Forthcoming in Journal of Econometrics

Semiparametric Quantile Regression Estimation in Dynamic Models with Partially Varying Coefficients*

Zongwu Cai[a,b] and Zhijie Xiao[c]

[a]Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA, E-mail: zcai@uncc.edu.

[b]Wang Yanan Institute for Studies in Economics, MOE Key Laboratory of Econometrics, and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China.

[c]Department of Economics, Boston College, Chestnut Hill, MA 02467, USA, E-mail: xiaoz@bc.edu.

ABSTRACT

We study quantile regression estimation for dynamic models with partially varying coefficients so that the values of some coefficients may be functions of informative covariates. Estimation of both parametric and nonparametric functional coefficients are proposed. In particular, we propose a three stage semiparametric procedure. Both consistency and asymptotic normality of the proposed estimators are derived. We demonstrate that the parametric estimators are root-n consistent and the estimation of the functional coefficients is oracle. In addition, efficiency of parameter estimation is discussed and a simple efficient estimator is proposed. A simple and easily implemented test for the hypothesis of varying-coefficient is proposed. A Monte Carlo experiment is conducted to evaluate the performance of the proposed estimators.

KEY WORDS: Efficiency; nonlinear time series; partially linear; partially varying coefficients; quantile regression; semiparametric.

*We thank two referees and the guest Editors for their very helpful comments and suggestions. We also thank the audiences at the Symposium on Econometric Theory and Applications (SETA2008), May 2008 in Seoul National University, South Korea, for their helpful comments. Finally, we thank Xiaoping Xu for his assistance for the early version of the paper. Cai's research was supported, in part, by the National Science Foundation of China grants #70871003 and #70971113, and funds provided by the University of North Carolina at Charlotte, the Cheung Kong Scholarship from Chinese Ministry of Education, the Minjiang Scholarship from Fujian Province, China and Xiamen University. Xiao thanks research support from Boston College.

1 Introduction

The quantile regression method, first introduced by Koenker and Bassett (1978), has been widely used in various disciplines, including finance, economics, medicine, and biology. For example, estimation of conditional quantiles is a common practice in risk management operations and many other financial applications. The literature on estimating conditional quantiles is large. Much of the study on quantile regression is based on linear parametric quantile regression models. More recently, nonparametric and semiparametric quantile regression models have attracted a great deal of research attention due to their greater flexibility than tightly specified parametric models. See, for example, Chaudhuri (1991), He and Shi (1996), Chaudhuri, Doksum and Samarov (1997), He, Ng and Portnoy (1998), Yu and Jones (1998), Koenker, Ng and Portnoy (1998), He and Ng (1999), He and Liang (2000), He and Portnoy (2000), Honda (2000, 2004), Khindanova and Rachev (2000), Cai (2002a), De Gooijer and Gannoun (2003), Kim (2007), Lee (2003), Yu and Lu (2004), Horowitz and Lee (2005), Cai and Xu (2008), Cai, Gu and Li (2009) and references therein for recent statistics and econometrics literature.

Let $\{Y_t, V_t\}_{t=-\infty}^{\infty}$ be a stationary sequence and $F(y\,|v)$ denote the conditional distribution of Y_t given $V_t = v$, where V_t is a vector of covariates, including possibly exogenous variables and lagged variables. The conditional quantile function of Y_t given $V_t = v$, $Q_{Y_t}(\tau|V_t = v)$, is defined as, for any $0 < \tau < 1$,

$$q_\tau(v) \equiv Q_{Y_t}(\tau|V_t = v) = \inf\{y \in \Re :\ F(y\,|v) \geq \tau\}.$$

Equivalently, $q_\tau(v)$ can be expressed as

$$q_\tau(v) = \operatorname{argmin}_{a\in\Re} E\left\{\rho_\tau(Y_t - a)\,|\,V_t = v\right\}, \tag{1}$$

where $\rho_\tau(y) = y\,[\tau - I\{y < 0\}]$ (with $y \in \Re$) is called the "check" (loss) function and $I\{A\}$ is the indicator function of any set A.

Given observed data $\{Y_t, V_t\}_{t=-\infty}^{\infty}$, our interest is to estimate $q_\tau(v)$. If we assume that $q_\tau(v) = \beta_\tau^T\, v$, where A^T denotes the transpose of a matrix or vector A, we obtain a linear quantile regression model as in Koenker and Bassett (1978). In some practical applications, a linear quantile regression model might not be flexible enough to capture the underlying complex dependence structure. However, a purely nonparametric quantile regression model may suffer from the so-called "curse of dimensionality" problem, the practical implementation might not be easy, and the visual display may not be useful for the exploratory purposes. To deal with the aforementioned problems, some dimension reduction modelling methods have been proposed in the literature. For example, He, Ng and Portnoy (1998), He and Ng (1999), He and Portnoy (2000), De Gooijer and Zerom (2003), Yu and Lu (2004), and Horowitz and Lee (2005) considered the additive quantile regression models for iid data, while Honda (2004) and Cai and Xu (2008) investigated the varying coefficient quantile regression models for time series processes. He and Shi (1996), He and Liang (2000), and Lee (2003) discussed the partially linear quantile regression models for iid samples.

In this paper, we consider another dimension-reduction modelling method - partially varying coefficient models. This approach allows appreciable flexibility on the structure of fitted models. The proposed model allows for linearity in coefficients in some variables and nonlinearity in other variables. In such a way, the model has an ability of capturing the individual variations and of easing the so-called “curse of dimensionality”.

By assuming that $V_t = (U_t^T,\, X_t^T)^T$, a partially varying coefficient quantile regression model for time series data takes the following form, which is a semiparametric form of model (1),

$$q_\tau(U_t,\, X_t) = \beta_\tau^T X_{t1} + \alpha_\tau(U_t)^T X_{t2}, \tag{2}$$

where $X_t = (X_{t1}^T,\, X_{t2}^T)^T \in \Re^{p+q}$, the first component of X_{t1} or X_{t2} might be 1, $U_t \in \mathcal{R}^d$ is called the smoothing variable, which might include some of X_t or other exogenous variables or lagged variables, $\alpha_\tau(\cdot) = (a_{1,\tau}(\cdot),\, \ldots,\, a_{q,\tau}(\cdot))^T$, and $\{a_{k,\tau}(\cdot)\}$ are smooth coefficient functions. Here, $\{a_{k,\tau}(\cdot)\}$ and β_τ are allowed to depend on τ. For simplicity, we may drop τ from $\alpha_\tau(\cdot)$ and β_τ whenever there is no confusion. Our interest here is to estimate the coefficient functions $\{a_{k,\tau}(\cdot)\}$, the parameter vector β_τ and the conditional quantile of Y_t given in (2).

The general setting in (2) is related to many familiar forms in quantile regression and semiparametric regression models. When X_{t1} are lagged dependent variables and $X_{t2} = 0$, we obtain the quantile autoregressive (QAR) model of Koenker and Xiao (2006). If there is no X_t (only U_t) in (2), then (2) reduces to the ordinary nonparametric quantile regression model which has been studied extensively; see Cai (2002a) and Cai, Gu and Li (2009). Further, if $X_{t2} = 1$ in (2), then model (2) includes a partially linear quantile model explored by He and Shi (1996), He and Liang (2000) and Lee (2003) as a special case. Finally, if there is no X_{t1} in (2), then model (2) becomes the varying coefficient quantile regression model

$$q_\tau(U_t,\, X_{t2}) = \alpha_\tau(U_t)^T\, X_{t2} \tag{3}$$

studied by Honda (2004) and Kim (2007) for iid data and Cai and Xu (2008) for time series contexts.

Comparing to the fully nonparametric models of Honda (2004) and Cai and Xu (2008) and the fully parametric models such as Koenker and Xiao (2006), the partially varying coefficient quantile regression model (2) serves as an intermediate class of models with good robustness by nonparametric treatment on certain covariates and relatively more precise estimation on the parametric effect of other variables. In this semiparametric approach, existing information concerning possible linearity of some of the components can be taken into account in such models to improve efficiency. Thus, root-n consistent estimation of the parametric coefficients are obtained. Engle, Granger, Rice and Weiss (1986) were among the first to study the partially linear model. It has since been studied extensively in both economics and statistics literature. With respect to developments in semiparametric dynamic modelling, various estimation and testing issues have been discussed for the case where data are strictly stationary (such as Gao (2007)) since the publication of Robinson (1988, 1989). Li, Huang, Li and Fu (2002), Zhang, Lee and Song (2002), Ahmad, Leelahanon and Li (2005), and Fan and Huang (2005) studied partially varying coefficient estimation for the conditional mean model. To the best of our knowledge, the semiparametric dynamic quantile modelling like (2) has not been studied in either econometrics or statistics literature.

In this paper, we propose a consistent semiparametric estimation procedure for the dynamic quantile regression model (2). Although the focus of our model is on the parameters β_τ, estimation of both $\alpha_\tau(\cdot)$ and β_τ and thus $q_\tau(U_t, X_t)$ are studied. To estimate both the parameter vector β_τ and the functional coefficients $\alpha_\tau(\cdot)$, we propose a three-stage approach as follows. First, β_τ is regarded as a function of U_t, $\beta_\tau(U_t)$. Thus, the model becomes a functional coefficient model and all coefficient functions can be estimated by a nonparametric fitting scheme. Second, an average method is used to obtain a root-n consistent estimator for β_τ. To estimate $\alpha_\tau(\cdot)$, for any $\sqrt{n}$-consistent estimate $\widehat{\beta}_*$ of β_τ, we construct the partial quantile residual $Y_{t*} = Y_t - \widehat{\beta}_*^T X_{t1}$, and a nonparametric approach can be applied to estimate $\alpha_\tau(\cdot)$ based on the partial quantile residuals. We show that our three-stage nonparametric estimator of $\alpha_\tau(\cdot)$ is asymptotically consistent and is indeed "oracle" in the sense that the asymptotic properties of this nonparametric estimator are not affected by knowing β or not. Further, we address the efficiency issue for the data observed from a martingale difference sequence and propose a simple efficient estimator to estimate β_τ by using the weighted average approach and choosing the optimal weighting function via minimizing the asymptotic variance. An important statistical question in fitting model (2) arises if the coefficient functions $\alpha_\tau(\cdot)$ are actually varying (namely, if a linear quantile regression model is adequate). This amounts to testing whether the coefficient functions are constant or in a certain parametric form. A simple and easily implemented testing procedure is proposed based on the asymptotic theory derived in this paper. Our simulation shows that the proposed estimators has reasonably good sampling properties and the testing procedure is indeed powerful. Finally, notice that the well known Robinson (1988) type approach or profile least squares type method of Speckman (1988) for classical semiparametric regression models (see Gao (2007)) might not be suitable to quantile setting; see Remark 1 later in Section 2 for more discussion on this issue.

The rest of this paper is as follows. Section 2 is devoted to the presentation of estimation procedures with some discussions. The asymptotic results are given in Section 3. In Section 4, we propose a simple and easily implemented testing method for testing the goodness-of-fit of a parametric model against model (2). Efficiency is discussed and an efficient estimator is proposed in Section 5. To illustrate the finite sample performance of the proposed estimators, we conduct a Monte Car lo simulation in Section 6. Concluding remarks are presented inSection 7. Finally, all theoretical proofs of the asymptotic results stated in Sections 3 and 4 are given in the Appendix.

2 Estimation Procedures

Throughout this section, we consider estimation of model (2) based on the observed data $\{(Y_t, U_t, X_t)\}_{t=1}^n$. Without loss of generality and for simplicity of exposition, we consider only the case when U_t in (2) is one-dimensional ($d = 1$). For multivariate U_t, the modeling procedure and the related theory for the univariate case continue to hold but more and complicated notations involve. In addition, since the rate of convergence of the nonparametric functional coefficient estimate is dependent on d, the conventional curse of dimensionality presents in estimation of $\alpha_\tau(\cdot)$; see Ruppert and Wand (1994) for related discussion. We apply a local polynomial fitting scheme to estimate the related functionals although other smoothing methods such as the Nadaraya-Watson kernel method and spline methods are applicable. The main reason for using a local polynomial fitting is due to the attractive properties of this method, such as high statistical efficiency in an asymptotic minimax sense, design adaptation, and automatic boundary corrections; see Fan and Gijbels (1996) for some discussions on advantages of using local polynomial method. For ease of notation, we may drop the subscript τ from β_τ and $\alpha_\tau(\cdot)$ and simply denote them as β and $\alpha(\cdot)$.

Given model (2), if β_τ were known, we could be able to construct the following partial quantile residual:

$$Y_{t1} = Y_t - \beta_\tau^T X_{t1}.$$

Using the above transformation, the quantile regression model in (2) can be re-written as

$$Q_{Y_{1t}}(\tau | U_t, X_{2t}) = q_\tau(U_t, X_{2t}) = \alpha_\tau(U_t)^T X_{t2}.$$

Under smoothness condition of coefficient functions $\alpha_\tau(\cdot)$ so that it has $(m+1)$th continuous derivative ($m \geq 1$), for any given point u_0, when U_t is in a neighborhood of u_0, $\alpha_\tau(U_t)$ can be approximated by a polynomial function as

$$\alpha_\tau(U_t) \approx \alpha_\tau(u_0) + \alpha_\tau'(u_0)\,(U_t - u_0) + \cdots + \alpha_\tau^{(m)}(u_0)\,(U_t - u_0)^m / m\,!,$$

where $\approx$ denotes the approximation by ignoring the higher orders, thus,

$$q_\tau(U_t, X_{2t}) \approx \sum_{j=0}^{m} \theta_{j\tau}^T\, X_{t2}\, (U_t - u_0)^j,$$

where $\theta_j = \theta_{j\tau} = \alpha_\tau^{(j)}(u_0)/j!$ for $0 \leq j \leq m$. Then, we may estimate $\alpha_\tau(u_0)$ based on the following nonparametric functional coefficient quantile regression estimation

$$\min_\theta \sum_{t=1}^{n} \rho_\tau \left(Y_{t1} - \sum_{j=0}^{m} \theta_j^T\, X_{t2}\, (U_t - u_0)^j \right) K_h(U_t - u_0), \tag{4}$$

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 and it controls the amount of smoothing used in estimation.

In practice, β_τ is unknown and thus the transformation $Y_{t1} = Y_t - \beta_\tau^T X_{t1}$ is infeasible. To estimate both the parameter vector β and the functional coefficients $\alpha(\cdot)$ in (2), we propose the following three-stage approach.

First, β is regarded as a function of U_t, $\beta(U_t)$. Then, the model becomes a functional coefficient model and all coefficient functions can be estimated by using the following local fitting

$$\min_{\beta,\theta} \sum_{t=1}^{n} \rho_\tau \left(Y_t - \beta^T X_{t1} - \sum_{j=0}^{m} \theta_j^T X_{t2} \, (U_t - u_0)^j \right) K_h(U_t - u_0). \tag{5}$$

We denote the above local polynomial estimator of β as $\widehat{\beta}(u_0)$. If we smooth locally around U_t and consider a local linear estimation, the objective function of the above estimation becomes

$$\min_{\beta,\theta} \sum_{s \neq t}^{n} \rho_\tau \left(Y_s - \beta^T X_{s1} - \theta_0^T X_{s2} - \theta_1^T X_{s2}(U_s - U_t) \right) K_h(U_s - U_t). \tag{6}$$

Notice that while β is a global parameter, the above estimation of β involves only local data points in a neighborhood of U_t so that it is not optimal. Indeed, it follows from Cai and Xu (2008) that under some regularity conditions, $\widehat{\beta}(\cdot) - \beta = O_p((n\,h)^{-1/2})$. An optimal estimation of the constant coefficients requires using all data points and the optimal convergence rate should be $\sqrt{n}$ instead of $\sqrt{n\,h}$. To obtain a $\sqrt{n}$-consistent estimator for β_τ, we use the following average method to obtain a second stage estimator of β that achieves the optimal rate of convergence:

$$\widetilde{\beta} = \widetilde{\beta}_\tau = \frac{1}{n} \sum_{t=1}^{n} \widehat{\beta}(U_t), \tag{7}$$

as we show in Theorem 1 (see later) that the above estimator[1] is $\sqrt{n}$-consistent and asymptotically normal.

To estimate the functional coefficients $\alpha(\cdot)$, we define the estimated partial quantile residual as $Y_{t*} = Y_t - \widetilde{\beta}^T X_{t1}$, where $\widetilde{\beta}$ is a $\sqrt{n}$-consistent estimate of β, and we consider the following feasible local polynomial functional coefficient estimation of (4)

$$\min_{\theta_*} \sum_{t=1}^{n} \rho_\tau \left(Y_{t*} - \sum_{j=0}^{m} \theta_{j*}{}^T X_{t2} \, (U_t - u_0)^j \right) K_{h_1}(U_t - u_0), \tag{8}$$

where h_1 is the bandwidth used for this step, which is different from the bandwidth used in (6); see Remark 6 later in Section 3 for more discussions. Solving the minimization problem in (8) gives $\widetilde{\alpha}(u_0) = \widehat{\theta}_{0*}$, the local polynomial estimate of $\alpha(u_0)$, and $\widetilde{\alpha}^{(j)}(u_0) = j\,!\,\widehat{\theta}_{j*}$ $(j \geq 1)$, the local polynomial estimate of the jth derivative $\alpha^{(j)}(u_0)$ of $\alpha(u_0)$.

[1]There are alternative ways of constructing root-n estimator of β.

We show in Theorem 2 in Section 3 that the above nonparametric estimator is "oracle" in the sense that the asymptotic properties of this nonparametric estimator are not affected by preliminary estimation of β_τ. All details about their asymptotic properties are presented in the next section.

Remark 1: It is worth to point out that the well known Robinson (1988) type or profile least squares type of Speckman (1988) estimation approach for classical semiparametric regression models (see Gao (2007)) might not be suitable to quantile setting. For example, for a profile least squares method, to estimate the parameters in the linear component under the least squares framework, one usually multiplies a projection matrix to remove the nonparametric component and then fit a linear model; see Fan and Huang (2005) for details. But this approach is not applicable to the quantile setting due to lack of explicit normal equations.

Remark 2: For exploratory purposes, one might use a simple global parametric method such as series-type or splines or sieve approximation to $\alpha_\tau(\cdot)$ and then estimate β_τ under a parametric framework. This approach is easily implemented. However, for such methods, it might not be easy to establish the asymptotic results for the estimator of β_τ without imposing strong assumptions like $E[X_t|U_t = u] = 0$ all u or even stronger. Under this type of assumptions, He and Shi (1996) and He and Liang (2000) studied a series-type method for a partially linear quantile model for iid data. This harsh assumption might not be appropriate for a dynamic model.

Remark 3: The programming involved in the above local polynomial quantile estimation can be modified with few efforts from the existing programs for a linear quantile model. For example, to obtain the nonparametric estimate of parameters for each grid point u_0 in (5), the local polynomial quantile estimation can be implemented by the function **rq()** in the package **quantreg** in the computing language **R**, due to Koenker (2004), by setting covariates as X_t , $X_t\,(U_t - u_0)$, $\cdots$, $X_t\,(U_t - u_0)^m$, and the weight as $K_h(U_t - u_0)$. Alternatively, one can use the function **lprq()** in the same package.

Remark 4: In various practical applications, Fan and Gijbels (1996) recommended using the local linear fit ($m = 1$). Therefore, for expositional purpose and without loss of generality, in this paper, we consider only the case when $m = 1$ (local linear fitting).

3 Asymptotic Properties

In this section, we develop the asymptotic theory for the proposed estimators $\widetilde{\beta}$ and $\widetilde{\alpha}(u_0)$ based on local linear estimation. To show that $\widetilde{\beta}$ is a $\sqrt{n}$ -estimator of β and to establish its asymptotic normality, we will employ the U-statistics technique. For the convenience of analyzing U-statistics, we introduce β-mixing (absolutely regular), which is defined as follows. A stationary process $\{(\xi_t, \mathcal{F}_t), -\infty < t < \infty\}$ is said to be absolutely regular if the mixing coefficient $\beta(n)$ defined by

$$\beta(n) = E\Big\{ \sup_{A \in \mathcal{F}_{t+n}^{\infty}} |P(A|\, \mathcal{F}_{-\infty}^{t}) - P(A)| \Big\}$$

converges to zero as $n \to \infty$. β-mixing includes many linear and nonlinear time series models as special cases; see Doukhan (1994) for the definition and Cai (2002a) for some examples in economics and finance.

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. Denote $f_u(\cdot)$ the marginal density of U_t and $f_{y|u,x}(\cdot|\cdot)$ the conditional density of Y_t given (U_t, X_t). In addition, let

$$\Omega(u) = E\left[X_t X_t^T | U_t = u\right] \quad \text{and} \quad \Omega^*(u) = E\left[X_t X_t^T f_{y|u,x}(q_\tau(U_t, X_t)) | U_t = u\right].$$

and define

$$\mu_j = \int u^j K(u) du, \quad \text{and} \quad \nu_0 = \int K^2(u) du. \tag{9}$$

Assumption A:

(A1) $\alpha(u)$ is $(m+1)$-th order continuously differentiable in a neighborhood of u_0 for any u_0. Further, $f_u(u)$ is continuous and $f_u(u) > 0$ on $\{u : 0 < F_u(u) < 1\}$, and $f_{y|u,x}(y)$ is bounded and satisfies the Lipschitz condition.

(A2) $\Omega(u_0)$ and $\Omega^*(u_0)$ are positive-definite and continuous in a neighborhood of u_0.

(A3) The kernel function $K(\cdot)$ is symmetric and has a compact support, say $[-1,\ 1]$.

(A4) The bandwidth h satisfies $h \to 0$ and $n\,h \to \infty$.

Assumption B:

(B1) $\{(V_t, Y_t)\}$ is a strictly β-mixing stationary process with mixing coefficient $\beta(n)$ satisfying $\sum_{n=1}^{\infty} n^2 \beta^{\delta/(1+\delta)}(n) < \infty$ for some $\delta > 0$.

(B2) $E\|X_t\|^{2(1+\delta)} < \infty$ for some $\delta > 0$. Further, functions $f_u(\cdot)$, $\Omega^*(\cdot)$, and $\Omega(\cdot)$ and their inverse functions are uniformly bounded.

(B3) The bandwidth $h = O(n^{-\lambda})$, where $1/4 < \lambda < 1/3$.

Clearly, Assumption B3 is about the undersmoothing at the first step for the nonparametric estimate and it is slightly stronger than $n\, h^4 \to 0$, which is commonly imposed for iid samples.

The main idea of establishing the consistency and asymptotic normality of $\widetilde{\beta}$ under the mixing setting is that first we give an explicit expression for $\widehat{\beta}(U_t)$ as a linear estimator plus a higher order term, and then, we can express $\widetilde{\beta}$ as a U-statistic form. Finally, we apply the U-statistic technique as in Dette and Spreckelsen (2004) to obtain the consistency and asymptotic normality for $\widetilde{\beta}$.

In Appendix A, we provide some useful lemmas. From Lemmas 1 - 3 in our Appendix A and Theorem 2 in Dette and Spreckelsen (2004), we can establish the following asymptotic normality for $\widetilde{\beta}$. The detailed proofs of the above lemmas and the following theorem are relegated to Appendix B. All limits will be taken as $n \to \infty$; this will not be mentioned explicitly in the body of the paper. Finally, we define,

$$B_1^* = e_1^T\, E\left[(\Omega^*(U_1))^{-1} \Omega^{*'}(U_1) \begin{pmatrix} 0 \\ \alpha'(U_1) \end{pmatrix} \right]$$

where $\Omega^{*'}(u)$ is the first order derivative of $\Omega^*(u)$, $e_1^T = (I_p, 0_{p\times q})$ with I_p being a $p \times p$ identity matrix and $0_{p\times q}$ being a $p \times q$ zero matrix, and

$$B_2^* = e_1^T\, E\left[(\Omega^*(U_1))^{-1} \Gamma(U_1) \right],$$

where

$$\Gamma(u_0) = E\left[f'_{y|u,x}(q_\tau(U_t, X_t)) X_t \left(\alpha'(U_t)^T X_{t2} \right)^2 | U_t = u_0 \right]$$

and $f'_{y|u,x}(y)$ denotes the derivative of $f_{y|u,x}(y)$ with respect to y.

Theorem 1: Under Assumptions A1 - A4 and B1 - B2,

$$\sqrt{n}\left[\widetilde{\beta}_\tau - \beta_\tau - B_\beta\right] \stackrel{d}{\longrightarrow} \mathcal{N}(0,\ \Sigma_\beta),$$

where the asymptotic bias term is $B_\beta = h^2\mu_2(B_1^* - B_2^*/2)$, μ_2 is defined as (9), and the asymptotic variance is

$$\begin{aligned}\Sigma_\beta &= \tau(1-\tau)\, E\left[e_1^T(\Omega^*(U_1))^{-1}\Omega(U_1)(\Omega^*(U_1))^{-1}e_1\right] \\ &\quad +2\sum_{s=1}^{\infty}\mathrm{Cov}(e_1^T\left(\Omega^*(U_1))^{-1}X_1\,\eta_1, e_1^T(\Omega^*(U_{s+1}))^{-1}X_{s+1}\,\eta_{s+1}\right).\end{aligned}$$

Here, $\eta_t = \tau - I\{Y_t \le q_\tau(U_t, X_t)\}$. In addition, under Assumptions A1 - A4 and B1 - B3, we have

$$\sqrt{n}\left[\widetilde{\beta}_\tau - \beta_\tau\right] \stackrel{d}{\longrightarrow} \mathcal{N}(0,\ \Sigma_\beta).$$

Remark 5: From Theorem 1, the estimator is root-n consistent because of the bandwidth Condition B3 so that $\sqrt{n}h^2 \to 0$. Therefore, we should under-smooth in the first stage to reduce the bias since the variance is averaged out in the second stage.

In many important time series models including the case when Y_t is a p-th order Markov process, η_t is a martingale difference sequence. In this case, we have the following corollary.

Corollary 1: Under the additional assumption that η_t is a martingale difference sequence, the results of Theorem 1 holds with the following asymptotic variance

$$\Sigma_{\beta,0} = \tau(1-\tau)\, E\left[e_1^T(\Omega^*(U_1))^{-1}\Omega(U_1)(\Omega^*(U_1))^{-1}e_1\right].$$

Of course, the above corollary implies that, in the special case when $\{(U_t, X_t, Y_t)\}_{t=1}^n$ are iid, the result is the same as that in Lee (2003) for iid data (see Theorem 2 in Lee (2003)) for a partially linear quantile regression model (that is $X_{t2} = 1$), while Lee (2003) did not use a kernel smoothing method and did not provide the asymptotic bias term.

Finally, we establish the asymptotic results for $\widetilde{\alpha}(u_0)$ given in (8). To this effect, we introduce the following additional definitions:

$$\Omega_{22}(u) = E\left[X_{t2}X_{t2}^T|U_t = u\right] \text{ and } \Omega_{22}^*(u) = E\left[X_{t2}X_{t2}^T f_{y|u,x}(q_\tau(U_t, X_t))|U_t = u\right].$$

and set $\Sigma_a(u_0) = \left[\Omega_{22}^*(u_0))\right]^{-1}\Omega_{22}(u_0)\left[\Omega_{22}^*(u_0)\right]^{-1}$. We need some sufficient conditions as follows.

Assumption C:

(C1) Same as (B1).

(C2) $E\|X_{t2}\|^{2(1+\delta^*)} < \infty$ with $\delta^* > \delta$. Further, functions $f_u(\cdot)$, $\Omega_{22}^*(\cdot)$, and $\Omega_{22}(\cdot)$ and their inverse functions are uniformly bounded.

(C3) $f(u, v|x_{02},\, x_{s2}; s) \le M < \infty$ for $s \ge 1$, where $f(u, v|x_{02},\, x_{s2}; s)$ is the conditional density of $(U_0,\, U_s)$ given $(X_{02} = x_{02},\, X_{s2} = x_{s2})$.

(C4) $n^{1/2-\delta/4}\, h_1^{\delta/\delta^*-1/2-\delta/4} = O(1)$, and $h/h_1 = o(1)$.

Clearly, (C4) allows the choice of a wide range of smoothing parameter values and is slightly stronger than the usual condition of $n\, h_1 \to \infty$. However, for the bandwidths of optimal size (i.e., $h_1 = O(n^{-1/5})$), (C4) is automatically satisfied for $\delta \ge 3$ and it is still fulfilled for $2 < \delta < 3$ if δ^* satisfies $\delta < \delta^* \le 1+1/(3-\delta)$. This assumption is also imposed by Cai, Fan and Yao (2000) for mean regression. Finally, if $X_{t2} = 1$ in model (2), (C1) can be replaced by (C1)$'$: $\beta(n) = O(n^{-\delta})$ for some $\delta > 2$ and (C4) can be substituted by (C4)$'$: $n\, h_1^{\delta/(\delta-2)} \to \infty$; see Cai (2002a) for related discussions on this issue.

Theorem 2: Under Assumptions A and C, the local linear estimator of $\alpha(u_0)$ has the following asymptotic distribution:

$$\sqrt{nh_1}\left[\widetilde{\alpha}(u_0) - \alpha(u_0) - \frac{h_1^2 \mu_2 \alpha''(u_0)}{2} + o_p(h_1^2)\right] \xrightarrow{d} \mathcal{N}\left(0, \Sigma_\alpha\right),$$

where $\Sigma_\alpha = \tau(1-\tau)\nu_0 \Sigma_a(u_0)/f_u(u_0)$, and μ_2 and ν_0 are defined in (9).

Remark 6: Notice that from Theorem 2, we see easily (by comparing Theorem 1 in Cai and Xu (2008)) that the asymptotic result is exactly same as that for the case where β would be known. This property is referred as "oracle" in the literature.

Remark 7: It is clear that the asymptotic mean squared error (AMSE) is of the order $O(n^{-4/5})$ if the bandwidth is taken to be the optimal one as $h_{1,opt} = O(n^{-1/5})$. Also, at the final step, any data-driven type bandwidth selection can be applied; see, for example, Cai and Xu (2008) for a rule-of-thumb bandwidth.

4 Inference

The asymptotic results derived in the previous sections facilitate statistical inference in conditional quantile models. An important inference problem is to test constancy of the coefficients $\alpha(\cdot)$, corresponding to

$$H_0 : \alpha_\tau(u) = \alpha_\tau, \quad \text{versus} \quad H_1 : \ \text{varying coefficients} \ \ \alpha_\tau(u). \tag{10}$$

This hypothesis can be tested in different ways. A natural approach to test constancy of $\alpha(\cdot)$ is to directly look at the variability of the estimated coefficient $\widetilde{\alpha}(u)$. For this purpose, in addition to the semiparametric functional coefficient estimator $\widetilde{\alpha}(u)$, we may consider the (null) restricted regression

$$q_\tau(U_t, X_t) = \beta_\tau^T X_{t1} + {\alpha_\tau}^T X_{t2}, \tag{11}$$

and compare $\widetilde{\alpha}(u)$ with the restricted quantile regression estimator $\breve{\alpha}$ from (11) over a range of u, based on $[\widetilde{\alpha}(u) - \breve{\alpha}]$.

Notice that under the null hypothesis and regularity conditions,

$$\sqrt{n}(\breve{\alpha} - \alpha) \ \xrightarrow{d} \ \mathcal{N}\left(0, \tau(1-\tau)D_{22}\right),$$

where D_{22} is the lower diagonal sub-matrix of $D = H^{-1}\Sigma_0 H^{-1}$, $H = E[\Omega(U_t)]$, and $\Sigma_0 = E[\Omega^*(U_t)]$; see Theorem 4.1 in Koenker (2005) for details. If we choose $nh_1^5 \to 0$, then under the null hypothesis,

$$\sqrt{nh_1}\left(\widetilde{\alpha}(u) - \breve{\alpha}\right) = \sqrt{nh_1}\left(\widetilde{\alpha}(u) - \alpha\right) + o_p(1) \ \xrightarrow{d} \ \mathcal{N}\left(0, \Sigma_\alpha(u)\right).$$

If we denote the consistent estimator of Σ_α as $\widehat{\Sigma}_\alpha(u)$, which may be obtained by the method proposed in Cai and Xu (2008), we have

$$\sqrt{nh_1}\widehat{\Sigma}_\alpha^{-1/2}(u)\left(\widetilde{\alpha}(u) - \breve{\alpha}\right) \ \xrightarrow{d} \ \mathcal{N}\left(0, I_q\right),$$

where I_q is an identity matrix and q is the dimension of X_{t2}, and then,

$$\left\|\sqrt{nh_1}\widehat{\Sigma}_\alpha(u)^{-1/2}\left(\widetilde{\alpha}(u) - \breve{\alpha}\right)\right\|^2 \ \xrightarrow{d} \ \chi^2(q).$$

where $\chi^2(q)$ is χ^2 random variable with q degrees of freedom.

In order to look at $\widetilde{\alpha}(u) - \breve{\alpha}$ over a range of u, and construct an asymptotically valid test, we need to find out the joint distribution of the estimated functional coefficients over a number of points. Let u_i^* $(i = 1, \cdots, m)$ be m distinct points. The joint distribution of $\widetilde{\alpha}(u_i^*)$ $(i = 1, \cdots, m\ )$ is given by the following Theorem.

Theorem 3. Under Assumptions A and C, the kernel estimators of the parameters has the following limiting joint distribution

$$\sqrt{n\,h_1}\begin{pmatrix} \widetilde{\alpha}(u_1^*) - \breve{\alpha} \\ \vdots \\ \widetilde{\alpha}(u_m^*) - \breve{\alpha} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \Sigma_\alpha(u_1^*) & & 0 \\ & \ddots & \\ 0 & & \Sigma_\alpha(u_m^*) \end{pmatrix}\right).$$

Now we can define the test statistic. For distinct $u_1^*, ..., u_m^*$, define

$$T_m = \max_{1 \le i \le m} \left\| \sqrt{nh_1}\widehat{\Sigma}_\alpha(u_i^*)^{-1/2} \left(\widetilde{\alpha}(u_i^*) - \breve{\alpha}\right) \right\|^2. \tag{12}$$

Then, we show easily that suppose Assumptions A and C hold, under H_0 given by (10), as $n \to \infty$,

$$T_m \xrightarrow{d} \max_{1 \le i \le m} \chi_i^2(q),$$

where $\chi_1^2(q), \cdots, \chi_m^2(q)$ are independent chi-square random variables with q degrees of freedom. Thus, one can reject the null if T_m is too large. The critical value of T_m can be easily tabulated since the limiting distribution of T_m is a functional of independent chi-square random variables (with q degrees of freedom) that is free of nuisance parameters and quantiles.

Remark 8: The testing procedure given by (12) and Theorem 2 is an asymptotic test. It has the advantage that its limiting distribution is free of nuisance parameter and quantiles. As an alternative, we may consider a bootstrap based test of (12), which may give some finite sample improvement. Another issue related to the proposed test is the choice of finite distinct points $\{u_i^*\}_{i=1}^m$. In practice, we may consider, say choosing lower quartile, median, and upper quartiles, or we may construct the test based on all deciles. In some applications, different choices of m and the points $\{u_i^*\}_{i=1}^m$ may potentially lead to different conclusions in finite sample, thus it would be desirable to consider all points u on the domain U_t, and treat $\widetilde{\alpha}(u)$ as a process in u, and Kolmogorov-Smirnov or Cramer-von-Mises type tests may be constructed. Of course, it would be warranted as a future research topic to investigate the properties of those test statistics.

5 More Efficient Estimation of β_τ

The estimator $\widetilde{\beta}$ proposed in the previous section has the advantage that it is easy to construct and also achieves the $\sqrt{n}$-rate of convergence. In addition to this simple estimator, other root-n consistent estimators of β can be constructed. To estimate the parameters β without being overly influenced by the tail behavior of the distribution of U_t, one might use a trimming function $w_t = I(U_t \in \mathcal{D})$ with a compact subset $\mathcal{D}$ of $\Re$; see Cai and Masry (2000) for details. Then, (7) becomes the weighted average estimator as

$$\widetilde{\beta}_w = \frac{\sum_{t=1}^n w_t \widehat{\beta}(U_t)}{\sum_{t=1}^n w_t}. \tag{13}$$

For simplification of presentation, our focus is on $\widehat{\beta}$ given in (7). Indeed, this type of estimator was considered by Lee (2003) for a partially linear quantile regression model.

To estimate β in a more efficient way, similar to Cai and Fan (2000), we can construct a general weighted average approach as follows

$$\check{\beta}_w = \left[\frac{1}{n}\sum_{t=1}^n W(U_t)\right]^{-1} \left[\frac{1}{n}\sum_{t=1}^n W(U_t)\widehat{\beta}(U_t)\right], \tag{14}$$

where $W(\cdot)$ is a weighting function (a symmetric matrix). This is a functional of $\widehat{\beta}(u)$ and possesses some good properties. The weighted averaging can significantly reduce the variance, but not bias, of the resulting estimate $\check{\beta}_w$. This enables us to obtain an optimal estimate of β via adjusting the bandwidths and choosing the optimal weighting function by minimizing the asymptotic variance. Of course, the trimming function w_t in (13) can be applied to (14) too.

Following a similar argument as the proof of Theorem 1, it can be shown that, when η_t is a martingale difference sequence, the weighted average estimate $\check{\beta}_w$ of β_τ defined in (14) has the following asymptotic distribution under some regularity conditions,

$$\sqrt{n}\,\left[\check{\beta}_w - \beta_\tau\right] \stackrel{d}{\longrightarrow} \mathcal{N}(0,\, \Sigma_w), \tag{15}$$

where

$$\Sigma_w = \tau(1-\tau) W_e^{-1} E\left[W(U_1) e_1^T (\Omega^*(U_1))^{-1} \Omega(U_1) (\Omega^*(U_1))^{-1} e_1 W(U_1)^T\right] W_e^{-1}, \tag{16}$$

where $W_e = E[W(U_1)]$. When the conditional density $f_{y|u,x}(q_\tau(u,x))$ is constant, denoted by $f_\tau(0)$, then Σ_w becomes

$$\Sigma_w = \frac{\tau(1-\tau)}{f_\tau^2(0)} W_e^{-1} E\left[W(U_1) e_1^T (\Omega(U_1))^{-1} e_1 W(U_1)\right] W_e^{-1}.$$

Clearly, the optimal choice of the weighting function is to set

$$W_{opt}(u) = \left[e_1^T(\Omega(u))^{-1}e_1\right]^{-1} = \Omega_{11}(u) - \Omega_{12}(u)\Omega_{22}^{-1}(u)\Omega_{21}(u) \tag{17}$$

based on the matrix theory, where $\Omega_{11}(u) = E[X_{t1}X_{t1}^T|U_t = u]$, $\Omega_{12}(u) = E[X_{t1}X_{t2}^T|U_t = u]$, $\Omega_{21} = \Omega_{12}^T$, and $\Omega_{22}(u) = E[X_{t2}X_{t2}^T|U_t = u]$, so that the optimal asymptotic variance is

$$\Sigma_{w,opt} = \frac{\tau(1-\tau)}{f_\tau^2(0)}[E(W_{opt}(U_1))]^{-1}.$$

Specifically, if $X_{t2} = 1$, then, $W_{opt}(u) = \mathrm{Var}(X_{t1}|U_t = u)$, which is the same as that in He and Shi (1996) and Lee (2003). By the same token, one may derive the optimal weighting function when the conditional density $f_{y|u,x}(q_\tau(u, x))$ is not constant, however it has a complex form. Finally, notice that if $\Omega(u)$ does not depend on u, then $W_{opt}(u)$ in (17) is constant so that the efficient estimator $\check{\beta}_w$ given in (14) is the same as the estimator given in (7).

In practical applications, $\Omega(u)$ is usually unknown. However, it can be estimated by using any nonparametric method, say local linear approach since it is a conditional mean function. For example, in view of (17), we can use the following estimated weighting function

$$\widehat{W}_{opt}(u) = \widehat{\Omega}_{11}(u) - \widehat{\Omega}_{12}(u)\widehat{\Omega}_{22}^{-1}(u)\widehat{\Omega}_{21}(u), \tag{18}$$

where $\widehat{\Omega}_{ij}(u)$ denotes a kernel estimate of the corresponding conditional expectation $\Omega_{ij}(u)$; see Fan and Gijbels (1996). One can show easily based on the kernel estimation theory in Fan and Gijbels (1996) that $\widehat{W}_{opt}(u)$ is a consistent estimate of $W_{opt}(u)$ given in (17).

Remark 9: Without assuming that η_t is a martingale difference sequence, it is still possible to establish the asymptotic result in (15) for $\check{\beta}_w$. In this general case, the asymptotic variance in (16) should be given by

$$\Sigma_w = \tau(1-\tau)W_e^{-1}E\left[W(U_1)e_1^T(\Omega^*(U_1))^{-1}\Omega(U_1)(\Omega^*(U_1))^{-1}e_1W(U_1)^T\right]W_e^{-1}$$
$$+2W_e^{-1}\sum_{s=1}^{\infty}\mathrm{Cov}(W(U_1)e_1^T\left(\Omega^*(U_1))^{-1}X_1\,\eta_1, W(U_{s+1})e_1^T(\Omega^*(U_{s+1}))^{-1}X_{s+1}\,\eta_{s+1}\right)W_e^{-1}.$$

But, due to the complex expression for Σ_w, it may not be easy to find an explicit mathematical expression for the optimal weighting function $W_{opt}(\cdot)$.

6 Monte Carlo Simulations

We conduct a Monte Carlo experiment to examine the finite sample performance of the proposed estimation procedures. For comparison purpose, we also estimate the conditional quantile model using the fully parametric quantile autoregression (QAR) model as Koenker and Xiao (2006) in Example 1 and the fully nonparametric model of Cai and Xu (2008) in Example 2. In Example 3, we examine the efficiency gain from using the information of linearity or nonlinearity in estimation of the coefficients and conditional quantiles. Finally, the simulation result in Example 4 shows that the resulting testing procedure is indeed powerful, and the proposed method does give the correct null distribution. The Monte Carlo simulations are repeated 1000 times for each sample size $n = 200$, 500 and 800. Also, we consider several values of τ as $\tau = 0.05$, $\tau = 0.25$, $\tau = 0.50$, $\tau = 0.75$, and $\tau = 0.95$. The bandwidth is selected based on a rule-of-thumb idea similar to the procedure in Cai (2002b) as follows. First, we use a data-driven bandwidth selector as suggested in Cai and Xu (2008) to obtain an initial bandwidth denoted by $\widehat{h}_0$ which should be $O(n^{-1/5})$. At the step of estimating β_1, the bandwidth should be under-smoothed. Therefore, by following the idea in Cai (2002b) for a two-step approach, we take the bandwidth as $\widehat{h}_1 = A_0 \times \widehat{h}_0$ with $A_0 = n^{-\alpha_a}$ with $\alpha_a = 1/10$ so that $\widehat{h}_1$ satisfies Assumption B3. To obtain $\widetilde{\alpha}_2(\cdot)$, we use a a rule-of-thumb bandwidth suggested in Cai and Xu (2008). Finally, the Gaussian kernel is used in the nonparametric estimation.

Example 1: The data generating process is given by:

$$Y_t = \beta_1 Y_{t-1} + \alpha_2(U_t)\, Y_{t-2} + e_t, \qquad t = 1, \ldots, n, \tag{19}$$

where $\beta_1 \equiv 0.5$, $\alpha_2(U_t) = -0.75 + 0.5\cos(\sqrt{2}\,\pi\, U_t)$, U_t is generated from uniform (0, 1) independently, and $e_t \sim N(0,\, 1)$. Clearly, the quantile regression is

$$q_\tau(U_t,\, Y_{t-1},\, Y_{t-2}) = \beta_{0,\tau} + \beta_1 Y_{t-1} + \alpha_2(U_t)\, Y_{t-2},$$

where $\beta_{0,\tau} \equiv \Phi^{-1}(\tau)$ and $\Phi^{-1}(\tau)$ is the τ-th quantile of the standard normal. Therefore, only $\beta_{0,\tau}$ is a function of τ. Notice that $\beta_{0,0.5} = 0$ when $\tau = 0.5$. Obviously, q_τ is a semi-parametric function. The estimators of $\beta_{0,\tau}$, $\beta_{1\tau}$, $\alpha_2(\cdot)$, and $q_\tau(\cdot)$ based on the proposed three stage estimation procedure are denoted as $\widetilde{\beta}_{0,\tau}$, $\widetilde{\beta}_{1\tau}$, $\widetilde{\alpha}_2(\cdot)$, and $\widetilde{q}_\tau(\cdot)$ respectively.

We first look at estimation of the parametric coefficient (i.e., β_1) and investigate the efficiency gain when we take into account the nonlinearity in α_2. We compare the proposed estimator with the estimator of β_1 that does not take into account the nonlinearity in α_2. For this purpose, we also estimate the conditional quantile model using the following fully parametric quantile autoregression (QAR) estimation as Koenker and Xiao (2006) (denote the corresponding estimator as $\check{\beta}_1$):

$$q_\tau(Y_{t-1}, Y_{t-2}) = \beta_{0,\tau} + \beta_1 Y_{t-1} + \alpha_2 Y_{t-2}, \qquad t = 1, \ldots, n. \tag{20}$$

Comparing to $\widetilde{\beta}_1$, $\check{\beta}_1$ ignores the information that α_2 is a function of u. The assessment is based on the absolute deviation error (ADE) as follows:

$$\mathrm{ADE}(\check{\beta}_{j,\tau}) = |\check{\beta}_{j,\tau} - \beta_{j,\tau}|, \quad \text{and} \quad \mathrm{ADE}(\widetilde{\beta}_{j,\tau}) = |\widetilde{\beta}_{j,\tau} - \beta_{j,\tau}|$$

for $j = 0$ and $j = 1$.

The median and standard deviation (in parentheses) of the 1000 values of ADE for both $\check{\beta}$ and $\widetilde{\beta}$ are summarized in the left part of Table 1 for the estimator $\beta_{0,\tau}$ and in the right part of Table 1 for the estimator β_1. From Table 1, one can observe clearly that the medians of the estimated values for $\check{\beta}_0$ are kept the same when the sample size increases - and this is because that $\check{\beta}_0$ is not consistent due to the model misspecification. Also, we can see that the medians and standard deviations of the ADEs for $\widetilde{\beta}_0$ and $\widetilde{\beta}_1$ as well as $\check{\beta}_1$ become smaller for all τ values when the sample size increases. This is along with the asymptotic theory. More importantly, one can see that $\widetilde{\beta}_{1,\tau}$ always outperforms $\check{\beta}_{1,\tau}$ for all settings. To gauge the efficiency gain for $\widetilde{\beta}_1$, we compute the ratios of ADE($\widetilde{\beta}_1$) over ADE($\check{\beta}_1$) and the medians of the 1000 values of the ratios are given in Table 1. One can observe from Table 1 that the ADE for $\widetilde{\beta}_{1,\tau}$ is less than that for $\check{\beta}_{1,\tau}$ up to 40%. Therefore, the efficiency gain for $\widetilde{\beta}_{1,\tau}$ is huge. Moreover, it is interesting but not surprising to see that the performance for $\widetilde{\beta}_0$ is not as good as $\widetilde{\beta}_1$ for all selected quantiles since $\widetilde{\beta}_0$ is a function of τ. Further, it is not surprising to see, due to the sparsity of data in the tailed regions, that the median quantile ($\tau = 0.50$) performance is slightly better than that for two tails ($\tau = 0.05$ and $\tau = 0.95$).

Next, estimation of the conditional quantiles $q_\tau(\cdot)$ is reported in Table 2. As a measure of performance in quantile estimates, we report the mean square error (MSE) of the various estimators. The mean square error of the conditional quantiles is measured as averaged over the sample, i.e. the MSE is calculated as

$$\mathrm{MSE}(\widetilde{q}_\tau) = \frac{1}{n} \sum_{t=1}^{n} \left[\widetilde{q}_\tau(U_t, X_t) - q_\tau(U_t, X_t)\right]^2.$$

Table 1: *Median and Standard Deviation (in parentheses) of 1000 ADE Values for* $\check{\beta}_{j,\tau}$ *and* $\widetilde{\beta}_{j,\tau}$ *as well as Ratios ADE*($\widetilde{\beta}$)/*ADE*($\check{\beta}$).

		Results for $\check{\beta}_{0,\tau}$			Results for $\widetilde{\beta}_{1,\tau}$		
τ	ADE	$n = 200$	$n = 500$	$n = 800$	$n = 200$	$n = 500$	$n = 800$
0.05	$\check{\beta}$	0.526(0.373)	0.533(0.220)	0.560(0.182)	0.059(0.054)	0.039(0.035)	0.030(0.028)
	$\widetilde{\beta}$	0.119(0.101)	0.067(0.060)	0.057(0.045)	0.040(0.038)	0.028(0.023)	0.019(0.017)
	$\widetilde{\beta}/\check{\beta}$				0.640	0.672	0.607
0.25	$\check{\beta}$	0.171(0.123)	0.168(0.085)	0.168(0.067)	0.032(0.033)	0.022(0.021)	0.018(0.016)
	$\widetilde{\beta}$	0.066(0.061)	0.039(0.036)	0.032(0.029)	0.024(0.026)	0.016(0.015)	0.013(0.012)
	$\widetilde{\beta}/\check{\beta}$				0.760	0.726	0.745
0.50	$\check{\beta}$	0.074(0.066)	0.047(0.042)	0.560(0.182)	0.030(0.030)	0.021(0.019)	0.030(0.028)
	$\widetilde{\beta}$	0.059(0.051)	0.037(0.033)	0.030(0.026)	0.023(0.025)	0.015(0.014)	0.011(0.010)
	$\widetilde{\beta}/\check{\beta}$				0.780	0.732	0.694
0.75	$\check{\beta}$	0.161(0.120)	0.174(0.088)	0.167(0.066)	0.037(0.032)	0.022(0.020)	0.018(0.016)
	$\widetilde{\beta}$	0.066(0.057)	0.041(0.035)	0.031(0.028)	0.026(0.027)	0.016(0.014)	0.013(0.012)
	$\widetilde{\beta}/\check{\beta}$				0.759	0.702	0.728
0.95	$\check{\beta}$	0.515(0.363)	0.568(0.223)	0.568(0.164)	0.061(0.053)	0.037(0.035)	0.032(0.027)
	$\widetilde{\beta}$	0.123(0.096)	0.068(0.058)	0.053(0.044)	0.041(0.041)	0.024(0.022)	0.018(0.018)
	$\widetilde{\beta}/\check{\beta}$				0.736	0.649	0.602

Similarly, we can define MSE($\check{q}_\tau$). It is evident that the semiparametric estimator provides much better estimator for the conditional quantiles than the fully parametric model in the presence of nonlinearity in some coefficients. Therefore, one can conclude that the proposed method performs very well comparing to the misspecified linear model.

Example 2: In this example, we consider the following data generating process:

$$Y_t = \beta_1 X_{t1} + \alpha_2(U_t) X_{t2} + \sigma(U_t) e_t, \qquad t = 1, \ldots, n, \tag{21}$$

where $\beta_1 \equiv 0.5$, $\alpha_2(U_t) = \cos(\sqrt{2}\,\pi\, U_t)$, $\sigma(U_t) = 3\exp(-4(U_t-1)^2) + 2\exp(-5(U_t-2)^2)$, X_{t1} is generated from $X_{t1} = 0.75\, X_{t-1,1} + v_{t1}$ with $v_{t1} \sim N(0,1)$ iid, X_{t2} is generated from $X_{t2} = -0.5\, X_{t-1,2} + v_{t2}$ with $v_{t2} \sim N(0, 1/4)$ iid, U_t is generated from $U_t = 0.5\, U_{t-1} + v_{t3}$ with $v_{t3} \sim N(0,1)$ iid, and $e_t \sim N(0,\, 1)$. Clearly, the quantile regression is

$$q_\tau(U_t,\, X_{t1},\, X_{t2}) = \beta_{0,\tau}(U_t) + \beta_1 X_{t1} + \alpha_2(U_t) X_{t2},$$

where $\beta_{0,\tau}(U_t) = \sigma(U_t)\Phi^{-1}(\tau)$.

Table 2: *Median and Standard Deviation (in parentheses) of 1000 MSE Values for* $\check{q}_\tau$ *and* $\widetilde{q}$.

τ	$\tau = 0.05$		$\tau = 0.50$		$\tau = 0.95$	
MSE	$\check{q}_\tau$	$\widetilde{q}_\tau$	$\check{q}_\tau$	$\widetilde{q}_\tau$	$\check{q}_\tau$	$\widetilde{q}_\tau$
$n = 200$	1.120(1.431)	0.175(0.092)	0.811(0.946)	0.057(0.031)	1.192(1.291)	0.171(0.093)
$n = 500$	1.192(0.702)	0.071(0.034)	0.888(0.698)	0.025(0.012)	1.191(0.750)	0.074(0.034)
$n = 800$	1.225(0.559)	0.049(0.022)	0.886(0.408)	0.016(0.008)	1.226(0.606)	0.047(0.023)

First, we compare estimator of the functional coefficient $\alpha_2(\cdot)$ based on the fully nonparametric estimation of Cai and Xu (2008), where $q_\tau(U_t, X_t)$ is

$$q_\tau(U_t, X_t) = \beta_{0,\tau}(U_t) + \beta_1(U_t)\, Y_{t-1} + \alpha_2(U_t)\, Y_{t-2}, \qquad t = 1, \ldots, n. \tag{22}$$

(denoted as $\check{\alpha}_2(\cdot)$) and the estimator based on the proposed procedure (denoted as $\widetilde{\alpha}_2(\cdot)$) in (8). Comparing to $\widetilde{\alpha}_2(\cdot)$, $\check{\alpha}_2(\cdot)$ does not utilize existing information concerning linearity of β_1. To obtain $\check{\alpha}_2(\cdot)$, we use the bandwidth as suggested in Cai and Xu (2008). To compare two estimators, we define the mean absolute deviation error (MADE) as

$$\text{MADE}(\check{\alpha}_2(\cdot)) = \frac{1}{n_u}\sum_{k=1}^{n_u} |\check{\alpha}_2(u_k) - \alpha_2(u_k)|, \quad \text{and} \quad \text{MADE}(\widetilde{\alpha}_2(\cdot)) = \frac{1}{n_u}\sum_{k=1}^{n_u} |\widetilde{\alpha}_2(u_k) - \alpha_2(u_k)|,$$

where $\{u_k\}_{k=1}^{n_u=100}$ are the grid points taken from the interval $(\min(U_t), \max(U_t))$, the domain of U_t with an equal increment. Table 3 reports the median and standard deviation

Table 3: *Median and Standard Deviation (in parentheses) of 1000 MADE Values for* $\check{\alpha}_2(\cdot)$ *and* $\widetilde{\alpha}_2(\cdot)$.

τ	$\tau = 0.05$		$\tau = 0.50$		$\tau = 0.95$	
MADE	MADE($\check{\alpha}_2$)	MADE($\widetilde{\alpha}_2$)	MADE($\check{\alpha}_2$)	MADE($\widetilde{\alpha}_2$)	MADE($\check{\alpha}_2$)	MADE($\widetilde{\alpha}_2$)
n=200	0.117(0.046)	0.110(0.046)	0.071(0.030)	0.064(0.028)	0.119(0.046)	0.114(0.044)
n=500	0.075(0.024)	0.073(0.023)	0.046(0.015)	0.043(0.014)	0.076(0.025)	0.073(0.023)
n=800	0.061(0.018)	0.058(0.017)	0.036(0.011)	0.032(0.011)	0.060(0.019)	0.058(0.018)

of 1000 values for MADE for different settings. One can observe from Table 3 that the finite sample performance of $\widetilde{\alpha}_2(\cdot)$ is better than $\check{\alpha}_2(\cdot)$, although the efficiency gain is not huge. Finally, although $\alpha_2(\cdot)$ does not depend on τ, due to the sparsity of data in the tailed regions, it is not surprising to see that the median quantile ($\tau = 0.50$) performance is better than that for two tails ($\tau = 0.05$ and $\tau = 0.95$).

Next, we compare estimators of $q_\tau(U_t, X_t)$ from these two different estimation procedures: the proposed semiparametric estimation (denoted as $\widetilde{q}_\tau$) and the fully nonparametric estimation (denoted as $\check{q}_\tau$). The results are displayed in Table 4. From Table 4,

Table 4: *Median and Standard Deviation (in parentheses) of 1000 MSE Values for $\check{q}_\tau$ and $\widetilde{q}_\tau$ as well as Ratios MSE($\widetilde{q}_\tau$)/MSE($\check{q}_\tau$).*

τ	MSE	$n = 200$	$n = 500$	$n = 800$
$\tau = 0.05$	MSE($\check{q}_\tau$)	0.979(0.394)	0.514(0.167)	0.364(0.113)
	MSE($\widetilde{q}_\tau$)	0.811(0.358)	0.400(0.145)	0.281(0.094)
	MSE($\widetilde{q}_\tau$)/MSE($\check{q}_\tau$)	0.852	0.804	0.790
$\tau = 0.50$	MSE($\check{q}_\tau$)	0.619(0.213)	0.299(0.089)	0.364(0.113)
	MSE($\widetilde{q}_\tau$)	0.521(0.308)	0.217(0.085)	0.151(0.050)
	MSE($\widetilde{q}_\tau$)/MSE($\check{q}_\tau$)	0.844	0.755	0.720
$\tau = 0.95$	MSE($\check{q}_\tau$)	0.975(0.377)	0.515(0.172)	0.363(0.116)
	MSE($\widetilde{q}_\tau$)	0.792(0.348)	0.402(0.144)	0.282(0.098)
	MSE($\widetilde{q}_\tau$)/MSE($\check{q}_\tau$)	0.860	0.802	0.787

it is evident that: First, the MSEs for $\check{q}_\tau$ and $\widetilde{q}_\tau$ are decreasing when the sample size becomes larger, corroborating the asymptotic theory. Second, we can see clearly that the MSE for $\widetilde{q}_\tau$ is less than the MSE for $\check{q}_\tau$ by 14% to 28% for all settings. Therefore, $\widetilde{q}_\tau$ always outperforms $\check{q}_\tau$. Comparing to the coefficient estimation, the efficiency gain of the proposed estimator in estimation of conditional quantile is more substantial. Finally, once again, the performance for both $\check{q}_\tau$ and $\widetilde{q}_\tau$ when $\tau = 0.50$ is better than that when $\tau = 0.05$ or $\tau = 0.95$.

Example 3: We compare the finite sample performance of the estimator given in Section 2 ((7) and (13)) with that for the efficient estimator in (14) proposed in Section 5. To this effect, we consider the following model

$$q_\tau(U_t,\, X_{t1},\, X_{t2}) = \Phi^{-1}(\tau) + \beta_1\, X_{t1} + \alpha_2(U_t)\, X_{t2},$$

where $\beta_1 = 0.5$, and X_{t1} and X_{t2} are generated as follows

$$X_{t1} = g_1(U_t) + v_{t5}, \quad \text{and} \quad X_{t2} = g_2(U_t) + v_{t6}$$

with $\{v_{t5}\}$ iid from $N(0,\sigma_5^2)$ and $\{v_{t6}\}$ iid from $N(0,\sigma_6^2)$. It is easy to check that the conditional density of Y_t given U_t and X_t is $f(Y_t|U_t,X_t) = \phi(Y_t - \beta_1 X_{t1} - \alpha_2(U_t) X_{t2})$, where $\phi(\cdot)$ is the standard normal density, so that $f(q_\tau(U_t,X_t)|U_t,X_t) = \phi(\Phi^{-1}(\tau)) = f_\tau(0)$. Therefore, $f(q_\tau(U_t,X_t)|U_t,X_t)$ does not depend on U_t. Then, the optimal weighting

Table 5: *Median and Standard Deviation (in parentheses) of 1000 Bias Values for $\widetilde{\beta}_1$ and $\check{\beta}_w$.*

	$\tau = 0.25$		$\tau = 0.75$	
	Bias($\widetilde{\beta}_1$)	Bias($\check{\beta}_w$)	Bias($\widetilde{\beta}_1$)	Bias($\check{\beta}_w$)
n=200	0.002(0.131)	-0.005(0.145)	0.003(0.127)	0.008(0.141)
n=500	-0.002(0.081)	-0.003(0.089)	0.011(0.078)	0.011(0.088)
n=800	0.004(0.064)	-0.003(0.073)	0.004(0.062)	0.005(0.070)

function in (18) has the following form,

$$W_{opt}(U_t) = \left[\sigma_5^2 g_2(U_t)^2 + \sigma_6^2 g_1(U_t)^2 + \sigma_5^2\sigma_6^2\right] / \left[g_2(U_t)^2 + \sigma_6^2\right],$$

although it can be estimated using (18).

We compare the estimator in (7) with the estimator in (14) by examining the biases and ADEs. For this purpose, we consider the following setting: $U_t \sim N(0,1)$, $\alpha_2(U_t) = \cos(\sqrt{2}\,\pi\, U_t)$, $\sigma_5^2 = 0.01$, $\sigma_6^2 = 1$, $g_1(U_t) = 6\,\exp(-4(U_t+1/2)^2) + 4\,\exp(-5\,(U_t-1/2)^2)$ and $g_2(U_t) = U_t/4$. $\widehat{W}_{opt}(u)$ is computed using (18). The median and standard deviation of 1000 values for bias and ADE are represented in Table 5 and in Table 6, respectively. From Table 5, we can observe that the biases for both estimators are almost same. This supports the theory in Section 5 that the efficient estimator can not reduce the bias.

Table 6: *Median and Standard Deviation (in parentheses) of 1000 ADE Values for $\widetilde{\beta}_1$ and $\check{\beta}_w$.*

	$\tau = 0.25$		$\tau = 0.75$	
	ADE($\widetilde{\beta}_1$)	ADE($\check{\beta}_w$)	ADE($\widetilde{\beta}_1$)	ADE($\check{\beta}_w$)
n=200	0.136(0.082)	0.092(0.063)	0.116(0.072)	0.093(0.058)
n=500	0.081(0.051)	0.068(0.037)	0.069(0.043)	0.055(0.034)
n=800	0.059(0.038)	0.039(0.023)	0.051(0.033)	0.037(0.020)

However, it is strongly evident from Table 6 that the ADE value for $\check{\beta}_w$ is much smaller than that for $\widetilde{\beta}_1$ for all cases. This implies that the efficiency gain for $\check{\beta}_w$ is very significant.

Example 4: Finally, we examine the finite sample performance of the test statistic proposed in (12). To this effect, we consider the following model

$$q_\tau(U_t,\, X_{t1},\, X_{t2}) = \Phi^{-1}(\tau) + \beta_1\, X_{t1} + \alpha_2(U_t)\, X_{t2},$$

where $\beta_1 = 0.5$, $U_t \sim U(0,1)$ iid, $X_{t1} \sim N(0,1)$ iid, and X_{t2} is generated from $X_{t2} = g_2(U_t) + v_{t7}$, where $g_2(U_t) = 2\cos(\sqrt{2}\pi\, U_t)$ and $\{v_{t7}\}$ is iid from $N(0,1)$. Clearly, the conditional density of Y_t given U_t and X_t is $f(Y_t|U_t, X_t) = \phi(Y_t - \beta_1\, X_{t1} - \alpha_2(U_t)\, X_{t2})$ and $f(q_\tau(U_t, X_t)|U_t, X_t) = f_\tau(0)$, which is independent of U_t. Also, it is easy to obtain that $\Omega_{22}(U_t) = E[X_{t2}^2|U_t] = g_2(U_t)^2 + 1$, which implies that $\Sigma_\alpha(U_t) = \tau(1-\tau)\nu_0\Omega_{22}(U_t)^{-1}/f_\tau^2(0) = \tau(1-\tau)\sqrt{\pi}\exp(\Phi^{-1}(\tau)^2)/[g_2(U_t)^2 + 1]$.

To demonstrate the power of the proposed test, we consider the null hypothesis

$$H_0 : \alpha_2(U_t) = \alpha_0, \tag{23}$$

namely a linear quantile model, versus the alternative

$$H_1 : \alpha_2(U_t) \neq \alpha_0.$$

The power function is evaluated under a sequence of the alternative models indexed by γ

$$H_1 : \alpha_2(U_t) = \alpha_0 + \gamma(\alpha_2^0(U_t) - \alpha_0), \quad 0 \le \gamma \le 1, \tag{24}$$

where $\alpha_2^0(U_t) = 3\exp(-4(U_t - 1/2)^2) + 4\exp(-5(U_t - 3/4)^2)$ and α_0 is the average height of $\alpha_2^0(U_t)$ which is about 3.9466. We take $u_i^* = 0.05i$ for $1 \le i \le m = 19$. The simulation is repeated 1000 times for each sample size ($n = 200$, $n = 500$ and $n = 800$) and for each quantile ($\tau = 0.2$, $\tau = 0.40$, $\tau = 0.60$ and $\tau = 0.80$). The power function $p(\gamma)$ is estimated based on the relative frequency of T_m over the critical value among 1000 simulations. Figure 1 (1(a) for $\tau = 0.20$, 1(b) for $\tau = 0.40$, 1(c) for $\tau = 0.60$ and 1(d) for $\tau = 0.80$) plots the simulated power function $p(\gamma)$ against γ for $n = 200$ (dashed line), $n = 500$ (solid line) and $n = 800$ (dashed-dotted line). When $\gamma = 0$, the specified alternative collapses into the null hypothesis and the power becomes the test size. Notice that for simplicity, the bandwidth suggested in Cai and Xu (2008) is used to compute the power although some sophisticated bandwidth selectors may be applicable. From Figure

1, it is clear that the power becomes larger and the test size is closer to the significance level of 5% when the sample size increases. Also, the power and size are almost the same for all four given quantiles though the performance for the middle two quantiles ($\tau = 0.40$ and $\tau = 0.60$) is a slightly better than that for the tailed two quantiles ($\tau = 0.20$ and $\tau = 0.80$). This indicates that the simulation results are indeed along with the line of the asymptotic theory given in Theorem 3. In particular, when $n = 800$, the empirical size is 0.051 for $\tau = 0.20$, 0.052 for $\tau = 0.40$, 0.049 for $\tau = 0.60$, and 0.056 for $\tau = 0.08$, and they are very close to the significant level of 5%. The power function shows that our test is indeed powerful. To appreciate why, consider the specific alternative with $\gamma = 0.3$. The functions $\alpha_2(\cdot)$ under H_0 and H_1 are shown in Figure 2. The null hypothesis is essentially the constant curve in Figure 2. Even with a small difference under our noise level, when $n = 800$, we can correctly detect the alternative over 80% (80.4% for $\tau = 0.20$, 88.3% for $\tau = 0.40$, 86.0% for $\tau = 0.60$ and 80.5% for $\tau = 0.60$ respectively) among 1000 simulations. The power increases rapidly to 1 when $\gamma = 0.5$ for $\tau = 0.40$ and $\tau = 0.60$ and when $\gamma = 0.6$ for $\tau = 0.20$ and $\tau = 0.80$.

7 Conclusion

We study quantile regression with partially varying coefficients. The proposed partially varying coefficient quantile regression model serves as an intermediate model between the fully nonparametric functional-coefficient model and the dynamic linear quantile regression model. Such a model provides a trade-off on robustness and precision, and suffers less from the so-called "curse of dimensionality" problem comparing to purely nonparametric models.

A simple and easily implemented three stage semiparametric procedure is proposed. In particular, we construct an estimator for the parameter component based on averaging (or weighted averaging) preliminary nonparametric functional coefficient estimates. The parametric estimators are root-n consistent and the estimation of the functional coefficients is oracle. The proposed estimators are asymptotically normal, which facilitates inference on the functional form of the coefficients. Our Monte Carlo experiment indicates that efficiency gain can be achieved when appropriately using information about the partial linear structure.

Important and interesting further studies can be conducted on inference problems based on the proposed partially varying coefficient quantile regression model. As remarked in Section 4, one may consider inference problems based on the stochastic process $\widetilde{\alpha}(u)$ in u, and one may also consider inference procedures of other forms. In addition, other types of inference problems can be studied. We hope to explore these issues in subsequent work.

Appendix: Proofs

Appendix A: Useful Lemmas

For convenience of the proof, we introduce some notations. Define $\xi_t = (U_t, X_t, Y_t)$, $B(\xi_t) = f_u(U_t)\,\Omega^*(U_t)$, $M(\xi_t) = X_t$, $K_h(u_0, \xi_t) = K_h(U_t - u_0)$, $\psi_\tau(x) = \tau - I(x \le 0)$, $\psi_\tau(u_0, \xi_j) = \tau - I\{Y_j \le \beta^T X_{j1} + \alpha(u_0)^T X_{j2}\} = \tau - I\{Y_j \le q_\tau(u_0, X_j)\}$, and $Z(u_0, \xi_j) = \psi_\tau(u_0, \xi_j) M(\xi_j) K_h(u_0, \xi_j)$. It is easy to show that $\eta_t = \psi_\tau(\xi_t, \xi_t)$, $E(\eta_t) = 0$, and $\mathrm{Var}(\eta_t) = \tau(1-\tau)$.

It follows from Theorem 1 in Cai and Xu (2008) that for any u_0,

$$\begin{aligned}
\sqrt{nh}\begin{pmatrix}\widehat{\beta}(u_0) - \beta(u_0)\\ \widehat{\alpha}(u_0) - \alpha(u_0)\end{pmatrix} &\approx \frac{1}{\sqrt{nh} f_u(u_0)}\,[\Omega^*(u_0)]^{-1}\sum_{j=1}^n \psi_\tau(\varepsilon_j) X_j K((U_j - u_0)/h)\\
&= \frac{h}{\sqrt{nh}}\sum_{j=1}^n B^{-1}(u_0) Z(u_0,\, \xi_j)\\
&\quad + \frac{h}{\sqrt{nh}}\sum_{j=1}^n B^{-1}(u_0)[\psi_\tau(\varepsilon_j) - \psi_\tau(u_0, \xi_j)] X_j K((U_j - u_0)/h),
\end{aligned}$$

where $\varepsilon_j = \varepsilon_{j\tau} = Y_j - \beta^T X_{j1} - [\alpha(u_0) + \alpha^{'}(u_0)(U_j - u_0)]^T X_{j2}$. In particular

$$\widehat{\beta}(u_0) - \beta(u_0) \approx \frac{1}{n}\sum_{j=1}^n e_1^T\, B^{-1}(u_0) Z(u_0,\, \xi_j) + B_n(u_0), \tag{25}$$

holds uniformly for all u_0 under Assumption B, where $a_n = (nh)^{-1/2}$ and

$$B_n(u_0) = \frac{1}{n}\sum_{j=1}^n e_1^T\, B^{-1}(u_0)[\psi_\tau(\varepsilon_j) - \psi_\tau(u_0, \xi_j)] X_j K((U_j - u_0)/h).$$

Indeed, one can show that (25) is true by applying Assumption B to the proofs of Lemmas 1 - 4 in Cai and Xu (2008). By using the leave-one-out method, the similar Bahadur representation for each design point U_i is

$$\widehat{\beta}(U_i) - \beta(U_i) \approx \frac{1}{n}\sum_{j \ne i}^n B^{-1}(\xi_i) Z(\xi_i,\, \xi_j) + B_n(U_i)$$

holds uniformly for all u_0. Thus,

$$\begin{aligned}
\widetilde{\beta}-\beta &= \frac{1}{n}\sum_{i=1}^{n} e_1^T\left[\widehat{\beta}(U_i)-\beta(U_i)\right] \\
&\approx \frac{2}{n^2}\sum_{1\le i<j\le n} e_1^T B^{-1}(\xi_i) Z(\xi_i,\, \xi_j)+B_n \\
&= \frac{1}{n^2}\sum_{1\le i<j\le n} [e_1^T B^{-1}(\xi_i) Z(\xi_i,\, \xi_j)+e_1^T B^{-1}(\xi_j) Z(\xi_j,\, \xi_i)]+B_n \\
&= \frac{n-1}{2n} V_n + B_n, \qquad (26)
\end{aligned}$$

where $T_n(\xi_i,\, \xi_j)=e_1^T B^{-1}(\xi_i) Z(\xi_i,\, \xi_j)+e_1^T B^{-1}(\xi_j) Z(\xi_j,\, \xi_i)$,

$$V_n=\frac{2}{n(n-1)}\sum_{1\le i<j\le n} T_n(\xi_i,\, \xi_j), \quad \text{and} \quad B_n=\frac{1}{n}\sum_{i=1}^{n} B_n(U_i).$$

Then, we show that V_n is a U-statistics with non-degenerate n -dependent kernel $T_n(\xi_i,\, \xi_j)$.

To derive the asymptotic properties for $\widetilde{\beta}$, we use the Hoeffding decomposition (see Lee (1990)) as follows. Let,

$$H_n^{(1)}=\frac{1}{n}\sum_{i=1}^{n} h_n^{(1)}(\xi_i), \quad \text{and} \quad H_n^{(2)}=\frac{2}{n(n-1)}\sum_{1\le i<j\le n} h_n^{(2)}(\xi_i,\xi_j),$$

where the kernels in the statistics $H_n^{(1)}$ and $H_n^{(2)}$ are defined respectively by

$$h_n^{(1)}(v)=E[T_n(v,\xi_j)]-\gamma_n, \quad \text{and} \quad h_n^{(2)}(v,w)=T_n(v,w)-E[T_n(v,\xi_j)]-E[T_n(\xi_i,w)]+\gamma_n$$

with $F(\cdot)$ being the distribution of ξ_i,

$$\gamma_n=\int\int T_n(\xi_i,\xi_j)dF(\xi_i)dF(\xi_j)\equiv E^{\otimes} T_n(\xi_i,\xi_j)$$

and $E^{\otimes}$ denoting the expectation with respect to the measure $P^{\xi_i}\otimes P^{\xi_j}$. Then,

$$V_n=\gamma_n+2\,H_n^{(1)}+H_n^{(2)}.$$

To establish the consistency and asymptotic normality of the proposed estimator, we use Theorem 2 in Dette and Spreckelsen (2004). To do so, we need to check the conditions in Theorem 2 of Dette and Spreckelsen (2004), which are provided by the following lemmas.

Lemma 1: Under Assumptions A and B1 - B2,

(1) $\gamma_n = h^2\mu_2(2\,B_1^* - B_2^*) + o(h^2)$;

(2) $h_n^{(1)}(v) = e_1^T B^{-1}(v)\psi_\tau(v,v)M(v)f_u(v) + o(h)$;

(3) $h_n^{(2)}(v,w) = T_n(v,w) - e_1^T B^{-1}(v)\psi_\tau(v,v)M(v)f_u(v) - e_1^T B^{-1}(w)\psi_\tau(w,w)M(w)f_u(w)$ $+o(h)$, where $f_u(\cdot)$ is the density distribution of U_1.

Lemma 2: Under Assumptions A and B1 - B2,

(1) $E[h_n^{(1)}(\xi_i)] = 0$;

(2) $\mathrm{Var}(h_n^{(1)}(\xi_i)) = \Sigma_{\beta,0} + o(1)$;

(3) $\mathrm{Cov}(h_n^{(1)}(\xi_1),\, h_n^{(1)}(\xi_{s+1})) = \mathrm{Cov}(W_1, W_{s+1}) \le C\,\beta(s)$ for some $C > 0$, where $W_s = e_1^T[\Omega^*(U_s)]^{-1}X_s\,\eta_s$.

Lemma 3: Under Assumptions A and B1 - B2,

(1) $E[H_n^{(1)}] = 0$;

(2) $n\,\mathrm{Var}(H_n^{(1)}) = \Sigma_\beta + o(1)$;

(3) $E|h_n^{(1)}(\xi_i)|^4 = O(1)$;

(4) $E|h_n^{(2)}(\xi_i,\xi_j)|^2 = O(h^{-2})$.

Lemma 4: Under Assumptions A and B1 - B2, we have

$$B_n = h^2\mu_2(-B_1^* + B_2^*/2) + o_p(h^2),$$

where B_1^* and B_2^* are given before Theorem 1.

The detailed proofs of the above lemmas and the following theorem are relegated to Appendix B.

Appendix B: Proofs of Lemmas and Theorems

Proof of Lemma 1: It is easy to see by the Taylor expansion that for U_j close to u_0,

$$
\begin{aligned}
& E\left[\psi_\tau(u_0, \xi_j) | X_j, U_j\right] \\
= \;& \tau - F_{y|u,x}(q_\tau(U_j, X_j) - X_{j2}^T(\alpha(U_j) - \alpha(u_0)) \\
\approx \;& f_{y|u,x}(q_\tau(U_j, X_j)) X_{j2}^T(\alpha(U_j) - \alpha(u_0)) - \frac{1}{2} f'_{y|u,x}(q_\tau(U_j, X_j)) \left[X_{j2}^T(\alpha(U_j) - \alpha(u_0))\right]^2 \\
= \;& f_{y|u,x}(q_\tau(U_j, X_j)) X_j^T \begin{pmatrix} 0 \\ \alpha(U_j) - \alpha(u_0) \end{pmatrix} - \frac{1}{2} f'_{y|u,x}(q_\tau(U_j, X_j)) \left([\alpha(U_j) - \alpha(u_0)]^T X_{j2}\right)^2,
\end{aligned}
$$

which implies that

$$
\begin{aligned}
& E\left[Z(u_0, \xi_j)\right] \\
= \;& E\left[X_j\{\tau - F_{y|u,x}(q_\tau(U_j, X_j) - X_{j2}^T(\alpha(U_j) - \alpha(u_0))\} K_h(U_j - u_0)\right] \\
\approx \;& E\left[f_{y|u,x}(q_\tau(U_j, X_j)) X_j X_j^T \begin{pmatrix} 0 \\ \alpha(U_j) - \alpha(u_0) \end{pmatrix} K_h(U_j - u_0)\right] \\
& -\frac{1}{2} E\left\{f'_{y|u,x}(q_\tau(U_j, X_j)) X_j \left([\alpha(U_j) - \alpha(u_0)]^T X_{j2}\right)^2 K_h(U_j - u_0)\right\} \\
\approx \;& \frac{h^2 \mu_2}{2} \Omega^*(u_0) f_u(u_0) \begin{pmatrix} 0 \\ \alpha_1(u_0) \end{pmatrix} + h^2 \mu_2 \Omega^{*\prime}(u_0) f_u(u_0) \begin{pmatrix} 0 \\ \alpha'(u_0) \end{pmatrix} - \frac{h^2 \mu_2}{2} \Gamma(u_0) f_u(u_0) \\
= \;& \frac{h^2 \mu_2}{2} B(u_0) \begin{pmatrix} 0 \\ \alpha_1(u_0) \end{pmatrix} + h^2 \mu_2 \Omega^{*\prime}(u_0) f_u(u_0) \begin{pmatrix} 0 \\ \alpha'(u_0) \end{pmatrix} - \frac{h^2 \mu_2}{2} \Gamma(u_0) f_u(u_0), \qquad (27)
\end{aligned}
$$

where $\alpha_1(u_0) = \alpha''(u_0) + 2\alpha'(u_0) f_u'(u_0)/f_u(u_0)$. It follows from the definition of $T_n(\xi_i,\, \xi_j)$ and (27) that

$$
\begin{aligned}
\gamma_n = \;& E^{\otimes} T_n(\xi_i, \xi_j) \\
= \;& \int\int [e_1^T B^{-1}(\xi_i) Z(\xi_i,\, \xi_j) + e_1^T B^{-1}(\xi_j) Z(\xi_j,\, \xi_i)] dF(\xi_i) dF(\xi_j) \\
= \;& 2 e_1^T \int\int B^{-1}(\xi_i) Z(\xi_i,\, \xi_j) dF(\xi_i) dF(\xi_j) \\
= \;& 2 e_1^T \int \frac{\mu_2 h^2}{2} \begin{pmatrix} 0 \\ \alpha_1(\xi_i) \end{pmatrix} dF(\xi_i) + 2\, h^2 \mu_2 B_1^* - h^2\, \mu_2 B_2^* + o(h^2), \\
= \;& h^2 \mu_2 (2\, B_1^* - B_2^*) + o(h^2),
\end{aligned}
$$

and

$$\begin{aligned}
E[T_n(v,\xi_j)] &= E\left[e_1^T B^{-1}(v) Z(v,\xi_j)\right] + E\left[e_1^T B^{-1}(\xi_j) Z(\xi_j, v)\right] \\
&= E e_1^T B^{-1}(\xi_j) Z(\xi_j, v) + o(h^2) \\
&= E\left[e_1^T B^{-1}(\xi_j) \psi_\tau(\xi_j, v) M(v) K_h(\xi_j, v)\right] + o(h^2) \\
&= E\left[e_1^T B^{-1}(\xi_j) \psi_\tau(\xi_j, v) K_h(v - U_j) M(v)\right] + o(h^2) \\
&= e_1^T B^{-1}(v) \psi_\tau(v, v) M(v) f_u(v) + o(h).
\end{aligned}$$

The lemma is established. $\square$

Proof of Lemma 2: It is easy to see that $E\left[h_n^{(1)}(\xi_i)\right] = 0$ holds. Similar to the proofs of Lemma 4 and Theorem 1 in Cai and Xu (2008), one has

$$\begin{aligned}
\mathrm{Var}(h_n^{(1)}(\xi_i)) &= E\left[e^T B^{-1}(\xi_i)\eta_i\, M(\xi_i) f(\xi_i)\right]^2 + o(h^2) \\
&= E\left[e_1^T (\Omega^*(U_i))^{-1} X_i X_i^T (\Omega^*(U_i))^{-1} e_1 \eta_i^2\right] + o(h^2) \\
&= E\left\{e_1^T (\Omega^*(U_i))^{-1} X_i X_i^T (\Omega^*(U_i))^{-1} e_1 E[\eta_i^2 | U_i,\, X_i]\right\} + o(h^2) \\
&= \Sigma_{\beta,0} + o(h^2).
\end{aligned}$$

Further, one has

$$\begin{aligned}
&\mathrm{Cov}(h_n^{(1)}(\xi_1),\, h_n^{(1)}(\xi_{s+1})) = E\left[h_n^{(1)}(\xi_1) h_n^{(1)}(\xi_{s+1})\right] \\
= \;& E\left[e_1^T (\Omega^*(U_1))^{-1} X_1 X_{s+1}^T (\Omega^*(U_{s+1}))^{-1} e_1 \eta_1\, \eta_{s+1}\right] + o(h^2) \\
= \;& \mathrm{Cov}(W_1, W_{s+1}) + o(1) \le C\,\beta(s).
\end{aligned}$$

This proves the lemma. $\square$

Proof of Lemma 3: The first assertion follows easily by Lemma 2. For the second result, similar to the method used in the proofs of Lemma 4 and Theorem 1 in Cai and Xu (2008), it follows from Lemma 2 that

$$\begin{aligned}
n\,\mathrm{Var}\left(H_n^{(1)}\right) &= \tau(1-\tau) E\left[e_1^T (\Omega^*(U_i))^{-1} \Omega(U_i) (\Omega^*(U_i))^{-1} e_1\right] \\
&\quad + 2\sum_{s=1}^{n-1} (1 - \frac{s}{n}) \mathrm{Cov}(h_n^{(1)}(\xi_1), h_n^{(1)}(\xi_{s+1})) \\
&= \Sigma_\beta + o(1).
\end{aligned}$$

Thirdly, by Lemma 1, it can be easily shown that

$$\begin{aligned}
E|h_n^{(1)}(\xi_i)|^4 &\le C\, E|e_1^T (\Omega^*(U_i))^{-1} \psi_\tau(\xi_i, \xi_i) X_i|^4 \\
&\le C\, E|e_1^T (\Omega^*(U_i))^{-1} X_i X_i^T (\Omega^*(U_i))^{-1} e_1|^2 \le C.
\end{aligned}$$

Finally, by Lemma 1 again, one has

$$
\begin{aligned}
E|h_n^{(2)}(\xi_i,\xi_j)|^2 &\le CE|T_n(\xi_i,\,\xi_j) - e_1^T B^{-1}(\xi_i)\psi_\tau(\xi_i,\xi_i)M(\xi_i)f(\xi_i) \\
&\quad - e_1^T B^{-1}(\xi_j)\psi_\tau(\xi_j)M(\xi_j,\xi_j)f(\xi_j)|^2 \\
&\le C\{E|T_n(\xi_i,\,\xi_j)|^2 + E[e_1^T B^{-1}(\xi_i)\psi_\tau(\xi_i,\xi_i)M(\xi_i)f(\xi_i)]^2 \\
&\quad + E[e_1^T B^{-1}(\xi_j)\psi_\tau(\xi_j)M(\xi_j,\xi_j)f(\xi_j)]^2\} \\
&\le C\{E|T_n(\xi_i,\,\xi_j)|^2 + E[e_1^T(\Omega^*(U_i))^{-1}\psi_\tau(\xi_i,\xi_i)X_i]^2 \\
&\quad + E[e_1^T(\Omega^*(U_j))^{-1}\psi_\tau(\xi_j,\xi_j)X_j]^2\} \\
&\le C\{E|T_n(\xi_i,\,\xi_j)|^2 + 2E[e_1^T(\Omega^*(U_i))^{-1}\psi_\tau(\xi_i,\xi_i)X_i]^2\} \\
&\le CE|T_n(\xi_i,\,\xi_j)|^2 + C_1,
\end{aligned}
$$

and

$$
\begin{aligned}
E|T_n(\xi_i,\,\xi_j)|^2 &= E\{e_1^T B^{-1}(\xi_i)Z(\xi_i,\,\xi_j) + e_1^T B^{-1}(\xi_j)Z(\xi_j,\,\xi_i)\}^2 \\
&\le C\;E|e_1^T B^{-1}(\xi_i)Z(\xi_i,\,\xi_j)|^2 \\
&\le CE\left[e_1^T B^{-1}(\xi_i)Z(\xi_i,\,\xi_j)Z^T(\xi_i,\,\xi_j)B^{-1}(\xi_i)e_1\right] \\
&= C\;e_1^T E\left\{E[B^{-1}(\xi_i)Z(\xi_i,\,\xi_j)Z^T(\xi_i,\,\xi_j)B^{-1}(\xi_i)|\xi_i]\right\}e_1. \\
&= O(h^{-1}).
\end{aligned}
$$

The last inequality is by Lemma 4 and Theorem 1 in Cai and Xu (2008). The proof of the lemma is complete. □

Proof of Lemma 4: Similar to the proofs in Lemma 1, we have, for U_j close to u_0,

$$
\begin{aligned}
&E\left[\{\psi_\tau(\varepsilon_j) - \psi_\tau(u_0,\xi_j)\}|X_j,U_j\right] \\
=\;& F_{y|u,x}(q_\tau(U_j,X_j) - X_{j2}^T\alpha(u_0)(U_j-u_0)) \\
&- F_{y|u,x}(q_\tau(U_j,X_j) - X_{j2}^T(\alpha(U_j) - \alpha(u_0) - \alpha'(u_0)(U_j-u_0))) \\
\approx\;& -f_{y|u,x}(q_\tau(U_j,X_j))X_{j2}^T\alpha'(u_0)(U_j-u_0) + \frac{1}{2}f'_{y|u,x}(q_\tau(U_j,X_j))\left[X_{j2}^T\alpha'(u_0)(U_j-u_0)\right]^2 \\
=\;& -f_{y|u,x}(q_\tau(U_j,X_j))X_j^T\begin{pmatrix} 0 \\ \alpha'(u_0)(U_j-u_0)\end{pmatrix} + \frac{1}{2}f'_{y|u,x}(q_\tau(U_j,X_j))\left(\alpha'(u_0)^T X_{j2}(U_j-u_0)\right)^2,
\end{aligned}
$$

which implies that

$$
\begin{aligned}
& E\left[\{\psi_\tau(\varepsilon_j) - \psi_\tau(u_0, \xi_j)\} X_j K_h(U_j - u_0)\right] \\
= \;& E\left[E\left[\{\psi_\tau(\varepsilon_j) - \psi_\tau(u_0, \xi_j)\} | X_j, U_j\right] X_j K_h(U_j - u_0)\right] \\
\approx \;& -E\left[f_{y|u,x}(q_\tau(U_j, X_j)) X_j X_j^T \begin{pmatrix} 0 \\ \alpha'(u_0)(U_j - u_0) \end{pmatrix} K_h(U_j - u_0)\right] \\
& + \frac{1}{2} E\left\{f'_{y|u,x}(q_\tau(U_j, X_j)) X_j \left(\alpha'(u_0)^T X_{j2}(U_j - u_0)\right)^2 K_h(U_j - u_0)\right\} \\
\approx \;& -h^2 \mu_2 \left[\Omega^*(u_0) f'_u(u_0) + \Omega^{*\prime}(u_0) f_u(u_0)\right] \begin{pmatrix} 0 \\ \alpha'(u_0) \end{pmatrix} + \frac{h^2 \mu_2}{2} \Gamma(u_0) f_u(u_0) \\
= \;& -h^2 \mu_2 \left[B(u_0) f'_u(u_0)/f_u(u_0) + \Omega^{*\prime}(u_0) f_u(u_0)\right] \begin{pmatrix} 0 \\ \alpha'(u_0) \end{pmatrix} + \frac{h^2 \mu_2}{2} \Gamma(u_0) f_u(u_0),
\end{aligned}
$$

so that

$$E[B_n(U_1)] = h^2 \mu_2(-B_1^* + B_2^*/2) + o(h^2).$$

Similarly, we can show that $\mathrm{Var}(B_n) = o(h^4)$. Therefore, $B_n = E[B_n(U_1)] + o_p(h^2) = h^2 \mu_2(-B_1^* + B_2^*/2) + o_p(h^2)$. This proves the lemma. □

Now we embark on the proof of Theorem 1 based on Lemmas 1 - 4.

Proof of Theorem 1: It suffices to check that the assumptions of Theorem 2 in Dette and Spreckelsen (2004) are satisfied for the kernel $T_n(\xi_i, \xi_j)$. Condition II of Theorem 2 in Dette and Spreckelsen (2004) is obviously satisfied by Lemmas 2 and 3. Thus, one only needs to check Condition I of Theorem 2 in Dette and Spreckelsen (2004). To this end, for $1 < \eta < 2/(1+\delta)$, ζ is chosen to satisfy $1/\zeta + 1/\eta = 1$. Then, by the Hölder's inequality, for $(i, j) \neq (k, l)$,

$$E|T_n(\xi_i, \xi_j) T_n(\xi_i, \xi_j)|^{1+\delta} \leq [E|T_n(\xi_i, \xi_j)|^{\zeta(1+\delta)}]^{\frac{1}{\zeta}} [E|T_n(\xi_i, \xi_j)|^{\eta(1+\delta)}]^{\frac{1}{\eta}}.$$

It follows by the C_r-inequality that

$$
\begin{aligned}
& E|T_n(\xi_i, \xi_j)|^{\zeta(1+\epsilon)} \\
= \;& E|e_1^T B^{-1}(\xi_i) Z(\xi_i, \xi_j) + e_1^T B^{-1}(\xi_j) Z(\xi_j, \xi_i)|^{\zeta(1+\delta)} \\
\leq \;& C E|e_1^T B^{-1}(\xi_i) Z(\xi_i, \xi_j)|^{\zeta(1+\delta)} + E|e_1^T B^{-1}(\xi_j) Z(\xi_j, \xi_i)|^{\zeta(1+\delta)} \\
\leq \;& C E|e_1^T B^{-1}(\xi_i) Z(\xi_i, \xi_j)|^{\zeta(1+\delta)} \\
= \;& C E|e_1^T B^{-1}(U_i) \psi_\tau(\xi_i, \xi_j) X_j K_h(U_j - U_i)|^{\zeta(1+\delta)} \\
= \;& O(h^{-\zeta(1+\delta)}).
\end{aligned}
$$

Similarly, $E|T_n(\xi_i,\xi_j)|^{\eta(1+\delta)} = O(h^{-\eta(1+\delta)})$. Thus, it follows that

$$\sup_{i\neq j,k\neq l,j\neq l} E|T_n(\xi_i,\xi_j)T_n(\xi_k,\xi_l)|^{1+\epsilon} = O(h^{-2(1+\delta)}).$$

For other cases, by the same token, one obtains

$$\sup_{i\neq j,k\neq l,j\neq l} E^{1\otimes}|T_n(\xi_i,\xi_j)T_n(\xi_k,\xi_l)|^{1+\epsilon} = O(h^{-2(1+\delta)}),$$

$$\sup_{i\neq j,k\neq l,j\neq l} E^{3\otimes}|T_n(\xi_i,\xi_j)T_n(\xi_k,\xi_l)|^{1+\epsilon} = O(h^{-2(1+\delta)}),$$

and

$$\sup_{i\neq j,i\neq l,j\neq l} E^{2\otimes}|T_n(\xi_i,\xi_j)T_n(\xi_i,\xi_l)|^{1+\epsilon}\} = O(h^{-d(1+\delta)}).$$

Therefore, $C_n = O(h^{-2(1+\epsilon)})$ so that Condition I of Theorem 2 in Dette and Spreckelsen (2004) is satisfied. Thus, by Lemma 2, one has

$$\frac{V_n - \gamma_n}{\sqrt{\frac{4}{n}\mathrm{Var}(h_n^{(1)}(\xi_1))}} \xrightarrow{d} \mathcal{N}(0,1).$$

By (26) and Assumption B2, finally, it shows the asymptotic normality:

$$\sqrt{n}\,\left[\widetilde{\beta} - \beta - \gamma_n - B_n\right] \xrightarrow{d} \mathcal{N}(0,\ \Sigma_\beta).$$

This, in conjunction with Lemmas 1 and 4, completes the proof of the theorem. □

Proof of Theorem 2: For a given $\sqrt{n}$-consistent estimator $\widehat{\beta}_*$ of β, similar to the proof of Theorem 1 in Cai and Xu (2008), we can show that

$$\sqrt{n\,h}\,[\widetilde{\alpha}(u_0) - \alpha(u_0)] = \frac{[\Omega_{22}^*(u_0)]^{-1}}{\sqrt{n\,h}\,f_u(u_0)}\sum_{t=1}^n \psi_\tau(Y_t^*)\,X_{t2}\,K(U_{th}) + o_p(1), \tag{28}$$

where $Y_t^* = Y_{t*} - X_{t2}^T[\alpha(u_0) + \alpha'(u_0)(U_t - u_0)]$ and $U_{th} = (U_t - u_0)/h$. From (28),

$$\begin{aligned}\sqrt{n\,h}\,[\widetilde{\alpha}(u_0) - \alpha(u_0)] \quad &\approx \quad \frac{[\Omega_{22}^*(u_0)]^{-1}}{\sqrt{n\,h}\,f_u(u_0)}\sum_{t=1}^n [\psi_\tau(Y_t^*) - \eta_t]\,X_{t2}\,K(U_{th}) \\ &\quad + \frac{[\Omega_{22}^*(u_0)]^{-1}}{\sqrt{n\,h}\,f_u(u_0)}\sum_{t=1}^n \eta_t\,X_{t2}\,K(U_{th}) \equiv A_{1n} + A_{2n},\end{aligned}$$

the definitions of $A_{jn} = A_{jn}(u_0)$ ($j = 1$ and 2) should be apparent from the context. Similar to the proof of Theorem 2 in Cai, Fan and Yao (2000) or Theorem 1 in Cai

(2002a), by using the small-block and large-block technique and the Cramér-Wold device, one can show (although lengthy and tediously) that

$$A_{2n} \xrightarrow{d} \mathcal{N}(0,\ \Sigma_\alpha). \tag{29}$$

By the stationarity and Lemma 4 in Cai and Xu (2008),

$$E[A_{1n}] = \frac{[\Omega_{22}^*(u_0)]^{-1}}{\sqrt{nh} f_u(u_0)} \, n \, E[\{\psi_\tau(Y_t^*) - \eta_t\} X_{t2} K(U_{th})] = a_n^{-1} \frac{h^2}{2} \alpha''(u_0) \mu_2 \{1 + o(1)\}. \tag{30}$$

Since $\psi_\tau(Y_t^*) - \eta_t = I(Y_t \le c_{1t}) - I(Y_t \le c_{2t})$, where $c_{1t} = \beta^T X_{t1} + \alpha(U_t)^T X_{t2}$ and $c_{2t} = \widehat{\beta}_*^T X_{t1} + [\alpha(u_0) + \alpha'(u_0)(U_t - u_0)]^T X_{t2}$, then, $[\psi_\tau(Y_t^*) - \varepsilon_t]^2 = I(d_{1t} < Y_t \le d_{2t})$, where $d_{1t} = \min(c_{1t}, c_{2t})$ and $d_{2t} = \max(c_{1t}, c_{2t})$. Further,

$$\begin{aligned} E\left[\{\psi_\tau(Y_t^*) - \eta_t\}^2 K^2(U_{th}) X_{t2} X_{t2}^T\right] &= E\left[\{F_{y|u,x}(d_{2t}) - F_{y|u,x}(d_{1t})\} K^2(U_{th}) X_{t2} X_{t2}^T\right] \\ &= O(h^3). \end{aligned}$$

Thus, $\mathrm{Var}(A_{1n}) = o(1)$. This, in conjunction with (29) and (30) and the Slutsky Theorem, proves the theorem. □

Proof of Theorem 3: It is clear from (29) that to establish the theorem, it suffices to show the following, for any $u_i^* \neq u_j^*$,

$$\mathrm{Cov}(A_{2n}(u_i^*), A_{2n}(u_j^*)) \ \to\ 0. \tag{31}$$

To this end, we define, for ant t and s,

$$M(U_t, U_s) = E[\eta_t \eta_s X_{t2} X_{s2} | U_t, U_s].$$

Then, it is easy to show that

$$\begin{aligned} & E[\eta_t \, X_{t2} \, K((U_t - u_i^*)/h) \eta_s \, X_{s2} \, K((U_s - u_j^*)/h))] \\ = \ & E[M(U_t, U_s) K((U_t - u_i^*)/h) K((U_t - u_i^*)/h)] = O(h^2). \end{aligned}$$

Thus, similar to the proof of Lemma A.1 in Cai, Fan and Yao (2000), we can show easily that (31) holds. This proves the theorem. □

References

Ahmad, I., Leelahanon, S., Li, Q., 2005. Efficient estimation of a semiparametric partially linear varying coefficient mode. The Annals of Statistics 33, 258-283.

Cai, Z., 2002a. Regression quantile for time series. Econometric Theory 18, 169-192.

Cai, Z., 2002b. Two-step likelihood estimation procedure for varying-coefficient models. Journal of Multivariate Analysis 82, 189-209.

Cai, Z., Fan, J., 2000. Average regression surface for dependent data. Journal of Multivariate Analysis 75,112-142.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941-956.

Cai, Z., Gu, J., Li, Q., 2009. Recent developments in nonparametric econometrics. Advances in Econometrics 25, 495-549.

Cai, Z., Masry, E., 2000. Nonparametric estimation of additive nonlinear ARX time series: Local linear fitting and projection. Econometric Theory 16, 465-501.

Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. Journal of the American Statistical Association 103, 1596-1608.

Chaudhuri, P., 1991. Nonparametric estimates of regression quantiles and their local Bahadur representation. The Annals of Statistics 19, 760-777.

Chaudhuri, P., Doksum, K., Samarov, A., 1997. On average derivative quantile regression. The Annuals of Statistics 25, 715-744.

De Gooijer, J., Zerom, D., 2003. On additive conditional quantiles with high dimensional covariates. Journal of the American Statistical Association 98, 135-146.

Dette, H., Spreckelsen, I., 2004. Some comments on specification tests in nonparametric absolutely regular processes. Journal of Time Series Analysis 25, 159-172.

Doukhan, P., 1994. Mixing. Lecture Notes in Statistics, Vol. 85. Springer-Verlag, Heielberg.

Engle, R., Granger, C.W.J., Rice, R., Weiss, A., 1986. Nonparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association 81, 310-320.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and Its Applications. Chapman and Hall, London.

Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11, 1031-1057.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman and Hall, London.

He, X., Liang, H. 2000. Quantile regression estimates for a class of linear and partially linear errors-in-variables models. Statistica Sinica 10, 129-140.

He, X., Ng, P., 1999. Quantile splines with several covariates. Journal of Statistical Planning and Inference 75, 343-352.

He, X., Ng, P., Portnoy, S., 1998. Bivariate quantile smoothing splines. Journal of the Royal Statistical Society, Series B 60, 537-550.

He, X., Portnoy, S., 2000. Some asymptotic results on bivariate quantile splines. Journal of Statistical Planning and Inference 91, 341-349.

He, X., Shi, P., 1996. Bivariate tensor-product B-splines in a partly linear model. Journal of Multivariate Analysis 58, 162-181.

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.

Horowitz, J.L., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. Journal of the American Statistical Association 100, 1238-1249.

Khindanova, I.N., Rachev, S.T., 2000. Value at risk: Recent advances. Handbook on Analytic-Computational Methods in Applied Mathematics, CRC Press LLC, Boca Raton, FL.

Kim, M.-O., 2007. Quantile regression with varying coefficients. The Annals of Statistics 35, 92-108.

Koenker, R., 2004. Quantreg: An R Package for Quantile Regression and Related Methods. http://cran.r-project.org.

Koenker, R., 2005. Quantile Regression. Econometric Society Monograph Series, Cambridge University Press, New York.

Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33-50.

Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81, 673-680.

Koenker, R., Xiao, Z., 2006. Quantile autoregression. Journal of the American Statistical Association 101, 980-990.

Lee, A.J., 1990. U-Statistics: Theory and Practice. Marcel Dekker, New York.

Lee, S., 2003. Efficient semiparametric estimation of partially linear quantile regression model. Econometric Theory 19, 1-31.

Li, Q., Huang, C.J., Li, D., Fu, T., 2002. Semiparametric smooth coefficient model. Journal of Business & Economics Statistics 20, 412-422.

Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931-954.

Robinson, P.M., 1989. Hypothesis testing in semiparametric and nonparametric models for econometric time series. Review of Economic Studies 56, 511-534.

Ruppert, D., Wand, M., 1994. Multivariate locally least squares regression. The Annals of Statistics 22, 1346-1370.

Speckman, P., 1988. Kernel smoothing partial linear models. The Journal of Royal Statistical Society, Series B 50, 413-426.

Yu, K., Jones, M.C., 1998. Local linear quantile regression. Journal of the American Statistical Association 93, 228-237.

Yu, K., Lu, Z., 2004. Local linear additive quantile regression. Scandinavian Journal of Statistics 31, 333-346.

Zhang, W., Lee, S.Y., Song, X., 2002. Local polynomial fitting in semivarying coefficient model. Journal of Multivariate Analysis 82, 166-188.

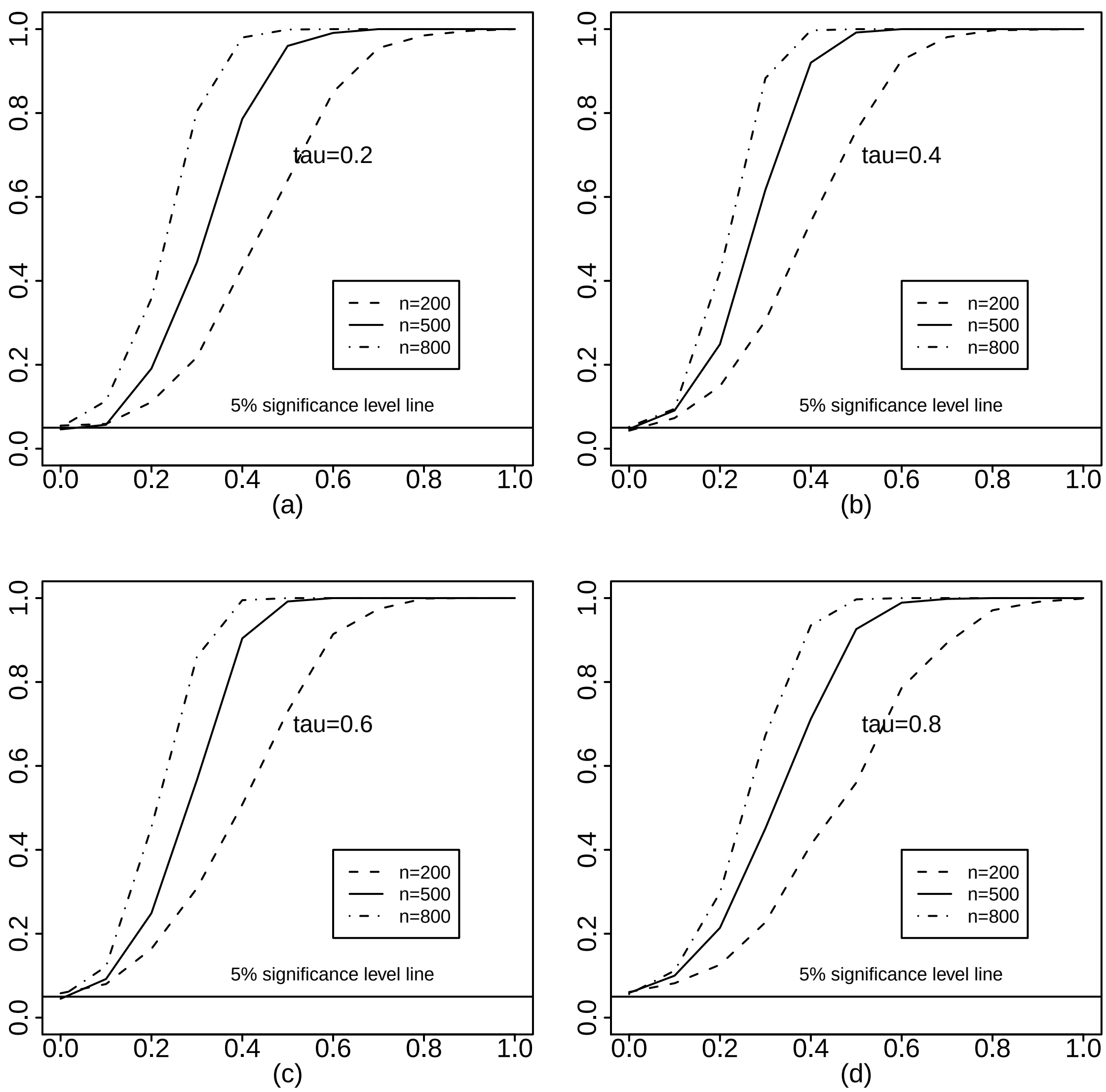

Figure 1: The plot of power curves against γ for the testing hypothesis. The dashed line is for $n = 200$, the solid line is for $n = 500$ and the dashed-dotted line is for $n = 800$. (a) $\tau = 0.20$; (b) $\tau = 0.40$; (c) $\tau = 0.60$; (d) $\tau = 0.80$.

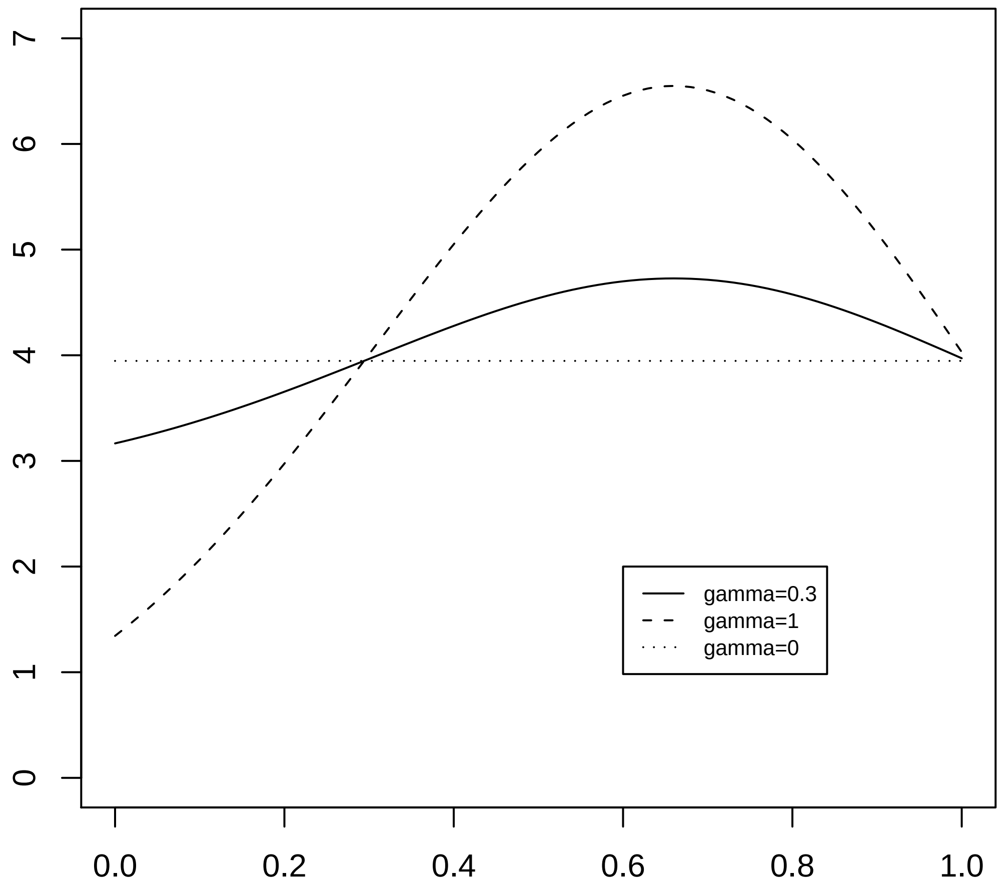

Figure 2: The coefficient function $\alpha_2(\cdot)$ under the null hypothesis (dotted line) with $\gamma = 0$ and specific alternative hypothesis with $\gamma = 0.3$ (solid line) and $\gamma = 1$ (dashed line).