

WISE WORKING PAPER SERIES

WISEWP0603

**Nonparametric Quantile Estimations For Dynamic Smooth
Coefficient Models**

Zongwu Cai And Xiaoping Xu

July 16, 2006



厦门大学王亚南经济研究院

The Wang Yanan Institute for Studies in Economics, Xiamen University

COPYRIGHT© WISE, XIAMEN UNIVERSITY, CHINA

Nonparametric Quantile Estimations For Dynamic Smooth Coefficient Models *

Zongwu Cai and Xiaoping Xu

This Version: July 16, 2006

In this paper, quantile regression methods are suggested for a class of smooth coefficient time series models. We employ a local polynomial fitting scheme to estimate the smooth coefficients in a quantile framework. The programming involved in the local polynomial quantile estimation is relatively simple and it can be modified with few efforts from existing programs for the linear quantile model. We establish the asymptotic normality of the local polynomial estimator for α -mixing time series. Also, a bandwidth selector based on the nonparametric version of the Akaike information criterion is suggested, together with a consistent estimate of the asymptotic covariance matrix. Further, the asymptotic behaviors of the estimator at boundaries are examined. A comparison of the local polynomial quantile estimator with the local constant estimator is presented. A simulation study is carried out to illustrate the performance of estimates. An empirical application of the model to the exchange rate time series data and the well-known Boston house price data further demonstrates the potential of the proposed modeling procedures.

KEY WORDS: Bandwidth selection; boundary effect; covariance estimation; kernel smoothing methods; nonlinear time series; quantile regression; value-at-risk; varying coefficients.

Revised for Journal of the American Statistical Association

*Zongwu Cai is Professor, Department of Mathematics & Statistics and Department of Economics, University of North Carolina, Charlotte, NC 28223, USA and Special-term Appointed Professor, Wang Yanan Institute for Studies in Economics, Xiamen University, China, and College of Economics and Management, Shanghai Jiaotong University, China (E-mail: zca@uncc.edu). Xiaoping Xu is Professor, Department of Statistics, College of Humanity Sciences and Economics, China University of Geosciences, Wuhan, China (E-mail: jxiaopxu@yahoo.com). Cai's research was supported, in part, by the National Science Foundation grant DMS-0404954 and funds provided by the University of North Carolina at Charlotte. The authors thank the editor, the associate editor and two referees as well as Zhijie Xiao for their constructive and detailed suggestions that led to significant improvements in the presentation of the article.

1 Introduction

Over the last three decades, quantile regression, also called conditional quantile or regression quantile, introduced by Koenker and Bassett (1978), has been used widely in various disciplines, such as finance, economics, medicine, and biology. It is well-known that when the distribution of data is typically skewed or data contains some outliers, the median regression, a special case of quantile regression, is more explicable and robust than the mean regression. Also, regression quantiles can be used to test heteroscedasticity formally or graphically (Koenker and Bassett, 1982; Efron, 1991; Koenker and Zhao, 1996; Koenker and Xiao, 2002). Although some individual quantiles, such as the conditional median, are sometimes of interest in practice, more often one wishes to obtain a collection of conditional quantiles which can characterize the entire conditional distribution. More importantly, another application of conditional quantiles is the construction of prediction intervals for the next value given a small section of recent past values in a stationary time series (Granger, White, and Kamstra, 1989; Koenker, 1994; Zhou and Portnoy, 1996; Koenker and Zhao, 1996; Taylor and Bunn, 1999). Also, Granger, White, and Kamstra (1989), Koenker and Zhao (1996), and Taylor and Bunn (1999) considered an interval forecasting for parametric autoregressive conditional heteroscedastic (ARCH) type models.

Recently, the quantile regression technique has been successfully applied to various applied fields. For example, in the 1992 presidential selection, the Democrats used the yearly Current Population Survey data to show that between 1980 and 1992 there was an increase in the number of people in the high-salary category as well as an increase in the number of people in the low-salary category. This phenomena could be characterized by using a quantile regression method. See Figure 6.4 in Fan and Gijbels (1996, p. 229). More importantly, by following the regulations of the Bank for International Settlements, many of financial institutions have begun to use a uniform measure of risk to measure market risks called Value-at-Risk (VaR), which can be defined as the maximum potential loss of a specific portfolio for a given horizon in finance. In essence, it is to compute an estimate of the lower tail quantile (with a small probability) of future portfolio returns, conditional on current information. Therefore, VaR can be regarded as a special application of the quantile regression. There is a vast amount of literature in this area; see, to name just a few, Khindanova and Rachev (2000), Bao, Lee and Saltoğlu (2001), Engle and Manganelli (2004), and Tsay (2005), and references therein.

In this paper, we assume that $\{\mathbf{X}_t, Y_t\}_{t=-\infty}^{\infty}$ is a stationary sequence. Denote $F(y|\mathbf{x})$ the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$, where $\mathbf{X}_t = (X_{t1}, \dots, X_{td})'$ with $'$ denoting the

transpose of a matrix or vector, is the associated covariate vector in \mathfrak{R}^d with $d \geq 1$, which might be a function of exogenous (covariate) variables or some lagged variables or time t . A regression quantile function $q_\tau(\mathbf{x})$ is defined as, for any $0 < \tau < 1$,

$$q_\tau(\mathbf{x}) = \inf\{y \in \mathfrak{R} : F(y|\mathbf{x}) \geq \tau\}, \quad \text{or} \quad q_\tau(\mathbf{x}) = \operatorname{argmin}_{a \in \mathfrak{R}} E\{\rho_\tau(Y_t - a) | \mathbf{X}_t = \mathbf{x}\}, \quad (1)$$

where $\rho_\tau(y) = y(\tau - I_{\{y < 0\}})$ with $y \in \mathfrak{R}$ is called the loss (“check”) function, and I_A is the indicator function of any set A . Clearly, the simplest form of (1) is $q_\tau(\mathbf{x}) = \beta'_\tau \mathbf{x}$, which is called the linear quantile regression model well studied by many authors. For details, see the papers by Koenker and Bassett (1978, 1982), Koenker, Ng and Portnoy (1994), Koenker and Zhao (1996), Khindanova and Rachev (2000), Bao, Lee and Saltoğlu (2001), Koenker and Xiao (2002, 2004), and Engle and Manganelli (2004), and references therein.

In many practical applications, however, the linear quantile regression model might not be “rich” enough to capture the underlying relationship between the quantile of response variable and its covariates. Indeed, some components may be highly nonlinear or some covariates may be interactive. To make quantile regression models more flexible, there is a swiftly growing literature on nonparametric quantile regression. Various smoothing techniques, such as kernel methods, splines, and their variants, have been used to estimate the nonparametric quantile regression for both independent and time series data. For recent developments and detailed discussions on theory, methodologies, and applications, see, for example, the papers by He, Ng, and Portnoy (1998), Yu and Jones (1998), He and Ng (1999), He and Portnoy (2000), Honda (2000, 2004), Khindanova and Rachev (2000), Bao, Lee and Saltoğlu (2001), Cai (2002a), De Gooijer and Gannoun (2003), Kim (2003), Yu and Lu (2004), Tsay (2005), and Horowitz and Lee (2005), and references therein. In particular, for the univariate case, Honda (2000) derived the asymptotic properties of the local linear estimator of the quantile regression function under α -mixing condition. For the high dimensional case, however, the aforementioned methods encounter some difficulties such as the so-called “curse of dimensionality” and their implementation in practice is not easy as well as the visual display is not so useful for the exploratory purposes.

To attenuate the above problems, De Gooijer and Zerom (2003), Yu and Lu (2004), and Horowitz and Lee (2005) considered an additive quantile regression model $q_\tau(\mathbf{X}_t) = \sum_{k=1}^d g_k(X_{tk})$. To estimate each component, for the time series case, De Gooijer and Zerom (2003) first estimated a high dimensional quantile function by inverting the conditional distribution function estimated by using a weighted Nadaraya-Watson approach, proposed by Cai (2002a), and then used a projection method to estimate each component, while Yu and Lu (2004) focused on independent data and used a back-fitting algorithm method to estimate

each component. On the other hand, to estimate each additive component for independent data, Horowitz and Lee (2005) used a two-stage approach consisting of the series estimation at the first step and a local polynomial fitting at the second step. For independent data, the above model was extended by He, Ng and Portony (1998), He and Ng (1999), and He and Portony (2000) to include interaction terms by using spline methods. Finally, Xiao (2006) investigated a new and robust approach of estimating conditional quantiles based on GARCH type models. Since quantile regression estimation of GARCH models is highly nonlinear, Xiao (2006) discussed the problem of estimating this type model using traditional recursive methods for nonlinear quantile regression and proposed two new methods of estimating quantiles of GARCH models.

In this paper, we adapt another dimension reduction modelling method to analyze dynamic time series data, termed as the smooth (functional or varying) coefficient modelling approach. This approach allows appreciable flexibility on the structure of fitted models. It allows for linearity in some continuous or discrete variables which can be exogenous or lagged and nonlinear in other variables in coefficients. In such a way, the model has the ability of capturing the individual variations. More importantly, it can ease the so-called ‘‘curse of dimensionality’’ and combines both additivity and interactivity. A smooth coefficient quantile regression model for time series data takes the following form

$$q_\tau(\mathbf{U}_t, \mathbf{X}_t) = \sum_{k=0}^d a_{k,\tau}(\mathbf{U}_t) X_{tk} = \mathbf{X}_t' \mathbf{a}_\tau(\mathbf{U}_t), \quad (2)$$

where \mathbf{U}_t is called the smoothing variable, which might be one part of X_{t1}, \dots, X_{td} or just time or other exogenous variables or lagged variables, $\mathbf{X}_t = (X_{t0}, X_{t1}, \dots, X_{td})'$ with $X_{t0} \equiv 1$, $\{a_{k,\tau}(\cdot)\}$ are smooth coefficient functions, and $\mathbf{a}_\tau(\cdot) = (a_{0,\tau}(\cdot), \dots, a_{d,\tau}(\cdot))'$. Here, some of $\{a_{k,\tau}(\cdot)\}$ are allowed to depend on τ . For simplicity, we drop τ from $\{a_{k,\tau}(\cdot)\}$ in what follows. Our interest here is to estimate coefficient functions $\mathbf{a}(\cdot)$ rather than the quantile regression surface $q_\tau(\cdot, \cdot)$ itself. Note that model (2) was studied by Kim (2003), Honda (2004), and Wei and He (2005) for an independent sample, but our focus here is on a dynamic model for nonlinear time series, which has more capacity of applications.

The general setting in (2) covers many familiar quantile regression models, including the quantile autoregressive model (QAR) proposed by Koenker and Xiao (2004) who applied QAR model for the unit root inference. In particular, it includes a specific class of ARCH models, such as heteroscedastic linear models considered by Koenker and Zhao (1996) and nonlinear models studied by Xiao (2006). Also, if there is no \mathbf{X}_t in the model ($d = 0$), $q_\tau(\mathbf{U}_t, \mathbf{X}_t)$ becomes $q_\tau(\mathbf{U}_t)$ so that model (2) reduces to the ordinary nonparametric quantile

regression model which has been studied extensively. For recent developments, refer to the papers by He, Ng and Portony (1998), Yu and Jones (1998), He and Ng (1999), He and Portony (2000), Honda (2000), Cai (2002a), De Gooijer and Zerom (2003), Yu and Lu (2004), and Horowitz and Lee (2005). If \mathbf{U}_t is just time, then the model is called the time-varying coefficient quantile regression model, which is potentially useful to see whether the quantile regression changes over time and in a case with a practical interest is, for example, the aforementioned political example and the analysis of the reference growth data by Cole (1994), Wei, Pere, Koenker and He (2003), and Wei and He (2005), and references therein. However, if \mathbf{U}_t is time, the observed time series might not be stationary. Therefore, the treatment for non-stationary case would require a different approach so that it is beyond the scope of this paper and deserves a further investigation. Kim (2003) and Wei and He (2005) considered the case when \mathbf{U}_t is time and data are iid by using spline estimation procedures. For more applications, see the paper by Xu (2005). Finally, note that the smooth coefficient mean regression model is one of the most popular nonlinear time series models in mean regression and has various applications. For more discussions, refer to the papers by Chen and Tsay (1993), Cai, Fan, and Yao (2000), Cai and Tiwari (2000), and Hong and Lee (2003), and the book by Tsay (2005), and references therein.

