[2] Chan, K.S. and Tong, H. 2001. Chaos: a Statistical Perspective. Springer-Verlag, New York.

[3] Fan, J. and Gijbels, I. 1996. Local polynomial modelling and its applications. Chapman and Hall, London.

[4] Jennings L.S., Wong K.H. and Teo, K.L. 1996. Optimal Control Computation to Account for Eccentric Movement, Journal of Australia Mathematical Society B-38: 182–193.

[5] Koenker, R. and Bassett, G. 1978. Regression quantiles. Econometrica 46: 33-50.

[6] Koenker, R. and d'Orey V. 1987. Computing regression quantiles. Applied Statistics 36: 383-390.

[7] Koenker, R. and Park, B.J. 1996. An interior point algorithm for nonlinear quantile regression. Journal of Econometrics 71: 265-283.

[8] Lawrence, C.T., Zhou J.L. and Tits, A.L. 1997. User's guide for CFSQP version 2.5: A C code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints. University of Maryland, Institute for Systems Research, Technical Report.

[9] Portnoy, S. and Koenker, R. 1997. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. Statistical Science 4: 279-296.

[10] Yao, Q. and Tong, H. 1996. Asymmetric least squares regression estimation: a nonparametric approach. Nonparametric Statistics 6: 273–92.

[11] Teo, K.L., Goh, C.J. and Wong, K.H. 1991. A Unified Computational Approach to Optimal Control Problems. Longman Scientific and Technical.

Figure Captions

Fig. 1. Empirical percents of Y_t that are less than the α-conditional quantiles given $X_t = (Y_{t-1}, \cdots, Y_{t-p})^T$ that are estimated from the log lynx data up to and including Y_{t-1}, with the estimation by the proposed (ZTC) method and the Yao-Tong (YT) method. The bandwidth $h = 0.57$. Empirical percents are on the y-axis and α on the x-axis.

Fig. 2. Empirical percents of Y_t that are less than the α-conditional quantiles given $X_t = (Y_{t-1}, \cdots, Y_{t-p})^T$ that are estimated from the log lynx data up to and including Y_{t-1}, with the estimation by the proposed (ZTC) method and the Yao-Tong (YT) method. M equals 60. Empirical percents are on the y-axis and α on the x-axis.

Table 1: Comparison of the empirical performance of the current (ZTC) method and the Yao-Tong (YT) method, based on data of size 200 that are generated according to (15) with t-distributed noise. The predictor $X_t = Y_{t-1}$. All experiments are replicated 1000 times.

d.f.	mean of average absolute deviation		mean of maximum absolute deviation		mean computing time	
	ZTC	YT	ZTC	YT	ZTC	YT
$\phi = 0.9$						
∞	0.03022	0.03110	0.06612	0.06708	18.14	2.467
20	0.03014	0.03071	0.06670	0.06678	19.28	2.722
10	0.03030	0.03173	0.06680	0.06760	17.58	2.392
5	0.03036	0.03484	0.06625	0.07381	18.30	2.512
3	0.03046	0.04418	0.06695	0.09229	17.50	2.410
$\phi = 0.5$						
∞	0.02852	0.02937	0.06382	0.06511	17.19	2.316
20	0.02823	0.02833	0.06334	0.06303	17.50	2.356
10	0.02845	0.03002	0.06367	0.06529	17.15	2.312
5	0.02915	0.03309	0.06450	0.07159	17.96	2.426
3	0.02884	0.04267	0.06399	0.09004	17.63	2.396
$\phi = 0.0$						
∞	0.02798	0.02956	0.06355	0.06508	17.02	2.261
20	0.02860	0.02951	0.06408	0.06486	17.35	2.316
10	0.02903	0.02977	0.06437	0.06480	18.06	2.417
5	0.02840	0.03256	0.06332	0.07071	18.06	2.407
3	0.02824	0.04211	0.06386	0.08934	19.28	2.613
$\phi = -0.5$						
∞	0.02865	0.02893	0.06464	0.06403	17.50	2.360
20	0.02883	0.02948	0.06409	0.06500	17.31	2.319
10	0.02849	0.02937	0.06405	0.06462	17.41	2.352
5	0.02853	0.03263	0.06308	0.07046	18.50	2.484
3	0.02842	0.04213	0.06410	0.08966	17.30	2.352
$\phi = -0.9$						
∞	0.02816	0.02930	0.06341	0.06448	18.03	2.441
20	0.02851	0.02900	0.06328	0.06347	18.38	2.535
10	0.02819	0.02901	0.06365	0.06433	17.53	2.407
5	0.02842	0.03217	0.06346	0.06987	19.23	2.620
3	0.02876	0.04196	0.06374	0.08914	19.22	2.605

