Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models

Zongwu CAI and Xiaoping XU

We suggest quantile regression methods for a class of smooth coefficient time series models. We use both local polynomial and local constant fitting schemes to estimate the smooth coefficients in a quantile framework. We establish the asymptotic properties of both the local polynomial and local constant estimators for α-mixing time series. We also suggest a bandwidth selector based on the nonparametric version of the Akaike information criterion, along with a consistent estimate of the asymptotic covariance matrix. We evaluate the asymptotic behaviors of the estimators at boundaries and compare the local polynomial quantile estimator and the local constant estimator. A simulation study is carried out to illustrate the performance of estimates. An empirical application of the model to real data further demonstrates the potential of the proposed modeling procedures.

KEY WORDS: Bandwidth selection; Boundary effect; Covariance estimation; Kernel smoothing method; Nonlinear time series; Quantile regression; Value-at-risk; Varying coefficients.

1. INTRODUCTION

Over the last three decades, quantile regression, also called conditional quantile or regression quantile (introduced in Koenker and Bassett 1978), has been widely used in various disciplines, including finance, economics, medicine, and biology. It is well known that when the distribution of data is typically skewed or the data contain some outliers, the median regression—a special case of quantile regression—is more explicable and robust than the mean regression. In addition, regression quantiles can be used to test heteroscedasticity formally or graphically (e.g., Koenker and Bassett 1982; 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 that can characterize the entire conditional distribution. Another, more important 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 (e.g., Koenker 1994; Koenker and Zhao 1996; Zhou and Portnoy 1996).

Recently, the quantile regression technique has been successfully applied to various applied fields. For example, by following the regulations of the Bank for International Settlements, many 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, the VaR computes an estimate of the lower-tail quantile (with a small probability) of future portfolio returns, conditional on current information. Therefore, VaR can be considered a special application of the quantile regression. There is a vast amount of literature on this topic (see Khindanova and Rachev 2000; Engle and Manganelli 2004 for further discussion).

In this article we assume that $\{\mathbf{X}_t, Y_t\}_{t=-\infty}^{\infty}$ is a stationary sequence. Let $F(y|\mathbf{x})$ denote 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 $\Re^d$ with $d \geq 1$, which may be a function of exogenous variables (covariates) or some lagged variables or time t. A regression quantile function $q_\tau(\mathbf{x})$ is defined as

$$q_\tau(\mathbf{x}) = \arg\min_{a \in \Re} E\{\rho_\tau(Y_t - a)|\mathbf{X}_t = \mathbf{x}\}, \tag{1}$$

for any $0 < \tau < 1$, where $\rho_\tau(y) = y(\tau - I_{\{y<0\}})$ with $y \in \Re$ is called the loss ("check") function and I_A is the indicator function of any set A. Clearly, the simplest form of (1) is $q_\tau(\mathbf{x}) = \boldsymbol{\beta}_\tau' \mathbf{x}$, which is known as the linear quantile regression model.

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. In a effort to make quantile regression models more flexible, there is a swiftly growing literature on nonparametric quantile regression. Various smoothing techniques (e.g., kernel methods, splines, and their variants) have been used to estimate the nonparametric quantile regression for both independent and time series data. Some recent developments and detailed discussions on theory, methodologies, and applications can be found in the literature. 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 conditions. 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, and the visual display is not so useful for exploratory purposes.

To address the foregoing 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 of Cai (2002) and then used a projection

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

Journal of the American Statistical Association
December 2008, Vol. 103, No. 484, Theory and Methods
DOI 10.1198/016214508000000977

method to estimate each component, whereas Yu and Lu (2004) focused on independent data and used a backfitting algorithm method to estimate each component. In contrast, to estimate each additive component for independent data, Horowitz and Lee (2005) used a two-stage approach consisting of the series estimation as the first step and a local polynomial fitting as the second step. For independent data, the foregoing model was extended by He, Ng, and Portnoy (1998), He and Ng (1999), and He and Portnoy (2000) to include interaction terms by using spline methods. Finally, Xiao (2006) investigated a new, robust approach for estimating conditional quantiles based on generalized autoregressive conditional heterscedasticity (GARCH)-type models. Because quantile regression estimation of GARCH models is highly nonlinear, Xiao (2006) discussed the problem of estimating this type of model using traditional recursive methods for nonlinear quantile regression and proposed two new methods for estimating quantiles of GARCH models.

In this article we adapt another dimension reduction modeling method to analyze dynamic time series data, termed the smooth (functional or varying) coefficient modeling approach. A smooth coefficient quantile regression model for time series data takes the form

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

where $\mathbf{U}_t$ is called the smoothing variable, which might be one part of $X_{t1}, \ldots, X_{td}$ or time or other exogenous variables or lagged variables; $\mathbf{X}_t = (X_{t0}, X_{t1}, \ldots, X_{td})'$ with $X_{t0} \equiv 1$ are covariates; $\{a_{k,\tau}(\cdot)\}$ are smooth coefficient functions; and $\mathbf{a}_\tau(\cdot) = (a_{0,\tau}(\cdot), \ldots, a_{d,\tau}(\cdot))'$. Here some of the $\{a_{k,\tau}(\cdot)\}$ are allowed to depend on τ. For simplicity, we drop τ from $\{a_{k,\tau}(\cdot)\}$ in what follows. Our interest here is in estimating coefficient functions $\mathbf{a}(\cdot)$ rather than the quantile regression surface $q_\tau(\cdot, \cdot)$ itself. Note that model (2) was studied by Honda (2004), Wei and He (2006), and Kim (2007) for an independent sample, but our focus here is on a dynamic model for nonlinear time series, which has more capacity for 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 it for unit root inference. In particular, it includes a specific class of autoregressive conditional heteroscedasticity (ARCH) models, such as heteroscedastic linear models, considered by Koenker and Zhao (1996) and nonlinear models, studied by Xiao (2006). In addition, if there is no $\mathbf{X}_t$ in the model ($d = 0$), then $q_\tau(\mathbf{U}_t, \mathbf{X}_t)$ becomes $q_\tau(\mathbf{U}_t)$, so that model (2) reduces to the ordinary nonparametric quantile regression model, which has been studied extensively (e.g., Chaudhuri, Doksum, and Samarov 1997; Yu and Jones 1998; Honda 2000; Cai 2002). If $\mathbf{U}_t$ is just time, then the model is called a time-varying coefficient quantile regression model, which is potentially useful for checking whether the quantile regression changes over time. A case of practical interest is the analysis of reference growth data by Cole (1994), Wei, Pere, Koenker, and He (2006), and Wei and He (2006).

The motivation of this study comes from analyzing the well-known Boston housing price data. The main interest lies in identifying factors affecting housing prices in the Boston area. As argued by Şentürk and Müller (2005), the correlation between housing prices and the crime rate can be adjusted by the confounding variable, the proportion of the population of lower educational status, through a varying-coefficient model, and the expected effect of increasing crime rate on declining housing prices seems to be observed only for lower educational status neighborhoods in Boston. The interesting features of this data set 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 covariates (including the confounding variable) are left-skewed. Therefore, quantile methods are suitable for analyzing this data set. This problem can be tackled using model (2). In another example, we are interested in exploring the possible nonlinearity, heteroscedasticity, and predictability of exchange rates, such as the Japanese Yen against the U.S. dollar. A detailed analysis of these data sets is reported in Section 3.

The rest of the article is organized as follows. In Section 2 we present both the local polynomial and local constant quantile estimations of coefficient functions. We also suggest an ad hoc data-driven fashioned bandwidth selector based on the nonparametric version of the Akaike information criterion (AIC) and provide a consistent estimator of the asymptotic covariance matrix. In Section 3 we illustrate the finite-sample performance of the proposed estimators with a Monte Carlo experiment and give an application to the exchange rate series and the Boston housing price data. We study the asymptotic properties of the proposed estimators in Section 4. We provide some concluding remarks in Section 5. Finally, in the Appendix we give brief derivations of the theorems with some lemmas.

2. MODELING PROCEDURES

2.1 Local Polynomial Quantile Estimate

Without loss of generality, we consider only the case in which $\mathbf{U}_t$ in (2) is one-dimensional, denoted by U_t in what follows. For multivariate $\mathbf{U}_t$, the modeling procedure and the related theory for the univariate case continue to hold, but further, more complicated notations are involved. To estimate the coefficient functions $\{a_k(\cdot)\}$, we use a local polynomial fitting because of its 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 that the coefficient functions $\{\mathbf{a}(\cdot)\}$ have the $(q+1)$th derivative $(q \geq 1)$, so that for any given grid point $u_0 \in \Re$, $a_k(\cdot)$ can be approximated by a polynomial function in a neighborhood of the given grid point u_0 as $\mathbf{a}(U_t) \approx \sum_{j=0}^{q} \boldsymbol{\beta}_j (U_t - u_0)^j$, where $\boldsymbol{\beta}_j = \mathbf{a}^{(j)}(u_0)/j!$ and $\mathbf{a}^{(j)}(u_0)$ is the jth derivative of $\mathbf{a}(u_0)$, so that $q_\tau(U_t, \mathbf{X}_t) \approx \sum_{j=0}^{q} \mathbf{X}_t' \boldsymbol{\beta}_j (U_t - u_0)^j$. Then the locally weighted loss function is

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

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 0, which controls the amount of smoothing used in estimation. Solving the minimization problem in (3) gives $\widehat{\mathbf{a}}(u_0) = \widehat{\boldsymbol{\beta}}_0$, the local polynomial estimate of $\mathbf{a}(u_0)$, and $\widehat{\mathbf{a}}^{(j)}(u_0) = j!\widehat{\boldsymbol{\beta}}_j$ $(j \geq 1)$, the local polynomial estimate of the jth derivative, $\mathbf{a}^{(j)}(u_0)$. By moving u_0 along with the real line, the estimate of the entire curve $\widehat{\mathbf{a}}(u_0)$ is obtained.

