Statistica Sinica **22** (2012), 1075-1101
doi:http://dx.doi.org/10.5705/ss.2010.220

NEW EFFICIENT AND ROBUST ESTIMATION IN VARYING-COEFFICIENT MODELS WITH HETEROSCEDASTICITY

Jie Guo[1], Maozai Tian[1] and Kai Zhu[2]

[1]*Renmin University of China and* [2]*Chinese Academy of Sciences*

Abstract: Varying-coefficient models with heteroscedasticity are considered in this paper. Based on local composite quantile regression, we propose a new estimation method to estimate the coefficient functions and heteroscedasticity simultaneously. Moreover, we can get the estimated conditional quantile curves of the error part. The conditional biases, variances, and asymptotic normalities of these estimators are studied explicitly. A simple and quick plug-in bandwidth selector is employed to select the optimal bandwidth. The estimators of the coefficient functions perform efficiently and robustly regardless of the error distributions. When the error ε follows a non-normal distribution, the proposed estimators of the coefficient functions are much more efficient than local polynomial weighted least squares estimators and almost as efficient for normal random errors. The estimator of heteroscedasticity also outperforms other classical estimators in the literature. A goodness-of-fit test based on a bootstrap procedure is proposed to test whether the coefficient functions are actually varying. Both simulations and data analysis are used to illustrate the proposed method.

Key words and phrases: Goodness-of-fit test, heteroscedasticity, local composite quantile regression, plug-in bandwidth selector, varying-coefficient models.

1. Introduction

Varying-coefficient models, see, Hastie and Tibshirani (1993), arise naturally when one wishes to examine how regression coefficients change over certain factors such as time. Their main appeal is that the modeling bias can significantly be reduced and the “curse of dimensionality” can be avoided. Heteroscedasticity, which occurs when the variance of the error varies across observations, exists in many cases, and varying-coefficient models with heteroscedasticity have been widely considered in the literature. Fan and Zhang (1999, 2000, 2008), among others, considered the model

$$Y_i = X_i^T \beta(T_i) + \sigma(T_i)\varepsilon_i, \tag{1.1}$$

where $\{(T_i, X_i, Y_i) : i = 1, \ldots, n\}$ is an i.i.d. random sample from (T, X, Y), $T_i \in \mathbb{R}$ is called smoothing variable, $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^T \in \mathbb{R}^p$ and $Y_i \in \mathbb{R}$,

$\beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_p(\cdot))^T$ is an $\mathbb{R}^p$-valued unknown smooth function, $\sigma(\cdot)$ is an unknown positive function of the smoothing variable T, the ε_i are i.i.d. and independent of (T_i, X_i), and $\mathrm{Var}(\varepsilon) = 1$.

There are many estimates of $\beta(\cdot)$ in model (1.1). Hastie and Tibshirani (1993) considered L_2 estimation and penalized least squares estimation. Fan and Zhang (1999) proposed a two-step local polynomial least squares estimation procedure. Chiang, Rice, and Wu (2001) proposed a componentwise smoothing spline method. Huang, Wu, and Zhou (2002) proposed a polynomial splines estimation method. All these methods are based on least squares, which suggests that error is homoscedastic and normal with finite variance. The works aside from Hastie and Tibshirani (1993) took the variance of the error to be an unknown function of the smoothing variable, but none considered the estimation of heteroscedasticity. Kim (2007) and Wang, Zhu, and Zhou (2009) considered varying-coefficient models in quantile regression using spline polynomials to approximate $\beta(\cdot)$.

Some papers consider the estimation of heteroscedasticity in (1.1), most of which are based on regression residuals. For example, Wu, Chiang, and Hoover (1998) and Fan and Zhang (2000) used the weighted residual sum of squares from the local polynomial least squares fit to estimate $\sigma(\cdot)$. Zhao (2001) proposed a k-NN method based on residuals. The estimator in Tian and Chan (2010) was also based on residuals. To our best knowledge, there is no literature considering the simultaneous estimation of $\beta(\cdot)$ and $\sigma(\cdot)$ in (1.1). The estimation accuracies of $\beta(\cdot)$ and $\sigma(\cdot)$ are relevant to each other, so it is meaningful to develop an estimation approach that can estimate $\beta(\cdot)$ and $\sigma(\cdot)$ simultaneously.

Quantile regression, proposed by Koenker and Bassett (1978), is a statistical technique designed to estimate and conduct inference about conditional quantile functions, a more complete statistical model than mean regression. Based on quantile regression, Zou and Yuan (2008) considered the linear model $Y = \sum_{j=1}^{p} X_j\beta_j + \varepsilon$, and proposed composite quantile regression (CQR) estimates of the coefficients. Let $\rho_{\tau_k}(s) = s(\tau_k - I(s < 0)), k = 1, 2, \ldots, q$, be q check loss functions at q quantile positions $\tau_k = k/(q+1)$. CQR estimates β by solving

$$(\hat{b}_1, \ldots, \hat{b}_q, \hat{\beta}^{CQR}) = \mathrm{argmin}_{b_1,\ldots,b_q,\beta} \sum_{k=1}^{q} \Big\{ \sum_{i=1}^{n} \rho_{\tau_k}(Y_i - b_k - X_i^T\beta) \Big\},$$

where b_k is the $100\tau_k\%$ quantile of ε. CQR can be more efficient and sometimes arbitrarily more efficient than least squares for non-normal random errors, and almost as efficient for normal random errors.

Based on CQR, Kai, Li, and Zou (2010) proposed local composite quantile regression (LCQR) smoothing that outperforms local polynomial regression for various non-normal random errors and is efficient for normal random errors. The theoretical properties and efficiency of LCQR motivate us to apply it to the estimation of model (1.1).

Here we propose a new estimation method called local Composite Quantile Regression with Averaged Quantile-Ratio estimation (CQR-AQR) that can estimate $\beta(\cdot)$ and $\sigma(\cdot)$ in model (1.1) simultaneously. We also get estimated conditional quantile curves of $\sigma(t)\varepsilon$ in (1.1) so as to give a thorough description of the error part. The local linear and quadratic CQR are employed to estimate $\beta(\cdot)$ and its derivative, respectively. To estimate $\sigma(\cdot)$, we propose an estimator called the Averaged Quantile-Ratio estimator (AQR), see (2.2) or (2.3). More generally, local m-polynomial CQR-AQR is studied. A plug-in method, which can be implemented easily, is employed to select the optimal bandwidth.

In a simulation study, we found the proposed CQR-AQR estimation to be very efficient and robust. By comparison with local polynomial weighted least squares estimation of $\beta(\cdot)$ in model (1.1), we find that local CQR estimation is more efficient and robust for non-normal distributed ε, especially for the Cauchy distribution, while it is almost as efficient for normal distributed ε. Thus the estimation of $\beta(\cdot)$ requires no specification of error distributions. By comparison with the triangular k-NN weight estimator (Zhao (2001)) of $\sigma(\cdot)$, the proposed AQR estimator is also more efficient and accurate.

One important inference question in model (1.1) is whether the coefficient functions are actually varying. A goodness-of-fit test procedure based on a resampling bootstrap is proposed. Simulation suggests that the test procedure is indeed powerful.

The rest of the paper is organized as follows. Section 2 and Section 3 study the local linear and quadratic CQR-AQR estimation of model (1.1), respectively. We discuss the plug-in bandwidth selector in Section 4. Section 5 presents a hypothesis testing procedure. Section 6 and Section 7 present the results of simulations and empirical study. In Section 8, we consider a more general estimation method for model (1.1), local m-polynomial CQR-AQR. Section 9 concludes the paper with discussion. Proofs are presented in the Appendix.

2. Local Linear CQR-AQR Estimation

2.1. Estimation

In model (1.1), if $\beta(\cdot)$ is second-order differentiable, it can be approximated locally as $\beta(T_i) \approx \beta(t) + \beta'(t)(T_i - t)$ for T_i in a neighborhood of any given grid point t. Similarly, $\sigma(T_i)$ can be approximated locally by a constant $\sigma(t)$. Let $a_{\tau_k}(t)$ denote the $100\tau_k\%$ quantile of $\sigma(t)\varepsilon$, for $k = 1, \ldots, q$, where $\tau_1, \tau_2, \ldots, \tau_q$ satisfy $0 < \tau_1 < \tau_2 < \cdots < \tau_q < 1$. Typically, we use equally spaced quantile positions: $\tau_k = k/(q+1)$ for $k = 1, \ldots, q$. The estimators $\hat{\beta}(t)$, $\hat{\beta}'(t)$ and $\hat{a}_{\tau_k}(t)$, for $k = 1, \ldots, q$, are obtained by minimizing

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\rho_{\tau_k}\left(Y_i - a_{\tau_k}(t) - X_i^T\left(\beta(t) + \beta'(t)(T_i - t)\right)\right)K\left(\frac{T_i - t}{h}\right) \tag{2.1}$$

with respect to $\beta(t)$, $\beta'(t)$, and $a_{\tau_k}(t)$ for $k = 1, \ldots, q$, where $K(\cdot)$ is a kernel function and h is a bandwidth.

For any grid point t, $\hat{a}_{\tau_k}(t)$ is the estimator of the $100\tau_k\%$ quantile of $\sigma(t)\varepsilon$, for $k = 1, \ldots, q$. Therefore, we can get q different estimated conditional quantile curves of $\sigma(t)\varepsilon$ at quantile positions τ_k, for $k = 1, \ldots, q$. The q different estimated conditional quantile curves reflect the properties of the error part $\sigma(t)\varepsilon$.

We denote the $100\tau_k\%$ quantile of ε by c_{τ_k}, for $k = 1, \ldots, q$. For brevity, we assume that the density function of ε is non-vanishing everywhere. Therefore c_{τ_k} is uniquely defined for any $0 < \tau_k < 1$. Obviously, $\sigma(t)$ can be easily estimated by $\hat{a}_{\tau_k}(t)/c_{\tau_k}$, which is called quantile-ratio estimator, for any $k = 1, \ldots, q$, as long as $c_{\tau_k} \neq 0$. However, based on the idea of CQR, we combine the strength of the quantile-ratio estimators ($\hat{a}_{\tau_k}(t)/c_{\tau_k}$ for $k = 1, \ldots, q$) and propose a new estimator of $\sigma(t)$ for any grid point t as follows.