Table 2: Deviation of the empirical relative frequencies from the nominal probabilities for the conditional α-quantiles of the nonlinear autoregressive model (16), computed by the proposed (ZTC) method and the Yao-Tong (YT) method. The bandwidth is set to $h = 0.57$.

	average absolute deviation		maximum absolute deviation		mean computing time	
	ZTC	YT	ZTC	YT	ZTC	YT
$p = 4$						
30	0.1443	0.1957	0.3030	0.3466	123	20
60	0.1215	0.1783	0.2227	0.2984	96	16
$p = 3$						
30	0.08545	0.1350	0.1853	0.2362	80	12
60	0.07081	0.1344	0.1500	0.2240	62	8
$p = 2$						
30	0.05728	0.1072	0.1206	0.1761	48	6
60	0.04977	0.1076	0.1318	0.1921	36	5
$p = 1$						
30	0.02353	0.1115	0.05294	0.2467	34	3
60	0.04450	0.1405	0.08182	0.3012	27	3

Table 3: Deviation of the empirical relative frequencies from the nominal probabilities for the conditional α-quantiles of the nonlinear autoregressive model (16), computed by the proposed (ZTC) method and the Yao-Tong (YT) method with monotonicity constraints. The bandwidth is set to $h = 0.57$.

M	mean abs. deviation		max abs. deviation		mean computing time		mean 90% interval length (empirical coverage %)	
	ZTC	YT	ZTC	YT	ZTC	YT	ZTC	YT
				$p = 4$				
30	0.1294	0.1883	0.2794	0.3348	186	52	0.7231 (49.41)	0.5778 (49.41)
60	0.09857	0.1668	0.1682	0.2803	142	35	0.6441 (58.18)	0.5796 (50.91)
				$p = 3$				
30	0.07833	0.1313	0.1500	0.2244	155	27	0.8110 (62.35)	0.7668 (63.53)
60	0.06412	0.1286	0.1500	0.2058	107	21	0.8060 (63.64)	0.7476 (58.18)
				$p = 2$				
30	0.05263	0.1034	0.1088	0.1643	99	21	1.072 (72.94)	1.026 (76.47)
60	0.04306	0.1018	0.1136	0.1739	81	14	1.074 (72.73)	1.034 (74.55)
				$p = 1$				
30	0.01889	0.1090	0.05294	0.2467	64	14	1.928 (84.71)	1.769 (82.35)
60	0.04115	0.1366	0.08182	0.3012	38	11	1.904 (81.82)	1.756 (80.00)

Table 4: Deviation of the empirical relative frequencies from the nominal probabilities for the conditional α-quantiles of the nonlinear autoregressive model (16), computed by the proposed (ZTC) method and the Yao-Tong (YT) method with monotonicity constraints.

h	mean abs. deviation		max abs. deviation		mean computing time		mean 90% interval length (empirical coverage %)	
	ZTC	YT	ZTC	YT	ZTC	YT	ZTC	YT
$p = 4$								
1	0.06938	0.1525	0.1500	0.2588	181	39	0.9347 (70.91)	0.8433 (65.45)
2	0.05550	0.1152	0.08636	0.1897	208	53	1.139 (80.00)	1.049 (76.36)
4	0.03780	0.1142	0.08636	0.2159	261	49	1.264 (83.64)	1.125 (83.64)
$p = 3$								
1	0.07177	0.1133	0.1409	0.1838	86	14	1.061 (72.73)	0.9701 (72.73)
2	0.05215	0.1085	0.1409	0.2103	88	15	1.177 (85.45)	1.140 (83.64)
4	0.03876	0.1095	0.07727	0.2103	94	16	1.267 (90.91)	1.205 (81.82)
$p = 2$								
1	0.04498	0.1028	0.08182	0.1861	66	8	1.208 (80.00)	1.139 (80.00)
2	0.03828	0.1066	0.09546	0.2103	59	8	1.254 (87.27)	1.233 (83.64)
4	0.03254	0.1018	0.06818	0.1921	51	9	1.335 (92.73)	1.254 (85.45)
$p = 1$								
1	0.04402	0.1338	0.1045	0.3194	37	7	1.9188 (83.64)	1.751 (78.18)
2	0.03541	0.1290	0.08636	0.2830	26	7	1.9392 (83.64)	1.830 (78.18)
4	0.03780	0.1223	0.08182	0.2830	26	8	1.9678 (81.82)	1.862 (78.18)