Note that the local constant (Nadaraya–Watson type) quantile estimation of $\mathbf{a}(u_0)$, denoted by $\widetilde{\mathbf{a}}(u_0)$, is $\widetilde{\boldsymbol{\beta}}$ minimizing the subjective function

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

which is a special case of (3) with $q = 0$. We compare $\widehat{\mathbf{a}}(u_0)$ and $\widetilde{\mathbf{a}}(u_0)$ theoretically at the end of Section 4 and empirically in Section 3.1. The comparisons lead us to suggest that the local polynomial approach should be used in practice.

Remark 1. Note that many other nonparametric methods can be used here, including spline approaches (e.g., He et al. 1998; He and Ng 1999; He and Portnoy 2000). As pointed out by the editor, local polynomial estimates of nonparametric quantile regressions might tend to be rough, particularly for small or large values of τ, because only a small number of data points are available in the regions. It this regard, a spline approach might be better, because it can be considered a global parametric method. But a spline method might not be rich enough to characterize the local properties of nonparametric functions.

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, almost nothing has been done about selecting the bandwidth in the context of estimating the coefficient functions in the quantile regression, even though there is a rich literature on this issue in the mean regression setting (see, e.g., Cai, Fan, and Yao 2000). Yu and Jones (1998) and Yu and Lu (2004) proposed a simple, convenient method for 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, because of (nonlinear) heteroscedasticity. Furthermore, the mean regression approach cannot directly estimate the variance function. To address these problems, we propose a method of selecting the bandwidth for the foregoing estimation procedure, based on the nonparametric version of the AIC, which can address the structure of time series data and the overfitting or underfitting tendency. The basic idea is motivated by its analog from Cai and Tiwari (2000) for nonlinear mean regression for time series models, as we describe briefly.

By recalling the classical AIC for linear models under the likelihood setting (i.e., the negative of twice of the maximized log-likelihood plus twice of the number of estimated parameters), we propose the following nonparametric version of the bias-corrected AIC, due to Hurvich, Simonoff, and Tsai (1998) and Cai and Tiwari (2000) for nonparametric regression models, to select h by minimizing

$$\mathrm{AIC}(h) = \log\{\widehat{\sigma}_\tau^2\} + 2(p_h + 1)/[n - (p_h + 2)], \tag{5}$$

where $\widehat{\sigma}_\tau^2$ and p_h are as defined later. This criterion may be interpreted as the AIC for the local quantile smoothing problem, and it seems to perform well in some limited applications. Note that, similar to (5), Koenker, Ng, and Portnoy (1994) considered the Schwarz information criterion (SIC) (Schwarz 1978) with the second term on the right 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. Machado (1993) studied similar criteria for parametric quantile regression models and more general M-estimators of regression.

We now turn to defining $\widehat{\sigma}_\tau^2$ and p_h in this setting. In the mean regression setting, $\widehat{\sigma}_\tau^2$ is just the mean squared error. In the quantile regression, we define $\widehat{\sigma}_\tau^2$ as $n^{-1}\sum_{t=1}^{t}\rho_\tau(Y_t - \mathbf{X}_t'\widehat{\mathbf{a}}(U_t))$, which may be interpreted as the mean squared error in the least squares setting and also was used by Koenker et al. (1994). In nonparametric models, p_h is the nonparametric version of degrees of freedom, called the effective number of parameters, which usually is based on the trace of various quasi-projection (hat) matrixes in the least squares theory (linear estimators). (See, e.g., Cai and Tiwari 2000 for a cogent discussion for nonparametric regression models for time series.) For the quantile smoothing setting, the explicit expression for the quasi-projection matrix does not exist, because of 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. Toward this end, set

$$U_{th} = (U_t - u_0)/h, \qquad \mathbf{X}_t^* = \begin{pmatrix} \mathbf{X}_t \\ U_{th}\mathbf{X}_t \end{pmatrix},$$

$$Y_t^* = Y_t - \mathbf{X}_t'\big[\mathbf{a}(u_0) + \mathbf{a}^{(1)}(u_0)(U_t - u_0)\big],$$

$$\mathbf{H} = \mathrm{diag}\{\mathbf{I}_d, h\mathbf{I}_d\},$$

$\mathbf{I}_d$ as the $d \times d$ identity matrix, and

$$\boldsymbol{\theta} = \sqrt{nh}\,\mathbf{H}\begin{pmatrix} \boldsymbol{\beta}_0 - \mathbf{a}(u_0) \\ \boldsymbol{\beta}_1 - \mathbf{a}^{(1)}(u_0) \end{pmatrix}.$$

Define $\mathbf{S}_n = \mathbf{S}_n(u_0) = a_n\sum_{t=1}^{n}\xi_t\mathbf{X}_t^*\mathbf{X}_t^{*\prime}K(U_{th})$, where $a_n = (nh)^{-1/2}$ and $\xi_t = I(Y_t \leq \mathbf{X}_t'\mathbf{a}(u_0) + a_n) - I(Y_t \leq \mathbf{X}_t'\mathbf{a}(u_0))$. In the Appendix we shown that

$$\mathbf{S}_n(u_0) = f_u(u_0)\boldsymbol{\Omega}_1^*(u_0) + o_p(1), \tag{6}$$

where $f_u(u)$ represents the marginal density of U, $\boldsymbol{\Omega}_1^*(u_0) = \mathrm{diag}\{1, \mu_2\} \otimes \boldsymbol{\Omega}^*(u_0)$, $\mu_2 = \int u^2 K(u)\,du$, $\boldsymbol{\Omega}^*(u_0) \equiv E[\mathbf{X}_t\mathbf{X}_t' \times f_{y|u,x}(q_\tau(u_0, \mathbf{X}_t))|U_t = u_0]$, and $f_{y|u,x}(y)$ is the conditional density of Y given U and $\mathbf{X}$. It is easy to see from (6) and (A.1) in the Appendix that $\widehat{\boldsymbol{\theta}} = a_n\mathbf{S}_n^{-1}\sum_{t=1}^{n}\psi_\tau(Y_t^*)\mathbf{X}_t^*K(U_{th}) + o_p(1)$, where $\psi_\tau(x) = \tau - I_{\{x<0\}}$. Then we have

$$\begin{aligned}\widehat{q}_\tau(U_t, \mathbf{X}_t) &- q_\tau(U_t, \mathbf{X}_t) \\ &= \frac{1}{n}\sum_{s=1}^{n}\psi_\tau(Y_s^*(U_t))K_h((U_s - U_t)/h)\mathbf{X}_t^{0\prime}\mathbf{S}_n^{-1}(U_t)\mathbf{X}_s^* \\ &\quad + o_p(a_n),\end{aligned}$$

where

$$\mathbf{X}_t^0 = \begin{pmatrix} \mathbf{X}_t \\ \mathbf{0} \end{pmatrix}.$$

The coefficient of $\psi_\tau(Y_s^*(U_s))$ on the right side of the foregoing expression is $\gamma_s = a_n^2 K(0)\mathbf{X}_s^{0\prime}\mathbf{S}_n^{-1}(U_s)\mathbf{X}_s^0$. Now we have that $p_h = \sum_{s=1}^{n}\gamma_s$, which can be considered an approximation to the trace of the quasi-projection (hat) matrix for linear estimators. In a practical implementation, we need to estimate $\mathbf{a}(u_0)$

first, because $\mathbf{S}_n(u_0)$ involves $\mathbf{a}(u_0)$. We recommend using a pilot bandwidth, which can be that proposed by Yu and Jones (1998). Similar to the least squares theory, as expected, the criterion proposed in (5) counteracts the overfitting tendency of the generalized cross-validation due to its relatively weak penalty and the underfitting of the SIC (Schwarz 1978) studied by Koenker et al. (1994) because of its heavy penalty.

2.2 Covariance Estimate

For the purpose of statistical inference, we next consider 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 $\mathbf{\Omega}(u_0)$ and $\mathbf{\Omega}^*(u_0)$. Toward this end, define $\widehat{\mathbf{\Omega}}_{n,0} = n^{-1}\sum_{t=1}^n \mathbf{X}_t\mathbf{X}_t' K_h(U_t - u_0)$ and $\widehat{\mathbf{\Omega}}_{n,1} = n^{-1}\sum_{t=1}^n w_t\mathbf{X}_t\mathbf{X}_t' \times K_h(U_t - u_0)$, where $w_t = I(\mathbf{X}_t'\widehat{\mathbf{a}}(u_0) - \delta_n < Y_t \le \mathbf{X}_t'\widehat{\mathbf{a}}(u_0) + \delta_n)/(2\delta_n)$ for any $\delta_n \to 0$ as $n \to \infty$. In the Appendix we shown that

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

where $\mathbf{\Omega}(u_0) \equiv E[\mathbf{X}_t\mathbf{X}_t'|U_t = u_0]$. Therefore, a consistent estimate of $\mathbf{\Sigma}(u_0) = [\mathbf{\Omega}^*(u_0)]^{-1}\mathbf{\Omega}(u_0)[\mathbf{\Omega}^*(u_0)]^{-1}/f_u(u_0)$ can be given by $\widehat{\mathbf{\Sigma}}(u_0) = \widehat{\mathbf{\Omega}}_{n,1}^{-1}\widehat{\mathbf{\Omega}}_{n,0}(u_0)\widehat{\mathbf{\Omega}}_{n,1}^{-1}$. Note that $\widehat{\mathbf{\Omega}}_{n,1}(u_0)$ may be close to singular for some sparse regions. To avoid this computational difficulty, two alternative methods are available for constructing a consistent estimate of $f_u(u_0)\mathbf{\Omega}^*(u_0)$ by estimating the conditional density of Y, $f_{y|u,x}(q_\tau(u,\mathbf{x}))$. The first of these methods is the Nadaraya–Watson type (or local linear) double-kernel method of Fan, Yao, and Tong (1996), defined as $\widehat{f}_{y|u,x}(q_\tau(u,\mathbf{x})) = \sum_{t=1}^n K_{h_2}(U_t - u, \mathbf{X}_t - \mathbf{x})L_{h_1}(Y_t - q_\tau(u,\mathbf{x}))/\sum_{t=1}^n K_{h_2}(U_t - u, \mathbf{X}_t - \mathbf{x})$, where $L(\cdot)$ is a kernel function. The second is the difference quotients method of Koenker and Xiao (2004), such that $\widehat{f}_{y|u,x}(q_\tau(u,\mathbf{x})) = (\tau_j - \tau_{j-1})/[q_{\tau_j}(u,\mathbf{x}) - q_{\tau_{j-1}}(u,\mathbf{x})]$ for some appropriately chosen sequence of $\{\tau_j\}$. Then, in view of the definition of $f_u(u_0)\mathbf{\Omega}^*(u_0)$, the estimator $\widetilde{\mathbf{\Omega}}_{n,1}$ can be constructed as $\widetilde{\mathbf{\Omega}}_{n,1} = n^{-1}\sum_{t=1}^n \widehat{f}_{y|u,x}(\widehat{q}_\tau(U_t,\mathbf{X}_t))\mathbf{X}_t\mathbf{X}_t' K_h(U_t - u_0)$. By an analog of (7), we can show that under some regularity conditions, both estimators are consistent.

3. EMPIRICAL APPLICATIONS

In this section we report a Monte Carlo simulation for examining the finite-sample property of the proposed estimator that we use to further explore the possible nonlinearity, heteroscedasticity, and predictability of the exchange rate of the Japanese Yen against the U.S. dollar and to identify factors affecting housing prices in the Boston area. In our computation we use the Epanechnikov kernel, $K(u) = .75(1 - u^2)I(|u| \le 1)$, and construct the pointwise confidence intervals based on the consistent estimate of the asymptotic covariance described in Section 2.3 without bias correction. In the examples that follow, we use the proposed data-driven bandwidth selection method proposed in Section 2.2 to choose the optimal h_{opt}. For a predetermined sequence of h's from a wide range (say from h_a to h_b with an increment of h_δ), based on the AIC bandwidth selector described in Section 2.2, we compute the AIC(h) for each h and choose h_{opt} to minimize the AIC(h).

3.1 A Simulated Example

Example 1. We consider the following data-generating process:

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

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 $e_t \sim \mathrm{N}(0, 1)$. 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 τth quantile of the standard normal. Therefore, only $a_0(\cdot)$ is a function of τ. Note that $a_0(\cdot) = 0$ when $\tau = .5$. To assess the performance of finite samples, we compute the mean absolute deviation error (MADE) for $\widehat{a}_j(\cdot)$, defined as $MADE_j = n_0^{-1}\sum_{k=1}^{n_0}|\widehat{a}_j(u_k) - a_j(u_k)|$, where $\widehat{a}_j(\cdot)$ is either the local linear or local constant quantile estimate of $a_j(\cdot)$ and $\{u_k = .1(k-1) + .2 : 1 \le k \le n_0 = 27\}$ are the grid points. The Monte Carlo simulation is repeated 500 times for each sample size $n = 200$, 500, and 1,000 and for each $\tau = .05$, .50, and .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 see that the MADE values for both the local linear and local constant quantile estimates decrease when

Table 1. Median and standard deviation of 500 MADE values

	$\tau = .05$			$\tau = .5$			$\tau = .95$		
n	MADE_0	MADE_1	MADE_2	MADE_0	MADE_1	MADE_2	MADE_0	MADE_1	MADE_2
Local linear estimator									
200	.911(.520)	.186(.041)	.177(.041)	.401(.091)	.092(.032)	.089(.032)	.920(.517)	.187(.042)	.175(.039)
500	.510(.414)	.085(.023)	.083(.02)	.311(.056)	.055(.019)	.055(.018)	.517(.390)	.085(.023)	.083(.023)
1,000	.419(.071)	.060(.018)	.059(.017)	.311(.051)	.050(.014)	.049(.014)	.416(.072)	.060(.017)	.059(.017)
Local constant estimator									
200	3.753(2.937)	.285(.050)	.290(.051)	.501(.115)	.144(.027)	.147(.028)	3.763(3.188)	.287(.052)	.287(.051)
500	2.201(3.025)	.147(.024)	.146(.025)	.355(.062)	.084(.016)	.085(.015)	2.223(3.320)	.147(.025)	.147(.025)
1,000	.883(.462)	.086(.015)	.086(.014)	.322(.054)	.060(.012)	.061(.011)	.882(.427)	.086(.015)	.087(.015)

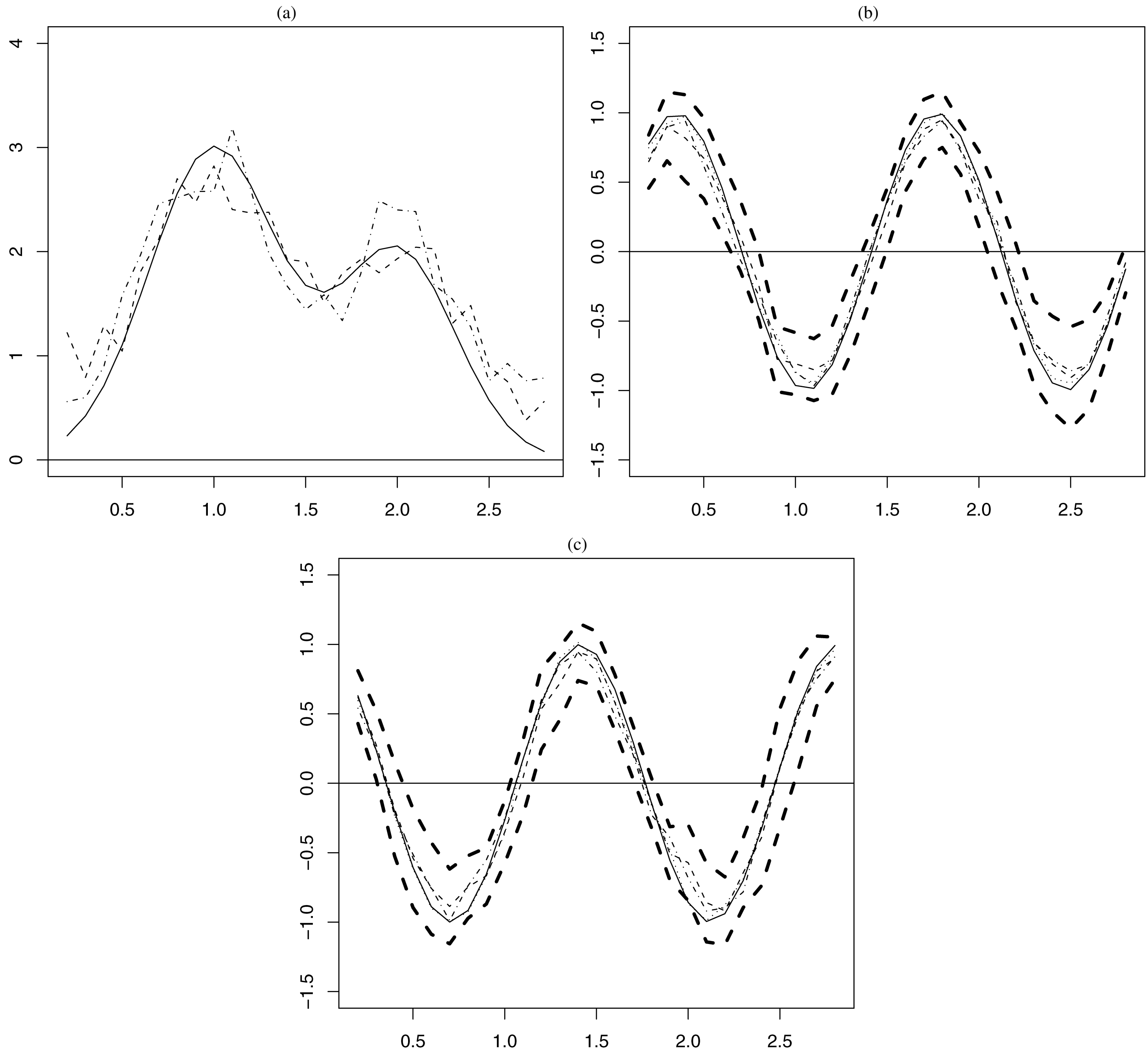

Figure 1. Simulated example. The plots of the estimated coefficient functions for three quantiles, $\tau = .05$ (- -), $\tau = .50$ (···), and $\tau = .95$ (·-·) with their true functions (—): $\sigma(u)$ versus u in (a), $a_1(u)$ versus u in (b), and $a_2(u)$ versus u in (c). The 95% pointwise confidence interval (**−**) with the bias ignored for the $\tau = .5$ quantile estimate are provided in (b) and (c).

n increases for all three values of τ, and that the local linear estimate outperforms the local constant estimate. This is another example that demonstrates that the local linear method is superior to the local constant even in the quantile setting. Moreover, the performance for the median quantile estimate is slightly better than that for two tails ($\tau = .05$ and .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 estimates of $a_1(\cdot)$ and $a_2(\cdot)$ (functions in the conditional mean), which, in contrast, avoids the need for a two-stage approach for estimating the variance function in the mean regression (see Fan and Yao 1998 for details). Nonetheless, it is interesting that because of the larger variation, the performance for $a_0(\cdot)$, although reasonably good, is not as good as that of $a_1(\cdot)$ and $a_2(\cdot)$. Further evidence of this is shown in Figure 1. The results of this simulated experiment demonstrate that the proposed procedure is reliable and follows along the lines 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 $\tau = .05$ (dashed line), .50 (dotted line), and .95 (dotted-dashed line), for $n = 500$ based on a typical sample chosen based on its MADE value equal to the median of the 500 MADE values. The selected optimal bandwidths are $h_{opt} = .10$ for $\tau = .05$, .075 for $\tau = .50$, and .10 for $\tau = .95$. Note that the estimate of $\sigma(\cdot)$ for $\tau = .50$ cannot be recovered from the estimate of $a_0(\cdot) = 0$, and it is not presented in Figure 1(a). The 95% pointwise confidence intervals without bias correction are represented in Figures 1(b) and (c) by thick lines for the $\tau = .50$ quantile estimate. Basically, all confidence intervals cover the true values. Similar plots were obtained for the local constant estimates (not shown due to space limitations). Overall, the proposed modeling procedure performed fairly well.

3.2 Real Examples

Example 2. We analyze a subset of the Boston housing price data set consisting of 14 variables collected on each of 506 different houses from a variety of locations. (This data set can be downloaded from *http://lib.stat.cmu.edu/datasets/boston*.) The dependent variable is Y, the median value of owner-occupied homes in \$1,000s (housing price). Some major factors possibly affecting the housing price used are U, the proportion of the 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 a complete description of all 14 variables, see Harrison and Rubinfeld 1978.) Recently, several articles have been devoted to the analysis of this data set; for example, Breiman and Friedman (1985), Chaudhuri et al. (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) used the additive quantile technique to analyze the data. Recently, Şentürk and Müller (2005) studied the correlation between the housing price, Y, and the crime rate, X_2, adjusted by the confounding variable U through a varying-coefficient model and concluded that the expected effect of increasing crime rate on declining housing prices seemed to be observed only for lower educational status neighborhoods in Boston. Some existing analyses (e.g., Breiman and Friedman 1985; Yu and Lu 2004) with both mean and quantile regressions concluded that most of the variation in housing prices seen 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 Y and X_1 are $-.7377$ and .6954. The scatterplots of Y versus U and versus X_1 are displayed in Figures 2(a) and (b). Interesting features of this

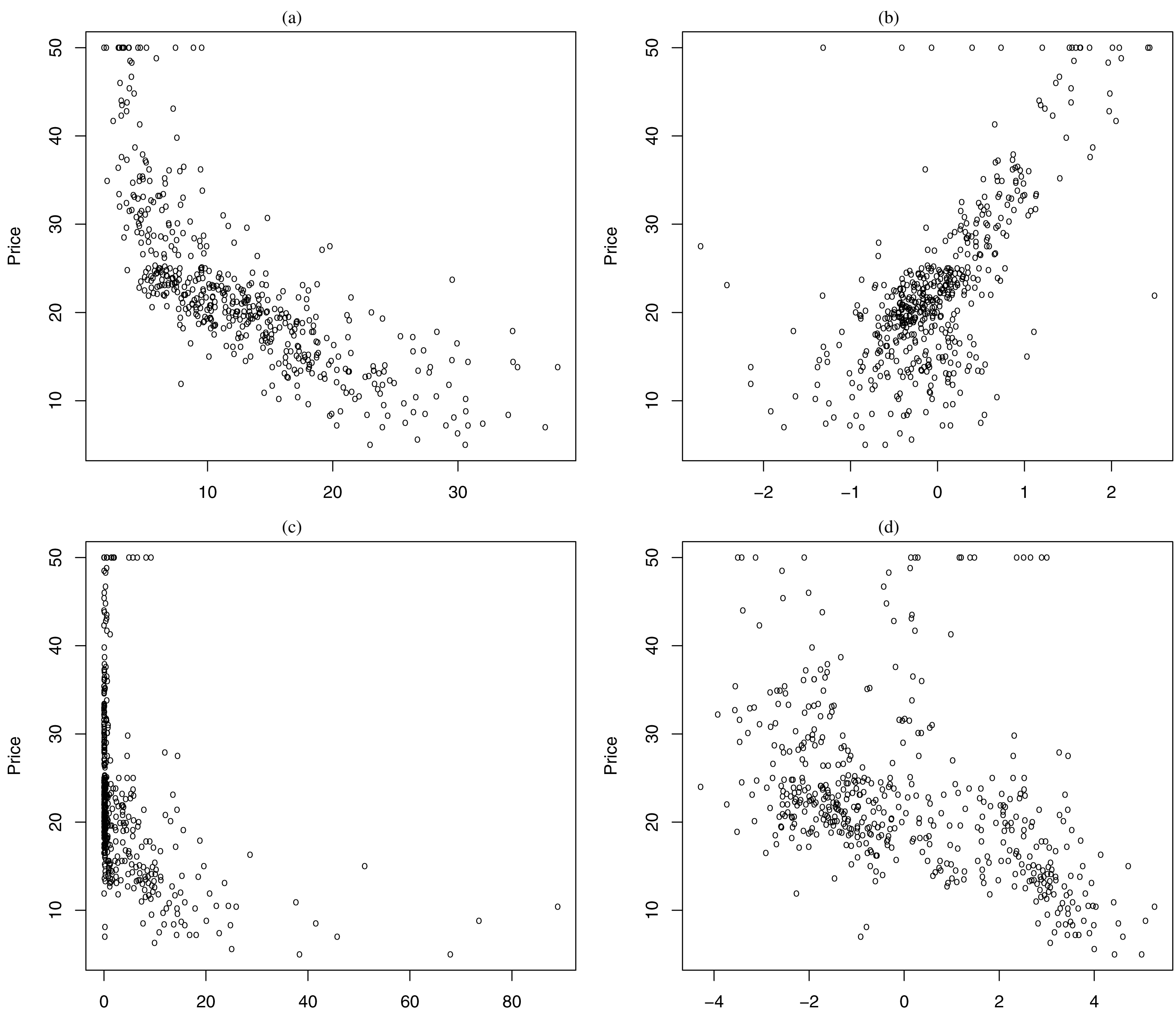

Figure 2. Boston housing price data. Scatterplots of house price versus the covariates U (a), X_1 (b), X_2 (c), and $\log(X_2)$ (d).

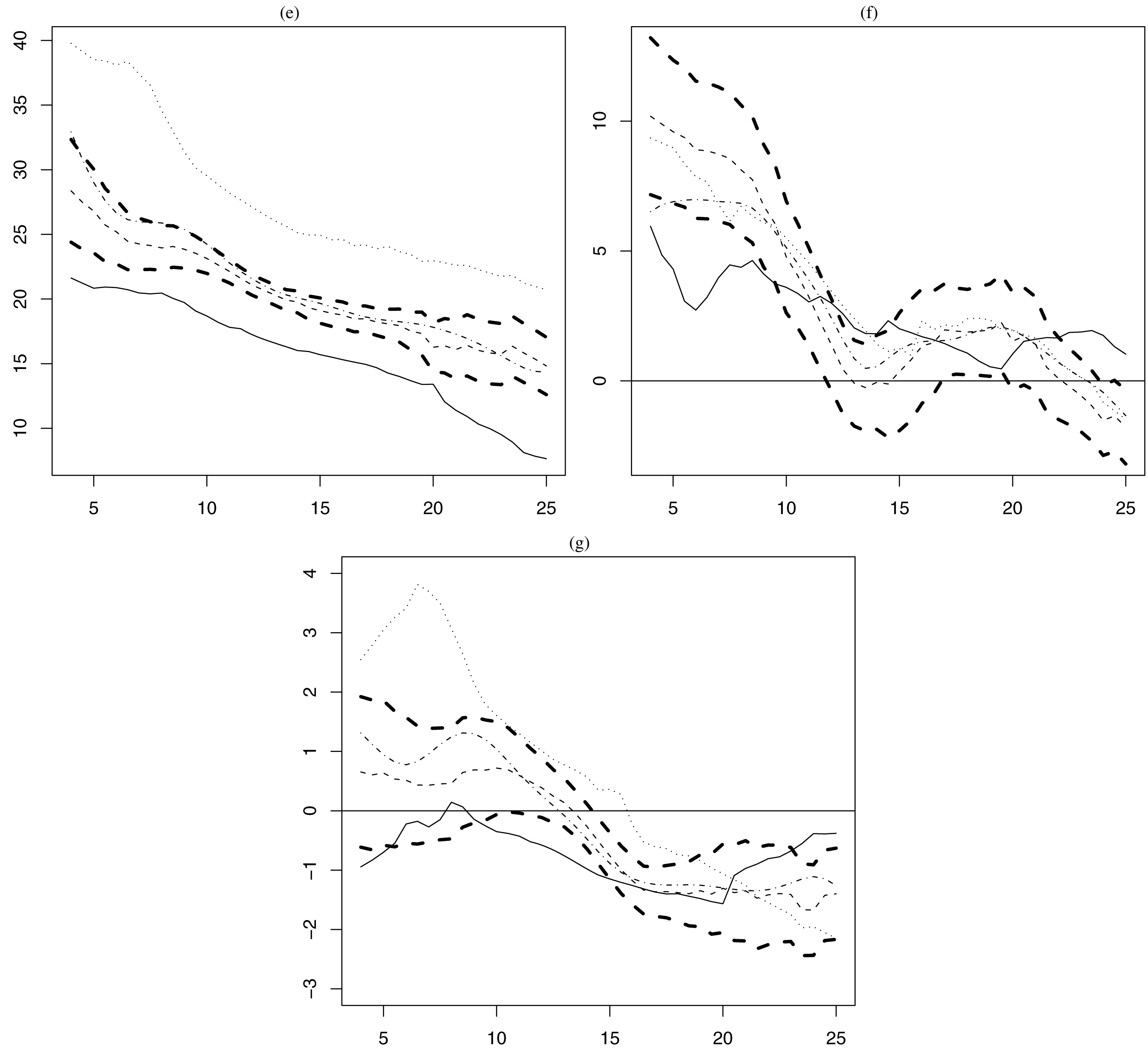

Figure 2. (Continued.) Plots of the estimated coefficient functions for three quantiles, $\tau = .05$ (—), $\tau = .50$ (- -), and $\tau = .95$ (⋯), and the mean regression (·-·): $\widehat{a}_{0,\tau}(u)$ and $\widehat{a}_0(u)$ versus u in (e), $\widehat{a}_{1,\tau}(u)$ and $\widehat{a}_1(u)$ versus u in (f), and $\widehat{a}_{2,\tau}(u)$ and $\widehat{a}_2(u)$ versus u in (g). The thick dashed lines indicate the 95% pointwise confidence interval for the median estimate with the bias ignored.

data set are that the response variable is the median price of a home in a given area and that the distributions of Y and the major covariate U are left-skewed. (The density estimates are not presented.) Finally, it is surprising that all of the nonparametric models mentioned earlier did not include the crime rate X_2, which may be an important factor affecting housing price, and did not consider the interaction terms, such as U and X_2.

Based on the foregoing discussion, we can conclude that the model studied in this work may be well suited to analyzing this data set. Therefore, we analyze this data set using the following quantile smooth coefficient model:

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

where $X_{t2}^* = \log(X_{t2})$. We obtain the quantile estimate of the coefficient functions $\{a_j(\cdot)\}$ using (3), denoted by $\{\widehat{a}_{j,\tau}(\cdot)\}$. We do not include the other variables such as X_3 and X_4 in model (8), because we found that the coefficient functions for these variables seem to be constant. Therefore, a semiparametric model would be appropriate if the model included these variables. But this is beyond the scope of this article and merits further investigation. 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}^* (correlation coefficient, $-.4543$) is slightly stronger than that between Y_t and X_{t2} ($-.3883$), which also can be seen in Figures 2(c) and (d). In the model fitting, covariates X_1 and X_2 are centralized. For comparison, we also consider the following functional coefficient model in the mean regression:

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

We use the local linear fitting technique to estimate the coefficient functions $\{a_j(\cdot)\}$, denoted by $\{\widehat{a}_j(\cdot)\}$ (see Cai et al. 2000 for details).

The coefficient functions are estimated through the local linear quantile approach using the bandwidth selector described in Section 2.2. As a result, the selected optimal bandwidths are $h_{opt} = 2.0$ for $\tau = .05$, 1.5 for $\tau = .50$, and 3.5 for $\tau = .95$. Figures 2(e), (f), and (g) present the estimated coefficient functions $\widehat{a}_{0,\tau}(\cdot)$, $\widehat{a}_{1,\tau}(\cdot)$, and $\widehat{a}_{2,\tau}(\cdot)$ for three quantiles, $\tau = .05$ (solid line), .50 (dashed line), and .95 (dotted line), together with the estimates $\{\widehat{a}_j(\cdot)\}$ from the mean regression model (dotted-dashed line). In addition, the 95% pointwise confidence intervals for the median estimate without bias correction are represented by thick dashed lines.

First, from these three figures, we can see that the median estimates are quite close to the mean estimates, and that the estimates based on the mean regression are always within the 95% confidence interval of the median estimates. We can conclude that the distribution of the measurement error e_t in (9) may be symmetric and that $\widehat{a}_{j,.5}(\cdot)$ in (8) is almost same as $\widehat{a}_j(\cdot)$ in (9). Also, from Figure 2(e), we can see that three quantile curves are parallel, implying that the intercept in $\widehat{a}_{0,\tau}(\cdot)$ depends on τ, and they decrease exponentially. More importantly, from Figures 2(f) and (g), we can see that heteroscedasticity might exist, due to the intersection of three quantile estimated coefficient curves.

From Figure 2(f), we can see that the expected effect of increasing the number of rooms can raise the housing price slightly in any low educational status neighborhood but much greater in relatively high educational status neighborhoods. Moreover, although the number of rooms has a positive effect on the median- and higher-priced houses in relatively high and low educational status neighborhoods, increasing the number of rooms might not increase the housing price in very low educational status neighborhoods. In other words, it is very difficult to sell high-priced houses with high numbers of rooms at a reasonable price in very low educational status neighborhoods.

From Figure 2(g), we can conclude that the positive correlation between the housing prices ($\tau = .50$ and .95) and the crime rate for relatively high educational status neighborhoods seems counterintuitive. However, the reason for this positive correlation is the existence of high educational status neighborhoods close to central Boston, where high housing prices and high crime rates occur simultaneously. Therefore, the expected effect of increasing crime rate on declining housing prices for $\tau = .50$ and .95 seems to occur only for lower educational status neighborhoods in Boston. Finally, it can be seen that the correlation between the housing prices for $\tau = .05$ and the crime rate is almost negative, although the degree depends on the value of U. This implies that an increasing crime rate slightly decreases the prices of the cheap houses ($\tau = .05$).

In summary, this example demonstrates that the factors U, X_1, and X_2 have different effects on the different quantiles of the conditional distribution of housing prices. Overall, housing price and the proportion of population of lower educational status have a strong negative correlation, the number of rooms has a mostly positive effect on housing prices, whereas the crime rate has the most negative effect on housing prices. 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 in terms of the U.S. 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 set 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 1,566 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 article, where ξ_t is an exchange rate level on the tth 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), and Figure 3(b) shows that there is almost no significant autocorrelation in $\{Y_t\}$, which was confirmed by Hong and Lee (2003) 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

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

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, then $a_j(\cdot)$ can be estimated by a local linear fitting denoted by $\widehat{a}_j(\cdot)$ (see Cai et al. 2000 for details). Here σ_t is the stochastic volatility that 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, or may be chosen using data-driven methods, as done by Fan et al. (2003), if no prior information is available. By following the analysis of Fan et al. (2003) and Hong and Lee (2003), the smooth variable U_t is chosen as an moving average technical trading rule in finance, defined as $U_t = \xi_{t-1}/M_t - 1$, where $M_t = \sum_{j=1}^{L}\xi_{t-j}/L$, the moving average and considered a proxy for the trend at the time $t-1$. Following Hong and Lee (2003), we choose $L = 26$ (a half year). The time series plot of $\{U_t\}$ is given in Figure 3(c).

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

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

(We also considered the models with more than two lagged variables and found that the conclusions were similar, and thus we do not report them here.) The first 1,540 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

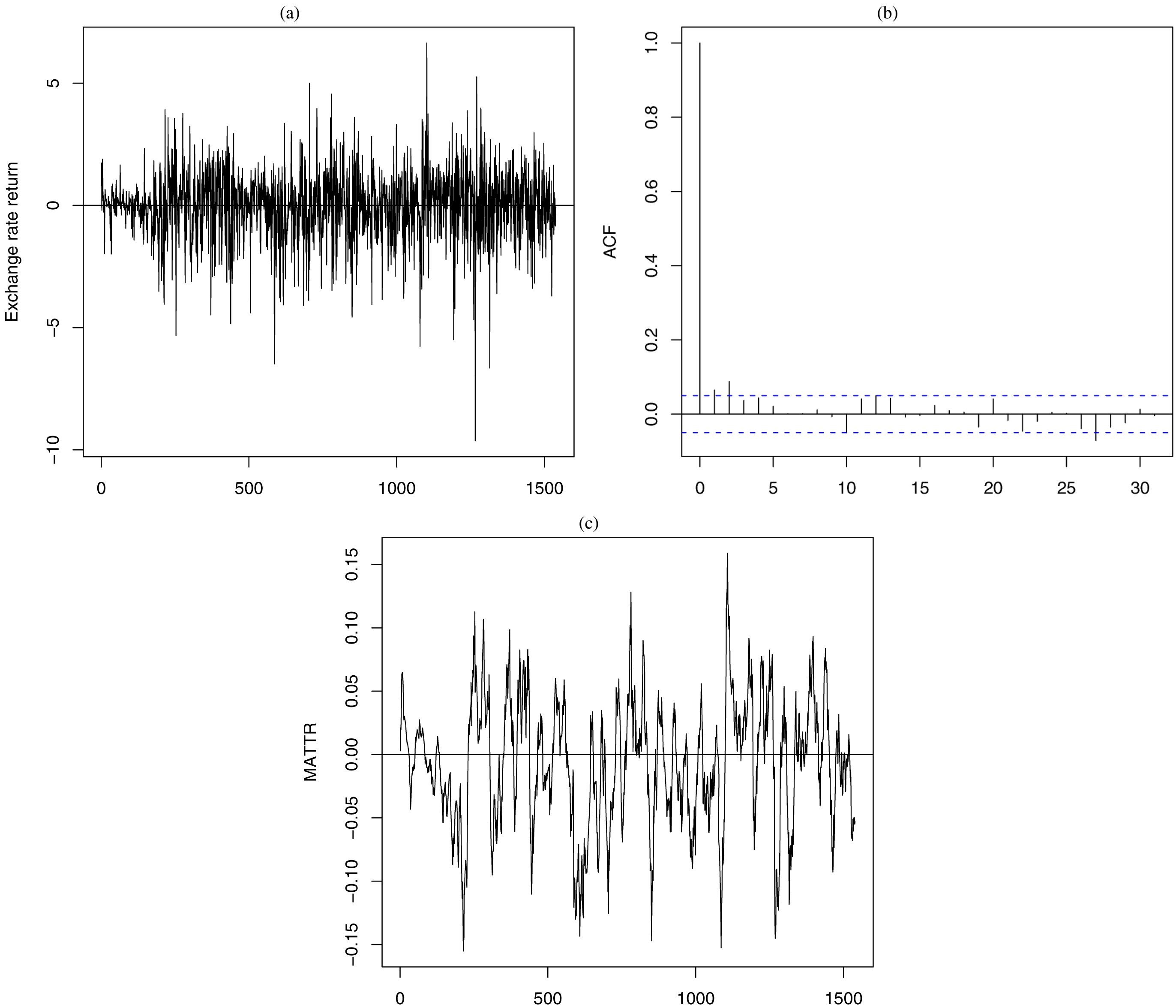

Figure 3. Exchange rate series: (a) Japanese Yen–U.S. dollar exchange rate return series $\{Y_t\}$; (b) autocorrelation function of $\{Y_t\}$; and (c) moving average trading technique rule.

$\{\widehat{a}_{j,\tau}(\cdot)\}$. Optimal bandwidths are $h_{opt} = .03$ for $\tau = .05$, .025 for $\tau = .50$, and .03 for $\tau = .95$. Figures 3(d)–(g) depict the estimated coefficient functions $\widehat{a}_{0,\tau}(\cdot)$, $\widehat{a}_{1,\tau}(\cdot)$, and $\widehat{a}_{2,\tau}(\cdot)$ for three quantiles $\tau = .05$ (solid line), .50 (dashed line), and .95 (dotted line), together with the estimates $\{\widehat{a}_j(\cdot)\}$ (dotted-dashed line) from the mean regression model in (10). The 95% pointwise confidence intervals for the median estimate are represented by the thick dashed lines without bias correction.

First, from Figures 3(d), (f), and (g), we clearly see that the median estimates $\widehat{a}_{j,.50}(\cdot)$ in (11) are almost parallel with or close to the mean estimates $\widehat{a}_j(\cdot)$ in (10) and that the mean estimates are almost within the 95% confidence interval of the median estimates. Second, $\widehat{a}_{0,.50}(\cdot)$ in Figure 3(d) shows a nonlinear pattern, and $\widehat{a}_{0,.05}(\cdot)$ and $\widehat{a}_{0,.95}(\cdot)$ in Figure 3(e) exhibit slightly U-shaped and symmetrically. More importantly, Figures 3(f) and (g) show that the lower and upper quantile estimated coefficient curves intersect and behave slightly differently. Particularly, from Figure 3(g), we may conclude that the distribution of the measurement error e_t in (10) might not be symmetric about 0, and that there exists a nonlinearity in $a_{j,\tau}(\cdot)$. This implies that a nonlinearity exists. We also note that the quantile has a complex structure and that heteroscedasticity exists. We conclude that the GARCH effects occur in the exchange rate time series (see Engle, Ito, and Lin 1990).

Finally, we consider the postsample forecasting for the last 25 observations based on the local linear quantile estimators computed using the same bandwidths as those used in the model fitting. The 95% nonparametric prediction interval is constructed as $(\widehat{q}_{.025}(\cdot), \widehat{q}_{.975}(\cdot))$. The prediction results are reported in Table 2, which shows that 24 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.

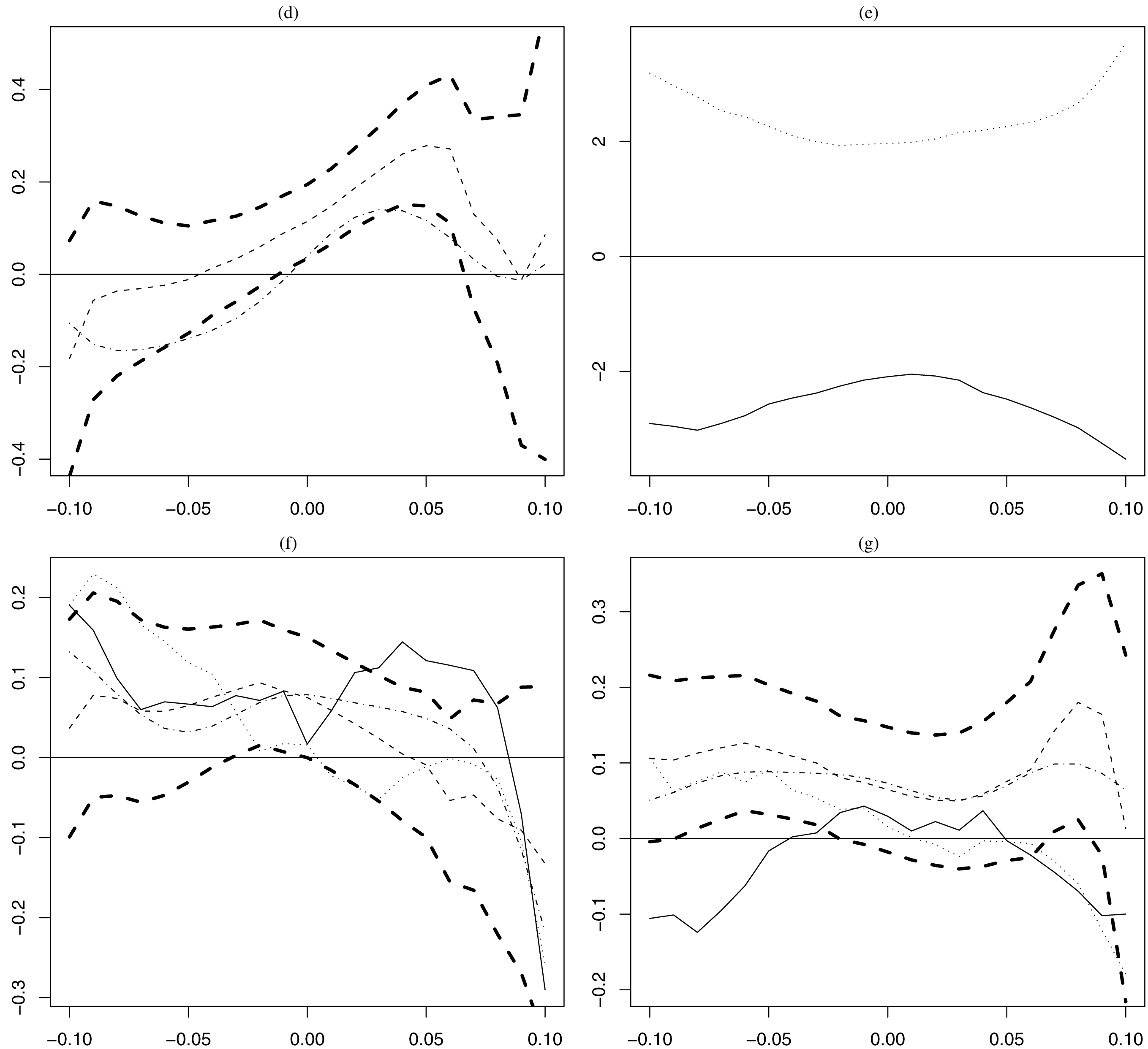

Figure 3. (Continued.) Plots of the estimated coefficient functions for three quantiles, $\tau = .05$ (—), $\tau = .50$ (- -), and $\tau = .95$ (···), and the mean regression (·-·): $\widehat{a}_{0,.50}(u)$ and $\widehat{a}_0(u)$ versus u in (d), $\widehat{a}_{0,.05}(u)$ and $\widehat{a}_{0,.95}(u)$ versus u in (e), $\widehat{a}_{1,\tau}(u)$ and $\widehat{a}_1(u)$ versus u in (f), and $\widehat{a}_{2,\tau}(u)$ and $\widehat{a}_2(u)$ versus u in (g). The thick dashed lines indicate the 95% pointwise confidence interval for the median estimate with the bias ignored.

4. ASYMPTOTIC RESULTS

The asymptotic results presented here were derived under the α-mixing assumption. (See Cai 2002 for the definition of α-mixing.) 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, including functional coefficient autoregressive processes with or without exogenous variables, ARCH- and GARCH-type processes, stochastic volatility models, and nonlinear additive autoregressive models with or without exogenous variables, are strong mixing under some mild conditions (see Cai 2002 for details).

We now give some regularity conditions that are sufficient for the consistency and asymptotic normality of the proposed estimators, although they might not be the weakest ones possible.

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) $\{(U_t, \mathbf{X}_t, Y_t)\}$ is a strictly α-mixing stationary process with mixing coefficient $\alpha(t)$ satisfying $\sum_{t\geq 1}^{\infty} t^l \times \alpha^{(\delta-2)/\delta}(t) < \infty$ for some positive real number $\delta > 2$ and $l > (\delta - 2)/\delta$.

Table 2. The postsample predictive intervals for exchange rate data

Observation number	True value	Prediction interval
1,541	.392	(−2.891, 2.412)
1,542	.509	(−3.099, 2.405)
1,543	1.549	(−2.943, 2.446)
1,544	−.121	(−2.684, 2.525)
1,545	−.991	(−2.677, 2.530)
1,546	−.646	(−3.110, 2.401)
1,547	−.354	(−3.178, 2.365)
1,548	−1.393	(−3.083, 2.372)
1,549	.997	(−3.110, 2.230)
1,550	−.916	(−3.033, 2.431)
1,551	−3.707	(−3.021, 2.286)
1,552	−.919	(−3.841, 2.094)
1,553	−.901	(−3.603, 2.770)
1,554	.071	(−3.583, 2.821)
1,555	−.497	(−3.351, 2.899)
1,556	−.648	(−3.436, 2.783)
1,557	1.648	(−3.524, 2.866)
1,558	−1.184	(−3.121, 2.810)
1,559	.530	(−3.529, 2.531)
1,560	.107	(−3.222, 2.648)
1,561	−.804	(−3.294, 2.651)
1,562	.274	(−3.419, 2.534)
1,563	−.847	(−3.242, 2.640)
1,564	−.060	(−3.426, 2.532)
1,565	−.088	(−3.300, 2.576)

(C6) $E\|\mathbf{X}_t\|^{2\delta^*} < \infty$ with $\delta^* > \delta$.
(C7) $\boldsymbol{\Omega}(u_0)$ is positive definite and continuous in a neighborhood of u_0.
(C8) $\boldsymbol{\Omega}^*(u_0)$ is continuous and positive definite in a neighborhood of u_0.
(C9) The bandwidth h satisfies $h \to 0$ and $nh \to \infty$.
(C10) $f(u, v|\mathbf{x}_0, \mathbf{x}_s; s) \le M < \infty$ for $s \ge 1$, 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 of the foregoing assumptions has been given by Cai (2002). Assumption (C6) is commonly required to ensure 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 $\mathbf{a}(u_0)$ is identified (uniquely determined) if and only if $\boldsymbol{\Omega}(u_0)$ 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:

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

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

Because the case where q is even leads to a more complicated derivation, we consider only the case where q is odd. For the case where q is even, we can obtain a similar result (see Fan and Gijbels 1996 for details). From Theorem 1, the asymptotic mean squared error (AMSE) of $\widehat{\mathbf{a}}(u_0)$ is

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

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

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

and the optimal AMSE is $\mathrm{AMSE}_{opt} = O(n^{-(2q+2)/(2q+3)})$. Further, note that results similar to 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, e.g., Cai et al. 2000). For various practical applications, Fan and Gijbels (1996) recommended using the local linear fit ($q = 1$). Therefore, for ease of notation, in what follows we consider only the case where $q = 1$ (local linear fitting).

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

Now we consider the comparison of the performance of the local linear estimation $\widehat{\mathbf{a}}(u_0)$ obtained in (3) with that of the local constant estimation $\widetilde{\mathbf{a}}(u_0)$ given in (4). First, we derive the asymptotic results for the local constant estimator, but omit the proof. Under some regularity conditions, it can be shown that

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

where

$$\widetilde{\mathbf{b}} = \frac{1}{2}h^2\mu_2\big[\mathbf{a}^{(2)}(u_0) + 2\mathbf{a}^{(1)}(u_0)f_u^{(1)}(u_0)/f_u(u_0) + 2\{\boldsymbol{\Omega}^*(u_0)\}^{-1}\boldsymbol{\Omega}^{*(1)}(u_0)\mathbf{a}^{(1)}(u_0)\big].$$

This implies that the asymptotic bias for $\widetilde{\mathbf{a}}(u_0)$ is different than that for $\widehat{\mathbf{a}}(u_0)$, but both $\widetilde{\mathbf{a}}(u_0)$ and $\widehat{\mathbf{a}}(u_0)$ have the same asymptotic variance. Therefore, the local constant quantile estimator does not adapt to nonuniform designs; the bias can be large when $f_u^{(1)}(u_0)/f_u(u_0)$ or $\{\boldsymbol{\Omega}^*(u_0)\}^{-1}\boldsymbol{\Omega}^{*(1)}(u_0)$ 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 the 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, but omit its proof. 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:

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

where $b_c = [\mu_{2,c}^2 - \mu_{1,c}\mu_{3,c}]/[\mu_{2,c}\mu_{0,c} - \mu_{1,c}^2]$ and $v_c = [\mu_{2,c}^2\nu_{0,c} - 2\mu_{1,c}\mu_{2,c}\nu_{1,c} + \mu_{1,c}^2\nu_{2,c}][\mu_{2,c}\mu_{0,c} - \mu_{1,c}^2]^{-2}$. Furthermore, we have the following asymptotic normality of the local constant quantile estimator at the left boundary point, $u_0 = ch$, for $0 < c < 1$:

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

where $\widetilde{\mathbf{b}}_c = [h\mu_{1,c}\mathbf{a}^{(1)}(0) + h^2\mu_{2,c}/2\{\mathbf{a}^{(2)}(0) + 2\mathbf{a}^{(1)}(0) \times f_u^{(1)}(0)/f_u(0) + 2\boldsymbol{\Omega}^{*-1}(0)\boldsymbol{\Omega}^{*(1)}(0)\mathbf{a}^{(1)}(0)\}]/\mu_{0,c}$. Similar results hold for the right boundary point, $u_0 = 1 - ch$.

We note that if the point 0 were an interior point, then Theorem 2 would hold with $c = 1$, which becomes Theorem 1. Moreover, it is easy to see that as $c \to 1$, $b_c \to \mu_2$ and $v_c \to \nu_0$, and these limits are exactly the constant factors appearing in the asymptotic bias and variance for an interior point. Therefore, Theorem 2 shows that the local linear estimation has automatic good behavior at boundaries without the need for boundary correction. Theorem 2 also shows that at the boundaries, the asymptotic bias term for the local constant quantile estimate is of order h by comparing it with order h^2 for the local linear quantile estimate. This demonstrates that the local linear quantile estimate does not suffer from boundary effects, but the local constant quantile estimate does.

5. CONCLUSION

Here we have studied a class of quantile regression models with functional coefficients for time series data. We have suggested using the local polynomial fitting scheme to estimate the nonparametric coefficient functions and derived the asymptotic properties of the proposed estimators. We proposed an ad hoc method for selecting the bandwidth and estimating the asymptotic covariance. We conducted a Monte Carlo simulation experiment to illustrate the proposed the methodology and analyzed two real data sets. We presented some new findings related to these two real examples based on the dynamic smooth coefficient quantile regression model. Some interesting future research topics related to this work should be mentioned. First, it would be very useful to discuss the bandwidth theoretically and empirically. Second, an important application of quantile regression is to measure how much the τth response quantile changes as one covariate is perturbed while the other covariates are held fixed (see Chaudhuri et al. 1997). Therefore, we can estimate $\nabla q_\tau(U, \mathbf{x})$ using the proposed methodology. Furthermore, the foregoing models and results can be extended to the cases where some $\mathbf{X}_t$'s might be nonstationary, such as I(1), and some $\mathbf{X}_t$'s might be endogenous. Finally, nonparametric quantile regression potentially can be applied to the analysis of financial data, such as GARCH-type models studied by Xiao (2006) and VaR and other types of risk models and their extensions (see Bassett, Koenker, and Kordas 2004).

APPENDIX: PROOFS OF THEOREMS

In this section, due to space limitations, we give only brief derivations of the main results based on some lemmas. For expositional purposes, we consider only the case where $q = 1$. First, we need the following two lemmas, the proofs of which were given by Koenker and Zhao (1996) and Ruppert and Carroll (1980).

Lemma A.1. Let $\mathbf{V}_n(\boldsymbol{\Delta})$ be a vector function that satisfies (a) $-\boldsymbol{\Delta}'\mathbf{V}_n(\lambda\boldsymbol{\Delta}) \ge -\boldsymbol{\Delta}'\mathbf{V}_n(\boldsymbol{\Delta})$ for $\lambda \ge 1$ and (b) $\sup_{\|\boldsymbol{\Delta}\| \le M}\|\mathbf{V}_n(\boldsymbol{\Delta}) + \mathbf{D}\boldsymbol{\Delta} - \mathbf{A}_n\| = o_p(1)$, where $\|\mathbf{A}_n\| = O_p(1)$, $0 < M < \infty$, and $\mathbf{D}$ is a positive-definite matrix. Suppose that $\boldsymbol{\Delta}_n$ is a vector such that $\|\mathbf{V}_n(\boldsymbol{\Delta}_n)\| = o_p(1)$. Then we have

$$\|\boldsymbol{\Delta}_n\| = O_p(1)$$

and

$$\boldsymbol{\Delta}_n = \mathbf{D}^{-1}\mathbf{A}_n + o_p(1).$$

Lemma A.2. Let $\widehat{\boldsymbol{\beta}}$ be the minimizer of the function $\sum_{t=1}^n w_t \times \rho_\tau(y_t - \mathbf{X}_t'\boldsymbol{\beta})$, where $w_t > 0$ and $\rho_\tau(\cdot)$ is the check function defined in Section 1. Then $\|\sum_{t=1}^n w_t\mathbf{X}_t\psi_\tau(y_t - \mathbf{X}_t'\widehat{\boldsymbol{\beta}})\| \le \dim(\mathbf{X})\max_{t\le n}\|w_t \times \mathbf{X}_t\|$, where $\psi_\tau(\cdot)$ is as defined in Section 2.2.

By the definition of $\boldsymbol{\theta}$ defined in Section 2.2,

$$\boldsymbol{\beta} = \begin{pmatrix}\mathbf{a}(u_0)\\ \mathbf{a}^{(1)}(u_0)\end{pmatrix} + a_n\mathbf{H}^{-1}\boldsymbol{\theta},$$

where a_n is as defined in Section 2.2. 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,

$$\widehat{\boldsymbol{\theta}} = \arg\min\sum_{t=1}^n \rho_\tau(Y_t^* - a_n\boldsymbol{\theta}'\mathbf{X}_t^*)K(U_{th}) \equiv \arg\min G(\boldsymbol{\theta}).$$

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

Lemma A.3. Under the assumptions of Theorem 1, we have

$$\|\mathbf{V}_n(\boldsymbol{\theta}) - \mathbf{V}_n(\mathbf{0}) - E[\mathbf{V}_n(\boldsymbol{\theta}) - \mathbf{V}_n(\mathbf{0})]\| = o_p(1)$$

and

$$\|E[\mathbf{V}_n(\boldsymbol{\theta}) - \mathbf{V}_n(\mathbf{0})] + \mathbf{D}\boldsymbol{\theta}\| = o(1)$$

uniformly over $\boldsymbol{\theta} \in \mathbf{A}_m$, where $\mathbf{D} = f_u(u_0)\boldsymbol{\Omega}_1^*(u_0)$.

Lemma A.4. Under the assumptions of Theorem 1, we have

$$E[\mathbf{Z}_1] = \frac{h^3 f(u_0)}{2}\begin{pmatrix}\mu_2\boldsymbol{\Omega}^*(u_0)\mathbf{a}^{(2)}(u_0)\\ \mathbf{0}\end{pmatrix} + o(h^3)$$

and

$$\mathrm{var}[\mathbf{Z}_1] = h\tau(1-\tau)f(u_0)\boldsymbol{\Omega}_1(u_0) + o(h),$$

where $\boldsymbol{\Omega}_1(u_0) = \mathrm{diag}\{\nu_0, \nu_2\} \otimes \boldsymbol{\Omega}(u_0)$. Furthermore, $\mathrm{var}[\mathbf{V}_n(\mathbf{0})] = \tau(1-\tau)f(u_0)\boldsymbol{\Omega}_1(u_0) + o(1)$; therefore, $\|\mathbf{V}_n(\mathbf{0})\| = O_p(1)$.

Proof of Theorem 1

Here we present only a sketch of the proof; details are available on request. By Lemmas A.3 and A.4, $\mathbf{V}_n(\boldsymbol{\theta})$ satisfies condition (b) of Lemma A.1; that is, $\|\mathbf{A}_n\| = O_p(1)$ and $\sup_{\boldsymbol{\theta}\in\mathbf{A}_m} \|\mathbf{V}_n(\boldsymbol{\theta}) + \mathbf{D}\boldsymbol{\theta} - \mathbf{A}_n\| = o_p(1)$, where $\mathbf{A}_n = \mathbf{V}_n(\mathbf{0})$. It follows from Lemma A.2 that $\|\mathbf{V}_n(\widehat{\boldsymbol{\theta}})\| = o_p(1)$, where $\widehat{\boldsymbol{\theta}}$ is the minimizer of $G(\boldsymbol{\theta})$. Finally, because $\psi_\tau(x)$ is an increasing function of x; then $-\boldsymbol{\theta}'\mathbf{V}_n(\lambda\boldsymbol{\theta}) = a_n \sum_{t=1}^n \psi_\tau[Y_t^* + \lambda a_n(-\boldsymbol{\theta}'\mathbf{X}_t^*)](-\boldsymbol{\theta}'\mathbf{X}_t^*)K(U_{th})$ is an increasing function of λ. Thus the condition (a) of Lemma A.1 is satisfied. Therefore,

$$\widehat{\boldsymbol{\theta}} = \mathbf{D}^{-1}\mathbf{A}_n + o_p(1) = \frac{(\boldsymbol{\Omega}_1^*)^{-1}}{\sqrt{nh}\, f_u(u_0)} \sum_{t=1}^n \psi_\tau(Y_t^*)\mathbf{X}_t^* K(U_{th}) + o_p(1). \tag{A.1}$$

Let $\varepsilon_t = \psi_\tau(Y_t - \mathbf{X}_t'\mathbf{a}(U_t))$. Clearly, $E(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) = \tau(1-\tau)$. From (A.1),

$$\widehat{\boldsymbol{\theta}} = \frac{(\boldsymbol{\Omega}_1^*)^{-1}}{\sqrt{nh}\, f_u(u_0)} \sum_{t=1}^n [\psi_\tau(Y_t^*) - \varepsilon_t]\mathbf{X}_t^* K(U_{th}) + \frac{(\boldsymbol{\Omega}_1^*)^{-1}}{\sqrt{nh}\, f_u(u_0)} \sum_{t=1}^n \varepsilon_t \mathbf{X}_t^* K(U_{th}) + o_p(1) \equiv \mathbf{B}_n + \boldsymbol{\xi}_n + o_p(1).$$

Similar to the proof of theorem 2 of Cai et al. (2000), using the small-block and large-block technique and the Cramér–Wold device, we can show (although lengthily and tediously) that

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

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

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

Because $\psi_\tau(Y_t^*) - \varepsilon_t = I(Y_t \le \mathbf{X}_t'\mathbf{a}(U_t)) - I(Y_t \le \mathbf{X}_t'(\mathbf{a}(u_0) + \mathbf{a}^{(1)}(u_0)(U_t - u_0)))$, we have $[\psi_\tau(Y_t^*) - \varepsilon_t]^2 = I(d_{1t} < Y_t \le d_{2t})$, 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 $c_{2t} = \mathbf{X}_t'[\mathbf{a}(u_0) + \mathbf{a}^{(1)}(u_0)(U_t - u_0)]$. Furthermore,

$$E\big[\{\psi_\tau(Y_t^*) - \varepsilon_t\}^2 K^2(U_{th})\mathbf{X}_t^*\mathbf{X}_t^{*\prime}\big] = E\big[\{F_{y|u,x}(d_{2t}) - F_{y|u,x}(d_{1t})\}K^2(U_{th})\mathbf{X}_t^*\mathbf{X}_t^{*\prime}\big] = O(h^3).$$

Thus $\mathrm{var}(\mathbf{B}_n) = o(1)$. This, in conjunction with (A.2) and (A.3) and Slutsky's theorem, proves the theorem.

Proof of (6) and (7)

By Taylor's expansion, we have

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

Therefore, $E[\mathbf{S}_n] = h^{-1}E[f_{y|u,x}(\mathbf{X}_t'\mathbf{a}(u_0))\mathbf{X}_t^*\mathbf{X}_t^{*\prime}K(U_{th})] + o(1) \to f_u(u_0)\boldsymbol{\Omega}_1^*(u_0)$. Similar to the proof of $\mathrm{var}[\mathbf{V}_n(\mathbf{0})]$ in Lemma A.4, we can show that $\mathrm{var}(\mathbf{S}_n) \to 0$. Therefore, $\mathbf{S}_n \to f_u(u_0)\boldsymbol{\Omega}_1^*(u_0)$ in probability. This proves (6). Clearly,

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

Similarly, one can show that $\mathrm{var}(\widehat{\boldsymbol{\Omega}}_{n,0}) \to 0$. This proves the first part of (7). By the same token, we can easily show that $E[\widehat{\boldsymbol{\Omega}}_{n,1}] \to f_u(u_0)\boldsymbol{\Omega}^*(u_0)$ and $\mathrm{var}(\widehat{\boldsymbol{\Omega}}_{n,1}) \to 0$. Thus $\widehat{\boldsymbol{\Omega}}_{n,1} \to f_u(u_0)\boldsymbol{\Omega}^*(u_0)$. We prove (7).

[Received January 2005. Revised August 2008.]

REFERENCES

Bassett, G. W., Koenker, R., and Kordas, G. (2004), "Pessimistic Portfolio Allocation and Choquet Expected Utility," *Journal of 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. (2002), "Regression Quantile for Time Series," *Econometric Theory*, 18, 169–192.

Cai, Z., and Tiwari, R. C. (2000), "Application of a Local Linear Autoregressive Model to BOD Time Series," *Environmetrics*, 11, 341–350.

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.

Chaudhuri, P., Doksum, K., and Samarov, A. (1997), "On Average Derivative Quantile Regression," *The Annals of Statistics*, 25, 715–744.

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., and Manganelli, S. (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantile," *Journal of Business and Economics Statistics*, 22, 367–381.

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.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman & Hall.

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*, Ser. 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.

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., and Portnoy, S. (2000), "Some Asymptotic Results on Bivariate Quantile Splines," *Journal of Statistical Planning and Inference*, 91, 341–349.

He, X., Ng, P., and Portnoy, S. (1998), "Bivariate Quantile Smoothing Splines," *Journal of the Royal Statistical Society*, Ser. B, 60, 537–550.

Honda, T. (2000), "Nonparametric Estimation of a Conditional Quantile for α-Mixing Processes," *Annals of the Institute of Statistical Mathematics*, 52, 459–470.

——— (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*, Ser. B, 60, 271–293.

Khindanova, I. N., and Rachev, S. T. (2000), "Value at Risk: Recent Advances," in *Handbook on Analytic-Computational Methods in Applied Mathematics*, ed. G. A. Anastassiou, Boca Raton, FL: CRC Press, pp. 801–858.

Kim, M. (2007), "Quantile Regression With Varying Coefficients," *The Annals of Statistics*, 35, 92–108.

Koenker, R. (1994), "Confidence Intervals for Regression Quantiles," in *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics*, eds. P. Mandl and M. Huskova, Heidelberg: Physica, pp. 349–359.

Koenker, R., and Bassett, G. W. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

——— (1982), "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43–61.

Koenker, R., and Xiao, Z. (2002), "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

——— (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.

Koenker, R., Ng, P., and Portnoy, S. (1994), "Quantile Smoothing Splines," *Biometrika*, 81, 673–680.

Machado, J. A. F. (1993), "Robust Model Selection and M-Estimation," *Econometric Theory*, 9, 478–493.

Opsomer, J. D., and Ruppert, D. (1998), "A Fully Automated Bandwidth Selection for Additive Regression Models," *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.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Şentürk, D., and Müller, H. G. (2005), "Covariate-Adjusted Correlation Analysis via Varying-Coefficient Models," *Scandinavian Journal of Statistics*, 32, 365–383.

Wei, Y., and He, X. (2006), "Conditional Growth Charts" (with discussion), *The Annals of Statistics*, 34, 2069–2097.

Wei, Y., Pere, A., Koenker, R., and He, X. (2006), "Quantile Regression Methods for Reference Growth Charts," *Statistics in Medicine*, 25, 1369–1382.

Xiao, Z. (2006), "Conditional Quantile Estimation for GARCH Models," working paper, Boston College, Dept. of Economics.

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.

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.