(a) If $c_{\tau_k} \neq 0$ for any $k = 1, \ldots, q$, then

$$\hat{\sigma}(t) = \frac{1}{q}\sum_{k=1}^{q}\frac{\hat{a}_{\tau_k}(t)}{c_{\tau_k}}. \tag{2.2}$$

(b) If $c_{\tau_j} = 0$ for $j \in \{1, \ldots, q\}$, then

$$\hat{\sigma}(t) = \frac{1}{q-1}\sum_{\substack{k=1\\k\neq j}}^{q}\frac{\hat{a}_{\tau_k}(t)}{c_{\tau_k}}. \tag{2.3}$$

From (2.2), $\hat{\sigma}(t)$ is the average of q different Quantile-Ratio estimators. Thus, we call the estimator of $\sigma(t)$ presented in (2.2) or (2.3) the AQR, and the new estimation method the CQR-AQR. By minimizing (2.1) and employing the AQR estimator, we get simultaneous estimates of $\beta(t)$ and $\sigma(t)$.

Note that when we apply (2.2) or (2.3), the values of c_{τ_k} for $k = 1, \ldots, q$, have to be known. If the distribution of ε is known, it is easy to get $\hat{\sigma}(t)$ using (2.2) or (2.3). If it is unknown, we assume that ε is normal, and take $c_{\tau_k} = \Phi^{-1}(\tau_k)$.

2.2. Asymptotic properties of CQR estimator

In this subsection, we establish the asymptotic properties of $\hat{\beta}(t)$. Let $f_T(\cdot)$ denote the marginal density function of the covariate T and $f(\cdot)$ denote the density function of ε. The observed covariates vector is written as $\mathcal{D} = (T_1, \ldots, T_n, X_{11}, \ldots, X_{1n}, \ldots, X_{p1}, \ldots, X_{pn})^T$, and we impose the following regularity conditions.

1. $\beta(t)$ is $(m+1)$ times continuously differentiable in a neighborhood of t.
2. $\sigma(t)$ is positive and continuous.
3. $f_T(\cdot)$ and $f(\cdot)$ are positive continuous functions.
4. The kernel function $K(\cdot)$ is symmetric with a compact support.
5. $E(XX^T \mid T = t)$ is positive-definite and continuous in a neighborhood of any given t.
6. The bandwidth $h \to 0$, as $n \to \infty$ and $nh \to \infty$.

We use the notation

$$\mu_j = \int u^j K(u)du, \quad \nu_j = \int u^j K^2(u)du, \qquad j = 0, 1, 2, \ldots,$$

$$R_1(q) = \frac{\sum_{k=1}^q \sum_{k'=1}^q \tau_{kk'}}{(\sum_{k=1}^q f(c_{\tau_k}))^2},$$

$$R_2(q) = \frac{1}{q^2} \sum_{k=1}^q \sum_{k'=1}^q \frac{1}{c_{\tau_k} c_{\tau_{k'}}} \frac{\tau_{kk'}}{f(c_{\tau_k}) f(c_{\tau_{k'}})},$$

$$R_3(q) = \frac{1}{q^2} \sum_{k=1}^q \sum_{k'=1}^q \frac{1}{c_{\tau_k} c_{\tau_{k'}}} \frac{\sum_{k=1}^q \tau_{kk'}}{\sum_{k=1}^q f(c_{\tau_k}) f(c_{\tau_{k'}})},$$

where $\tau_{kk'} \equiv \min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}$. Let $\mathbf{c} = (1/q) \sum_{k=1}^q (1/c_{\tau_k})$, $\Xi = E(X^T \mid T = t)$, $\Sigma_X = \mathrm{Var}(X \mid T = t)$, $\Psi = E(XX^T \mid T = t)$, and $\Omega = \mathrm{E}(X^T \mid T = t) \left(\mathrm{Var}(X^T \mid T = t)\right)^{-1} \mathrm{E}(X \mid T = t)$.

Theorem 1. *Under Conditions* $1-6$*,*

$$bias\{\hat{\beta}(t) \mid \mathcal{D}\} = \frac{1}{2}\beta''(t)\mu_2 h^2 + o_p(h^2),$$

$$cov\{\hat{\beta}(t) \mid \mathcal{D}\} = \frac{\nu_0 \Sigma_X^{-1} \sigma^2(t)}{nhf_T(t)} R_1(q) + o_p(\frac{1}{nh}).$$

Furthermore,

$$\sqrt{nh} \left\{ \hat{\beta}(t) - \beta(t) - \frac{1}{2}\beta''(t)\mu_2 h^2 \right\} \xrightarrow{d} N\left(\mathbf{0}, \frac{\nu_0 \Sigma_X^{-1} \sigma^2(t)}{f_T(t)} R_1(q)\right),$$

where $\xrightarrow{d}$ *is convergence in distribution.*

By a simple calculation, we have

$$\mathrm{MSE}(\hat{\beta}(t)) = \frac{1}{4}\beta''^T(t)\Psi\beta''(t)\mu_2^2 h^4 + \frac{\nu_0\sigma^2(t)\mathrm{tr}\{\Psi\Sigma_X^{-1}\}}{nhf_T(t)}R_1(q),$$

and global optimal bandwidth for $\hat{\beta}(t)$, obtained by minimizing the mean integrated squared error, is

$$h_{opt} = \left(\frac{\nu_0 R_1(q)\mathrm{tr}\{\Psi\Sigma_X^{-1}\}}{\mu_2^2}\frac{\int\sigma^2(t)dt}{\int\beta''^T(t)\Psi\beta''(t)f_T(t)dt}\right)^{1/5} n^{-1/5} \sim n^{-1/5}. \tag{2.4}$$

2.3. Asymptotic properties of AQR estimator

In this subsection, we state the asymptotic properties of $\hat{\sigma}(t)$ at (2.2); the statistical properties of (2.3) are similar.

Theorem 2. *Under Conditions* $1-6$,

$$bias\{\hat{\sigma}(t) \mid \mathcal{D}\} = \frac{1}{2}\Xi\beta''(t)\mu_2 h^2\mathbf{c}(1-\Omega) + o_p(h^2)$$
$$Var\{\hat{\sigma}(t) \mid \mathcal{D}\} = \frac{1}{nh}\frac{\nu_0\sigma^2(t)}{f_T(t)}\left(R_2(q) - R_3(q)\Omega\right) + o_p\left(\frac{1}{nh}\right)$$

Furthermore,

$$\sqrt{nh}\left\{\hat{\sigma}(t)-\sigma(t)-\frac{1}{2}\Xi\beta''(t)\mu_2 h^2\mathbf{c}(1-\Omega)\right\} \xrightarrow{d} N\left(\mathbf{0}, \frac{\nu_0\sigma^2(t)}{f_T(t)}\left(R_2(q)-R_3(q)\Omega\right)\right).$$

3. Local Quadratic CQR-AQR Estimation

3.1. Estimation

In many situations we are interested in estimating the derivatives of the coefficient functions; local quadratic regression is often preferred, although we can get estimators by employing local linear regression. The estimation of $\beta'(t)$ can be improved minimizing

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\rho_{\tau_k}\left(Y_i-a_{\tau_k}(t)-X_i^T\left(\beta(t)+\beta'(t)(T_i-t)+\frac{1}{2}\beta''(t)(T_i-t)^2\right)\right)K\left(\frac{T_i-t}{h}\right) \tag{3.1}$$

with respect to $a_{\tau_k}(t)$, for $k=1,\ldots,q$, $\beta(t)$, $\beta'(t)$ and $\beta''(t)$.

3.2. Asymptotic properties

We state the properties of the improved estimate $\hat{\beta}'(t)$.

Theorem 3. *Under Conditions* $1-6$,

$$bias\{\hat{\beta}'(t) \mid \mathcal{D}\} = \frac{1}{6}\beta'''(t)h^2\frac{\mu_4}{\mu_2} + o_p(h^2),$$
$$cov\{\hat{\beta}'(t) \mid \mathcal{D}\} = \frac{1}{nh^3}\frac{\nu_2\Psi^{-1}\sigma^2(t)}{\mu_2^2 f_T(t)}R_1(q) + o_p\left(\frac{1}{nh^3}\right).$$

Furthermore,

$$\sqrt{nh^3}\left\{\hat{\beta}'(t) - \beta'(t) - \frac{1}{6}\beta'''(t)h^2\frac{\mu_4}{\mu_2}\right\} \xrightarrow{d} N\left(\mathbf{0}, \frac{\nu_2\Psi^{-1}\sigma^2(t)}{\mu_2^2 f_T(t)}R_1(q)\right).$$

4. Bandwidth Selection

Bandwidth selection is an important issue in local smoothing problems and there are existing techniques, such as plug-in bandwidth selector (Ruppert, Sheather, and Wand (1995)), and cross-validation (Wu, Chiang, and Hoover (1998)). In practice, leave-one-out cross-validation is quite computationally expensive, although it is natural and data-driven. We employ a plug-in method to select the optimal bandwidth.

To deal with (2.4), let

$$\Gamma_1 = \int \beta''^T(t)\Psi\beta''(t)f_T(t)dt, \quad \Gamma_2 = \int \sigma^2(t)dt.$$

The estimator $\hat{\beta}''(t)$ can be obtained by local cubic CQR fitting with an appropriate pilot bandwidth h_*, so a natural estimator of Γ_1 is

$$\hat{\Gamma}_1 = n_{grid}^{-1}\sum_{i=1}^{n_{grid}} \hat{\beta}''^T(t_i)\Psi\hat{\beta}''(t_i), \tag{4.1}$$

where $\{t_i : i = 1, \ldots, n_{grid}\}$ are grid points in the support of T. The estimators $\hat{a}_{\tau_k}(t)$, for $k = 1, \ldots, q$, can be obtained as byproducts when we use local cubic CQR fitting with a pilot bandwidth h_* to estimate $\hat{\beta}''(t)$. Then employing (2.2) or (2.3), we can obtain the estimator $\hat{\sigma}^2(t)$. The natural estimator of Γ_2 is

$$\hat{\Gamma}_2 = n_{grid}^{-1}\sum_{i=1}^{n_{grid}} \hat{\sigma}^2(t_i). \tag{4.2}$$

By replacing Γ_1 and Γ_2 in (2.4), respectively, with (4.1) and (4.2), we have the selected optimal bandwidth for estimating $\beta(t)$.

In the calculation of $\hat{\sigma}^2(t)$ and $R_1(q)$ at (2.4), we take ε to be normal if its true distribution is unknown.

5. Hypothesis Testing

In model (1.1), it is often of practical interest to test whether coefficient functions are actually varying. We consider the testing problem

$$H_0 : \beta_p(t) = \beta_p \leftrightarrow H_1 : \beta_p(t) \neq \beta_p, \tag{5.1}$$

where β_p is an unknown constant. Cai, Fan, and Yao (2000) and Huang, Wu, and Zhou (2002) developed goodness-of-fit tests of (5.1) based on the comparison of the residual sum of squares under the null and alternative models. In this paper, we propose a goodness-of-fit test based on the comparison of the residual sums of quantiles (RSQ) from local linear CQR fits under both the null hypothesis and the alternative. RSQ is an analog of residual sum of squares, see below. Under the null hypothesis, model (1.1) can be written as

$$Y_i = \sum_{j=1}^{p-1} X_{ij}^T \beta_j(T_i) + \beta_p X_{ip} + \sigma(T_i)\varepsilon_i. \tag{5.2}$$

Following Fan and Zhang (2000), we propose a method to estimate β_p in (5.2) under the null hypothesis. First, we ignore the fact that β_p is a constant, and treat it as an unknown function $\beta_p(t)$. Based on local linear CQR estimation, we obtain an estimator $\hat{\beta}_p(t)$. Each of $\{\hat{\beta}_p(T_i)\}$ is an estimator of the unknown parameter β_p under the null hypothesis, and we average them to obtain the estimator

$$\hat{\beta}_p = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_p(T_i).$$

The RSQ under H_0 is

$$RSQ_0 = \sum_{k=1}^{q}\sum_{i=1}^{n} \rho_{\tau_k}\Big(Y_i - \sum_{j=1}^{p-1} X_{ij}^T \hat{\beta}_j(T_i) - \hat{\beta}_p X_{ip} - \hat{a}_{\tau_k}(T_i)\Big).$$

The RSQ under H_1 is

$$RSQ_1 = \sum_{k=1}^{q}\sum_{i=1}^{n} \rho_{\tau_k}\Big(Y_i - \sum_{j=1}^{p} X_{ij}^T \hat{\beta}_j(T_i) - \hat{a}_{\tau_k}(T_i)\Big).$$

The goodness-of-fit test statistic is

$$\mathbf{Q}_n = \frac{RSQ_0 - RSQ_1}{RSQ_1} = \frac{RSQ_0}{RSQ_1} - 1, \tag{5.3}$$

and we reject H_0 for large value of $\mathbf{Q}_n$. Let

$$\hat{\eta}_i = Y_i - \sum_{j=1}^{p} X_{ij}^T \hat{\beta}_j(T_i)$$

and define

$$Y_i^* = \sum_{j=1}^{p-1} X_{ij}^T \hat{\beta}_j(T_i) + \hat{\beta}_p X_{ip} + \hat{\eta}_i.$$

We use a bootstrap procedure to evaluate the null distribution of $\mathbf{Q}_n$ and the p-values of the test.

Step 1. Resample n subjects with replacement from $\{(Y_i^*, X_i, T_i) : i = 1, \ldots, n\}$ and repeat the sampling procedure J times.

Step 2. From each bootstrap sample, calculate the test statistic $\mathbf{Q}_n^*$ and the empirical distribution of $\mathbf{Q}_n^*$ based on the J independent bootstrap samples.

Step 3. Reject the null hypothesis H_0 at level α if the observed test statistic $\mathbf{Q}_n$ is greater than the upper-α point of the empirical distribution of $\mathbf{Q}_n^*$.

The p-value of the test is the relative frequency of the event $\{\mathbf{Q}_n^* \geq \mathbf{Q}_n\}$ in the J replications of the bootstrap sampling. For simplicity, we use the same bandwidth in calculating $\mathbf{Q}_n^*$ as for $\mathbf{Q}_n$.

6. Simulation Study

In this section, we illustrate the performance of the proposed local CQR and AQR estimators through simulation studies. The proposed estimators were found using the majorization-minimization (MM) algorithm of Hunter and Lange (2000).

6.1. Examples

The model, as in Fan and Zhang (2000) is

$$Y = \beta_1(T)X_1 + \beta_2(T)X_2 + \sigma(T)\varepsilon, \tag{6.1}$$

where X_1 and X_2 are standard normal with correlation coefficient $2^{-1/2}$, T follows a uniform distribution on $[0,1]$. Further, T, ε and (X_1, X_2) are independent. The coefficient functions were $\beta_1(t) = \cos(2\pi t)$, $\beta_2(t) = 4t(1-t)$. To illustrate the robustness and efficiency of the proposed estimators, we considered five distributions of ε ($N(0,1)$, t-distribution with 3 degrees of freedom, Lognormal(0,1), mixed normal distribution $0.9N(0,1) + 0.1N(0,10^2)$ and Cauchy(0,1)) and three kinds of heteroscedasticity similar to the examples in Tian and Chan (2010):

Example 1. (*Homoscedasticity*)

$$\sigma(t) = \Big(0.2\mathrm{var}\{E(Y \mid U, X_1, X_2)\}\Big)^{1/2}.$$

Example 2. (*Small jumps heteroscedasticity*)

$$\sigma(t) = \begin{cases} 0.1, & 0 \le t \le 0.3, \\ 0.2, & 0.3 < t \le 0.6, \\ 0.1, & 0.6 < t \le 1. \end{cases}$$

Example 3. (*High frequency heteroscedasticity*)

$$\sigma(t) = \begin{cases} |50t^2 - 2|, & 0 \le t \le 0.3, \\ |3t|, & 0.3 < t \le 0.6, \\ |8t|, & 0.6 < t \le 1. \end{cases}$$

To assess the performance of local linear CQR estimator, we compared it with local linear least squares and local linear weighted least squares estimators solved, respectively, by minimizing

$$\sum_{i=1}^{n} \Big\{ Y_i - \sum_{j=1}^{2} X_{ij}^T \left(\beta_j(t) + \beta_j'(t)(T_i - t) \right) \Big\}^2 K\left(\frac{T_i - t}{h} \right), \tag{6.2}$$

$$\sum_{i=1}^{n} \frac{1}{\sigma^2(T_i)} \Big\{ Y_i - \sum_{j=1}^{2} X_{ij}^T \left(\beta_j(t) + \beta_j'(t)(T_i - t) \right) \Big\}^2 K\left(\frac{T_i - t}{h} \right). \tag{6.3}$$

In (6.3), we used the true value of $\sigma(T_i)$ as if the values of heteroscedasticity were known. We computed the ratio of averaged square errors

$$\mathrm{RASE1}(\hat{g}(t_j)) = \frac{\mathrm{ASE}(\hat{g}^{LS}(t_j))}{\mathrm{ASE}(\hat{g}^{CQR}(t_j))},$$
$$\mathrm{RASE2}(\hat{g}(t_j)) = \frac{\mathrm{ASE}(\hat{g}^{WLS}(t_j))}{\mathrm{ASE}(\hat{g}^{CQR}(t_j))},$$

where $\mathrm{ASE}(\hat{g}(t_j)) = n^{-1} \sum_{j=1}^{n_{grid}} (\hat{g}(t_j) - g(t_j))^2$, with $g(\cdot)$ either $\beta(\cdot)$ or $\beta'(\cdot)$, and $\{t_j, j = 1, \ldots, n_{grid}\}$ the grid points at which the functions $\{\hat{g}(\cdot)\}$ were evaluated. $\hat{g}^{LS}$ and $\hat{g}^{WLS}$ are the minimizers of (6.2) and (6.3), respectively. $\hat{g}^{CQR}$ is the local linear CQR estimator.

To assess the performance of AQR estimator, we compared it with the triangular k-NN weight estimator (see, Stone (1997), Zhao (2001))

$$\hat{\sigma}_i = \sum_{j=1}^{n} w_{ij} |\hat{\varepsilon}_j|, \quad i = 1, \ldots, n,$$

where $(w_{i1}, \ldots, w_{in})$ is the set of Stone's triangular k-NN weights corresponding to the ith observation, and $\hat{\varepsilon}_j$, for $j = 1, \ldots, n$, are the residuals of local linear least squares fitting.

To study how the performance of the proposed estimators varies with the number of quantiles q, we considered $q = 5, 9, 19$, respectively. The results are summarized in the following tables. CQR_5, CQR_9 and CQR_{19} correspond to the local linear CQR estimates with $q = 5, 9, 19$, respectively.

In the simulation, we set $n_{grid} = 200$ and grid points evenly distributed over the interval for which $\beta(\cdot)$ and $\beta'(\cdot)$ were estimated. We conducted 200 simulations for each case with sample size $n = 200$. The kernel used is Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$. The bandwidths were selected using the plug-in method of Section 4.

6.2. Simulation results and findings

From Tables 1, 2 and 3, we can see that, for all types of heteroscedasticity with ε normal distributed, the values of RASE1 and RASE2 are slightly less than 1; it is clear that local linear weighted least squares is the best estimation method. For all non-normal errors, the values of RASE1 and RASE2 are greater than 1, indicating the CQR estimator's huge gain in efficiency, especially for the Cauchy distribution.

When error is homoscedastic, as in Table 1, CQR_9 and CQR_{19} perform better than CQR_5, but when error is heteroscedastic, as in Tables 2 and 3, CQR_{19} performs better than CQR_5 and CQR_9. When error is heteroscedastic, combining the strength of relatively more quantile regressions at different quantile positions describe the data structures more thoroughly. However, the CQR estimator is not very sensitive to the number of quantiles, q, and a moderate q, such as $q = 9$, is enough.

Figure 1 is the plot of estimated coefficient functions when ε is Cauchy$(0, 1)$ and $q = 9$ in Example 1. We can see that the local CQR estimator performs much better than local LS estimator; The least squares method fails when the variance of ε is infinite, while CQR method does not.

Figure 2 is the plot of estimated conditional quantile curves of $\sigma(t)\varepsilon$ at at quantile positions $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ of Examples 1, 2, and 3, with ε Lognormal(0,1) and $q = 9$. From the left panel of Figure 2, we can see that the five estimated conditional quantile curves are parallel and horizontal, indicating that the model is homoscedastic, which is consistent with $\sigma(t)$ in Example 1 being constant. From the right panel of Figure 2, we can see that the five estimated quantile curves are not parallel or horizontal. Moreover, with the increase of t, the divergence of the quantile curves becomes more pronounced, indicating that the model is heteroscedastic and the heteroscedasticity increases sharply with

Table 1. Comparisons between local linear least squares (LS) and the local CQR based on the RASE1 criterion for Example 1 (Homoscedasticity).

		$\hat{\beta}_1(t)$	$\hat{\beta}_1'(t)$	$\hat{\beta}_2(t)$	$\hat{\beta}_2'(t)$
ε	method	RASE1	RASE1	RASE1	RASE1
	CQR_5	0.7960	0.6976	0.7609	0.7232
$N(0,1)$	CQR_9	0.8140	0.7108	0.7724	0.7401
	CQR_{19}	0.8231	0.7287	0.7826	0.7526
	CQR_5	1.4213	1.3712	1.4125	1.3399
$t(3)$	CQR_9	1.4367	1.3865	1.4201	1.3526
	CQR_{19}	1.4399	1.3927	1.4222	1.3572
	CQR_5	3.9290	3.9245	3.9097	4.1920
$lognormal$	CQR_9	4.0348	4.0540	3.9861	4.2845
	CQR_{19}	4.0720	4.1237	4.0057	4.3362
	CQR_5	5.2659	4.8938	4.6181	4.3297
$mixnormal$	CQR_9	5.2033	4.7991	4.5320	4.2606
	CQR_{19}	5.1107	4.7096	4.4733	4.2099
	CQR_5	64.551	32.156	166.57	127.59
$Cauchy$	CQR_9	32.023	16.169	86.699	66.483
	CQR_{19}	31.836	16.223	86.041	66.821

Table 2. Comparisons between local linear LS, local linear weighted LS, and the local CQR based on the RASE1 and RASE2 criteria for Example 2 (Small jumps Heteroscedasticity).

		$\hat{\beta}_1(t)$		$\hat{\beta}_1'(t)$		$\hat{\beta}_2(t)$		$\hat{\beta}_2'(t)$	
ε	method	RASE1	RASE2	RASE1	RASE2	RASE1	RASE2	RASE1	RASE2
	CQR_5	0.8382	0.8172	0.7909	0.7483	0.8497	0.8188	0.8147	0.7686
$N(0,1)$	CQR_9	0.8551	0.8337	0.8133	0.7695	0.8709	0.8392	0.8377	0.7903
	CQR_{19}	0.8613	0.8397	0.8211	0.7769	0.8758	0.844	0.8481	0.8002
	CQR_5	1.5998	1.5467	1.4655	1.4266	1.6306	1.532	1.4628	1.3817
$t(3)$	CQR_9	1.5936	1.5407	1.4773	1.4381	1.6328	1.534	1.4716	1.3901
	CQR_{19}	1.5879	1.5351	1.4805	1.4413	1.6232	1.5249	1.4727	1.3911
	CQR_5	4.8572	4.6092	4.3506	4.0801	4.9943	4.7645	5.1367	4.8335
$lognormal$	CQR_9	4.9729	4.7190	4.4606	4.1832	5.1485	4.9116	5.2816	4.9699
	CQR_{19}	5.0064	4.7508	4.5250	4.2436	5.1885	4.9497	5.3755	5.0583
	CQR_5	3.6068	3.4957	3.4910	3.2647	4.2954	4.1432	3.9197	3.6984
$mixnormal$	CQR_9	3.5819	3.4716	3.4713	3.2463	4.2515	4.1009	3.8404	3.6236
	CQR_{19}	3.5511	3.4417	3.4613	3.2369	4.2103	4.0611	3.8107	3.5956
	CQR_5	1227.6	1221.7	1540.3	1536.0	249.70	247.00	372.80	359.40
$Cauchy$	CQR_9	1027.0	1022.0	1287.4	1283.7	208.40	206.20	313.80	302.50
	CQR_{19}	1089.4	1084.1	1372.7	1368.8	223.10	220.70	337.40	325.20

the increase of t if t is bigger than 0.6, also consistent with the variation of $\sigma(t)$ in Example 3.

Table 3. Comparisons between local linear LS, local weighted LS, and the local CQR based on the RASE1 and RASE2 criteria for Example 3 (High frequency Heteroscedasticity).

		$\hat{\beta}_1(t)$		$\hat{\beta}_1'(t)$		$\hat{\beta}_2(t)$		$\hat{\beta}_2'(t)$	
ε	method	RASE1	RASE2	RASE1	RASE2	RASE1	RASE2	RASE1	RASE2
$N(0,1)$	CQR_5	0.8288	0.8030	0.8021	0.7600	0.8245	0.7973	0.8201	0.7762
	CQR_9	0.8388	0.8127	0.8137	0.7710	0.8477	0.8197	0.8504	0.8049
	CQR_{19}	0.8492	0.8227	0.8311	0.7874	0.8598	0.8314	0.8630	0.8167
$t(3)$	CQR_5	1.4344	1.3836	1.3256	1.2271	1.2008	1.1505	1.1785	1.0819
	CQR_9	1.4117	1.3617	1.3198	1.2217	1.1943	1.1442	1.1842	1.0872
	CQR_{19}	1.4113	1.3613	1.3221	1.2238	1.1951	1.1450	1.1911	1.0935
lognormal	CQR_5	5.0945	4.8374	4.7750	4.3203	3.1577	3.0029	3.5190	3.1936
	CQR_9	5.1421	4.8826	4.8802	4.4155	3.1479	2.9935	3.5583	3.2294
	CQR_{19}	5.1651	4.9044	4.9014	4.4347	3.1440	2.9898	3.5737	3.2433
mixnormal	CQR_5	2.9348	2.8232	2.9681	2.7779	3.0945	2.9715	3.1311	2.9123
	CQR_9	2.8358	2.7279	2.8835	2.6987	2.9779	2.8595	3.0390	2.8267
	CQR_{19}	2.8108	2.7038	2.8209	2.6400	2.9517	2.8344	2.9719	2.7642
Cauchy	CQR_5	638.29	701.70	421.38	317.92	308.82	308.71	249.48	230.71
	CQR_9	588.07	646.49	372.47	281.01	285.38	285.28	227.12	210.04
	CQR_{19}	578.52	635.99	367.47	277.25	281.32	281.21	223.77	206.94

Figure 3 is the plot of estimated $\sigma(t)$, with ε Lognormal(0,1) and $q = 9$. From Figure 3, we can see that the AQR estimator performs much better than the triangular k-NN weight estimator. In the simulation study, we can conclude that CQR-AQR estimation is an efficient and robust method to simultaneously estimate the coefficient functions and the heteroscedasticity.

6.3. Hypothesis testing

To demonstrate the power of the goodness-of-fit test in Section 5, we considered the null hypothesis that $\beta_2(t)$ in model (6.1) is constant versus the alternative that it is varying. We considered the case that $\sigma(t)$ in (6.1) is the same as that of Example 1 and ε is normal. The power was evaluated under a sequence of the alternative models indexed by λ:

$$\beta_2(t;\lambda) = c + \lambda\{\beta_2(t) - c\} \quad (0 \le \lambda \le 1),$$

where $c = \int_0^1 \beta_2(t)dt$. For each λ in $\{0, 0.1, 0.2, \ldots, 1.0\}$, we applied the goodness-of-fit test with 100 replications of sample size $n = 200$. For each replication, we repeated the bootstrap resampling 100 times. The significance level α was 0.05. Figure 4 shows the simulated power against λ, indicating the hypothesis testing procedure worked appropriately.

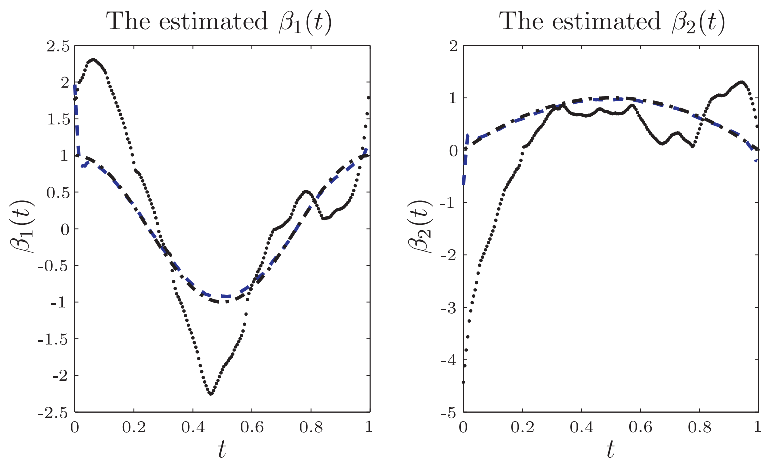

Figure 1. Plots of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ in Example 1, where ε is Cauchy. Dot-dashed curves: true functions; Long-dashed curves: local linear CQR estimated functions; Dot curves: local linear LS estimated functions.

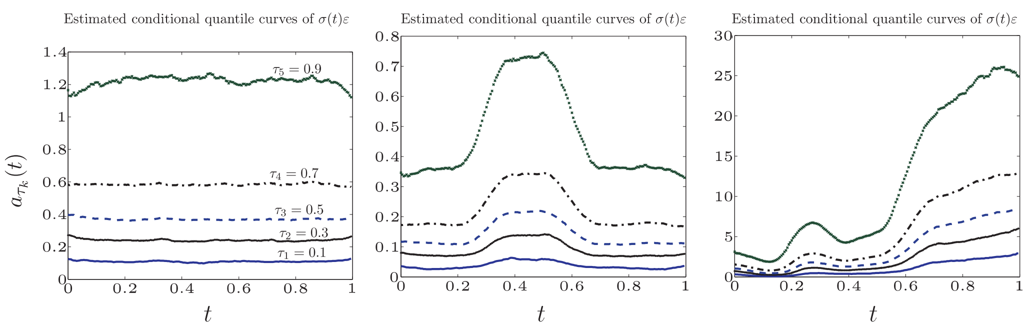

Figure 2. Plots of five estimated conditional quantile curves of $\sigma(t)\varepsilon$ of Examples 1, 2, and 3 (from the left panel to the right), with ε Lognormal(0,1).

7. Empirical Application

We illustrate the methodology of this paper via an application to an air pollution data set. The data were a subsample of 500 observations from a study conducted by the Norwegian Public Roads Administration, investigating how the air pollution at a road relates to the traffic volume and wind speed. The data was obtained from StatLib. It is of interest to study the association between the levels of pollutants and the traffic volume and wind speed, and to examine the extent to which the association varies over time. We consider the relation among the hourly values of the logarithm of the concentration of NO_2 (Y), measured at Alnabru in Oslo, Norway, between October 2001 and August 2003, and the logarithm of the number of cars per hour (X_1), wind speed X_2 (meters per

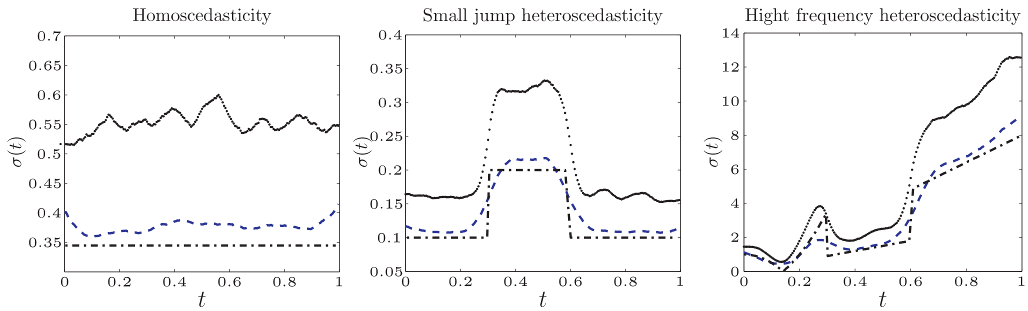

Figure 3. Plots of $\hat{\sigma}(t)$ in Examples 1, 2, and 3 (from the left panel to the right), with ε Lognormal(0,1). Dot-dashed curves: true functions; Long-dashed curves: AQR estimated functions; Dot curves: triangular k-NN weight estimated functions.

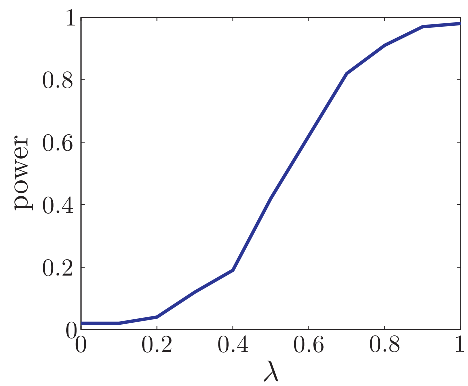

Figure 4. *Plot of power against λ for the goodness-of-fit test.*

second), and the hour of day (T). All covariates except T are standardized. We considered the following varying-coefficient model with heteroscedasticity

$$Y = \beta_1(T)X_1 + \beta_2(T)X_2 + \sigma(T)\varepsilon. \tag{7.1}$$

to fit the given data.

The proposed CQR-AQR estimation was employed to estimate $\beta_1(t)$, $\beta_2(t)$, and $\sigma(t)$. In our applications, we chose $q = 9$. The Epanechnikov kernel was employed and the plug-in selected bandwidth was 3.9357. The estimated coefficient functions are depicted in Figure 5. The figure shows that there is a strong time effect on the coefficient functions $\beta_1(\cdot)$ and $\beta_2(\cdot)$. Moreover, the association between the number of cars per hour and the concentration of NO_2 is always positive, and the association between the wind speed and the concentration of NO_2 is always negative.

When employing the AQR estimator to obtain $\hat{\sigma}(t)$, we took ε to be normal; $\hat{\sigma}(t)$ is presented in Figure 6 (left panel). We can see that the values of $\hat{\sigma}(t)$

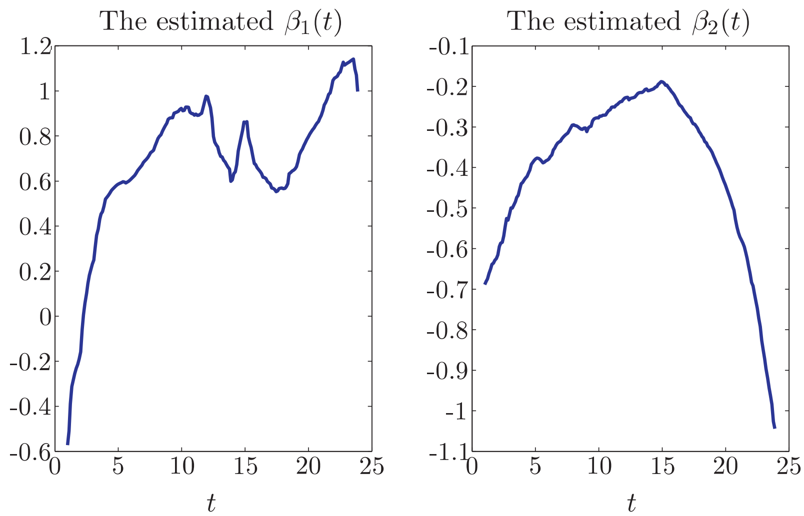

Figure 5. The estimated coefficient functions $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ of model (7.1).

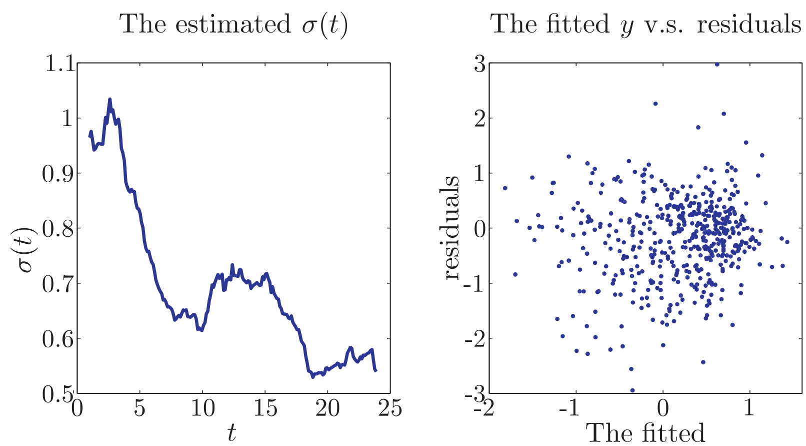

Figure 6. The plot of $\hat{\sigma}(t)$ in model (7.1) (left panel); The plot of fitted y to corresponding residuals of model (7.1) (right panel).

varies with time. Moreover, the plot of the fitted values of Y (that is, $\hat{Y} = \hat{\beta}_1(T)X_1 + \hat{\beta}_2(T)X_2$) to the residuals (that is, $Y - \hat{Y}$), presented in Figure 6 (right panel), also demonstrates that there is heteroscedasticity in model (7.1).

A natural question is whether the coefficient functions are really time varying. We used the hypothesis testing procedure in Section 5 to answer this question. The number of bootstrap replicates was $J = 500$. The observed statistics and their p-values are summarized in Table 4. From Table 4, at the 0.05 significance level, there is sufficient evidence to reject H_{01} and H_{02}.

Table 4. Test statistics and p-values for testing whether coefficient functions are actually time-varying.

Null hypothesis	Values of test statistic	p-value
$H_{01} : \beta_1(\cdot) = \beta_1$	0.1966	0.0260
$H_{02} : \beta_2(\cdot) = \beta_2$	0.0494	0.0120

8. Local m-Polynomial CQR-AQR Estimation

As a generalization of local linear and quadratic CQR-AQR estimation, we consider local m-polynomial CQR-AQR and establish the asymptotic theory for local m-polynomial CQR-AQR estimators in model (1.1). For each given t, we can get $\hat{a}_{\tau_k}(t)$, $\hat{b}_j$ for $k = 1, \ldots, q$, $j = 0, \ldots, m$, by minimizing

$$\sum_{k=1}^{q}\sum_{i=1}^{n}\rho_{\tau_k}\Big(Y_i - a_{\tau_k}(t) - \sum_{j=0}^{m} X_i^T b_j (T_i - t)^j\Big) K\Big(\frac{T_i - t}{h}\Big), \tag{8.1}$$

where $b_j = \beta^{(j)}(t)/j!$.

Let $u_k = \sqrt{nh}(a_{\tau_k}(t) - \sigma(t)c_{\tau_k})$, $v_j = h^j\sqrt{nh}(j!b_j - \beta^{(j)}(t))/j!$, and $\theta = (u_1, u_2, \ldots, u_q; v_0, v_1, \ldots, v_m)^T$.

Define

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

where S_{11} is a $q\times q$ diagonal matrix with diagonal elements $f(c_{\tau_k})$, $k = 1, 2, \ldots, q$, S_{12} is a $q \times (m + 1)$ matrix with (k, j)-element $f(c_{\tau_k})E(X^T \mid T = t)\mu_j$, $k = 1, 2, \ldots, q$ and $j = 0, 1, \ldots, m$, $S_{21} = S_{12}^T$, and S_{22} is a $(m+1)\times(m+1)$ matrix with (j, j')-element $E(XX^T \mid T = t)\mu_{j+j'}\sum_{k=1}^{q} f(c_{\tau_k})$, $j, j' = 0, 1, \ldots, m$. Partition S^{-1} into submatrices as follows:

$$S^{-1} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{-1} = \begin{pmatrix} (S^{-1})_{11} & (S^{-1})_{12} \\ (S^{-1})_{21} & (S^{-1})_{22} \end{pmatrix}$$

where here and hereafter $(\cdot)_{11}$ denotes the top left-hand $q\times q$ submatrix and $(\cdot)_{22}$ denotes the bottom right-hand $(m + 1) \times (m + 1)$ submatrix.

Define

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where Σ_{11} is a $q\times q$ matrix with (k, k')-element $\nu_0\tau_{kk'}$, $k, k' = 1, 2, \ldots, q$, Σ_{12} is a $q\times(m+1)$ matrix with (k, j)-element $E(X^T \mid T = t)\nu_j\sum_{k'=1}^{q}\tau_{kk'}$, $k = 1, 2, \ldots, q$, $j = 0, 1, \ldots, m$, $\Sigma_{21} = \Sigma_{12}^T$, Σ_{22} is a $(m + 1) \times (m + 1)$ matrix with (j, j')-element $E(XX^T \mid T = t)\nu_{j+j'}\sum_{k,k'=1}^{q}\tau_{kk'}$, $j, j' = 0, 1, \ldots, m$.

Let $d_{i,k} = c_{\tau_k}\{\sigma(T_i) - \sigma(t)\} + r_{i,m}$, $r_{i,m} = X_i^T(\beta(T_i) - \sum_{j=0}^{m}\beta^{(j)}(t)(T_i - t)^j/j!)$, $s_i = (T_i - t)/h$ and $K(T_i - t)/h = K_i$, and take $\eta_{i,k}^*$ to be $I(\varepsilon_i \leq c_{\tau_k} -$

$d_{i,k}/\sigma(T_i)) - \tau_k$. Let $\mathbf{W}_n^* = (w_{11}^*, w_{12}^*, \ldots, w_{1q}^*; w_{20}^*, w_{21}^*, \ldots, w_{2m}^*)^T$, where $w_{1k}^* = (\sqrt{nh})^{-1} \sum_{i=1}^n K_i \eta_{i,k}^*$, $w_{2j}^* = (\sqrt{nh})^{-1} \sum_{k=1}^q \sum_{i=1}^n X_i^T K_i s_i^j \eta_{i,k}^*$.

Theorem 4. *Under Conditions* 1–6*, for* $k = 1, \ldots, q$*,* $j = 0, \ldots, m$*,*

$$\hat{\theta} + \frac{\sigma(t)}{f_T(t)} S^{-1} E(\mathbf{W}_n^* \mid \mathcal{D}) \xrightarrow{d} N\Big(\mathbf{0}, \frac{\sigma^2(t)}{f_T(t)} S^{-1} \Sigma S^{-1}\Big). \tag{8.2}$$

9. Discussion

In this section, we discuss some directions to further extend this work.

(a) The proposed CQR-AQR method is not very sensitive to the number of quantiles q but, practically speaking, we can choose q by some tuning methods such as K-fold cross-validation. To circumvent the problem of selecting q, we can consider, instead of (2.1),

$$\int_0^1 \sum_{i=1}^n \rho_\gamma \left(Y_i - a_\gamma(t) - X_i^T \left(\beta(t) + \beta'(t)(T_i - t)\right)\right) K\left(\frac{T_i - t}{h}\right) \omega(\gamma) d\gamma, \tag{9.1}$$

where the weight function $\omega(\gamma)$ is a density function over $(0,1)$. In practice, we need to discretize the integral in order to numerically compute the estimators. Thus, a discrete distribution density $\omega(\gamma)$ is used to construct the weights. The proposed method uses a discrete uniform distribution on $\{1/(q+1), \ldots, q/(q+1)\}$.

(b) The model (1.1) can be extended to

$$Y_i = X_i^T \beta(T_i) + \sigma(X_i, T_i)\varepsilon_i,$$

where $\sigma(X_i, T_i)$ is an unknown form positive function. We can estimate $\beta(t)$ and $\sigma(\mathbf{x}, t)$ by minimizing

$$\sum_{k=1}^q \sum_{i=1}^n \rho_{\tau_k} \left(Y_i - a_{\tau_k}(\mathbf{x}, t) - X_i^T \left(\beta(t) + \beta'(t)(T_i - t)\right)\right) \mathcal{K}_{i,H}, \tag{9.2}$$

where $\mathbf{x} = (x_1, \ldots, x_p)$, $a_{\tau_k}(\mathbf{x}, t) = \sigma(\mathbf{x}, t) c_{\tau_k}$, $\mathcal{K}_{i,H} = K((T_i - t)/h_0, (X_{i1} - x_1)/h_1, \ldots, (X_{ip} - x_p)/h_p)$ is a multivariate kernel function, and $H = (h_0, h_1, \ldots, h_p)^T$ is a vector of bandwidths. Denote the minimizers of (9.2) by $\hat{\beta}(t)$ and $\hat{a}_{\tau_k}(\mathbf{x}, t)$, for $k = 1, \ldots, q$. If the AQR estimator is employed, we can easily get $\hat{\sigma}(\mathbf{x}, t) = (1/q) \sum_{k=1}^q \hat{a}_{\tau_k}(\mathbf{x}, t)/c_{\tau_k}$, if $c_{\tau_k} \neq 0$ for any $k = 1, \ldots, q$; or $\hat{\sigma}(\mathbf{x}, t) = (1/(q-1)) \sum_{\substack{k=1 \\ k \neq j}}^q \hat{a}_{\tau_k}(\mathbf{x}, t)/c_{\tau_k}$, if $c_{\tau_j} = 0$ for certain $j \in \{1, \ldots, q\}$.

Note that estimating $\sigma(\mathbf{x}, t)$ nonparametrically encounters the 'curse of dimensionality' and how to estimate $\beta(t)$ is an open problem.

Finally, we would like to point out that the proposed method is efficiently implemented by using the MM algorithm. For the simulation example in the case that $q = 9$ and the sample size is $n = 5,000$, the proposed method fit at a given location t was computed within $0.3280s$ on an Intel 2.8-GHz machine.

Acknowledgement

The authors gratefully acknowledge the helpful comments of an associate editor and the referees. The research is partially supported by the Major Project of Humanities Social Science Foundation of Ministry of Education (No. 08JJD910247), Key Project of Chinese Ministry of Education (No. 108120), National Natural Science Foundation of China (No. 10871201), Beijing Natural Science Foundation (No. 1102021), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 10XNL018).

Appendix

Proof of Theorem 4. Let

$$\Delta_{i,k} = \frac{u_k + \sum_{j=0}^m X_i^T v_j s_i^j}{\sqrt{nh}}.$$

Because $Y_i - a_{\tau_k}(t) - X_i^T \sum_{j=0}^m b_j(T_i - t)^j = \sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} - \Delta_{i,k}$, $\hat{\theta} = (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_q; \hat{v}_0, \hat{v}_1, \ldots, \hat{v}_m)^T$ is the minimizer of

$$L_n = \sum_{k=1}^q \sum_{i=1}^n \Big[\rho_{\tau_k}(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} - \Delta_{i,k}) - \rho_{\tau_k}(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k})\Big] K_i.$$

Applying the identity

$$\rho_\tau(r - s) - \rho_\tau(r) = s(I(r \le 0) - \tau) + \int_0^s [I(r \le z) - I(r \le 0)] dz,$$

we write L_n as

$$\begin{aligned} L_n = &\sum_{i=1}^n \sum_{k=1}^q \frac{u_k + \sum_{j=0}^m X_i^T v_j s_i^j}{\sqrt{nh}} \Big(I(\varepsilon_i \le c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - \tau_k \Big) K_i \\ &+ \sum_{i=1}^n \sum_{k=1}^q K_i \int_0^{\Delta_{i,k}} \Big(I(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} \le z) \\ &- I(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} \le 0) \Big) dz \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{q}\Big\{\sum_{i=1}^{n}\frac{K_i\eta^*_{i,k}}{\sqrt{nh}}\Big\}u_k+\sum_{j=0}^{m}\Big\{\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{X_i^TK_is_i^j\eta^*_{i,k}}{\sqrt{nh}}\Big\}v_j\\
&\quad +\sum_{k=1}^{q}\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}\Big(I(\sigma(T_i)(\varepsilon_i-c_{\tau_k})+d_{i,k}\le z)-c_{\tau_k})+d_{i,k}\le z)\\
&\quad -I(\sigma(T_i)(\varepsilon_i-c_{\tau_k})+d_{i,k}\le 0)\Big)dz\\
&= \mathbf{W}_n^{*T}\theta+\sum_{k=1}^{q}B_{n,k}(\theta),
\end{aligned}
$$

where $B_{n,k}(\theta)=\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}\Big(I(\sigma(T_i)(\varepsilon_i-c_{\tau_k})+d_{i,k}\le z)-I(\sigma(T_i)(\varepsilon_i-c_{\tau_k})+d_{i,k}\le 0)\Big)dz$.

Take

$$
S_n=\begin{pmatrix} S_{n,11} & S_{n,12}\\ S_{n,21} & S_{n,22}\end{pmatrix},
$$

where $S_{n,11}$ is a $q\times q$ diagonal matrix with diagonal elements $(1/nh)f(c_{\tau_k})\sum_{i=1}^{n}K_i/\sigma(T_i)$, $k=1,2,\ldots,q$, $S_{n,12}$ is a $q\times(m+1)$ matrix with (k,j)-elements $(1/nh)f(c_{\tau_k})\sum_{i=1}^{n}X_i^TK_is_i^j/\sigma(T_i)$, $j=0,1,\ldots,m$, $S_{n,21}=S_{n,12}^T$, $S_{n,22}$ is a $(m+1)\times(m+1)$ matrix with (j,j')-elements $(1/nh)\sum_{k=1}^{q}f(c_{\tau_k})\sum_{i=1}^{n}X_iX_i^TK_is_i^{j+j'}/\sigma(T_i)$, $j,j'=0,1,\ldots,m$.

$$
\begin{aligned}
E[B_{n,k}(\theta)\mid\mathcal{D}] &= E\Big[\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}\{I(\varepsilon_i\le c_{\tau_k}+\frac{z-d_{i,k}}{\sigma(T_i)})\\
&\qquad -I(\varepsilon_i\le c_{\tau_k}-\frac{d_{i,k}}{\sigma(T_i)})\}dz\mid\mathcal{D}\Big]\\
&= \sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}\{F(c_{\tau_k}-\frac{d_{i,k}}{\sigma(T_i)}+\frac{z}{\sigma(T_i)})-F(c_{\tau_k}-\frac{d_{i,k}}{\sigma(T_i)})\}dz\\
&= \sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}\{\frac{z}{\sigma(T_i)}f(c_{\tau_k}-\frac{d_{i,k}}{\sigma(T_i)})+o(z)\}dz\\
&= \sum_{i=1}^{n}K_i\Delta_{i,k}^2\frac{f(c_{\tau_k}-d_{i,k}/\sigma(T_i))}{2\sigma(T_i)}+o_p(1)\\
&= \sum_{i=1}^{n}K_i\Delta_{i,k}^2\frac{f(c_{\tau_k})}{2\sigma(T_i)}+o_p(1).
\end{aligned}
$$

As well,

$$
\mathrm{Var}[B_{n,k}(\theta)\mid\mathcal{D}]=\mathrm{Var}\Big[\sum_{i=1}^{n}K_i\int_0^{\Delta_{i,k}}\{I(\sigma(T_i)(\varepsilon_i-c_{\tau_k})+d_{i,k}\le z)
$$

$$
\begin{aligned}
& -I(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} \le 0)\}dz \mid \mathcal{D}\Big] \\
&\le \sum_{i=1}^{n} E\Big[\Big(K_i \int_0^{\Delta_{i,k}} \{I(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} \le z) \\
& -I(\sigma(T_i)(\varepsilon_i - c_{\tau_k}) + d_{i,k} \le 0)\}dz\Big)^2 \mid \mathcal{D}\Big] \\
&\le \sum_{i=1}^{n} K_i^2 \int_0^{|\Delta_{i,k}|} \int_0^{|\Delta_{i,k}|} \{F(c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)} + \frac{|\Delta_{i,k}|}{\sigma(T_i)}) \\
& -F(c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)})\}dz_1 dz_2 \\
&\le o(\sum_{i=1}^{n} K_i^2 \Delta_{i,k}^2). \\
&= o_p(1).
\end{aligned}
$$

Hence $\sum_{k=1}^{q} B_{n,k}(\theta) \xrightarrow{p} \sum_{k=1}^{q} \sum_{i=1}^{n} K_i \Delta_{i,k}^2 f(c_{\tau_k})/2\sigma(T_i)$, and we can write the limit, $\sum_{k=1}^{q} \sum_{i=1}^{n} K_i \Delta_{i,k}^2 f(c_{\tau_k})/2\sigma(T_i)$, into the matrix $\frac{1}{2}\theta^T S_n \theta$. That is,

$$
\sum_{k=1}^{q} B_{n,k}(\theta) \xrightarrow{p} \frac{1}{2}\theta^T S_n \theta.
$$

As in Parzen (1962), we have

$$
\begin{aligned}
\frac{1}{nh}\sum_{i=1}^{n} K_i &\xrightarrow{p} f_T(t), \\
\frac{1}{nh}\sum_{i=1}^{n} X_i K_i s_i^j &\xrightarrow{p} f_T(t)E(X \mid T = t)\mu_j, \\
\frac{1}{nh}\sum_{i=1}^{n} X_i X_i^T K_i s_i^j &\xrightarrow{p} f_T(t)E(XX^T \mid T = t)\mu_j,
\end{aligned}
$$

where $\xrightarrow{p}$ stands for convergence in probability. Thus,

$$
S_n \xrightarrow{p} \frac{f_T(t)}{\sigma(t)} S.
$$

This leads to

$$
L_n(\theta) \xrightarrow{p} \frac{1}{2}\frac{f_T(t)}{\sigma(t)}\theta^T S \theta + \mathbf{W}_n^{*T}\theta.
$$

Since L_n is a convex function, following Knight (1998) we have

$$
\hat{\theta} \xrightarrow{p} -\frac{\sigma(t)}{f_T(t)} S^{-1}\mathbf{W}_n^*.
$$

Let $\eta_{i,k} = I(\varepsilon_i \leq c_{\tau_k}) - \tau_k$, and $\mathbf{W}_n = (w_{11}, \ldots, w_{1q}, w_{20}, \ldots, w_{2m})^T$ with

$$w_{1k} = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K_i \eta_{i,k} \quad \text{and} \quad w_{2j} = \frac{1}{\sqrt{nh}} \sum_{k=1}^{q} \sum_{i=1}^{n} X_i^T K_i s_i^j \eta_{i,k}.$$

By the Cramer-Wald Theorem and the Central Limit Theorem, we have

$$\frac{\mathbf{W}_n \mid \mathcal{D} - E[\mathbf{W}_n \mid \mathcal{D}]}{\sqrt{\mathrm{cov}(\mathbf{W}_n \mid \mathcal{D})}} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{(m+q+1)\times(m+q+1)}).$$

Note that $\mathrm{cov}(\eta_{i,k}, \eta_{i,k'}) = \tau_{kk'}$ and $\mathrm{cov}(\eta_{i,k}, \eta_{j,k'}) = 0$, if $i \neq j$. As in Parzen (1962), we have

$$\begin{aligned}
\frac{1}{nh} \sum_{i=1}^{n} K_i^2 &\xrightarrow{p} f_T(t)\nu_0, \\
\frac{1}{nh} \sum_{i=1}^{n} K_i^2 s_i^j X_i &\xrightarrow{p} f_T(t)\nu_j E(X \mid T = t), \\
\frac{1}{nh} \sum_{i=1}^{n} K_i^2 s_i^{j+j'} X_i X_i^T &\xrightarrow{p} f_T(t)\nu_{j+j'} E(XX^T \mid T = t).
\end{aligned}$$

Therefore, $\mathrm{cov}(\mathbf{W}_n \mid \mathcal{D}) \xrightarrow{p} f_T(t)\Sigma$ and

$$\mathbf{W}_n \mid \mathcal{D} \xrightarrow{d} N(\mathbf{0}, f_T(t)\Sigma).$$

Moreover,

$$\begin{aligned}
\mathrm{Var}(w_{1k}^* - w_{1k} \mid \mathcal{D}) &= \mathrm{Var}\Big(\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K_i(\eta_{i,k}^* - \eta_{i,k}) \mid \mathcal{D}\Big) \\
&= \frac{1}{nh} \sum_{i=1}^{n} K_i^2 \mathrm{Var}\Big(I(\varepsilon_i \leq c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - I(\varepsilon_i \leq c_{\tau_k}) \mid \mathcal{D}\Big) \\
&\leq \frac{1}{nh} \sum_{i=1}^{n} K_i^2 E\Big|[I(\varepsilon_i \leq c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - I(\varepsilon_i \leq c_{\tau_k})] \mid \mathcal{D}\Big| \\
&\leq \frac{1}{nh} \sum_{i=1}^{n} K_i^2 [F(c_{\tau_k} + \frac{|d_{i,k}|}{\sigma(T_i)}) - F(c_{\tau_k})] \\
&= o_p(1),
\end{aligned}$$

$$\mathrm{Var}(w_{2j}^* - w_{2j} \mid \mathcal{D}) = \mathrm{Var}\Big(\frac{1}{\sqrt{nh}} \sum_{k=1}^{q} \sum_{i=1}^{n} X_i^T K_i s_i^j (\eta_{i,k}^* - \eta_{i,k}) \mid \mathcal{D}\Big)$$

$$
\begin{aligned}
&= \frac{1}{nh}\sum_{i=1}^{n} X_i^T X_i K_i^2 s_i^{2j} \mathrm{Var}\Big(\sum_{k=1}^{q}(\eta_{i,k}^{*}-\eta_{i,k}) \mid \mathcal{D}\Big)\\
&\le \frac{q^2}{nh}\sum_{i=1}^{n} X_i^T X_i K_i^2 s_i^{2j} \max_k E\Big|[I(\varepsilon_i \le c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)})\\
&\quad -I(\varepsilon_i \le c_{\tau_k})] \mid \mathcal{D}\Big|\\
&\le \frac{q^2}{nh}\sum_{i=1}^{n} X_i^T X_i K_i^2 s_i^{2j} \max_k [F(c_{\tau_k} + \frac{|d_{i,k}|}{\sigma(T_i)}) - F(c_{\tau_k})]\\
&= o_p(1).
\end{aligned}
$$

Thus, we have

$$\mathbf{W}_n^* \mid \mathcal{D} \xrightarrow{p} \mathbf{W}_n \mid \mathcal{D}.$$

By Slutsky's theorem, conditioning on $\mathcal{D}$, we have $\mathbf{W}_n^*|\mathcal{D} - E[\mathbf{W}_n^*|\mathcal{D}] \xrightarrow{d} N(\mathbf{0}, f_T(t)\Sigma)$, and so

$$\hat{\theta} + \frac{\sigma(t)}{f_T(t)} S^{-1} E(\mathbf{W}_n^* \mid \mathcal{D}) \to_d N\Big(\mathbf{0}, \frac{\sigma^2(t)}{f_T(t)} S^{-1}\Sigma S^{-1}\Big).$$

This completes the proof of Theorem 4.

Proof of Theorem 1. The asymptotic normality follows Theorem 4 with $m = 1$. We calculate the conditional bias and variance of $\hat{\beta}(t)$ in this section. When $m = 1$,

$$
\begin{aligned}
&S_{11} = \mathrm{diag}\{f(c_{\tau_1}), \ldots, f(c_{\tau_q})\},\\
&S_{12} = \{(f(c_{\tau_1})E(X \mid T = t), \ldots, f(c_{\tau_q})E(X \mid T = t))^T, \mathbf{0}_{q\times 1}\},\\
&S_{22} = \mathrm{diag}\{E(XX^T \mid T = t)\sum_{k=1}^{q} f(c_{\tau_k}), E(XX^T \mid T = t)\mu_2 \sum_{k=1}^{q} f(c_{\tau_k})\}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&(S^{-1})_{22} = (S_{22} - S_{21}S_{11}^{-1}S_{12})^{-1} = \mathrm{diag}\Big\{\frac{(\mathrm{Var}(X \mid T = t))^{-1}}{\sum_{k=1}^{q} f(c_{\tau_k})}, \frac{(E(XX^T \mid T = t))^{-1}}{\mu_2\sum_{k=1}^{q} f(c_{\tau_k})}\Big\},\\
&(S^{-1})_{21} = -(S^{-1})_{22}S_{21}S_{11}^{-1} = \Big(-\Big\{\frac{E(X \mid T = t)(\mathrm{Var}(X \mid T = t))^{-1}}{\sum_{k=1}^{q} f(c_{\tau_k})}\Big\}\mathbf{1}_{q\times 1}, \mathbf{0}_{q\times 1}\Big)^T,
\end{aligned}
$$

$$
\begin{aligned}
E(w_{1k}^* \mid \mathcal{D}) &= \sum_{i=1}^{n} \frac{K_i}{\sqrt{nh}}\Big(F(c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - F(c_{\tau_k})\Big),\\
E(w_{2j}^* \mid \mathcal{D}) &= \sum_{i=1}^{n} \frac{X_i^T K_i s_i^j}{\sqrt{nh}} \sum_{k=1}^{q}\Big(F(c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - F(c_{\tau_k})\Big),
\end{aligned}
$$

Write $e_1 = (1,0)^T$. By Theorem 4,

$$\begin{aligned}\text{bias}\{\hat{\beta}(t) \mid \mathcal{D}\} &= -\frac{1}{\sqrt{nh}}\frac{\sigma(t)}{f_T(t)} e_1^T\{(S^{-1})_{21}E[\mathbf{W}_{1n}^* \mid \mathcal{D}] + (S^{-1})_{22}E[\mathbf{W}_{2n}^* \mid \mathcal{D}]\} \\ &= -\frac{\sigma(t)}{nhf_T(t)}\frac{(\text{Var}(X \mid T=t))^{-1}}{\sum_{k=1}^q f(c_{\tau_k})}\sum_{i=1}^n (X_i - E(X \mid T=t)) \\ &\quad \cdot K_i \sum_{k=1}^q \Big(F(c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - F(c_{\tau_k})\Big).\end{aligned}$$

It is easy to check that

$$F(c_{\tau_k} - \frac{d_{i,k}}{\sigma(T_i)}) - F(c_{\tau_k}) = -\frac{r_{i,1}}{\sigma(T_i)} f(c_{\tau_k})\{1 + o_p(1)\},$$

and therefore

$$\begin{aligned}&\text{bias}(\hat{\beta}(t) \mid \mathcal{D}) \\ &= \frac{1}{nh}\frac{\sigma(t)}{f_T(t)}(\text{Var}(X \mid T=t))^{-1}\sum_{i=1}^n (X_i - E(X \mid T=t))K_i\frac{r_{i,1}}{\sigma(T_i)}\{1+o_p(1)\} \\ &= \frac{1}{2nh}\frac{\sigma(t)}{f_T(t)}\sum_{i=1}^n K_i \frac{\beta''(t)(T_i - t)^2}{\sigma(T_i)}\{1+o_p(1)\} = \frac{1}{2}\beta''(t)\mu_2 h^2 + o_p(h^2).\end{aligned}$$

Furthermore, the conditional covariance of $\hat{\beta}(t)$ is

$$\begin{aligned}\text{cov}(\hat{\beta}(t) \mid \mathcal{D}) &= \frac{\sigma^2(t)}{nhf_T(t)} e_1^T (S^{-1}\Sigma S^{-1})_{22}\, e_1 \\ &= \frac{\sigma^2(t)}{nhf_T(t)} \frac{\nu_0 \sum_{k=1}^q \sum_{k'=1}^q \tau_{kk'}}{(\sum_{k=1}^q f(c_{\tau_k}))^2}(\text{Var}(X \mid T=t))^{-1} + o_p(\frac{1}{nh}) \\ &= \frac{\nu_0 (\text{Var}(X \mid T=t))^{-1}\sigma^2(t)}{nhf_T(t)} R_1(q) + o_p(\frac{1}{nh}).\end{aligned}$$

This completes the proof of Theorem 1.

Proof of Theorem 2. Here we only calculate the conditional bias and variance of $\hat{\sigma}(t)$ as presented in (2.2). The proof of the asymptotic normality of (2.3) is similar. Let e_k be a $q \times 1$ vector, whose kth element is 1 and all other elements are 0. According to Theorem 4,

$$\begin{aligned}&\text{bias}\{\hat{a}_{\tau_k}(t) \mid \mathcal{D}\} \\ &= -\frac{1}{\sqrt{nh}}\frac{\sigma(t)}{f_T(t)} e_k^T\left\{(S^{-1})_{11}E(\mathbf{W}_{1n}^* \mid \mathcal{D}) + (S^{-1})_{12}E(\mathbf{W}_{2n}^* \mid \mathcal{D})\right\}\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{\sqrt{nh}}\frac{\sigma(t)}{f_T(t)}\Big[\frac{1}{f(c_{\tau_k})}\sum_{i=1}^{n}\frac{K_i}{\sqrt{nh}}\left(F(c_{\tau_k}-\frac{d_{i,k}}{\sigma(T_i)})-F(c_{\tau_k})\right)\\
&\quad -\frac{E(X^T\,|\,T\!=\!t)(\mathrm{Var}(X^T\,|\,T\!=\!t))^{-1}}{\sum_{k=1}^{q}f(c_{\tau_k})}\sum_{i=1}^{n}\frac{X_iK_i}{\sqrt{nh}}\sum_{k=1}^{q}\left(F(c_{\tau_k}-\frac{d_{i,k}}{\sigma(T_i)})-F(c_{\tau_k})\right)\Big]\\
&= \frac{1}{nh}\frac{\sigma(t)}{f_T(t)}\sum_{i=1}^{n}K_i\frac{r_{i,1}}{\sigma(T_i)}\left(1-E(X^T\mid T=t)\left(\mathrm{Var}(X\mid T=t)\right)^{-1}X_i\right)\{1+o_p(1)\}\\
&= \frac{1}{nh}\frac{\sigma(t)}{f_T(t)}\sum_{i=1}^{n}K_i\frac{X_i^T\beta''(t)(T_i-t)^2}{2\sigma(T_i)}\\
&\quad \cdot\left(1-E(X^T\mid T=t)\left(\mathrm{Var}(X\mid T=t)\right)^{-1}X_i\right)\{1+o_p(1)\}\\
&= \frac{1}{2}E(X^T\mid T=t)\beta''(t)\mu_2h^2(1-\Omega)+o_p(h^2).
\end{aligned}
$$

Because $\hat{\sigma}(t)=\frac{1}{q}\sum_{k=1}^{q}\hat{a}_{\tau_k}(t)/c_{\tau_k}$, we can get

$$
\begin{aligned}
\mathrm{bias}\{\hat{\sigma}(t)\mid\mathcal{D}\} &= \frac{1}{q}\sum_{k=1}^{q}\frac{1}{c_{\tau_k}}\mathrm{bias}\{\hat{a}_{\tau_k}(t)\mid\mathcal{D}\}\\
&= \frac{1}{2}E(X^T\mid T=t)\beta''(t)\mu_2h^2\left(\frac{1}{q}\sum_{k=1}^{q}\frac{1}{c_{\tau_k}}\right)(1-\Omega)+o_p(h^2)\\
&= \frac{1}{2}E(X^T\mid T=t)\beta''(t)\mu_2h^2\mathbf{c}(1-\Omega)+o_p(h^2).
\end{aligned}
$$

The conditional variance of $\hat{\sigma}(t)$ is

$$
\begin{aligned}
&\mathrm{Var}(\hat{\sigma}(t)\mid\mathcal{D})\\
&= \frac{1}{q^2}\sum_{k=1}^{q}\sum_{k'=1}^{q}\frac{1}{c_{\tau_k}c_{\tau_{k'}}}\mathrm{cov}(\hat{a}_{\tau_k}(t),\hat{a}_{\tau_{k'}}(t))\\
&= \frac{1}{nhq^2}\frac{\sigma^2(t)}{f_T(t)}\sum_{k=1}^{q}\sum_{k'=1}^{q}\frac{1}{c_{\tau_k}c_{\tau_{k'}}}\left(S^{-1}\Sigma S^{-1}\right)_{kk'}\\
&= \frac{1}{nh}\frac{\nu_0\sigma^2(t)}{f_T(t)}\Big[\frac{1}{q^2}\sum_{k=1}^{q}\sum_{k'=1}^{q}\frac{1}{c_{\tau_k}c_{\tau_{k'}}}\frac{\tau_{kk'}}{f(c_{\tau_k})f(c_{\tau_{k'}})}\\
&\quad -\frac{1}{q^2}\sum_{k=1}^{q}\sum_{k'=1}^{q}\frac{1}{c_{\tau_k}c_{\tau_{k'}}}\frac{\sum_{k=1}^{q}\tau_{kk'}}{\sum_{k=1}^{q}f(c_{\tau_k})}E(X^T\,|\,T\!=\!t)\left(\mathrm{Var}(X^T\,|\,T\!=\!t)\right)^{-1}E(X\,|\,T\!=\!t)\Big]\\
&\quad +o_p(\frac{1}{nh})\\
&= \frac{1}{nh}\frac{\nu_0\sigma^2(t)}{f_T(t)}\left(R_2(q)-R_3(q)\Omega\right)+o_p\left(\frac{1}{nh}\right).
\end{aligned}
$$

This completes the proof of Theorem 2.

Proof of Theorem 3. The asymptotic normality follows from Theorem 4 with $m = 2$, and the proof of Theorem 3 is very similar to that of Theorem 1. Let $e_2 = (0, 1, 0)^T$. We calculate the conditional bias and variance of $\hat{\beta}'(t)$

$$\begin{aligned}
\operatorname{bias}(\hat{\beta}'(t) \mid \mathcal{D}) &= -\frac{\sigma(t)}{h\sqrt{nh} f_T(t)} e_2^T \{(S^{-1})_{21} E(\mathbf{W}_{1n}^* \mid \mathcal{D}) + (S^{-1})_{22} E(\mathbf{W}_{2n}^* \mid \mathcal{D})\} \\
&= \frac{\sigma(t)}{nh^2 f_T(t)} \frac{1}{E(XX^T \mid T = t)\mu_2} \sum_{i=1}^n X_i K_i s_i \frac{r_{i,2}}{\sigma(T_i)} \{1 + o_p(1)\} \\
&= \frac{\sigma(t)}{nh^2 f_T(t)} \frac{1}{E(XX^T \mid T = t)\mu_2} \sum_{i=1}^n X_i K_i s_i \frac{X_i^T \beta'''(t) s_i^3 h^3}{6\sigma(T_i)} \{1 + o_p(1)\} \\
&= \frac{1}{6} \beta'''(t) h^2 \frac{\mu_4}{\mu_2} + o_p(h^2).
\end{aligned}$$

Furthermore, we can easily get the conditional covariance of $\hat{\beta}'(t)$,

$$\begin{aligned}
\operatorname{cov}(\hat{\beta}'(t) \mid \mathcal{D}) &= \frac{\sigma^2(t)}{nh^3} \frac{1}{f_T(t_0)} e_2^T (S^{-1} \Sigma S^{-1})_{22} e_2 \\
&= \frac{\sigma^2(t)}{nh^3} \frac{\nu_2 (E(XX^T \mid T = t))^{-1}}{\mu_2^2 f_T(t)} \frac{\sum_{k=1}^q \sum_{k'=1}^q \tau_{kk'}}{(\sum_{k=1}^q f(c_{\tau_k}))^2} + o_p\left(\frac{1}{nh^3}\right) \\
&= \frac{\sigma^2(t)}{nh^3} \frac{\nu_2 (E(XX^T \mid T = t))^{-1}}{\mu_2^2 f_T(t_0)} R_1(q) + o_p\left(\frac{1}{nh^3}\right).
\end{aligned}$$

This completes the proof of Theorem 3.

References

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **95**, 941-956.

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.* **96**, 605-619

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27**, 1491-1518.

Fan, J. and Zhang, W. (2000). Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models. *Scand. J. Statist.* **27**, 715-731.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface* **1**, 179-195.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient Models. *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.

Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *J. Comput. Graph. Statist.* **9**, 60-77.

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111-128.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Knight, K. (1998). Limiting distribution for regression estimators under general conditions. *Ann. Statist.* **26**, 755-770.

Kim, M. O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35**, 92-108.

Kai, B., Li, R. and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. Roy. Statist. Soc. Ser. B* **72**, 49-69.

Parzen, E. (1962). On estimation of a probability density function and model. *Ann. Math. Statist.* **33**, 1065-1076.

Ruppert, D., Sheather, S. J. and Wand, M. P.(1995). An effective bandwidth selector for local least least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257-1270.

Stone, C. J. (1997). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

Tian, M. Z. and Chan, N. H. (2010). Saddle point approximation and volatility estimation of value-at-risk. *Statist. Sinica* **20**.

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388-1402.

Wang, H. J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying-coefficient models. *Ann. Statist.* **37**, 3841-3866.

Zhao, Q. (2001). Asymptotically efficient median regression in the presence of heteroscedasticity of unknown form. *Econom. Theory* **17**, 765-784.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36**, 1108-1126.

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.

E-mail: guojieruc2009@ruc.edu.cn

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.

E-mail: mztian@ruc.edu.cn

National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100864 China.

E-mail: zhukai@bao.ac.cn

(Received October 2010; accepted August 2011)