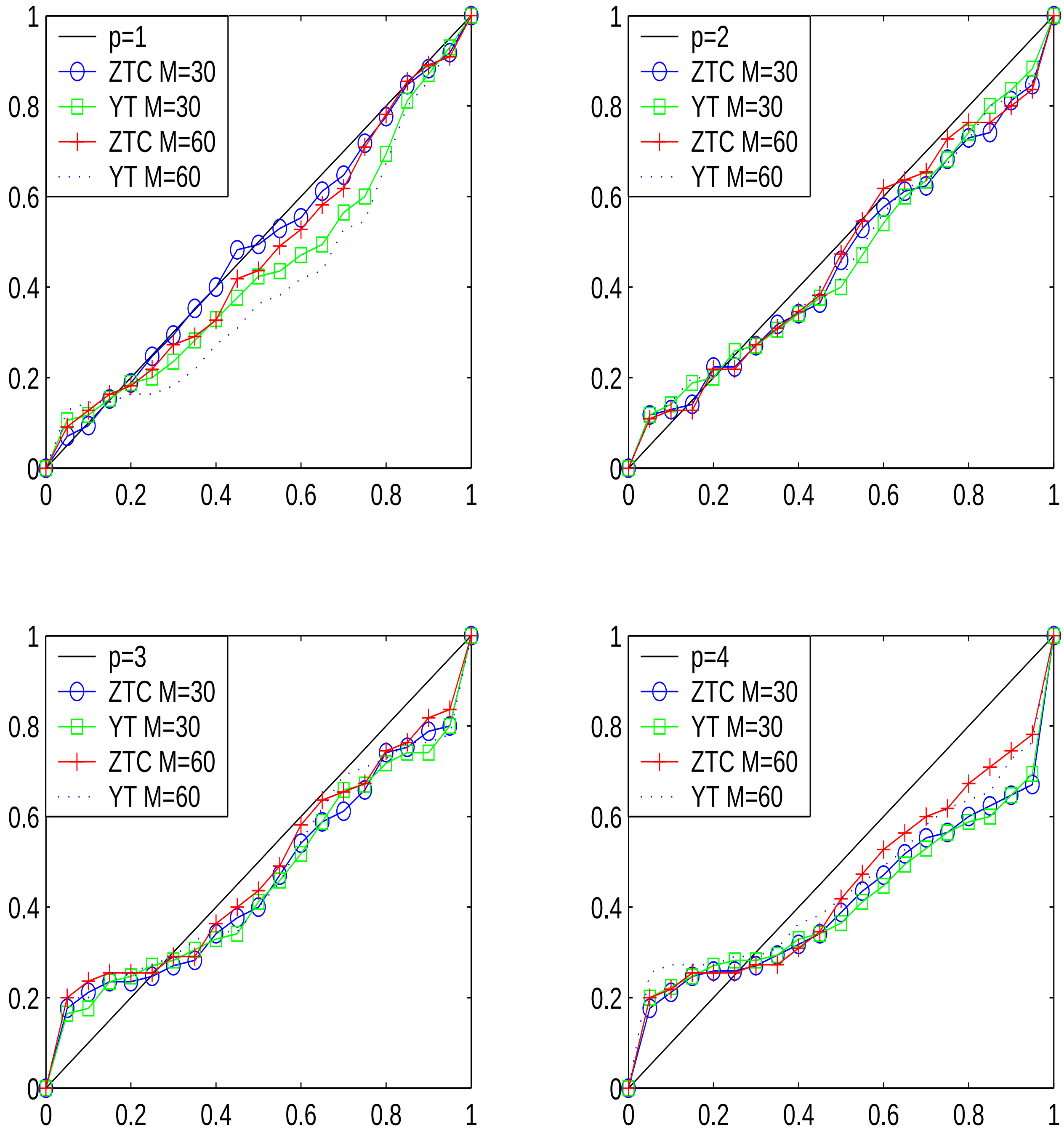
p=1
ZTC M=30
YT M=30
ZTC M=60
YT M=60
p=2
ZTC M=30
YT M=30
ZTC M=60
YT M=60
p=3
ZTC M=30
YT M=30
ZTC M=60
YT M=60
p=4
ZTC M=30
YT M=30
ZTC M=60
YT M=60