The motivation of this study comes from analyzing the well known Boston housing price data. The main interest is to identify factors affecting the house price in Boston area. As argued by Şentürk and Müller (2003), the correlation between the house price and the crime rate can be adjusted by the confounding variable which is the proportion of population of lower educational status through a varying coefficient model and the expected effect of increasing crime rate on declining house prices seems to be only observed for lower educational status neighborhoods in Boston. The interesting features of this dataset are that the response variable is the median price of a home in a given area and the distributions of the price and the major covariate (the confounding variable) are left skewed. Therefore, quantile methods are suitable for analyzing this dataset. Therefore, such a problem can be tackled by using model (2). In another example, one is interested in exploring the possible nonlinearity, heteroscedasticity, and predictability of the exchange rates such as the Japanese Yen against US dollar. The detailed analysis of these data sets is reported in Section 3.

The plan of this paper is as follows. In Section 2, we present the local polynomial quantile estimation of coefficient functions and its asymptotic normality of the estimators. Also, we discuss the local constant quantile estimation and its asymptotic properties. Moreover, the asymptotic behaviors of both estimators at boundaries are examined. A comparison of two estimators is presented. Finally, an ad hoc data-driven fashioned bandwidth selector is

suggested based on the nonparametric version of the Akaike information criterion and the consistent estimator of the asymptotic covariance matrix is provided in the same section. In Section 3, we illustrate the finite sample performance of the estimators with a Monte Carlo experiment and also we give an application to the exchange rate series and the Boston house price data. A concluding remark in Section 4 concludes the paper. Finally, the brief derivations of the theorems are given in Appendix with some lemmas.

2 Modeling Procedures

2.1 Local Polynomial Quantile Estimate

Without loss of generality, we consider only the case when \mathbf{U}_t in (2) is one-dimensional, denoted by U_t in what follows. For multivariate \mathbf{U}_t , the theory and procedure for the univariate case continue to hold but more and complicated notations involve; see Rupert and Wand (1994). A local polynomial fitting has several nice properties such as high statistical efficiency in an asymptotic minimax sense, design-adaptation, and automatic edge correction; see Fan and Gijbels (1996).

Now, we estimate $\{a_k(\cdot)\}$ using the local polynomial method based on observations $\{(U_t, \mathbf{X}_t, Y_t)\}_{t=1}^n$. We assume throughout the paper that the coefficient functions $\mathbf{a}(\cdot)$ have the $(q+1)$ th derivative ($q \geq 1$), so that for any given grid point u_0 , $a_k(\cdot)$ can be approximated by a polynomial function in a neighborhood of the given grid point u_0 as $\mathbf{a}(U_t) \approx \mathbf{a}(u_0) + \mathbf{a}'(u_0)(U_t - u_0) + \cdots + \mathbf{a}^{(q)}(u_0)(U_t - u_0)^q/q!$ and $q_\tau(U_t, \mathbf{X}_t) \approx \sum_{j=0}^q \mathbf{X}_t' \boldsymbol{\beta}_j (U_t - u_0)^j$, where $\boldsymbol{\beta}_j = \mathbf{a}^{(j)}(u_0)/j!$. Then, the locally weighted loss function is

$$\sum_{t=1}^n \rho_\tau \left(Y_t - \sum_{j=0}^q \mathbf{X}_t' \boldsymbol{\beta}_j (U_t - u_0)^j \right) K_h(U_t - u_0), \quad (3)$$

where $K(\cdot)$ is a kernel function, $K_h(x) = K(x/h)/h$, and $h = h_n$ is a sequence of positive numbers tending to zero, which controls the amount of smoothing used in estimation. Solving the minimization problem in (3) gives $\hat{\mathbf{a}}(u_0) = \hat{\boldsymbol{\beta}}_0$, the local polynomial estimate of $\mathbf{a}(u_0)$, and $\hat{\mathbf{a}}^{(j)}(u_0) = j! \hat{\boldsymbol{\beta}}_j$ ($j \geq 1$), the local polynomial estimate of the j th derivative $\mathbf{a}^{(j)}(u_0)$ of $\mathbf{a}(u_0)$. By moving u_0 along with the real line, one obtains the estimate for the entire curve.

The programming involved in the local polynomial quantile estimation is relatively simple and can be modified with few efforts from the existing programs for a linear quantile model. For example, for each grid point u_0 , the local polynomial quantile estimation can be implemented in the **R** package **quantreg**, of Koenker (2004) by setting covariates as \mathbf{X}_t and $\mathbf{X}_t(U_t - u_0)$ and the weight as $K_h(U_t - u_0)$.

Although some modifications are needed, the method developed here for the local polynomial quantile estimation is applicable to a general local polynomial quantile estimation. In particular, we note that the local constant (Nadaraya-Watson type) quantile estimation of $\mathbf{a}(u_0)$, denoted by $\tilde{\mathbf{a}}(u_0)$, is $\tilde{\boldsymbol{\beta}}$ minimizing the following subjective function

$$\sum_{t=1}^n \rho_{\tau}(Y_t - \mathbf{X}_t' \boldsymbol{\beta}) K_h(U_t - u_0), \quad (4)$$

which is a special case of (3) with $q = 0$. We compare $\hat{\mathbf{a}}(u_0)$ and $\tilde{\mathbf{a}}(u_0)$ theoretically at the end of Section 2.2 and empirically in Section 3.1 and the comparison leads to suggest that one should use the local polynomial approach in practice.

2.2 Asymptotic Results

Although our interest in conditional quantile estimation is motivated by the statistical inferences for time series data, we introduce our methods in a more general setting (α -mixing; see Hall and Heyde (1980) for the definition) which includes many time series models as a special case. The asymptotic results here are derived under the α -mixing assumption. In fact, under very mild assumptions linear autoregressive and more generally bilinear time series models are α -mixing with mixing coefficients decaying exponentially. Many nonlinear time series models, such as functional coefficient autoregressive processes with/without exogenous variables, ARCH and GARCH type processes, stochastic volatility models, and nonlinear additive autoregressive models with/without exogenous variables, are strong mixing under some mild conditions. See, for example, Auestad and Tjøstheim (1990), Chen and Tsay (1993), Masry and Tjøstheim (1995, 1997), Lu (1998), Cai and Masry (2000), and Carrasco and Chen (2002).

We first give some regularity conditions that are sufficient for the consistency and asymptotic normality of the proposed estimators, although they might not be the weakest possible. We introduce the following notations. Denote $\Omega(u_0) \equiv E[\mathbf{X}_t \mathbf{X}_t' | U_t = u_0]$ and $\Omega^*(u_0) \equiv E[\mathbf{X}_t \mathbf{X}_t' f_{y|u,x}(q_{\tau}(u_0, \mathbf{X}_t)) | U_t = u_0]$, where $f_{y|u,x}(y)$ is the conditional density of Y given U and \mathbf{X} . Let $f_u(u)$ present the marginal density of U .

Assumptions:

(C1) $\mathbf{a}(u)$ is $(q + 1)$ th continuously differentiable in a neighborhood of u_0 for any u_0 .

(C2) $f_u(u)$ is continuous and $f_u(u_0) > 0$.

(C3) $f_{y|u,x}(y)$ is bounded and satisfies the Lipschitz condition.

- (C4) The kernel function $K(\cdot)$ is symmetric and has a compact support, say $[-1, 1]$.
- (C5) $\{(\mathbf{X}_t, Y_t, U_t)\}$ is a strictly α -mixing stationary process with mixing coefficient $\alpha(t)$ satisfying $\sum_{t \geq 1} t^l \alpha^{(\delta-2)/\delta}(t) < \infty$ for some positive real number $\delta > 2$ and $l > (\delta-2)/\delta$.
- (C6) $E\|\mathbf{X}_t\|^{2\delta^*} < \infty$ with $\delta^* > \delta$.
- (C7) $\Omega(u_0)$ is positive-definite and continuous in a neighborhood of u_0 .
- (C8) $\Omega^*(u_0)$ is continuous and positive-definite in a neighborhood of u_0 .
- (C9) The bandwidth h satisfies $h \rightarrow 0$ and $nh \rightarrow \infty$.
- (C10) $f(u, v|\mathbf{x}_0, \mathbf{x}_s; s) \leq M < \infty$ for $s \geq 1$, where $f(u, v|\mathbf{x}_0, \mathbf{x}_s; s)$ is the conditional density of (U_0, U_s) given $(\mathbf{X}_0 = \mathbf{x}_0, \mathbf{X}_s = \mathbf{x}_s)$.
- (C11) $n^{1/2-\delta/4} h^{\delta/\delta^*-1/2-\delta/4} = O(1)$.

A similar discussion on the above assumptions can be found in Cai (2002a). Assumption (C6) is commonly required for ensuring the convergence of $n^{-1} \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t'$ to $E(\mathbf{X}_t \mathbf{X}_t')$ when \mathbf{X}_t is mixing. It is clear from (2) that $\Omega(u_0) \mathbf{a}(u_0) = E[q_\tau(u_0, \mathbf{X}_t) \mathbf{X}_t | U_t = u_0]$. Then, $\mathbf{a}(u_0)$ is identified (uniquely determined) if and only if $\Omega(u_0)$ is positive definite for any u_0 . Therefore, Assumption (C7) is the necessary and sufficient condition for the model identification. To establish the asymptotic normality of the proposed estimator, define $\mu_j = \int u^j K(u) du$ and $\nu_j = \int u^j K^2(u) du$.

Theorem 1: *Under Assumptions (C1)- (C11), we have the following asymptotic normality for q odd,*

$$\sqrt{nh} \left[\hat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \frac{h^{q+1}}{(q+1)!} \mathbf{a}^{(q+1)}(u_0) \mu_{q+1} + o_p(h^{q+1}) \right] \rightarrow N \{0, \tau(1-\tau) \nu_0 \boldsymbol{\Sigma}(u_0)\},$$

where $\boldsymbol{\Sigma}(u_0) = [\Omega^*(u_0)]^{-1} \Omega(u_0) [\Omega^*(u_0)]^{-1} / f_u(u_0)$.

The case when q is even leads to a more complicated derivation, we consider only the case for q odd. For the case when q is even, we can obtain a similar result; see Rupert and Wand (1994) and Fan and Gijbels (1996). See Xu (2005) for the detailed technical derivations. From Theorem 1, the asymptotic mean squares error (AMSE) of $\hat{\mathbf{a}}(u_0)$ is

$$AMSE = \frac{h^{2q+2} \mu_{q+1}^2}{[(q+1)!]^2} \|\mathbf{a}^{(q+1)}(u_0)\|^2 + \frac{\tau(1-\tau) \nu_0}{nh f_u(u_0)} \text{tr}(\boldsymbol{\Sigma}(u_0)),$$

which gives the optimal bandwidth h_{opt} by minimizing the AMSE

$$h_{opt} = \left(\frac{\tau(1-\tau)\nu_0(q+1)[q!]^2 \text{tr}(\boldsymbol{\Sigma}(u_0))}{2f_u(u_0)\|\mathbf{a}^{(q+1)}(u_0)\|^2\mu_{q+1}^2} \right)^{1/(2q+3)} n^{-1/(2q+3)}$$

and the optimal AMSE is $AMSE_{opt} = O(n^{-(2q+2)/(2q+3)})$. Further, notice that the similar results in Theorem 1 were obtained by Honda (2004) for independent data. Finally, it is interesting to note that the asymptotic bias in Theorem 1 is the same as that for the mean regression case but the two asymptotic variances are different; see, for example, Cai, Fan and Yao (2000). For various practical applications, Fan and Gijbels (1996) recommended using the local linear fit ($q = 1$). Therefore, for ease notation, in what follows, we consider only the case when $q = 1$ (local linear fitting).

If model (2) does not have \mathbf{X} ($d = 0$), it becomes the nonparametric quantile regression model $q_\tau(\cdot)$. Then, Theorem 1 covers the result in Yu and Jones (1998), Honda (2000), and Cai (2002a) for both independent and time series data.

Now we consider the comparison of the performance of the local linear estimation $\hat{\mathbf{a}}(u_0)$ obtained in (3) with that of the local constant estimation $\tilde{\mathbf{a}}(u_0)$ given in (4). To this effect, first, we derive the asymptotic results for the local constant estimator but the proof is omitted; see Xu (2005) for the detailed technical proofs. Under some regularity conditions, it can be shown that

$$\sqrt{nh} \left[\tilde{\mathbf{a}}(u_0) - \mathbf{a}(u_0) - \tilde{\mathbf{b}} + o_p(h^2) \right] \rightarrow N \{0, \tau(1-\tau)\nu_0 \boldsymbol{\Sigma}(u_0)\},$$

where

$$\tilde{\mathbf{b}} = \frac{h^2\mu_2}{2} \left[\mathbf{a}''(u_0) + 2\mathbf{a}'(u_0)f'_u(u_0)/f_u(u_0) + 2\{\Omega^*(u_0)\}^{-1}\Omega^{*'}(u_0)\mathbf{a}'(u_0) \right],$$

which implies that the asymptotic bias for $\tilde{\mathbf{a}}(u_0)$ is different from that for $\hat{\mathbf{a}}(u_0)$ but both have the same asymptotic variance. Therefore, the local constant quantile estimator does not adapt to nonuniform designs: the bias can be large when $f'_u(u_0)/f_u(u_0)$ or $\{\Omega^*(u_0)\}^{-1}\Omega^{*'}(u_0)$ is large even when the true coefficient functions are linear. It is surprising that to the best of our knowledge, this finding seems to be new for the nonparametric quantile regression setting although it is well documented in literature for the ordinary regression case; see Fan and Gijbels (1996) for details.

Finally, to examine the asymptotic behaviors of the local linear and local constant quantile estimators at the boundaries, we offer Theorem 2 below but its proofs are omitted; see Xu (2005) for the detailed technical proofs. Without loss of generality, we consider only the

left boundary point $u_0 = ch$, $0 < c < 1$, if U_t takes values only from $[0, 1]$. A similar result in Theorem 2 holds for the right boundary point $u_0 = 1 - ch$. Define $\mu_{j,c} = \int_{-c}^1 u^j K(u) du$ and $\nu_{j,c} = \int_{-c}^1 u^j K^2(u) du$.

Theorem 2: *Under the assumptions of Theorem 1, we have the following asymptotic normality of the local linear quantile estimator at the left boundary point,*

$$\sqrt{nh} \left[\hat{\mathbf{a}}(ch) - \mathbf{a}(ch) - \frac{h^2 b_c}{2} \mathbf{a}''(0+) + o_p(h^2) \right] \rightarrow N \{0, \tau(1 - \tau) v_c \boldsymbol{\Sigma}(0+)\},$$

where

$$b_c = \frac{\mu_{2,c}^2 - \mu_{1,c} \mu_{3,c}}{\mu_{2,c} \mu_{0,c} - \mu_{1,c}^2} \quad \text{and} \quad v_c = \frac{\mu_{2,c}^2 \nu_{0,c} - 2 \mu_{1,c} \mu_{2,c} \nu_{1,c} + \mu_{1,c}^2 \nu_{2,c}}{[\mu_{2,c} \mu_{0,c} - \mu_{1,c}^2]^2}.$$

Further, we have the following asymptotic normality of the local constant quantile estimator at the left boundary point $u_0 = ch$ for $0 < c < 1$,

$$\sqrt{nh} \left[\tilde{\mathbf{a}}(ch) - \mathbf{a}(ch) - \tilde{\mathbf{b}}_c + o_p(h^2) \right] \rightarrow N \left\{0, \tau(1 - \tau) \nu_{0,c} \boldsymbol{\Sigma}(0+) / \mu_{0,c}^2 \right\}.$$

where

$$\tilde{\mathbf{b}}_c = \left[h \mu_{1,c} \mathbf{a}'(0+) + \frac{h^2 \mu_{2,c}}{2} \left\{ \mathbf{a}''(0+) + \frac{2 \mathbf{a}'(0+) f'_u(0+)}{f_u(0+)} + 2 \Omega^{*-1}(0+) \Omega'^*(0+) \mathbf{a}'(0+) \right\} \right] / \mu_{0,c}.$$

Similar results hold for the right boundary point $u_0 = 1 - ch$.

We remark that if the point 0 were an interior point, then, Theorem 2 would hold with $c = 1$, which becomes Theorem 1. Also, as $c \rightarrow 1$, $b_c \rightarrow \mu_2$, and $v_c \rightarrow \nu_0$ and these limits are exactly the constant factors appearing respectively in the asymptotic bias and variance for an interior point. Therefore, Theorem 2 shows that the local linear estimation has the automatic good behavior at boundaries without the need of boundary correction. Further, one can see from Theorem 2 that at boundaries, the asymptotic bias term for the local constant quantile estimate is of the order h by comparing to the order h^2 for the local linear quantile estimate. This shows that the local linear quantile estimate does not suffer from boundary effects but the local constant quantile estimate does, which is another advantage of the local linear quantile estimator over the local constant quantile estimator. This suggests that one should use the local linear approach in practice.

As a special case, Theorem 2 includes the asymptotic properties for the local constant quantile estimator of the nonparametric quantile function $q_\tau(\cdot)$ at both the interior and boundary points; see Xu (2005) for details.

2.3 Bandwidth Selection

It is well known that the bandwidth plays an essential role in the trade-off between reducing bias and variance. To the best of our knowledge, there has been almost nothing done about selecting the bandwidth in the context of estimating the coefficient functions in the quantile regression even though there is a rich amount of literature on this issue in the mean regression setting; see, for example, Cai, Fan and Yao (2000). Yu and Jones (1998) or Yu and Lu (2004) proposed a simple and convenient method for the nonparametric quantile estimation. Their approach assumes that the second derivatives of the quantile function are parallel. However, this assumption might not be valid for many applications due to (nonlinear) heteroscedasticity. Further, the mean regression approach can not directly estimate the variance function. To attenuate these problems, we propose a method of selecting bandwidth for the foregoing estimation procedure, based on the nonparametric version of the Akaike information criterion (AIC), which can attend to the structure of time series data and the over-fitting or under-fitting tendency. The basic idea is motivated by its analogue of Cai and Tiwari (2000) and Cai (2002b) for nonlinear mean regression for time series models and it is briefly described below.

By recalling the classical AIC for linear models under the likelihood setting; that is -2 (maximized log likelihood) $+ 2$ (number of estimated parameters), we propose the following nonparametric version of the bias-corrected AIC, due to Hurvich, Simonoff and Tsai (1998) for nonparametric regression models, to select h by minimizing

$$\text{AIC}(h) = \log \left\{ \hat{\sigma}_\tau^2 \right\} + 2(p_h + 1)/[n - (p_h + 2)], \quad (5)$$

where $\hat{\sigma}_\tau^2$ and p_h are defined later. This criterion may be interpreted as the AIC for the local quantile smoothing problem and seems to perform well in some limited applications. Note that similar to (5), Koenker, Ng and Portnoy (1994) considered the Schwarz information criterion (SIC) of Schwarz (1978) with the second term on the right-hand side of (5) replayed by $2n^{-1}p_h \log n$, where p_h is the number of “active knots” for the smoothing spline quantile setting, and Machado (1993) studied similar criteria for parametric quantile regression models and more general M-estimators of regression.

Now we turn to define $\hat{\sigma}_\tau^2$ and p_h in this setting. In the mean regression setting, $\hat{\sigma}_\tau^2$ is just the estimate of the variance σ^2 . In the quantile regression, we define $\hat{\sigma}_\tau^2$ as $n^{-1} \sum_{t=1}^t \rho_\tau(Y_t - \mathbf{X}'_t \hat{\mathbf{a}}(U_t))$, which may be interpreted as the mean square error in the least square setting and was also used by Koenker, Ng and Portnoy (1994). In nonparametric models, p_h is the nonparametric version of degrees of freedom, called the effective number of parameters, and it is usually based on the trace of various quasi-projection (hat)

matrices in the least square theory (linear estimators); see, for example, Cai and Tiwari (2000) for a cogent discussion for nonparametric regression models and nonlinear time series models. For the quantile smoothing setting, the explicit expression for the quasi-projection matrix does not exist due to its nonlinearity. However, we can use the first order approximation (the local Bahadur representation) to derive an explicit expression, which may be interpreted as the quasi-projection matrix in this setting. To this end, set $U_{th} = (U_t - u_0)/h$, $\mathbf{X}_t^* = \begin{pmatrix} \mathbf{X}_t \\ U_{th} \mathbf{X}_t \end{pmatrix}$, $Y_t^* = Y_t - \mathbf{X}_t'[\mathbf{a}(u_0) + \mathbf{a}'(u_0)(U_t - u_0)]$, $\mathbf{H} = \text{diag}\{\mathbf{I}, h\mathbf{I}\}$, and $\boldsymbol{\theta} = \sqrt{nh} \mathbf{H} \begin{pmatrix} \boldsymbol{\beta}_0 - \mathbf{a}(u_0) \\ \boldsymbol{\beta}_1 - \mathbf{a}'(u_0) \end{pmatrix}$. Define $\mathbf{S}_n = \mathbf{S}_n(u_0) = a_n \sum_{t=1}^n \xi_t \mathbf{X}_t^* \mathbf{X}_t^{*'} K(U_{th})$, where $a_n = (nh)^{-1/2}$ and $\xi_t = I(Y_t \leq \mathbf{X}_t' \mathbf{a}(u_0) + a_n) - I(Y_t \leq \mathbf{X}_t' \mathbf{a}(u_0))$. It is shown in Appendix that

$$\mathbf{S}_n(u_0) = f_u(u_0) \Omega_1^*(u_0) + o_p(1), \quad (6)$$

where $\Omega_1^*(u_0) = \text{diag}\{\Omega^*(u_0), \mu_2 \Omega^*(u_0)\}$. It is not difficult to verify that

$$\hat{\boldsymbol{\theta}} \approx a_n \mathbf{S}_n^{-1} \sum_{t=1}^n \psi_\tau(Y_t^*) \mathbf{X}_t^* K(U_{th}),$$

where $\psi_\tau(x) = \tau - I_{\{x < 0\}}$. Then, we have

$$\hat{q}_\tau(U_t, \mathbf{X}_t) - q_\tau(U_t, \mathbf{X}_t) \approx \frac{1}{n} \sum_{s=1}^n \psi_\tau(Y_s^*(U_t)) K_h((U_s - U_t)/h) \mathbf{X}_t^{0'} \mathbf{S}_n^{-1}(U_t) \mathbf{X}_s^*$$

where $\mathbf{X}_t^0 = \begin{pmatrix} \mathbf{X}_t \\ \mathbf{0} \end{pmatrix}$. The coefficient of $\psi_\tau(Y_s^*(U_t))$ on the right-hand side of the above expression is $\gamma_s = a_n^2 K(0) \mathbf{X}_s^{0'} \mathbf{S}_n^{-1}(U_t) \mathbf{X}_s^0$. Now, we have that $p_h = \sum_{s=1}^n \gamma_s$, which can be regarded as an approximation to the trace of the quasi-projection (hat) matrix for linear estimators. In the practical implementation, we need to estimate $\mathbf{a}(u_0)$ first since $\mathbf{S}_n(u_0)$ involves $\mathbf{a}(u_0)$. We recommend using a pilot bandwidth which can be chosen as the one proposed by Yu and Jones (1998). Similar to the least square theory, as expected, the criterion proposed in (5) counteracts the over-fitting tendency of the generalized cross-validation due to its relatively weak penalty and the under-fitting of the SIC of Schwarz (1978) studied by Koenker, Ng and Portnoy (1994) because of the heavy penalty.

2.4 Covariance Estimate

For the purpose of statistical inferences, we next consider the estimation of the asymptotic covariance matrix to construct the pointwise confidence intervals. The explicit expression of the asymptotic covariance provides a direct estimator. Therefore, we can use the so-called ‘‘sandwich’’ method. In other words, we need to obtain a consistent estimate for both $\Omega(u_0)$

and $\Omega^*(u_0)$. To this effect, define,

$$\widehat{\Omega}_{n,0} = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t' K_h(U_t - u_0) \quad \text{and} \quad \widehat{\Omega}_{n,1} = \frac{1}{n} \sum_{t=1}^n w_t \mathbf{X}_t \mathbf{X}_t' K_h(U_t - u_0),$$

where $w_t = I(\mathbf{X}_t' \widehat{\mathbf{a}}(u_0) - \delta_n < Y_t \leq \mathbf{X}_t' \widehat{\mathbf{a}}(u_0) + \delta_n) / (2\delta_n)$ for any $\delta_n \rightarrow 0$ as $n \rightarrow \infty$. It is shown in Appendix that

$$\widehat{\Omega}_{n,0} = f_u(u_0) \Omega(u_0) + o_p(1) \quad \text{and} \quad \widehat{\Omega}_{n,1} = f_u(u_0) \Omega^*(u_0) + o_p(1). \quad (7)$$

Therefore, the consistent estimate of $\Sigma(u_0)$ is given by

$$\widehat{\Sigma}(u_0) = [\widehat{\Omega}_{n,1}(u_0)]^{-1} \widehat{\Omega}_{n,0}(u_0) [\widehat{\Omega}_{n,1}(u_0)]^{-1}.$$

Note that $\widehat{\Omega}_{n,1}(u_0)$ might be close to singular for some sparse regions. To avoid this computational difficulty, there are two alternative ways to construct a consistent estimate of $f_u(u_0) \Omega^*(u_0)$ through estimating the conditional density of Y , $f_{y|u,x}(q_\tau(u, \mathbf{x}))$. The first method is the Nadaraya-Watson type (or local linear) double kernel method of Fan, Yao and Tong (1996) defined as,

$$\widehat{f}_{y|u,x}(q_\tau(u, \mathbf{x})) = \frac{\sum_{t=1}^n K_{h_2}(U_t - u, \mathbf{X}_t - \mathbf{x}) L_{h_1}(Y_t - q_\tau(u, \mathbf{x}))}{\sum_{t=1}^n K_{h_2}(U_t - u, \mathbf{X}_t - \mathbf{x})},$$

where $L(\cdot)$ is a kernel function, and the second one is the difference quotients method of Koenker and Xiao (2004) such as

$$\widehat{f}_{y|u,x}(q_\tau(u, \mathbf{x})) = (\tau_j - \tau_{j-1}) / [q_{\tau_j}(u, \mathbf{x}) - q_{\tau_{j-1}}(u, \mathbf{x})],$$

for some appropriately chosen sequence of $\{\tau_j\}$; see Koenker and Xiao (2004) for more discussions. Then, in view of the definition of $f_u(u_0) \Omega^*(u_0)$, the estimator $\widetilde{\Omega}_{n,1}$ can be constructed as,

$$\widetilde{\Omega}_{n,1} = \frac{1}{n} \sum_{t=1}^n \widehat{f}_{y|u,x}(\widehat{q}_\tau(U_t, \mathbf{X}_t)) \mathbf{X}_t \mathbf{X}_t' K_h(U_t - u_0).$$

By an analogue of (7), one can show that under some regularity conditions, both estimators are consistent.

3 Empirical Applications

In this section we report a Monte Carlo simulation to examine the finite sample property of the proposed estimator and to further explore the possible nonlinearity, heteroscedasticity, and predictability of the exchange rate of the Japanese Yen against US dollar and to identify factors affecting the house price in Boston area. In our computation, we use the Epanechnikov

kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ and construct the pointwise confidence intervals based on the consistent estimate of the asymptotic covariance described in Section 2.4 without the bias correction. For a predetermined sequence of h 's from a wide range, say from h_a to h_b with an increment h_δ , based on the AIC bandwidth selector described in Section 2.3, we compute $AIC(h)$ for each h and choose h_{opt} to minimize $AIC(h)$.

3.1 A Simulated Example

Example 1: We consider the following data generating process

$$Y_t = a_1(U_t)Y_{t-1} + a_2(U_t)Y_{t-2} + \sigma(U_t)e_t, \quad t = 1, \dots, n,$$

where $a_1(U_t) = \sin(\sqrt{2}\pi U_t)$, $a_2(U_t) = \cos(\sqrt{2}\pi U_t)$, and $\sigma(U_t) = 3 \exp(-4(U_t - 1)^2) + 2 \exp(-5(U_t - 2)^2)$. U_t is generated from uniform $(0, 3)$ independently and $e_t \sim N(0, 1)$. The quantile regression is $q_\tau(U_t, Y_{t-1}, Y_{t-2}) = a_0(U_t) + a_1(U_t)Y_{t-1} + a_2(U_t)Y_{t-2}$, where $a_0(U_t) = \Phi^{-1}(\tau)\sigma(U_t)$ and $\Phi^{-1}(\tau)$ is the τ -th quantile of the standard normal. Therefore, only $a_0(\cdot)$ is a function of τ . Note that $a_0(\cdot) = 0$ when $\tau = 0.5$. To assess the performance of finite samples, we compute the mean absolute deviation errors (MADE) for $\hat{a}_j(\cdot)$, which is defined as $MAD E_j = n_0^{-1} \sum_{k=1}^{n_0} |\hat{a}_j(u_k) - a_j(u_k)|$, where $\hat{a}_j(\cdot)$ is either the local linear or local constant quantile estimate of $a_j(\cdot)$ and $\{u_k = 0.1(k - 1) + 0.2 : 1 \leq k \leq n_0 = 27\}$ are the grid points. The Monte Carlo simulation is repeated 500 times for each sample size $n = 200, 500$, and 1000 and for each $\tau = 0.05, 0.50$ and 0.95. We compute the optimal bandwidth for each replication, sample size, and τ . We compute the median and standard deviation (in parentheses) of 500 MADE values for each scenario and summarize the results in Table 1.

From Table 1, we can observe that the MADE values for both the local linear and local constant quantile estimates decrease when n increases for all three values of τ and the local linear estimate outperforms the local constant estimate. This is another example to show that the local linear method is superior over the local constant even in the quantile setting. Also, the performance for the median quantile estimate is slightly better than that for two tails ($\tau = 0.05$ and 0.95). This observation is not surprising because of the sparsity of data in the tailed regions. Moreover, another benefit of using the quantile method is that we can obtain the estimate of $a_0(\cdot)$ (conditional standard deviation) simultaneously with the estimation of $a_1(\cdot)$ and $a_2(\cdot)$ (functions in the conditional mean), which, in contrast, avoids a two-stage approach needed to estimate the variance function in the mean regression; see Fan and Yao (1998) for details. However, it is interesting to see that due to the larger variation, the performance for $a_0(\cdot)$, although it is reasonably good, is not as good as that of $a_1(\cdot)$ and

Table 1: The Median and Standard Deviation of 500 MADE Values

| The Local Linear Estimator | | | | | | | | | |
|----------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | $\tau = 0.05$ | | | $\tau = 0.5$ | | | $\tau = 0.95$ | | |
| n | MADE ₀ | MADE ₁ | MADE ₂ | MADE ₀ | MADE ₁ | MADE ₂ | MADE ₀ | MADE ₁ | MADE ₂ |
| 200 | 0.911 | 0.186 | 0.177 | 0.401 | 0.092 | 0.089 | 0.920 | 0.187 | 0.175 |
| | (0.520) | (0.041) | (0.041) | (0.091) | (0.032) | (0.032) | (0.517) | (0.042) | (0.039) |
| 500 | 0.510 | 0.085 | 0.083 | 0.311 | 0.055 | 0.055 | 0.517 | 0.085 | 0.083 |
| | (0.414) | (0.023) | (0.02) | (0.056) | (0.019) | (0.018) | (0.390) | (0.023) | (0.023) |
| 1000 | 0.419 | 0.060 | 0.059 | 0.311 | 0.050 | 0.049 | 0.416 | 0.060 | 0.059 |
| | (0.071) | (0.018) | (0.017) | (0.051) | (0.014) | (0.014) | (0.072) | (0.017) | (0.017) |

| The Local Constant Estimator | | | | | | | | | |
|------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | $\tau = 0.05$ | | | $\tau = 0.5$ | | | $\tau = 0.95$ | | |
| n | MADE ₀ | MADE ₁ | MADE ₂ | MADE ₀ | MADE ₁ | MADE ₂ | MADE ₀ | MADE ₁ | MADE ₂ |
| 200 | 3.753 | 0.285 | 0.290 | 0.501 | 0.144 | 0.147 | 3.763 | 0.287 | 0.287 |
| | (2.937) | (0.050) | (0.051) | (0.115) | (0.027) | (0.028) | (3.188) | (0.052) | (0.051) |
| 500 | 2.201 | 0.147 | 0.146 | 0.355 | 0.084 | 0.085 | 2.223 | 0.147 | 0.147 |
| | (3.025) | (0.024) | (0.025) | (0.062) | (0.016) | (0.015) | (3.320) | (0.025) | (0.025) |
| 1000 | 0.883 | 0.086 | 0.086 | 0.322 | 0.060 | 0.061 | 0.882 | 0.086 | 0.087 |
| | (0.462) | (0.015) | (0.014) | (0.054) | (0.012) | (0.011) | (0.427) | (0.015) | (0.015) |

$a_2(\cdot)$. This can be further evidenced from Figure 1. The results in this simulated experiment show that the proposed procedure is reliable and they are along the line of the asymptotic theory.

Finally, Figure 1 plots the local linear estimates for all three coefficient functions with their true values (solid line): $\sigma(\cdot)$ in Figure 1(a), $a_1(\cdot)$ in Figure 1(b), and $a_2(\cdot)$ in Figure 1(c), for three quantiles $\tau = 0.05$ (dashed line), 0.50 (dotted line) and 0.95 (dotted-dashed line), for $n = 500$ based on a typical sample which is chosen based on its MADE value equal to the median of the 500 MADE values. The selected optimal bandwidths are $h_{opt} = 0.10$ for $\tau = 0.05$, 0.075 for $\tau = 0.50$, and 0.10 for $\tau = 0.95$. Note that the estimate of $\sigma(\cdot)$ for $\tau = 0.50$ can not be recovered from the estimate of $a_0(\cdot) = 0$ and it is not presented in Figure 1(a). The 95% point-wise confidence intervals without the bias correction are depicted in Figures 1(b) and 1(c) in thick lines for the $\tau = 0.05$ quantile estimate. Basically, all confidence intervals cover the true values. Also, we can see that the confidence interval for $\hat{a}_0(\cdot)$ is wider than that for $\hat{a}_1(\cdot)$ and $\hat{a}_2(\cdot)$ due to the larger variation. Similar plots are obtained (not shown here) for the local constant estimates due to the space limitations. Overall, the proposed modeling procedure performs fairly well.

3.2 Real Examples

Example 2: We analyze a subset of the Boston house price data¹ consisting of 14 variables, collected on each of 506 different houses from a variety of locations. The dependent variable is Y , the median value of owner-occupied homes in \$1,000's (house price). Some major factors possibly affecting the house prices used are: U =proportion of population of lower educational status X_1 =the average number of rooms per house, X_2 =the per capita crime rate, X_3 =the full property tax rate, and X_4 =the pupil/teacher ratio. For the complete description of all 14 variables, see Harrison and Rubinfeld (1978). Recently, there have been several papers devoted to the analysis of this dataset. For example, Breiman and Friedman (1985), Chaudhuri, Doksum and Samarov (1997), and Opsomer and Ruppert (1998) used four covariates: X_1 , X_3 , X_4 and U or their transformations to fit the data through a mean additive regression model whereas Yu and Lu (2004) employed the additive quantile technique to analyze the data. Recently, Şentürk and Müller (2003) studied the correlation between the house price Y and the crime rate X_2 adjusted by the confounding variable U through a varying coefficient model and they concluded that the expected effect of increasing crime rate on declining house prices seems to be only observed for lower educational status neighborhoods in Boston. Some existing analyses (e.g., Breiman and Friedman, 1985; Yu and Lu, 2004) in both mean and quantile regressions concluded that most of the variation seen in housing prices in the restricted data set can be explained by two major variables: X_1 and U . Indeed, the correlation coefficients between Y and U and X_1 are -0.7377 and 0.6954 respectively. The scatter plots of Y versus U and X_1 are displayed in Figures 2(a) and 2(b) respectively. The interesting features of this dataset are that the response variable is the median price of a home in a given area and the distributions of Y and the major covariate U are left skewed (the density estimates are not presented). Finally, it is surprising that all the existing nonparametric models aforementioned above did not include the crime rate X_2 , which may be an important factor affecting the housing price, and did not consider the interaction terms such as U and X_2 .

Based on the above discussions, it concludes that the model studied in this paper might be well suitable to analyzing this dataset. Therefore, we analyze this dataset by the following quantile smooth coefficient model²

$$q_\tau(U_t, \mathbf{X}_t) = a_{0,\tau}(U_t) + a_{1,\tau}(U_t) X_{t1} + a_{2,\tau}(U_t) X_{t2}^*, \quad 1 \leq t \leq n = 506, \quad (8)$$

¹This dataset can be downloaded from the web site at <http://lib.stat.cmu.edu/datasets/boston>.

²We do not include the other variables such as X_3 and X_4 in model (8), since we found that the coefficient functions for these variables seem to be constant. Therefore, a semiparametric model would be appropriate if the model includes these variables. But it is beyond the scope of this paper and deserves a further investigation.

where $X_{t2}^* = \log(X_{t2})$. The reason for using the logarithm of X_{t2} in (8), instead of X_{t2} itself, is that the correlation between Y_t and X_{t2}^* (the correlation coefficient is -0.4543) is slightly stronger than that for Y_t and X_{t2} (-0.3883), which can be witnessed as well from Figures 2(c) and 2(d). In the model fitting, covariates X_1 and X_2 are centralized. For the purpose of comparison, we also consider the following functional coefficient model in the mean regression

$$Y_t = a_0(U_t) + a_1(U_t) X_{t1} + a_2(U_t) X_{t2}^* + e_t \quad (9)$$

and we employ the local linear fitting technique to estimate the coefficient functions $\{a_j(\cdot)\}$, denoted by $\{\hat{a}_j(\cdot)\}$; see Cai, Fan and Yao (2000) for details.

The coefficient functions are estimated through the local linear quantile approach by using the bandwidth selector described in Section 2.3. As a result, the selected optimal bandwidths are $h_{opt} = 2.0$ for $\tau = 0.05$, 1.5 for $\tau = 0.50$, and 3.5 for $\tau = 0.95$. Figures 2(e), 2(f) and 2(g) present the estimated coefficient functions $\hat{a}_{0,\tau}(\cdot)$, $\hat{a}_{1,\tau}(\cdot)$, and $\hat{a}_{2,\tau}(\cdot)$ respectively, for three quantiles $\tau = 0.05$ (solid line), 0.50 (dashed line) and 0.95 (dotted line), together with the estimates $\{\hat{a}_j(\cdot)\}$ from the mean regression model (dot-dashed line). Also, the 95% point-wise confidence intervals for the median estimate are displayed by the thick dashed lines without the bias correction.

First, from these three figures, one can see that the median estimates are quite close to the mean estimates and the estimates based on the mean regression are always within the 95% confidence interval of the median estimates. It can be concluded that the distribution of the measurement error e_t in (9) might be symmetric and $\hat{a}_{j,0.5}(\cdot)$ in (8) is almost same as $\hat{a}_j(\cdot)$ in (9). Also, from Figure 2(e), it can be seen that three quantile curves are parallel, which implies that the intercept in $\hat{a}_{0,\tau}(\cdot)$ depends on τ , and they decrease exponentially. More importantly, one can observe from Figures 2(f) and 2(g) that due to the intersection of three quantile estimated coefficient curves, the heteroscedasticity might exist. But unfortunately, this phenomenon was not observed in any previous analyses in the aforementioned papers.

From Figure 2(f), we can observe that the expected effect of increasing the number of rooms can make the house price slightly higher in any low educational status neighborhoods but much higher in relatively high educational status neighborhoods. Also, the number of room has a positive effect on the median and/or higher price houses in relatively high and low educational status neighborhoods but increasing the number of rooms might not increase the house price in very low educational status neighborhoods. In other words, it is very difficult to sell high price houses with high number of rooms at a reasonable price in very low educational status neighborhoods.

From Figure 2(g), one can conclude that the positive correlation between the housing prices ($\tau = 0.50$ and 0.95) and the crime rate for relatively high educational status neighborhoods seems against intuitive. However, the reason for this positive correlation is the existence of high educational status neighborhoods close to central Boston where high house prices and crime rate occur simultaneously. Therefore, the expected effect of increasing crime rate on declining house prices for $\tau = 0.50$ and 0.95 seems to be observed only for lower educational status neighborhoods in Boston. Finally, it can be seen that the correlation between the housing prices for $\tau = 0.05$ and the crime rate is almost negative although the degree depends on the value of U . This implies that increasing crime rate slightly decreases relatively the house prices for the cheap houses ($\tau = 0.05$).

In summary, it concludes that the factors U , X_1 and X_2 do have different effects on the different quantiles of the conditional distribution of the housing price. Overall, the housing price and the proportion of population of lower educational status have a strong negative correlation, and the number of rooms has a mostly positive effect on the housing price whereas the crime rate has the most negative effect on the housing price. In particular, by using the proportion of population of lower educational status U as the confounding variable, we demonstrate the substantial benefits obtained by characterizing the affecting factors X_1 and X_2 on the housing price based on the neighborhoods.

Example 3: This example concerns the closing bid prices of the Japanese Yen (JPY) in terms of US dollar. Here we use the proposed model and its modeling approaches to explore the possible nonlinearity feature, heteroscedasticity, and predictability of the exchange rate series. The data is a weekly series from January 1, 1974 to December 31, 2003. The weekly series is generated by selecting the Wednesdays series (if a Wednesday is a holiday then the following Thursday is used), which has 1566 observations. We model the return series $Y_t = 100 \log(\xi_t/\xi_{t-1})$, plotted in Figure 3(a), using the techniques developed in this paper, where ξ_t is an exchange rate level on the t -th week. Typically the classical financial theory would treat $\{Y_t\}$ as a martingale difference process. Therefore, Y_t would be unpredictable. But this assumption was strongly rejected by Hong and Lee (2003) by examining five major currencies and applying several testing procedures. Figure 3(b) shows that there exists almost no significant autocorrelation in $\{Y_t\}$, which also was confirmed by Tsay (2005) and Hong and Lee (2003) by using several statistical testing procedures.

Fan, Yao and Cai (2003) and Hong and Lee (2003) concluded that the exchange rate series is partially predictable by using the functional coefficient autoregressive model

$$Y_t = a_0(U_t) + \sum_{j=1}^d a_j(U_t) Y_{t-j} + \sigma_t e_t, \quad (10)$$

where U_t is the smooth variable defined later and σ_t is a function of U_t and the lagged variables. If $\{U_t\}$ is observable, $a_j(\cdot)$ can be estimated by a local linear fitting; see Cai, Fan and Yao (2000) for details, denoted by $\hat{a}_j(\cdot)$. Here, σ_t is the stochastic volatility which may depend on U_t and the lagged variables $\{Y_{t-j}\}$. Usually, U_t can be chosen based on the knowledge of data or economic theory. However, if no prior information is available, U_t may be chosen as a function of explanatory vector $\{\xi_{t-j}\}$ or through the use of data-driven methods such as AIC or cross-validation. Recently, Fan, Yao and Cai (2003) proposed a data-driven method to the choice of U_t by a linear combination of $\{\xi_{t-j}\}$ and the lagged variables $\{Y_{t-j}\}$. By following the analysis of Fan, Yao and Cai (2003) and Hong and Lee (2003), the smooth variable U_t is chosen as an moving average technical trading rule (MATTR) in finance so that the autoregressive coefficients vary with investment positions. U_t is defined as $U_t = \xi_{t-1}/M_t - 1$, where $M_t = \sum_{j=1}^L \xi_{t-j}/L$, which is the moving average and can be regarded as a proxy for the trend at the time $t - 1$. Following Hong and Lee (2003), we choose $L = 26$ (half a year). The time series plot of $\{U_t\}$ is given in Figure 3(c). As pointed out by Hong and Lee (2003), U_t is expected to reveal some useful information on the direction of changes. Note that model (8) was studied by Fan, Yao and Cai (2003) for the daily data and Hong and Lee (2003) for the weekly data under the homogenous assumption ($\sigma_t = \sigma$) based on the least square theory. In particular, Hong and Lee (2003) provided some empirical evidences to conclude that model (10) outperforms the martingale model and autoregressive models.

We analyze this exchange rate series by using the smooth coefficient model under the quantile regression framework with only two lagged variables³ as follows

$$q_\tau(U_t, Y_{t-1}, Y_{t-2}) = a_{0,\tau}(U_t) + a_{1,\tau}(U_t) Y_{t-1} + a_{2,\tau}(U_t) Y_{t-2}. \quad (11)$$

The first 1540 observations of $\{Y_t\}$ are used for estimation and the last 25 observations are left for prediction. The coefficient functions $\{a_{j,\tau}(\cdot)\}$ are estimated through the local linear quantile approach, denoted by $\{\hat{a}_{j,\tau}(\cdot)\}$. optimal bandwidths are $h_{opt} = 0.03$ for $\tau = 0.05$, 0.025 for $\tau = 0.50$, and 0.03 for $\tau = 0.95$. Figures 3(d) - 3(g) depict the estimated coefficient functions $\hat{a}_{0,\tau}(\cdot)$, $\hat{a}_{1,\tau}(\cdot)$, and $\hat{a}_{2,\tau}(\cdot)$ respectively, for three quantiles $\tau = 0.05$ (solid line), 0.50 (dashed line) and 0.95 (dotted line), together with the estimates $\{\hat{a}_j(\cdot)\}$ (dot-dashed line) from the mean regression model in (10). Also, the 95% point-wise confidence intervals for the median estimate are displayed by the thick dashed lines without the bias correction.

First, from Figures 3(d), 3(f) and 3(g), we see clearly that the median estimates $\hat{a}_{j,0.50}(\cdot)$ in (11) are almost parallel with or close to the mean estimates $\hat{a}_j(\cdot)$ in (10) and the mean

³We also considered the models with more than two lagged variables and we found that the conclusions are similar and not reported here.

estimates are almost within the 95% confidence interval of the median estimates. Secondly, $\hat{a}_{0,0.50}(\cdot)$ in Figure 3(d) shows a nonlinear pattern and $\hat{a}_{0,0.05}(\cdot)$ and $\hat{a}_{0,0.95}(\cdot)$ in Figure 3(e) exhibit slightly U -shape and symmetrically. More importantly, one can observe from Figures 3(f) and 3(g) that the lower and upper quantile estimated coefficient curves are intersect and they behave slightly differently. Particularly, from Figure 3(g), one might conclude that the distribution of the measurement error e_t in (10) might not be symmetric about 0 and there exists a nonlinearity in $a_{j,\tau}(\cdot)$. This implies that a nonlinearity exists. Also, our findings are: the quantile has a complex structure and the heteroscedasticity exists. This finding concludes that the GARCH effects occur in the exchange rate time series; see Engle, Ito and Lin (1990) and Tsay (2005).

Finally, we consider the post-sample forecasting for the last 25 observations based on the local linear quantile estimators which are computed by using the same bandwidths as

Table 2: The Post-Sample Predictive Intervals For Exchange Rate Data

| Observation | True Value | Prediction Interval |
|-------------|------------|---------------------|
| Y_{1541} | 0.392 | (-2.891, 2.412) |
| Y_{1542} | 0.509 | (-3.099, 2.405) |
| Y_{1543} | 1.549 | (-2.943, 2.446) |
| Y_{1544} | -0.121 | (-2.684, 2.525) |
| Y_{1545} | -0.991 | (-2.677, 2.530) |
| Y_{1546} | -0.646 | (-3.110, 2.401) |
| Y_{1547} | -0.354 | (-3.178, 2.365) |
| Y_{1548} | -1.393 | (-3.083, 2.372) |
| Y_{1549} | 0.997 | (-3.110, 2.230) |
| Y_{1550} | -0.916 | (-3.033, 2.431) |
| Y_{1551} | -3.707 | (-3.021, 2.286) |
| Y_{1552} | -0.919 | (-3.841, 2.094) |
| Y_{1553} | -0.901 | (-3.603, 2.770) |
| Y_{1554} | 0.071 | (-3.583, 2.821) |
| Y_{1555} | -0.497 | (-3.351, 2.899) |
| Y_{1556} | -0.648 | (-3.436, 2.783) |
| Y_{1557} | 1.648 | (-3.524, 2.866) |
| Y_{1558} | -1.184 | (-3.121, 2.810) |
| Y_{1559} | 0.530 | (-3.529, 2.531) |
| Y_{1560} | 0.107 | (-3.222, 2.648) |
| Y_{1561} | -0.804 | (-3.294, 2.651) |
| Y_{1562} | 0.274 | (-3.419, 2.534) |
| Y_{1563} | -0.847 | (-3.242, 2.640) |
| Y_{1564} | -0.060 | (-3.426, 2.532) |
| Y_{1565} | -0.088 | (-3.300, 2.576) |

those used in the model fitting. The 95% nonparametric prediction interval is constructed as $(\hat{q}_{0.025}(\cdot), \hat{q}_{0.975}(\cdot))$ and the prediction results are reported in Table 2, which shows that 24 out of 25 predictive intervals contain the corresponding true values. The average length of the intervals is 5.77, which is about 35.5% of the range of the data. Therefore, we can

conclude that under the dynamic smooth coefficient quantile regression model assumption, the prediction intervals based on the proposed method work reasonably well.

4 Conclusion

In this paper, we studied a class of quantile regression models with functional coefficients for time series data. We suggested using the local polynomial fitting scheme to estimate the nonparametric coefficient functions and derived the asymptotic properties of the proposed estimators. An ad hoc method is proposed to select the bandwidth and estimate the asymptotic covariance. A Monte Carlo simulation experiment was conducted to illustrate the proposed methodology and two real data sets were analyzed. There are some new findings to these two real example based on the dynamic smooth coefficient quantile regression model. Some interesting future research topics related to this paper should be mentioned. First, it would be very useful to discuss the bandwidth theoretically and empirically. Secondly, an important application of quantile regression is to measure how much the τ th response quantile changes as one covariate is perturbed while the other covariates are held fixed; see Chaudhuri, Doksum and Samarov (1997). Therefore, we can estimate $\nabla q_\tau(U, \mathbf{x})$ by using the proposed methodology. Further, the above models and results can be extended to the following cases: some of \mathbf{X}_t might be non-stationary such as I(1) and some of \mathbf{X}_t might be endogenous. Finally, the nonparametric quantile regression can be potentially applied to the analysis of the financial data such as GARCH type models studied by Xiao (2006) and VaR and other type risk models and their extensions; see Bassett, Koenker and Kordas (2004).

Appendix

In this section, due to the limited space, we give only the brief derivations of the main results based on some lemmas. The detailed technical proofs of those lemmas can be found in Xu (2005). For the expositional purpose, we consider only the case for $q = 1$. First, we need the following two lemmas and their proofs can be found in Koenker and Zhao (1996) and Ruppert and Carroll (1980).

Lemma 1: *Let $V_n(\Delta)$ be a vector function that satisfies*

$$(i) \quad -\Delta' V_n(\lambda \Delta) \geq -\Delta' V_n(\Delta) \text{ for } \lambda \geq 1$$

and

(ii) $\sup_{\|\Delta\| \leq M} \|V_n(\Delta) + \mathbf{D} \Delta - \mathbf{A}_n\| = o_p(1)$, where $\|\mathbf{A}_n\| = O_p(1)$, $0 < M < \infty$, and \mathbf{D} is a positive-definite matrix. Suppose that Δ_n is a vector such that $\|V_n(\Delta_n)\| = o_p(1)$, then, we have

$$(1) \quad \|\Delta_n\| = O_p(1) \quad \text{and} \quad (2) \quad \Delta_n = \mathbf{D}^{-1} \mathbf{A}_n + o_p(1).$$

Lemma 2: Let $\hat{\boldsymbol{\beta}}$ be the minimizer of the function $\sum_{t=1}^n w_t \rho_\tau(y_t - \mathbf{X}'_t \boldsymbol{\beta})$, where $w_t > 0$ and $\rho_\tau(\cdot)$ is the check function defined in Section 1. Then, $\|\sum_{t=1}^n w_t \mathbf{X}_t \psi_\tau(y_t - \mathbf{X}'_t \hat{\boldsymbol{\beta}})\| \leq \dim(\mathbf{X}) \max_{t \leq n} \|w_t \mathbf{X}_t\|$, where $\psi_\tau(\cdot)$ is defined in Section 2.3.

By the definition of $\boldsymbol{\theta}$ defined in Section 2.3, $\boldsymbol{\beta} = \begin{pmatrix} \mathbf{a}(u_0) \\ \mathbf{a}'(u_0) \end{pmatrix} + a_n \mathbf{H}^{-1} \boldsymbol{\theta}$, where a_n is defined in Section 2.3. Thus, $Y_t - \sum_{j=0}^q \mathbf{X}'_t \boldsymbol{\beta}_j (U_t - u_0)^j = Y_t^* - a_n \boldsymbol{\theta}' \mathbf{X}_t^*$. Therefore,

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin} \sum_{t=1}^n \rho_\tau [Y_t^* - a_n \boldsymbol{\theta}' \mathbf{X}_t^*] K(U_{th}) \equiv \operatorname{argmin} G(\boldsymbol{\theta}).$$

Now, define $V_n(\boldsymbol{\theta}) = a_n \sum_{t=1}^n \psi_\tau [Y_t^* - a_n \boldsymbol{\theta}' \mathbf{X}_t^*] \mathbf{X}_t^* K(U_{th})$. To establish the asymptotic properties of $\hat{\boldsymbol{\theta}}$, we show in the next two lemmas that $V_n(\boldsymbol{\theta})$ satisfies Lemma 1 so that we can derive the local Bahadur representation for $\hat{\boldsymbol{\theta}}$. The results are stated here and their detailed proofs can be found in Xu (2005). For the notational convenience define $A_m = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| \leq M\}$ for some $0 < M < \infty$ and $\mathbf{Z}_t = \psi_\tau(Y_t^*) \mathbf{X}_t^* K(U_{th})$.

Lemma 3: Under assumptions of Theorem 1, we have

$$\|V_n(\boldsymbol{\theta}) - V_n(0) - E[V_n(\boldsymbol{\theta}) - V_n(0)]\| = o_p(1) \quad \text{and} \quad \|E[V_n(\boldsymbol{\theta}) - V_n(0)] + \mathbf{D} \boldsymbol{\theta}\| = o(1)$$

uniformly over $\boldsymbol{\theta} \in A_m$, where $\mathbf{D} = f_u(u_0) \Omega_1^*(u_0)$.

Lemma 4: Under assumptions of Theorem 1, we have

$$E[\mathbf{Z}_1] \approx \frac{h^3 f(u_0)}{2} \begin{pmatrix} \mu_2 \Omega^*(u_0) \mathbf{a}''(u_0) \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad \operatorname{Var}[\mathbf{Z}_1] \approx h \tau(1 - \tau) f(u_0) \Omega_1(u_0),$$

where $\Omega_1(u_0) = \operatorname{diag}\{\nu_0 \Omega(u_0), \nu_2 \Omega(u_0)\}$. Further, $\operatorname{Var}[V_n(0)] \approx \tau(1 - \tau) f(u_0) \Omega_1(u_0)$. Therefore, $\|V_n(0)\| = O_p(1)$.

Proof of Theorem 1: Here we present only the sketch of proofs and details can be found in Xu (2005). By Lemmas 3 and 4, $V_n(\boldsymbol{\theta})$ satisfies the condition (ii) of Lemma 1; that is, $\|\mathbf{A}_n\| = O_p(1)$ and $\sup_{\boldsymbol{\theta} \in A_m} \|V_n(\boldsymbol{\theta}) + \mathbf{D} \boldsymbol{\theta} - \mathbf{A}_n\| = o_p(1)$, where $\mathbf{A}_n = V_n(0)$. It follows from Lemma 2 that $\|V_n(\hat{\boldsymbol{\theta}})\| = o_p(1)$, where $\hat{\boldsymbol{\theta}}$ is the minimizer of $G(\boldsymbol{\theta})$. Finally, since $\psi_\tau(x)$ is an increasing function of x , then, $-\boldsymbol{\theta}' V_n(\lambda \boldsymbol{\theta}) = a_n \sum_{t=1}^n \psi_\tau[Y_t^* + \lambda a_n (-\boldsymbol{\theta}' \mathbf{X}_t^*)] (-\boldsymbol{\theta}' \mathbf{X}_t^*) K(U_{th})$ is an increasing function of λ . Thus, the condition (i) of Lemma 1 is satisfied. Therefore,

$$\hat{\boldsymbol{\theta}} = \mathbf{D}^{-1} \mathbf{A}_n + o_p(1) = \frac{(\Omega_1^*)^{-1}}{\sqrt{n h} f_u(u_0)} \sum_{t=1}^n \psi_\tau(Y_t^*) \mathbf{X}_t^* K(U_{th}) + o_p(1). \quad (\text{A.1})$$

Let $\varepsilon_t = \psi_\tau(Y_t - \mathbf{X}'_t \mathbf{a}(U_t))$. Then, it is easy to show that $E(\varepsilon_t) = 0$ and $\operatorname{Var}(\varepsilon_t) = \tau(1 - \tau)$. From (A.1),

$$\hat{\boldsymbol{\theta}} \approx \frac{(\Omega_1^*)^{-1}}{\sqrt{n h} f_u(u_0)} \sum_{t=1}^n [\psi_\tau(Y_t^*) - \varepsilon_t] \mathbf{X}_t^* K(U_{th}) + \frac{(\Omega_1^*)^{-1}}{\sqrt{n h} f_u(u_0)} \sum_{t=1}^n \varepsilon_t \mathbf{X}_t^* K(U_{th}) \equiv \mathbf{B}_n + \boldsymbol{\xi}_n.$$

Similar to the proof of Theorem 2 in Cai, Fan and Yao (2000), by using the small-block and large-block technique and the Cramér-Wold device, one can show (although lengthly and tediously) that

$$\boldsymbol{\xi}_n \rightarrow N(\mathbf{0}, \tau(1-\tau)\nu_0 \boldsymbol{\Sigma}_\theta(u_0)), \quad (\text{A.2})$$

where $\boldsymbol{\Sigma}_\theta(u_0) = \text{diag}\{\nu_0 \boldsymbol{\Sigma}(u_0), \nu_2 \boldsymbol{\Sigma}(u_0)\}$. By the stationarity and Lemma 4,

$$E[\mathbf{B}_n] = \frac{(\Omega_1^*)^{-1}}{\sqrt{n} h f_u(u_0)} n E[\mathbf{Z}_1] \{1 + o(1)\} = a_n^{-1} \frac{h^2}{2} \begin{pmatrix} \mathbf{a}''(u_0) \mu_2 \\ \mathbf{0} \end{pmatrix} \{1 + o(1)\}. \quad (\text{A.3})$$

Since $\psi_\tau(Y_t^*) - \varepsilon_t = I(Y_t \leq \mathbf{X}_t' \mathbf{a}(U_t)) - I(Y_t \leq \mathbf{X}_t' (\mathbf{a}(u_0) + \mathbf{a}'(u_0)(U_t - u_0)))$, then,

$$[\psi_\tau(Y_t^*) - \varepsilon_t]^2 = I(d_{1t} < Y_t \leq d_{2t}),$$

where $d_{1t} = \min(c_{1t}, c_{2t})$ and $d_{2t} = \max(c_{1t}, c_{2t})$ with $c_{1t} = \mathbf{X}_t' \mathbf{a}(U_t)$ and $c_{2t} = \mathbf{X}_t' [\mathbf{a}(u_0) + \mathbf{a}'(u_0)(U_t - u_0)]$. Further,

$$E[\{\psi_\tau(Y_t^*) - \varepsilon_t\}^2 K^2(U_{th}) \mathbf{X}_t^* \mathbf{X}_t^{*'}] = E[\{F_{y|u,x}(d_{2t}) - F_{y|u,x}(d_{1t})\} K^2(U_{th}) \mathbf{X}_t^* \mathbf{X}_t^{*'}] = O(h^3).$$

Thus, $\text{Var}(\mathbf{B}_n) = o(1)$. This, in conjunction with (A.2) and (A.3) and the Slutsky Theorem, proves the theorem. \square

Proof of (6) and (7): By the Taylor expansion,

$$E[\xi_t | U_t, \mathbf{X}_t] = F_{y|u,x}(\mathbf{X}_t' \mathbf{a}(u_0) + a_n) - F_{y|u,x}(\mathbf{X}_t' \mathbf{a}(u_0)) \approx f_{y|u,x}(\mathbf{X}_t' \mathbf{a}(u_0)) a_n.$$

Therefore,

$$E[\mathbf{S}_n] \approx h^{-1} E[f_{y|u,x}(\mathbf{X}_t' \mathbf{a}(u_0)) \mathbf{X}_t^* \mathbf{X}_t^{*'} K(U_{th})] \approx f_u(u_0) \Omega_1^*(u_0).$$

Similar to the proof of $\text{Var}[V_n(0)]$ in Lemma 4, one can show that $\text{Var}(\mathbf{S}_n) \rightarrow 0$. Therefore, $\mathbf{S}_n \rightarrow f_u(u_0) \Omega_1^*(u_0)$ in probability. This proves (6). Clearly,

$$E[\widehat{\Omega}_{n,0}] = E[\mathbf{X}_t \mathbf{X}_t' K_h(U_t - u_0)] = \int \Omega(u_0 + hv) f_u(u_0 + hv) K(v) dv \approx f_u(u_0) \Omega(u_0).$$

Similarly, one can show that $\text{Var}(\widehat{\Omega}_{n,0}) \rightarrow 0$. This proves the first part of (7). By the same token, one can show that $E[\widehat{\Omega}_{n,1}] \approx f_u(u_0) \Omega^*(u_0)$ and $\text{Var}(\widehat{\Omega}_{n,1}) \rightarrow 0$. Thus, $\widehat{\Omega}_{n,1} = f_u(u_0) \Omega^*(u_0) + o_p(1)$. We prove (7). \square

References

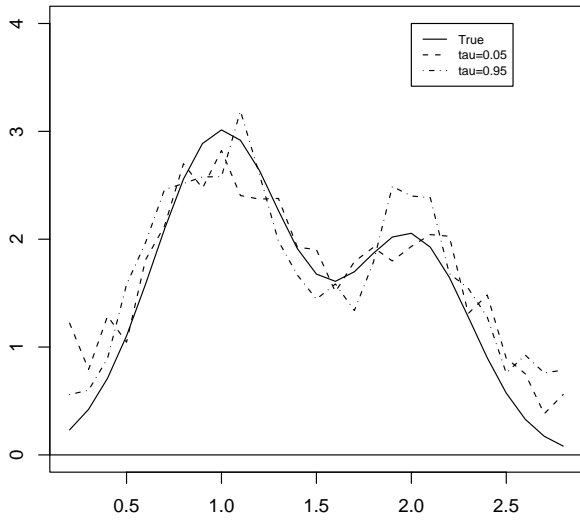
Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika*, **77**, 669-687.

- Bao, Y., Lee, T.-H. and Saltoğlu, B. (2001). Evaluating predictive performance of value-at-risk models in emerging markets: a reality check. Forthcoming in *Journal of Forecasting*.
- Bassett, G.W., Koenker, R. and Kordas, G. (2004). Pessimistic portfolio allocation and Choquet expected utility. *Journal Financial Econometrics*, **2**, 477-492.
- Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformation for multiple regression and correlation. *Journal of the American Statistical Association*, **80**, 580-619.
- Cai, Z. (2002a). Regression quantile for time series. *Econometric Theory*, **18**, 169-192.
- Cai, Z. (2002b). A two-stage approach to additive time series models. *Statistica Neerlandica*, **56**, 415-433.
- Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, **95**, 941-956.
- Cai, Z. and Masry, E. (2000). Nonparametric estimation in nonlinear ARX time series models: Projection and linear fitting. *Econometric Theory*, **16**, 465-501.
- Cai, Z. and Tiwari, R.C. (2000). Application of a local linear autoregressive model to BOD time series. *Environmetrics*, **11**, 341-350.
- Carrasco, M. and Chen, X. (2002). Mixing and moments properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17-39.
- Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics*, **19**, 760-777.
- Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, **25**, 715-744.
- Chen, R. and Tsay, R.S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, **88**, 298-308.
- Cole, T.J. (1994). Growth charts for both cross-sectional and longitudinal data. *Statistics in Medicine*, **13**, 2477-2492.
- De Gooijer, J. and Zerom, D. (2003). On additive conditional quantiles with high dimensional covariates. *Journal of the American Statistical Association*, **98**, 135-146.
- Engle, R.F., Ito, T. and Lin, W. (1990). Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica*, **58**, 525-542.
- Engle, R.F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantile. *Journal of Business and Economics Statistics*, **22**, 367-381.

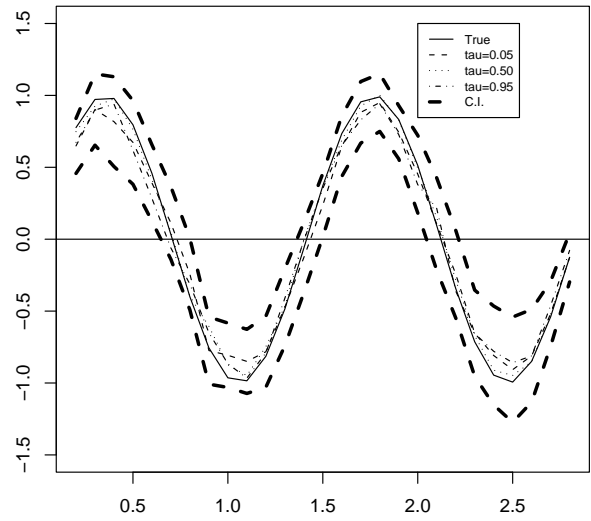
- Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, **1**, 93-125.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London.
- Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85**, 645-660.
- Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society, Series B*, **65**, 57-80.
- Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83**, 189-206.
- Granger, C.W.J., White, H. and Kamstra, M. (1989). Interval forecasting: an analysis based upon ARCH-quantile estimators. *Journal of Econometrics*, **40**, 87-96.
- Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Applications*. Academic Press, New York.
- Harrison, D. and Rubinfeld, D.L. (1978). Hedonic housing prices and demand for clean air. *Journal of Environmental Economics and Management*, **5**, 81-102.
- He, X. and Ng, P. (1999). Quantile splines with several covariates. *Journal of Statistical Planning and Inference*, **75**, 343-352.
- He, X., Ng, P. and Portnoy, S. (1998). Bivariate quantile smoothing splines. *Journal of the Royal Statistical Society, Series B*, **60**, 537-550.
- He, X. and Portnoy, S. (2000). Some asymptotic results on bivariate quantile splines. *Journal of Statistical Planning and Inference*, **91**, 341-349.
- Honda, T. (2000). Nonparametric estimation of a conditional quantile for α -mixing processes. *Annals of the Institute of Statistical Mathematics*, **52**, 459-470.
- Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inferences*, **121**, 113-125.
- Hong, Y. and Lee, T.-H. (2003). Inference on via generalized spectrum and nonlinear time series models. *The Review of Economics and Statistics*, **85**, 1048-1062.
- Horowitz, J.L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, **100**, 1238-1249.
- Hurvich, C.M., Simonoff, J.S. and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society, Series B*, **60**, 271-293.

- Khindanova, I.N. and Rachev, S.T. (2000). Value at risk: Recent advances. *Handbook on Analytic-Computational Methods in Applied Mathematics*, CRC Press LLC.
- Kim, M. (2003). *Quantile Regression in a Varying Coefficient Model*. Ph.D. Dissertation, Department of Statistics, University of Illinois at Urbana and Champaign.
- Koenker, R. (1994). Confidence intervals for regression quantiles. In *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Huskova, eds.), 349-359. Physica, Heidelberg.
- Koenker, R. (2004). *Quantreg: An R package for quantile regression and related methods* <http://cran.r-project.org>.
- Koenker, R. and Bassett, G.W. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.
- Koenker, R. and Bassett, G.W. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, **50**, 43-61.
- Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673-680.
- Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, **70**, 1583-1612.
- Koenker, R. and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association*, **99**, 775-787.
- Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, **12**, 793-813.
- Lu, Z. (1998). On the ergodicity of non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica*, **8**, 1205-1217.
- Machado, J.A.F. (1993). Robust model selection and M -estimation. *Econometric Theory*, **9**, 478-493.
- Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of non-linear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory*, **11**, 258-289.
- Masry, E. and Tjøstheim, D. (1997). Additive nonlinear ARX time series and projection estimates. *Econometric Theory*, **13**, 214-252.
- Opsomer, J.D. and Ruppert, D. (1998). A fully automated bandwidth selection for additive regression model. *Journal of The American Statistical Association*, **93**, 605-618.
- Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, **75**, 828-838.

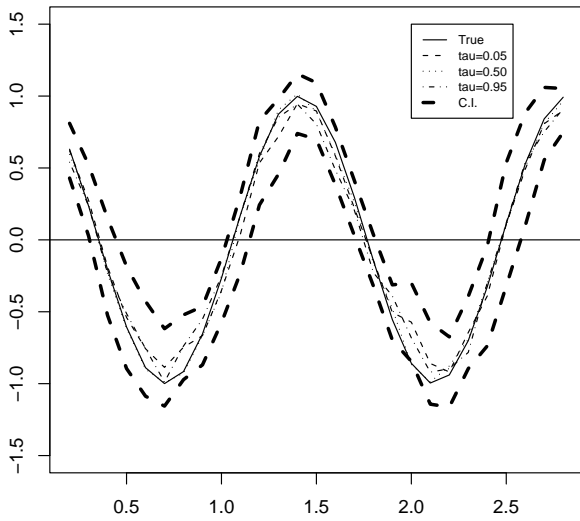
- Ruppert, D. and Wand, M. (1994). Multivariate locally least squares regression. *The Annals of Statistics*, **22**, 1346-1370.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.
- Şentürk, D. and Müller, H.G. (2003). Inference for covariate adjusted regression via varying coefficient models. Forthcoming in *Scandinavian Journal of Statistics*.
- Taylor, J.W. and Bunn, D.W. (1999). A quantile regression approach to generating prediction intervals. *Management Science*, **45**, 225-237.
- Tsay, R.S. (2005). *Analysis of Financial Time Series*, 2nd Edition. John Wiley & Sons, New York.
- Wei, Y. and He, X. (2005). Conditional growth charts. Forthcoming in *The Annals of Statistics*.
- Wei, Y., Pere, A., Koenker, R. and He, X. (2003). Quantile regression methods for reference growth charts. Forthcoming in *Statistics in Medicine*.
- Xiao, Z. (2006). Conditional quantile estimation for GARCH models. *Working Paper*, Department of Economics, Boston College.
- Xu, X. (2005). *Semi/Nonparametric Quantile Dynamic Time Series Models and Their Applications*. Ph.D. Dissertation, University of North Carolina at Charlotte.
- Zhou, K.Q. and Portnoy, S.L. (1996). Direct use of regression quantiles to construct confidence sets in linear models. *The Annals of Statistics*, **24**, 287-306.
- Yu, K. and Jones, M.C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228-237.
- Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, **31**, 333-346.



(a)

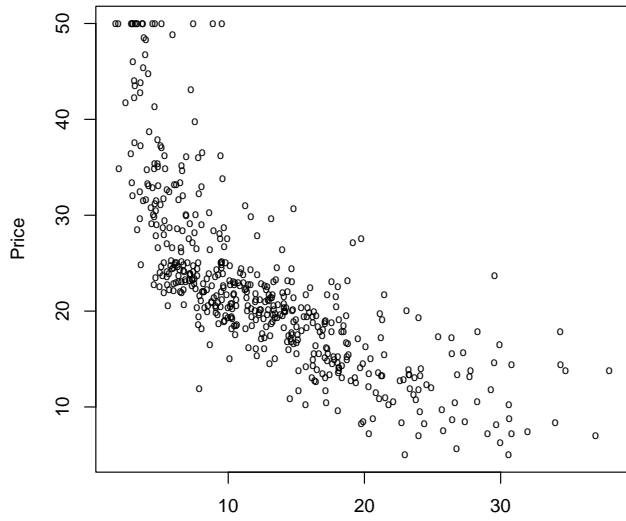


(b)

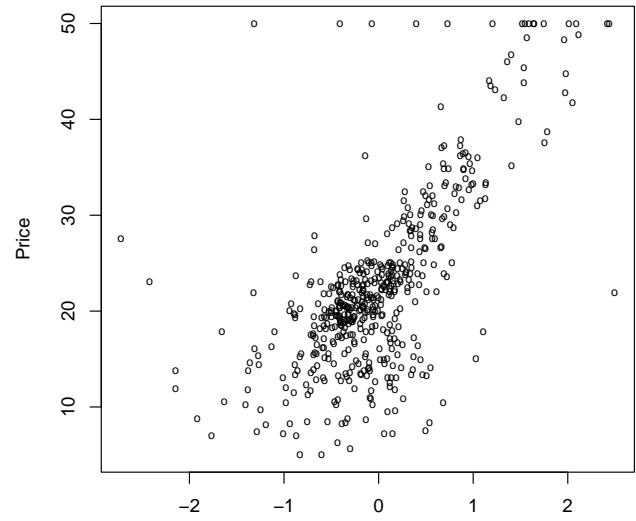


(c)

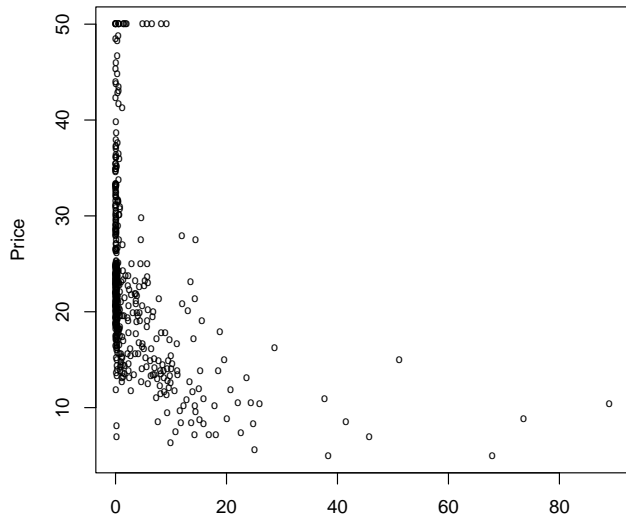
Figure 1: Simulated Example: The plots of the estimated coefficient functions for three quantiles $\tau = 0.05$ (dashed line), $\tau = 0.50$ (dotted line), and $\tau = 0.95$ (dot-dashed line) with their true functions (solid line): $\sigma(u)$ versus u in (a), $a_1(u)$ versus u in (b), and $a_2(u)$ versus u in (c). The 95% point-wise confidence interval (thick line) with the bias ignored for the $\tau = 0.5$ quantile estimate are provided in (b) and (c).



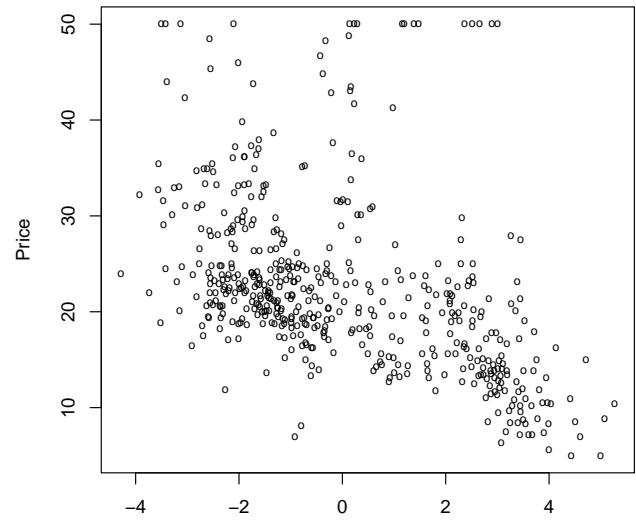
(a)



(b)

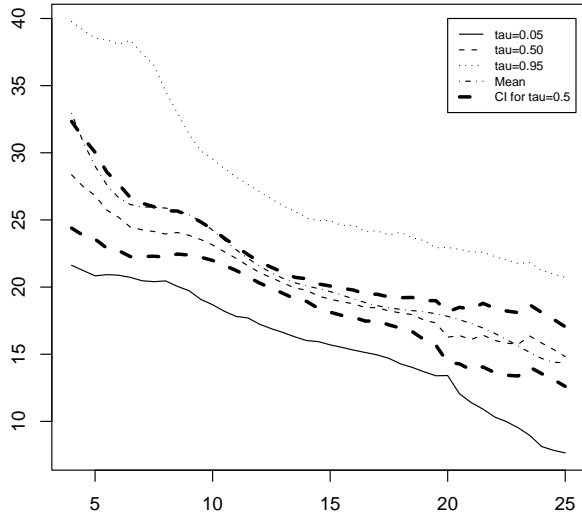


(c)

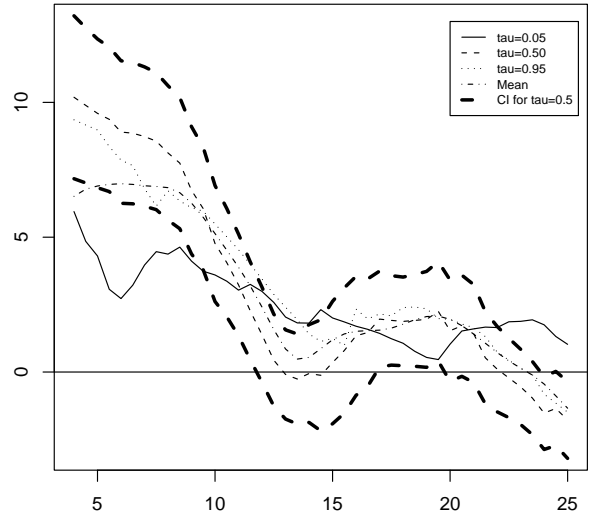


(d)

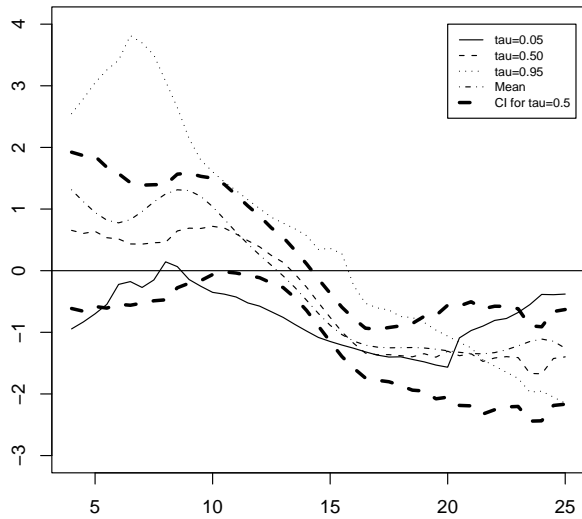
Figure 2: Boston Housing Price Data: Displayed in (a)-(d) are the scatter plots of the house price versus the covariates U , X_1 , X_2 and $\log(X_2)$, respectively.



(e)

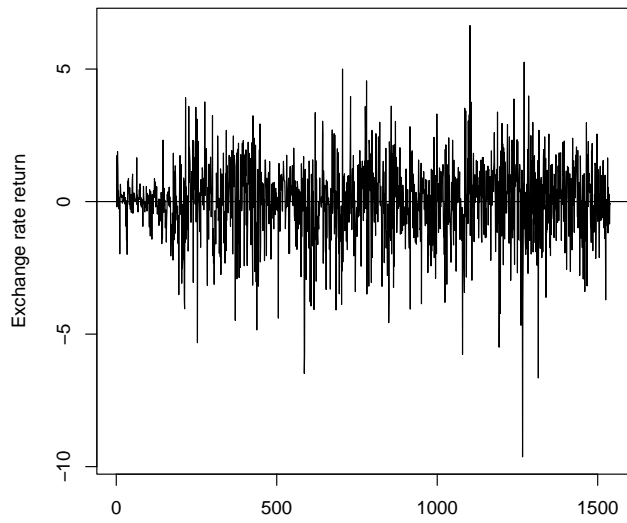


(f)

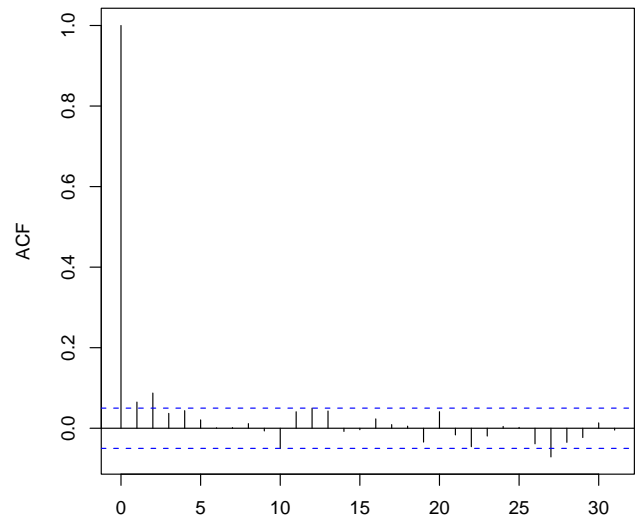


(g)

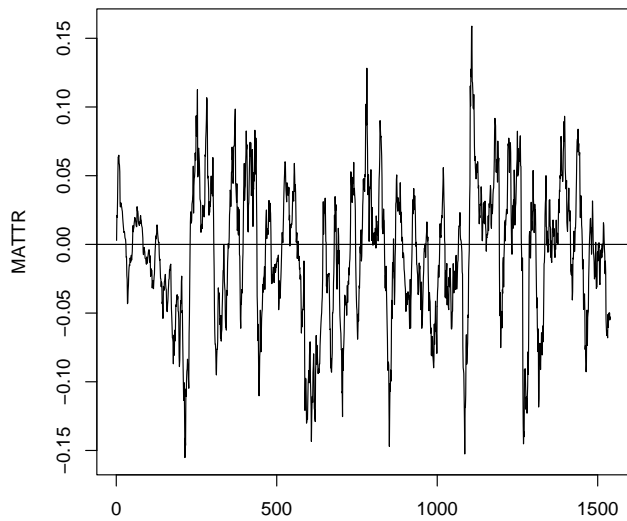
Figure 2: Boston Housing Price Data: The plots of the estimated coefficient functions for three quantiles $\tau = 0.05$ (solid line), $\tau = 0.50$ (dashed line), and $\tau = 0.95$ (dotted line), and the mean regression (dot-dashed line): $\hat{a}_{0,\tau}(u)$ and $\hat{a}_0(u)$ versus u in (e), $\hat{a}_{1,\tau}(u)$ and $\hat{a}_1(u)$ versus u in (f), and $\hat{a}_{2,\tau}(u)$ and $\hat{a}_2(u)$ versus u in (g). The thick dashed lines indicate the 95% point-wise confidence interval for the median estimate with the bias ignored.



(a)

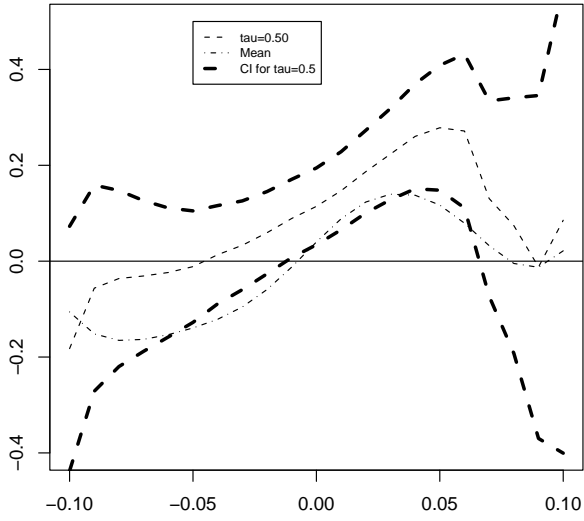


(b)

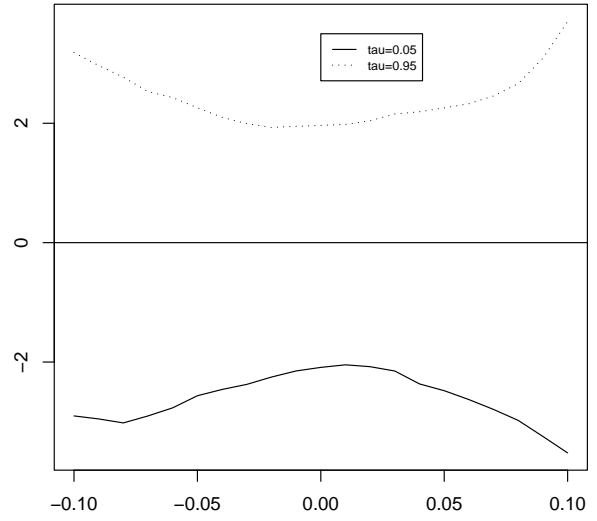


(c)

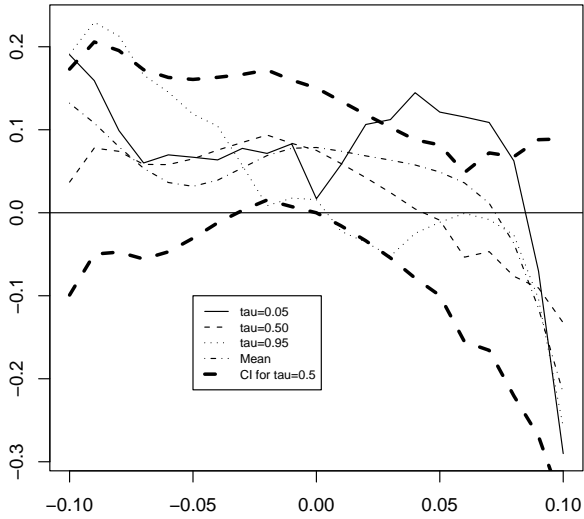
Figure 3: Exchange Rate Series: (a) Japanese-dollar exchange rate return series $\{Y_t\}$; (b) autocorrelation function of $\{Y_t\}$; (c) moving average trading technique rule.



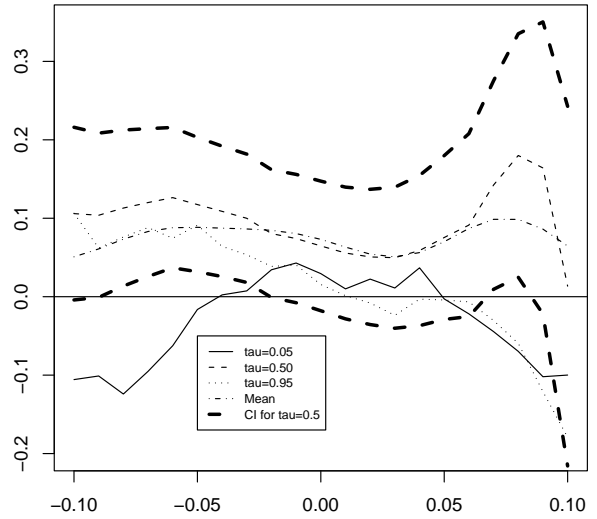
(d)



(e)



(f)



(g)

Figure 3: Exchange Rate Series: The plots of the estimated coefficient functions for three quantiles $\tau = 0.05$ (solid line), $\tau = 0.50$ (dashed line), and $\tau = 0.95$ (dotted line), and the mean regression (dot-dashed line): $\hat{a}_{0,0.50}(u)$ and $\hat{a}_0(u)$ versus u in (d), $\hat{a}_{0,0.05}(u)$ and $\hat{a}_{0,0.95}(u)$ versus u in (e), $\hat{a}_{1,\tau}(u)$ and $\hat{a}_1(u)$ versus u in (f), and $\hat{a}_{2,\tau}(u)$ and $\hat{a}_2(u)$ versus u in (g). The thick dashed lines indicate the 95% point-wise confidence interval for the median estimate with the bias ignored.