Figure 1:

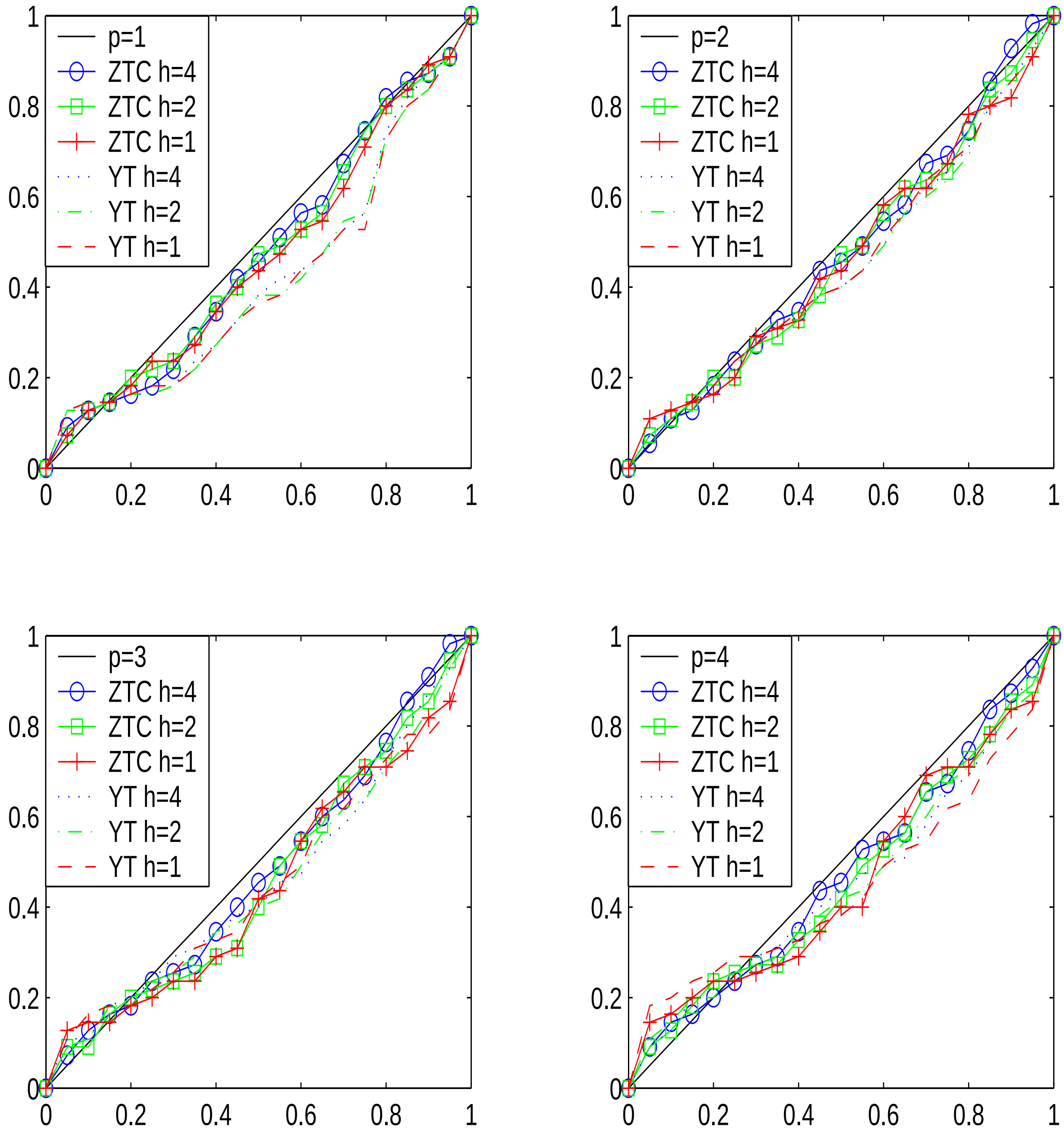

Figure 2: