Statistics and Its Interface Volume 10 (2017) 255–265

Identifying interaction effects via additive quantile regression models

Qianqian Zhu, Yanan Hu, and Maozai Tian*

As additive quantile regression (AQR) models possess the properties of robustness and flexibility, they become increasingly popular in many applications. However, such models may fail when predictors reflect interaction effects in the response. In fact, we often encounter such a problem that the main effects are not significant but the pairwise interactions are in regression. The existence of such a situation is neither accidental nor ignorable. Overlooking the interaction effects may render many of the traditional statistical techniques used for studying data relationships invalid. In these situations, it is necessary to consider more reasonable models such as AQR model with pairwise interactions. This paper mainly studies estimation and testing for the AQR model with pairwise interactions. To estimate the unknown functions in the model, local linear fitting and ordinary backfitting methods are applied. The generalized likelihood ratio (GLR) type test statistic is constructed to test the overall significance of pairwise interactions, and bootstrap method is utilized to approximate the asymptotic distribution of the test statistic. Theoretical properties of estimators and GLR type test statistic are derived. Bandwidth selection based on plug-in method for pairwise interactions is discussed as well. Finally, simulation study and a simple empirical analysis are presented to illustrate the performance of the proposed model.

Keywords and phrases: Additive quantile models, Backfitting algorithm, Bandwidth selection, Generalized likelihood ratio type testing, Pairwise interaction.

1. INTRODUCTION

Suppose that $\{(\mathbf{X}_i^T, Y_i)\}_{i=1}^n$ is an independent and identically distributed (i.i.d.) random sample satisfying

$$Y = m(\mathbf{X}) + \epsilon, \qquad \mathbf{X} \in \mathcal{R}^d, \tag{1}$$

where Y is response, $\mathbf{X} = (X_1, \ldots, X_d)^T$ is a vector of regressors, and ϵ is i.i.d. random error. Consider the following additive structure

$$m(\mathbf{X}) = c + \sum_{\alpha=1}^{d} g_\alpha(X_\alpha), \tag{2}$$

*Corresponding author.

where c is a constant and $g_\alpha(\cdot)$'s are unknown real-valued functions. Additive model (2) could reduce the "curse of dimensionality" in nonparametric modeling and allow for simple interpretation of marginal changes of each regressor. These attractive advantages of additive model make it widely studied and applied in many fields, see Hastie and Tibshirani (1990), Linton and Nielsen (1995), Mammen *et al.* (1999) and Derbort *et al.* (2002) for additive mean regression model. Furthermore, additive quantile regression (AQR) model has attracted more and more attentions since the introduction of quantile regression in Koenker and Bassett (1978), see De Gooijer and Zerom (2003), Yu and Lu (2004), Horowitz and Lee (2005) and Dette and Scheder (2011). Among these literature, ordinary or smoothed backfitting algorithms and marginal integration are commonly used estimation approaches for additive model (2).

Model (2) assumes the underlying model is purely additive, however, this assumption may not be valid. If the additivity assumption does not hold, pairwise interactions could be introduced and then a possibly more reasonable model is the additive model with pairwise interactions

$$m(\mathbf{X}) = c + \sum_{\alpha=1}^{d} g_\alpha(X_\alpha) + \sum_{1\le\alpha<j\le d} g_{\alpha j}(X_\alpha, X_j), \tag{3}$$

where unknown real-valued functions $g_{\alpha j}(X_\alpha, X_j)$'s represent the pairwise interactions between X_α and X_j. Sperlich *et al.* (2002) introduced model (3) and studied its mean regression estimation using marginal integration. As far as we know, few literature have considered quantile regression estimation for model (3). Therefore, in this paper, we consider the following AQR model with pairwise interactions

$$\theta_\tau(\mathbf{x}) = C_\tau + \sum_{\alpha=1}^{d} g_\alpha(x_\alpha) + \sum_{1\le\alpha<j\le d} g_{\alpha j}(x_\alpha, x_j), \tag{4}$$

where $\mathbf{x} \in \mathcal{R}^d$, $\theta_\tau(\mathbf{x}) = \theta_Y(\tau|\mathbf{X}=\mathbf{x})$ and $C_\tau = \theta_\epsilon(\tau|\mathbf{X}=\mathbf{x})$ are the τ-th conditional quantiles of Y and ϵ given $\mathbf{X}$, respectively. It can be noted that, to ensure the identification of C_τ, we assume that the constant term c in model (3) is zero. In this paper, ordinary backfitting algorithm is applied to estimate model (4) for its simple implementation and popular application in additive models.

It is of vital importance to test the existence of pairwise interactions. Many researchers have proposed several test statistics to check the additivity, see Eubank and Hart (1995), Fan *et al.* (2001), Derbort *et al.* (2002) and Fan and Jiang (2005). Among these test statistics, the generalized likelihood ratio (GLR) test statistic introduced by Fan *et al.* (2001) is general for nonparametric testing problems based on function estimation. Goh (2005) further studied GLR type test on conditional quantile processes. In this paper, we apply GLR type test statistic to check the overall significance of pairwise interactions in model (4).

The remainder of this paper is organized as follows. Section 2 introduces the estimation procedure for model (4). We construct the GLR test statistic to check significance of pairwise interactions and apply bootstrap method to approximate its asymptotic distribution in Section 3. In Section 4, theoretical properties of estimators and test statistic are established, and bandwidth selection is also discussed for functions of pairwise interactions. Simulation is conducted in Section 5 and an empirical study of mathematical achievement of senior high school students in Alberta is presented in Section 6 to illustrate the proposed model. Section 7 gives concluding remarks. All proofs are contained in the Appendix.

2. ESTIMATION FOR AQR MODEL WITH PAIRWISE INTERACTIONS

In this section, we study the estimation for model (4). To ensure the model identification, we assume τ-th quantiles of $g_\alpha(X_\alpha)$ and $g_{\alpha j}(X_\alpha, X_j)$ for $1 \le \alpha < j \le d$ to be zero. Otherwise, subtract each function by its τ-th quantile. Then the estimate of τ-th conditional quantile of Y has the form of

$$\hat{\theta}_\tau(\mathbf{x}) = \hat{C}_\tau + \sum_{\alpha=1}^{d} \hat{g}_\alpha(x_\alpha) + \sum_{1\le\alpha<j\le d} \hat{g}_{\alpha j}(x_\alpha, x_j), \tag{5}$$

where $\hat{C}_\tau$, $\hat{g}_\alpha(x_\alpha)$'s and $\hat{g}_{\alpha j}(x_\alpha, x_j)$'s are estimates obtained via the following backfitting algorithm:

1. Step 1, initial estimation.

$$\hat{C}_\tau^{(0)} = \arg\min_{\mu} \sum_{i=1}^{n} \rho_\tau(Y_i - \mu),$$

$$g_\alpha^{(0)}(x_\alpha) = \arg\min_{a} \sum_{i=1}^{n} \rho_\tau\Big(Y_i - \hat{C}_\tau^{(0)} - a - b(X_{\alpha i} - x_\alpha)\Big) \times K\left(\frac{X_{\alpha i} - x_\alpha}{h_\alpha}\right),$$

$$g_{\alpha j}^{(0)}(x_\alpha, x_j) = \arg\min_{a} \sum_{i=1}^{n} \rho_\tau\Big(Y_i - \hat{C}_\tau^{(0)} - \sum_{\alpha=1}^{d} g_\alpha^{(0)}(X_{\alpha i}) - a - b(X_{\alpha i} - x_\alpha) - c(X_{ji} - x_j)\Big) \times \mathcal{K}\left(\frac{X_{\alpha i} - x_\alpha}{h_{\alpha j}}, \frac{X_{ji} - x_j}{h_{\alpha j}}\right),$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, $I(\cdot)$ is the indicator function, $K(\cdot)$ and $\mathcal{K}(\cdot,\cdot)$ are one-dimensional and two-dimensional kernel functions respectively, h_α and $h_{\alpha j}$ are the bandwidths for $g_\alpha(X_\alpha)$ and $g_{\alpha j}(X_\alpha, X_j)$ respectively. Denote $Q_\tau\{g_\alpha^{(0)}(X_{\alpha i})_{i=1}^n\}$ and $Q_\tau\{g_{\alpha j}^{(0)}(X_{\alpha i}, X_{ji})_{i=1}^n\}$ as the τ-th sample quantiles of $g_\alpha^{(0)}(X_{\alpha i})$ and $g_{\alpha j}^{(0)}(X_{\alpha i}, X_{ji})$ for $i = 1, \ldots, n$, respectively. For model identification, center $g_\alpha^{(0)}(x_\alpha)$ and $g_{\alpha j}^{(0)}(x_\alpha, x_j)$ as follows

$$g_\alpha^{*(0)}(x_\alpha) = g_\alpha^{(0)}(x_\alpha) - Q_\tau\{g_\alpha^{(0)}(X_{\alpha i})_{i=1}^n\},$$
$$g_{\alpha j}^{*(0)}(x_\alpha, x_j) = g_{\alpha j}^{(0)}(x_\alpha, x_j) - Q_\tau\{g_{\alpha j}^{(0)}(X_{\alpha i}, X_{ji})_{i=1}^n\}.$$

2. Step 2, iteration.

$$\hat{C}_\tau^{(t)} = \arg\min_{\mu} \sum_{i=1}^{n} \rho_\tau\Big(Y_i - \sum_{\alpha=1}^{d} g_\alpha^{*(t-1)}(X_{\alpha i}) - \sum_{1\le\alpha<j\le d} g_{\alpha j}^{*(t-1)}(X_{\alpha i}, X_{ji}) - \mu\Big),$$

$$g_\alpha^{(t)}(x_\alpha) = \arg\min_{a} \sum_{i=1}^{n} \rho_\tau\Big(Y_i - \hat{C}_\tau^{(t)} - \sum_{l\neq\alpha}^{d} g_l^{*(t-1)}(X_{li}) - a - b(X_{\alpha i} - x_\alpha) - \sum_{1\le\alpha<j\le d} g_{\alpha j}^{*(t-1)}(X_{\alpha i}, X_{ji})\Big) \times K\left(\frac{X_{\alpha i} - x_\alpha}{h_\alpha}\right),$$

$$g_{\alpha j}^{(t)}(x_\alpha, x_j) = \arg\min_{a} \sum_{i=1}^{n} \rho_\tau\Big(Y_i - \hat{C}_\tau^{(t)} - \sum_{\alpha=1}^{d} g_\alpha^{(t)}(X_{\alpha i}) - \sum_{1\le l<m\le d, l\neq\alpha, m\neq j} g_{lm}^{*(t-1)}(X_{li}, X_{mi}) - a - b(X_{\alpha i} - x_\alpha) - c(X_{ji} - x_j)\Big) \times \mathcal{K}\left(\frac{X_{\alpha i} - x_\alpha}{h_{\alpha j}}, \frac{X_{ji} - x_j}{h_{\alpha j}}\right).$$

Then obtain $g_\alpha^{*(t)}(x_\alpha)$ and $g_{\alpha j}^{*(t)}(x_\alpha, x_j)$ as in Step 1.

3. Step 3, update. Iterate Step 2 for $t = 1, 2, 3, \ldots$ until all the values of $\hat{C}_\tau^{(t)}$, $g_\alpha^{*(t)}(x_\alpha)$ and $g_{\alpha j}^{*(t)}(x_\alpha, x_j)$ for $1 \le \alpha < j \le d$ converge.

Note that local linear fitting method (Fan and Gijbels, 1996) is used in the above procedure to approximate each nonparametric function. Local linear estimator outperforms local constant estimator in boundary regions and its fitting results are comparable among local polynomial approximations.

3. TEST STATISTIC FOR INTERACTIONS

In this section, we construct a GLR statistic to test the overall significance of pairwise interactions in model (4) and use bootstrap method to approximate its null distribution.

3.1 GLR type statistic

Consider the following hypothesis test for model (4)

$$H_0 : g_{\alpha j}(x_\alpha, x_j) = 0 \;\; \forall\, 1 \le \alpha < j \le d, \text{ v.s.}$$
$$H_1 : g_{\alpha j}(x_\alpha, x_j) \neq 0 \;\; \exists\, 1 \le \alpha < j \le d.$$

Denote the average fitting residuals under H_0 and H_1 as follows

$$ARS_0(\tau) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\Big(Y_i - \tilde{C}_\tau - \sum_{\alpha=1}^{d} \tilde{g}_\alpha(X_{\alpha i})\Big),$$
$$ARS_1(\tau) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\Big(Y_i - \hat{C}_\tau - \sum_{\alpha=1}^{d} \hat{g}_\alpha(X_{\alpha i}) - \sum_{1\le\alpha<j\le d} \hat{g}_{\alpha j}(X_{\alpha i}, X_{ji})\Big),$$

where $\tilde{C}_\tau$ and $\tilde{g}_\alpha(x_\alpha)$ for $\alpha = 1, 2, \ldots, d$ are estimates obtained for model (4) under H_0 while $\hat{C}_\tau$, $\hat{g}_\alpha(x_\alpha)$ and $\hat{g}_{\alpha j}(x_\alpha, x_j)$ for $1 \le \alpha < j \le d$ are estimates obtained for model (4) under H_1. The GLR type statistic $\lambda(\tau)$ is then constructed as follows

$$\lambda(\tau) = ARS_0(\tau) - ARS_1(\tau). \tag{6}$$

The larger the value of $\lambda(\tau)$ is, the more significant the pairwise interactions will be. In practice, it may be more interesting to study the existence of a specific pairwise interaction between any two regressors. Note that the above test is a nesting test which can be applied to test the significance of a specific pairwise interaction. Therefore, (6) can also be used in a simple test where only one pairwise interaction function involved in the model.

3.2 Bootstrap for testing

It should be noted that it is difficult to calculate critical values of $\lambda(\tau)$ due to unknown quantities in its asymptotic distribution, and the critical values obtained based on small or moderate sample size may not be effective. In these situations, it is reasonable to use bootstrap method to approximate the asymptotic critical values of $\lambda(\tau)$. According to Härdle and Mammen (1993), the standard bootstrap procedure fails in bootstrapping statistic which measures the deviation between two nonparametric fits, however, the wild bootstrap still works in this case. Therefore, we consider the wild bootstrap method which consists of the following steps:

1. Estimate model (4) under H_0 to obtain estimates $\tilde{\theta}_\tau(\mathbf{x})$, and then calculate the residuals $\tilde{u}_i = Y_i - \tilde{\theta}_\tau(\mathbf{X}_i)$;
2. Draw bootstrap residuals $\tilde{u}_i^* = \tilde{u}_i v_i^*$ for $i = 1, 2, \cdots, n$, where v_i^* is randomly sampled from a two-point distribution $F_{(a,b)}$ with $a = -(\sqrt{5}-1)/2$ and $b = (\sqrt{5}+1)/2$, and $v_i^* = a$ with probability $p = (\sqrt{5}+1)/(2\sqrt{5})$ and $v_i^* = b$ with probability $1-p$;
3. Generate a bootstrap sample $(\mathbf{X}_i^T, Y_i^*)$ for $i = 1, 2, \cdots, n$ with $Y_i^* = \tilde{\theta}_\tau(\mathbf{X}_i) + \tilde{u}_i^*$;
4. Calculate the value of test statistic $\lambda^*(\tau)$ based on the bootstrap sample in the same way as defined in (6);
5. Repeat steps 2–4 B times and determine critical values based on B generated $\lambda^*(\tau)$.

4. THEORETICAL PROPERTIES

In this section, we derive theoretical properties for estimators and GLR type test statistic, and discuss the bandwidth selection for pairwise interaction functions.

4.1 Asymptotic distributions of estimators

Denote $F(\cdot|\mathbf{x})$ as the conditional cumulative distribution function (cdf) of ϵ given $\mathbf{X} = \mathbf{x}$ and $f(\cdot|\mathbf{x})$ as the corresponding conditional probability density function (pdf). For $1 \le \alpha < j \le d$, let $f_\alpha(x_\alpha)$ be the marginal pdf of X_α, $f_{\alpha j}(x_\alpha, x_j)$ be the joint pdf of X_α and X_j, $f_\alpha(y|x_\alpha)$ be the conditional pdf of $Y - C_\tau - \sum_{l\neq\alpha} g_l(X_l) - \sum_{1\le\alpha<j\le d} g_{\alpha j}(X_\alpha, X_j)$ given $X_\alpha = x_\alpha$, $f_\alpha^*(y|x_\alpha)$ be the conditional pdf of $Y - C_\tau - \sum_{l\neq\alpha} g_l(X_l)$ given $X_\alpha = x_\alpha$, and $f_{\alpha j}(y|(x_\alpha, x_j))$ be the conditional pdf of $Y - C_\tau - \sum_{l=1}^{d} g_l(X_l) - \sum_{\substack{1\le l<m\le d,\\ l\neq\alpha, m\neq j}} g_{lm}(X_l, X_m)$ given $(X_\alpha, X_j) = (x_\alpha, x_j)$. To obtain asymptotic properties for estimators, assumptions are given below.

Assumption 1. *(i). $f(\cdot|\mathbf{x})$, $f_\alpha(\cdot)$ and $f_{\alpha j}(\cdot,\cdot)$ are continuous and bounded away from 0.*

(ii). $f_\alpha(y|x_\alpha) > 0$. Given $X_\alpha = \tilde{x}_\alpha$, for $\tilde{x}_\alpha$ in a neighborhood of x_α, $f_\alpha(y|\tilde{x}_\alpha)$ is uniformly continuous with respect to y in a neighborhood of $g_\alpha(x_\alpha)$, and is also continuous with respect to $\tilde{x}_\alpha$ for all y in a neighborhood of $g_\alpha(x_\alpha)$.

(iii). $f_{\alpha j}(y|(x_\alpha, x_j)) > 0$. Given $(X_\alpha, X_j) = (\tilde{x}_\alpha, \tilde{x}_j)$, for $(\tilde{x}_\alpha, \tilde{x}_j)$ in a neighborhood of (x_α, x_j), $f_{\alpha j}(y|(\tilde{x}_\alpha, \tilde{x}_j))$ is uniformly continuous with respect to y in a neighborhood of $g_{\alpha j}(x_\alpha, x_j)$, and is also continuous with respect to $(\tilde{x}_\alpha, \tilde{x}_j)$ for all y in a neighborhood of $g_{\alpha j}(x_\alpha, x_j)$.

Assumption 2. *Kernel function $K(\cdot)$ is a bounded density function such that*

$$\int \mathbf{u}K(\mathbf{u})d\mathbf{u} = 0, \quad \mu_2(K) = \int \mathbf{u}^T\mathbf{u}K(\mathbf{u})d\mathbf{u} > 0.$$

Assumption 3. *The bandwidth h_n satisfies, as $n \to \infty$,*

$$h_n \to 0, \quad nh_n^2 \to \infty.$$

Remark. Above assumptions are similar to those in Yu and Lu (2004). Assumption 1 (iii) and $nh_n^2 \to \infty$ in Assumption 3 are assumed for pairwise interaction functions. Subscript n of bandwidth h_n in Assumption 3 reflects the dependence of bandwidth on sample size. In Section 2 and the following sections, we omit the subscript n in h_n for simplicity and instead use subscripts α and αj to denote the corresponding bandwidths for component X_α and pairwise interaction between X_α and X_j, respectively.

Theorem 4.1. *Under Assumptions 1–3, for* $\alpha = 1, \ldots, d$, $\sqrt{nh_\alpha}(\hat{g}_\alpha(x_\alpha) - g_\alpha(x_\alpha) - B_\alpha(x_\alpha)h_\alpha^2)$ *is asymptotically distributed as* $N(0, V_\alpha(x_\alpha))$, *where*

$$B_\alpha(x_\alpha) = \frac{1}{2}g_\alpha^{''}(x_\alpha)\mu_2(K),$$
$$V_\alpha(x_\alpha) = \frac{R(K)}{f_\alpha(x_\alpha)}\frac{\tau(1-\tau)}{[f_\alpha(g_\alpha(x_\alpha)|x_\alpha)]^2},$$

and $g_\alpha(x_\alpha)$ *is the* τ*-th quantile function of* $Y - C_\tau - \sum_{l \neq \alpha} g_l(X_l) - \sum_{1 \le \alpha < j \le d} g_{\alpha j}(X_\alpha, X_j)$ *given* $X_\alpha = x_\alpha$, $g_\alpha^{''}(x_\alpha)$ *is the second derivative of* $g_\alpha(x_\alpha)$, *and* $R(K) = \int K^2(u)du$.

Theorem 4.2. *Under Assumptions 1–3, for* $1 \le \alpha < j \le d$, $\sqrt{nh_{\alpha j}^2}(\hat{g}_{\alpha j}(x_\alpha, x_j) - g_{\alpha j}(x_\alpha, x_j) - B_{\alpha j}(x_\alpha, x_j)h_{\alpha j}^2)$ *is asymptotically distributed as* $N(0, V_{\alpha j}(x_\alpha, x_j))$, *where*

$$B_{\alpha j}(x_\alpha, x_j) = \frac{1}{2}tr\{H_{g_{\alpha j}}(x_\alpha, x_j)\}\mu_2(K),$$
$$V_{\alpha j}(x_\alpha, x_j) = \frac{R(K)}{f_{\alpha j}(x_\alpha, x_j)}\frac{\tau(1-\tau)}{[f_{\alpha j}(g_{\alpha j}(x_\alpha, x_j)|(x_\alpha, x_j))]^2},$$

and $g_{\alpha j}(x_\alpha, x_j)$ *is the* τ*-th quantile function of* $Y - C_\tau - \sum_{l=1}^{d} g_l(X_l) - \sum_{\substack{1 \le l < m \le d,\\ l \neq \alpha, m \neq j}} g_{lm}(X_l, X_m)$ *given* $(X_\alpha, X_j) = (x_\alpha, x_j)$, $H_{g_{\alpha j}}(x_\alpha, x_j)$ *is the Hessian matrix of function* $g_{\alpha j}(x_\alpha, x_j)$, *and* $tr\{H_{g_{\alpha j}}(x_\alpha, x_j)\}$ *is the trace of this Hessian matrix.*

Remark. Theorems 4.1–4.2 illustrate the asymptotic distributions of one-dimensional estimators $\hat{g}_\alpha(x_\alpha)$ for $\alpha = 1, \ldots, d$ and two-dimensional estimators $\hat{g}_{\alpha j}(x_\alpha, x_j)$ for $1 \le \alpha < j \le d$ of model (4). For simplicity, we use equal bandwidth $h_{\alpha j}$ in kernel function $\mathcal{K}(\cdot,\cdot)$. Note that two-dimensional functions $\hat{g}_{\alpha j}(x_\alpha, x_j)$ have slower convergence rate than that of one-dimensional functions $\hat{g}_\alpha(x_\alpha)$.

4.2 Asymptotic distribution of GLR test statistic

To obtain the asymptotic distribution of $\lambda(\tau)$ under H_0, we further assume

Assumption 4. *Estimators* $\hat{g}_\alpha(x_\alpha)$ *for* $\alpha = 1, \ldots, d$ *have the same convergence rate under the alternative hypothesis* H_1 *with that under the null hypothesis* H_0.

Theorem 4.3. *Based on Assumptions 1–4, Theorems 4.1–4.2, for* $1 \le \alpha < j \le d$, *under* H_0,

$$\left[V_\lambda^{-1/2}(\tau)\left(\lambda(\tau) - B_\lambda(\tau)\right)|\chi_n\right] \to N(0,1)$$

in distribution as $n \to \infty$, *where*

$$B_\lambda(\tau) = \frac{1}{2n}\sum_{i=1}^{n} f(C_\tau|\mathbf{X}_i)\left[\left(\tilde{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\right)^2 - \left(\hat{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\right)^2\right],$$
$$V_\lambda(\tau) = \frac{\tau(1-\tau)}{n^2}\sum_{i=1}^{n}\left(\hat{\theta}_\tau(\mathbf{X}_i) - \tilde{\theta}_\tau(\mathbf{X}_i)\right)^2,$$

with $\chi_n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, $\theta_\tau(\mathbf{X}_i) = C_\tau + \sum_{\alpha=1}^{d} g_\alpha(X_{\alpha i})$, $\tilde{\theta}_\tau(\mathbf{X}_i) = \tilde{C}_\tau + \sum_{\alpha=1}^{d}\tilde{g}_\alpha(X_{\alpha i})$, *and* $\hat{\theta}_\tau(\mathbf{X}_i) = \hat{C}_\tau + \sum_{\alpha=1}^{d}\hat{g}_\alpha(X_{\alpha i}) + \sum_{1 \le \alpha < j \le d}\hat{g}_{\alpha j}(X_{\alpha i}, X_{ji})$.

4.3 Bandwidth selection

As we know, bandwidth selection has much more influence on nonparametric estimation than that of kernel selection. For one-dimensional functions in AQR model, Yu and Lu (2004) have proposed a heuristic rule to select the bandwidth. In this paper, we mainly consider the bandwidth selection for pairwise interaction functions.

From Theorem 4.2, the AMSE of $\hat{g}_{\alpha j}(x_\alpha, x_j)$ has the form of

$$AMSE_{\alpha j} = B_{\alpha j}^2(x_\alpha, x_j)h_{\alpha j}^4 + \frac{1}{nh_{\alpha j}^2}V_{\alpha j}(x_\alpha, x_j). \tag{7}$$

Minimize $AMSE_{\alpha j}$ with respect to $h_{\alpha j}$, we have

$$h_{\alpha j,opt}(\tau) = \left\{\frac{2\tau(1-\tau)R(K)}{n\mu_2^2(K)[tr\{H_{g_{\alpha j,\tau}}(x_\alpha, x_j)\}]^2} \times \frac{1}{f_{\alpha j}(x_\alpha, x_j)[f_{\alpha j}(g_{\alpha j,\tau}(x_\alpha, x_j)|(x_\alpha, x_j))]^2}\right\}^{\frac{1}{6}}. \tag{8}$$

Similar to Yu and Jones (1998) and Yu and Lu (2004), the optimal bandwidths for different quantiles have the following relationship

$$\left[\frac{h_{\alpha j,opt}(\tau_1)}{h_{\alpha j,opt}(\tau_2)}\right]^6 = \frac{\tau_1(1-\tau_1)[tr\{H_{g_{\alpha j,\tau_2}}(x_\alpha, x_j)\}]^2}{\tau_2(1-\tau_2)[tr\{H_{g_{\alpha j,\tau_1}}(x_\alpha, x_j)\}]^2} \times \frac{[f_{\alpha j}(g_{\alpha j,\tau_2}(x_\alpha, x_j)|(x_\alpha, x_j))]^2}{[f_{\alpha j}(g_{\alpha j,\tau_1}(x_\alpha, x_j)|(x_\alpha, x_j))]^2}. \tag{9}$$

According to Yu and Lu (2004), the second derivatives of any two quantiles will often be very similar, so it is reasonable to set the traces of Hessian matrices to be equal for any two quantiles. Assume that $f_{\alpha j}(g_{\alpha j}(x_\alpha, x_j)|(x_\alpha, x_j))$ is the joint density of two independently and normally distributed variables with means μ_α, μ_j and variances σ_α^2, σ_j^2, respectively. Then we have

$$\left[\frac{h_{\alpha j,opt}(\tau_1)}{h_{\alpha j,opt}(\tau_2)}\right]^6 = \frac{\tau_1(1-\tau_1)[\phi(\Phi^{-1}(\tau_2))]^4}{\tau_2(1-\tau_2)[\phi(\Phi^{-1}(\tau_1))]^4}. \tag{10}$$

Consider $\tau_2 = 0.5$, then

$$h^6_{\alpha j,opt}(\tau) = \frac{\tau(1-\tau)}{\pi^2[\phi(\Phi^{-1}(\tau))]^4} h^6_{\alpha j,opt}(0.5). \tag{11}$$

If we can obtain the estimate for optimal median bandwidth $h_{\alpha j,opt}(0.5)$, then the optimal bandwidth for pairwise interaction functions at any quantile can be calculated via (11). According to Scott's rule (Scott, 1992), the optimal bandwidth can be approximated by $\hat{h} = n^{-1/(d+4)}\hat{\sigma}$, where d is the dimension. In our case, $d = 2$ and then we can use $\hat{h}_{\alpha j,opt}(0.5) = n^{-1/6}\hat{\sigma}_{\alpha j}$ to approximate the optimal median bandwidth, where $\hat{\sigma}_{\alpha j} = (\hat{\sigma}_\alpha + \hat{\sigma}_j)/2$ for simplicity.

Remark. The above discussion provides a possible approach to select the optimal bandwidth for pairwise interactions. However, bandwidth selection is much more complex in practice and the obtained optimal bandwidth may not lead to the global optimal estimation. It would be better to try different bandwidths in practice. In the simulation and real data analysis, we use the canonical bandwidth of Gaussian kernel for simplicity.

5. SIMULATION

In this section, we implement three simulation experiments to illustrate the finite sample performance of the proposed model, estimation procedure and the GLR test statistic. Consider the following data generating model with two regressors ($d = 2$) and one pairwise interaction

$$Y = g_1(X_1) + g_2(X_2) + g_{12}(X_1, X_2) + \epsilon, \tag{12}$$

where $g_1(X_1) = 0.75X_1$, $g_2(X_2) = 1.5\sin(0.5\pi X_2)$, and $g_{12}(X_1, X_2) = aX_1X_2$ with X_1 and X_2 independently following uniform distribution $U(-2, 2)$ and a being a constant. Its corresponding AQR model with pairwise interaction is

$$\theta_\tau(\mathbf{x}) = C_\tau + g_1(x_1) + g_2(x_2) + g_{12}(x_1, x_2). \tag{13}$$

The first experiment is conducted to compare the performance of AQR models without and with pairwise interaction. Data is generated via (12) with $a = 2$ and ϵ following standard normal distribution $N(0, 1)$ for sample sizes $n = 80, 200$ and 500, respectively. Simulation is implemented for 100 replications and quantiles $\tau = 0.25, 0.5$ and 0.75 are considered in estimation. We apply AQR model with pairwise interaction and AQR model without pairwise interaction to the data respectively. Gaussian kernel $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and product kernel $\mathcal{K}(u_1, u_2) = K(u_1) \cdot K(u_2)$ are used with the canonical bandwidth $h_0 = 0.7764$ corresponding to Gaussian kernel. To assess the performance of the estimation procedure,

Table 1. MAAE under H_1 and H_0

		$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
	MAAE_1	0.683	0.281	0.905
$n = 80$	MAAE_0	2.962	2.947	3.388
	MAAE_1	0.627	0.263	0.659
$n = 200$	MAAE_0	2.376	2.337	3.098
	MAAE_1	0.614	0.181	0.611
$n = 500$	MAAE_0	2.202	2.191	2.430

we calculate the average absolute error (AAE) $AAE = n^{-1}\sum_{i=1}^n |\theta_\tau(x_i) - \hat{\theta}_\tau(x_i)|$ for each replication and then obtain the mean average absolute error (MAAE) $MAAE = 100^{-1}\sum_{j=1}^{100} AAE_j$ in 100 replications.

Table 1 summarizes the values of $MAAE$ for AQR model without pairwise interaction ($MAAE_0$) and AQR model with pairwise interaction ($MAAE_1$) for different values of τ and different sample sizes. From Table 1, it is noted that for the same sample size, both values of $MAAE_1$ and $MAAE_0$ are the smallest when $\tau = 0.5$, which illustrates the best performance of median regression. For the same τ, the values of $MAAE_1$ are significantly smaller than those of $MAAE_0$, which indicates that the proposed model outperforms the AQR model without pairwise interactions if pairwise interactions do exist. Moreover, it is indicated that, for the same τ, the values of $MAAE$ decrease as the sample size increases, which is reasonable since the estimated model is more accurate for larger sample size.

To illustrate estimation performance of the proposed model, we choose estimates of each function with respect to the 25%, 50% and 75% percentiles of AAE's for $\tau = 0.5$ and $n = 200$, see Figure 1. From Figure 1, it can be seen that the performance gets worse as the value of AAE increases, which is reasonable since the smaller value of AAE indicates more accurate estimation results. Although the boundary effects which are common in kernel regression also occur in our estimation, the estimation procedure has desirable results for both one-dimensional functions and pairwise function.

The second experiment is to illustrate the power of the GLR test statistic in Section 3. Data is generated via (12) with $a = 0.25, 0.5, 0.75$ and 1 for $n = 80$ and ϵ following $N(0, 1)$. Consider the following test for model (13)

$$H_0 : g_{12}(x_1, x_2) = 0 \text{ v.s. } H_1 : g_{12}(x_1, x_2) \neq 0.$$

As in Section 3, we construct $\lambda(\tau)$ and use bootstrap method with 100 bootstrap samples to approximate its asymptotic distribution under H_0. The empirical rejection rates (ERR) of $\lambda(\tau)$ calculated based on 100 replications under significance level $\alpha = 5\%$ are summarized in Table 2 for different values of a and $\tau = 0.25, 0.5$ and 0.75. From Table 2, we can see that, for the same τ, the values of ERR become larger as a increases, which is consistent with the fact that there is greater power to detect larger deviation from H_0. Secondly, the values of ERR approach 1 when $a = 0.75$ and $a = 1$,

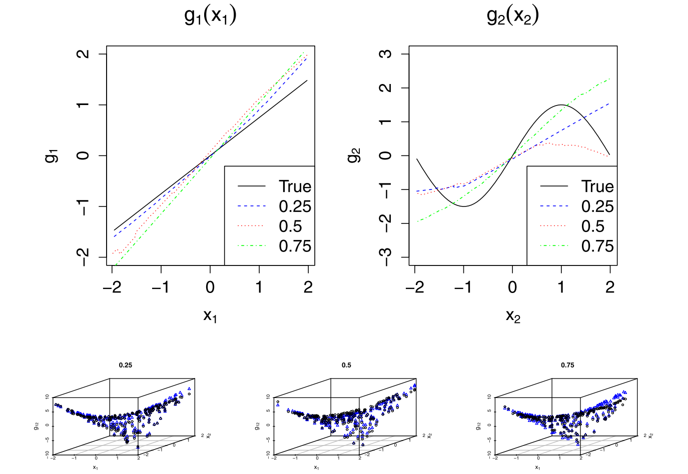

Figure 1. True functions and estimates at quantile $\tau = 0.5$. In the first row, the black solid lines denote the true functions of $g_1(x_1)$ and $g_2(x_2)$, the blue dashed, red dotted and green dotdashed lines denote their estimates corresponding to $25\%, 50\%, 75\%$ percentiles of AAE's. In the second row, the blue triangles denote the true function of $g_{12}(x_1, x_2)$ while the black circles denote the estimates. The titles of 3-d graphs denote the $25\%, 50\%, 75\%$ percentiles of AAE's.

Table 2. ERR of $\lambda(\tau)$ for $\alpha = 5\%$

	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
$a = 0.25$	0.68	0.85	0.63
$a = 0.5$	0.90	0.97	0.88
$a = 0.75$	0.97	1.00	0.97
$a = 1$	0.99	1.00	0.99

which shows the desirable performance of $\lambda(\tau)$. Moreover, for the same a the values of ERR are largest when $\tau = 0.5$, which indicates that $\lambda(\tau)$ has the greatest power in median regression.

In the third experiment, data is generated via (12) with $a = 2$ and ϵ following $N(0, 1)$ and standard Cauchy distribution denoted as $C(0, 1)$ for sample size $n = 80, 200$ and 500, respectively. To compare performance of median regression and mean regression methods for additive model with pairwise interactions, we apply these two methods to the data. The $MAAE$'s are listed in Table 3. Table 3 indicates that median regression has much better performance than that of mean regression when errors are from Cauchy distribution. When errors are normal, it can be seen that mean regression outperforms median regression, however, the difference will narrow down as the sample size increases.

6. AN EMPIRICAL DATA ANALYSIS

In this section, we apply the proposed model to a data set of mathematical achievement of senior high school students in Alberta, Canada. The data is from Canadian Center for Advanced Studies of National Databases. Researchers mainly used mean regression method to study the relationship between mathematical achievement of students and other factors including social, economic, and cultural ones. Tian (2006) analyzed this data via quantile regression approach to investigate the effects of family background factors on mathematical achievement. In this paper, we aim to check whether there are pairwise interactions among family background factors on mathematical achievement of students. We use a subset of $n = 180$ observations of the original data set. The response is the mathematical achievement of students. The regressors of interest are gender, mother's socioeconomic status and father's socioeconomic status denoted by x_1, x_2 and x_3, respectively. Gender takes value 1 if student is female and 0 if male, and parents' socioeconomic status is measured by the International Socioeconomic Index (ISEI). We first consider the following AQR model with pairwise interaction between x_1 and x_2

$$\theta_\tau(\mathbf{x}) = C_\tau + g_1(x_1) + g_2(x_2) + g_3(x_3) + g_{12}(x_1, x_2),$$

and the corresponding hypotheses

$$H_0 : g_{12}(x_1, x_2) = 0 \text{ v.s. } H_1 : g_{12}(x_1, x_2) \neq 0,$$

which can be used to test the significance of pairwise interaction between x_1 (gender) and x_2 (mother's socioeconomic status).

We estimate the model for $\tau = 0.5$ and use Gaussian kernel and product kernel with canonical bandwidths as in Section 5. The 95% confidence intervals are constructed for one-dimensional functions via bootstrap method. The estimates of $g_1(x_1)$ are 0 and -3 for male and female students respectively, which indicates that male students tend to outperform female students in mathematical achievement. For functions $g_2(x_2)$, $g_3(x_3)$ and $g_{12}(x_1, x_2)$, their estimation results are shown in Figure 2. From Figure 2, it can be seen that the rise of parents' socioeconomic status may not bring benefits for the mathematical achievement of their children. Moreover, the estimated interaction function illustrates that, as mother's socioeconomic status enhances, girls' mathematical achievement improves on the whole while boys' mathematical achievement improves first and then declines.

As in Section 3, we construct the GLR test statistic for the above hypothesis test and use wild bootstrap with 100 bootstrap samples to approximate its asymptotic critical

Table 3. MAAE for median regression and mean regression

		$n = 80$	$n = 200$	$n = 500$
	QR	0.416	0.327	0.270
$\epsilon \sim N(0, 1)$	OLS	0.366	0.315	0.288
	QR	0.592	0.385	0.304
$\epsilon \sim C(0, 1)$	OLS	7.473	4.604	11.387

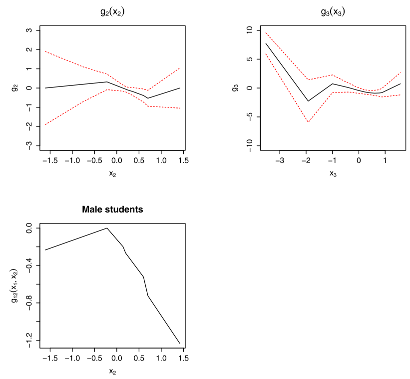

Figure 2. Estimates of functions at quantile $\tau = 0.5$. The first two graphs demonstrate the estimates of one-dimensional functions $g_2(x_2)$ and $g_3(x_3)$, and the last graph illustrates the estimates of $g_{12}(x_1, x_2)$ given $x_1 = 0$.

values. The value of the GLR test statistic is 0.080, which is larger than its critical value 0.056 under 5% significance level, which means that $g_{12}(x_1, x_2)$ is significant under 5% significance level. Therefore, we may conclude that there is pairwise interaction between gender and mother's socioeconomic status on students' mathematical achievement.

Similarly, we can consider the AQR model with pairwise interaction $g_{13}(x_1, x_3)$ or $g_{23}(x_2, x_3)$ respectively and test their statistical significance. The value of the GLR test statistic is 0.102 for $g_{13}(x_1, x_3)$, which is larger than its critical value 0.082 under 5% significance level. And the value of the GLR test statistic is 0.121 for $g_{23}(x_2, x_3)$, which is also larger than its critical value 0.112 under 5% significance level. Hence, it can be seen that there are pairwise interactions between gender and father's socioeconomic status or among parents' socioeconomic status on the mathematical achievement of students.

7. CONCLUDING REMARKS

AQR model with pairwise interactions could combine advantages of quantile regression and additive model with pairwise interactions. We use backfitting algorithm for model estimation and discuss bandwidth selection for pairwise interactions. GLR type test statistic is constructed to check the significance of pairwise interactions and wild bootstrap is applied to approximate its asymptotic distribution. Simulation study and empirical analysis illustrate good performance of the proposed estimation and testing methods.

As functional data analysis becomes increasingly popular, it is meaningful and interesting to apply AQR models to functional data. The difficulty is to deal with the continuum of function values over the entire time domain, see Müller and Yao (2008) for details. They used additive functional principle components (FPC) to overcome this difficulty in mean regression framework. However, for quantile regression, the estimation procedure will be much more complicated due to the difficulty in transforming the objective additive functional model into an additive FPC model. Moreover, extending AQR model with pairwise interactions to functional data will be even more complex since the interactions of functional predictors involved over the entire time domain. In the future, application of AQR models to functional data will be our potential research.

ACKNOWLEDGEMENTS

We are grateful to the Co-Editor Prof. Yuedong Wang and two anonymous referees for their valuable comments and constructive suggestions that led to the substantial improvement of this paper.

The work was partially supported by the major research projects of philosophy and social science of the Chinese Ministry of Education (No.15JZD015), National Natural Science Foundation of China (No.11271368), Project supported by the Major Program of Beijing Philosophy and Social Science Foundation of China (No. 15ZDA17), Project of Ministry of Education supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20130004110007), The Key Program of National Philosophy and Social Science Foundation Grant (No.13AZD064). The work was partially supported by the Major Project of Humanities Social Science Foundation of Ministry of Education (No.15JJD910001), The Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 15XNL008), The Project of Flying Apsaras Scholar of Lanzhou University of Finance & Economics, and The Project of Tianshan Mountain Scholar of Xinjiang University of Finance & Economics.

APPENDIX

In the Appendix, we first give Lemmas 1–2 which are useful to the proofs of Theorems 4.1–4.3 and then prove Theorems 4.1–4.3 in detail.

Lemma 1. *Assume Assumptions 1–3 are satisfied and quantile function $q_\tau(x)$ is continuous at x, then for any $x = x_\alpha$ and $q_\tau(x) = g_\alpha(x_\alpha)$ with $Y_i^{(\alpha)} = Y_i - C_\tau - \sum_{l \neq \alpha} g_l(X_{li}) - \sum_{1 \leq \alpha < j \leq d} g_{\alpha j}(X_{\alpha i}, X_{ji})$, the following Bahadur representation holds*

$$\sqrt{nh_\alpha}(\hat{q}_\tau(x) - q_\tau(x)) = \frac{\phi_\tau}{\sqrt{nh_\alpha}} \sum_{i=1}^{n} \psi_\tau(Y_i^{(\alpha)*}) \times K\left(\frac{X_{\alpha i} - x}{h_\alpha}\right) + o_p(1), \quad (1)$$

as $n \to \infty$, where $\psi_\tau(y) = \tau - I(y < 0)$, $Y_i^{(\alpha)*} = Y_i^{(\alpha)} - q_\tau(x) - \frac{\partial q_\tau(x)}{\partial x_\alpha}(X_{\alpha i} - x_\alpha)$ and $\phi_\tau \equiv \phi_\tau(x) = (f_\alpha(q_\tau(x)|x) f_\alpha(x))^{-1}$.

Lemma 2. *Assume Assumptions 1–3 are satisfied and quantile function $q_\tau(\mathbf{x})$ is continuous at $\mathbf{x}$, then for any $\mathbf{x} = (x_\alpha, x_j)$ and $q_\tau(\mathbf{x}) = g_{\alpha j}(x_\alpha, x_j)$ with $Y_i^{(\alpha,j)} = Y_i - C_\tau - \sum_{l=1}^{d} g_l(X_{li}) - \sum_{\substack{1\le l<m\le d,\\ l\neq\alpha, m\neq j}} g_{lm}(X_{li}, X_{mi})$ and $h = h_{\alpha j}$, the following Bahadur representation holds*

$$\sqrt{nh^2}(\hat{q}_\tau(\mathbf{x}) - q_\tau(\mathbf{x})) = \frac{\phi_\tau}{\sqrt{nh^2}} \sum_{i=1}^{n} \psi_\tau(Y_i^{(\alpha,j)*}) \times K\left(\frac{X_{\alpha i} - x_\alpha}{h}, \frac{X_{ji} - x_j}{h}\right) + o_p(1), \tag{2}$$

as $n \to \infty$, where $\psi_\tau(y) = \tau - I(y < 0)$, $Y_i^{(\alpha,j)*} = Y_i^{(\alpha,j)} - q_\tau(\mathbf{x}) - \frac{\partial q_\tau(\mathbf{x})}{\partial x_\alpha}(X_{\alpha i} - x_\alpha) - \frac{\partial q_\tau(\mathbf{x})}{\partial x_j}(X_{ji} - x_j)$ and $\phi_\tau \equiv \phi_\tau(\mathbf{x}) = (f_{\alpha j}(q_\tau(\mathbf{x})|\mathbf{x}) f_{\alpha j}(\mathbf{x}))^{-1}$.

Lemma 1 and Lemma 2 are special cases of Theorem 3.2 in Lu *et al.* (2000), so we omit the detailed proof here.

To clearly illustrate the proofs of Theorems 4.1–4.2, consider the following AQR model with one pairwise interaction

$$Q_\tau(\mathbf{x}) = C_\tau + g_1(X) + g_2(Z) + g_{12}(X, Z). \tag{3}$$

Under identification conditions that all τ-th quantiles of $g_\alpha(X_\alpha)$, $\alpha = 1, 2, \ldots, d$, and $g_{\alpha j}(X_\alpha, X_j)$, $1 \le \alpha < j \le d$, are zero, the following equations hold

$$Q_\tau(Y - C_\tau - g_2(Z) - g_{12}(X, Z)|X) = g_1(X), \tag{4}$$
$$Q_\tau(Y - C_\tau - g_1(X) - g_{12}(X, Z)|Z) = g_2(Z), \tag{5}$$
$$Q_\tau(Y - C_\tau - g_1(X) - g_2(Z)|(X, Z)) = g_{12}(X, Z). \tag{6}$$

Proof of Theorem 4.1. Let $Y_i^{(1)} = Y_i - C_\tau - g_2(Z_i) - g_{12}(X_i, Z_i)$, $Y_i^{(2)} = Y_i - C_\tau - g_1(X_i) - g_{12}(X_i, Z_i)$ and $Y_i^{(1,2)} = Y_i - C_\tau - g_1(X_i) - g_2(Z_i)$. Denote $g'(x)$ and $g''(x)$ as the first and second derivatives of function $g(x)$, respectively. For $h = h_1$, set $X_{hi} = \frac{X_i - x}{h}$, $K_i = K(X_{hi})$ and $T_{ni} = g_1'(x) X_{hi} h$. Lemma 1 and (4) entail that

$$\begin{aligned}
&\hat{g}_1(x) - g_1(x)\\
&= \frac{1}{nh}\phi_\tau^{(1)} \sum_{i=1}^{n} \psi_\tau(Y_i^{(1)*}) K_i + o_p\left(\frac{1}{\sqrt{nh}}\right)\\
&= \frac{1}{nh}\phi_\tau^{(1)} \sum_{i=1}^{n} \left[\psi_\tau(Y_i^{(1)*}) K_i - E\psi_\tau(Y_i^{(1)*}) K_i\right]\\
&\quad + \frac{1}{nh}\phi_\tau^{(1)} \sum_{i=1}^{n} E\psi_\tau(Y_i^{(1)*}) K_i + o_p\left(\frac{1}{\sqrt{nh}}\right)\\
&= Q_1 + Q_2 + o_p\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned} \tag{7}$$

where $Y_i^{(1)*} = Y_i^{(1)} - g_1(x) - g_1'(x)(X_i - x)$ and $\phi_\tau^{(1)} = \phi_\tau^{(1)}(x) = (f_1(g_1(x)|x) f_1(x))^{-1}$.

From (4), note that $\tau = F_1(g_1(X_i)|X_i)$ and when $K_i > 0$, there exists $0 < \theta < 1$ such that

$$\begin{aligned}
\Delta_i(x) &= g_1(X_i) - g_1(x) - T_{ni}\\
&= \frac{1}{2}[g_1''(x + \theta X_{hi} h)](X_{hi} h)^2.
\end{aligned} \tag{8}$$

Using (8), it follows that there exists some $0 < \xi < 1$ such that

$$\begin{aligned}
Q_2 &= \frac{\phi_\tau^{(1)}}{h} E(\tau - I(Y_i^{(1)*} < 0)) K_i\\
&= \frac{\phi_\tau^{(1)}}{h} E\Big\{[F_1(g_1(X_i)|X_i) - F_1(g_1(x) + T_{ni}|X_i)] K_i\Big\}\\
&= \frac{\phi_\tau^{(1)}}{h} E\Big[f_1(g_1(x) + T_{ni} + \xi\Delta_i(x)|X_i)\Delta_i(x) K_i\Big]\\
&= (1 + o(1))\left[\frac{1}{2} g_1''(x) h^2 \phi_\tau^{(1)} f_1(g_1(x)|x) f_1(x) \mu_2(K)\right]\\
&= (1 + o(1)) h^2 \left[\frac{1}{2} g_1''(x) \mu_2(K)\right].
\end{aligned} \tag{9}$$

Set $\nu_i = \phi_\tau^{(1)} \psi_\tau(Y_i^{(1)*}) K_i$ and $I_i(\tau) = I(Y_i^{(1)*}(\tau) < 0)$, then

$$\begin{aligned}
EQ_1^2 &= \frac{1}{(nh)^2} E\left[\sum_{i=1}^{n} (\nu_i - E\nu_i)^2 + \sum_{i\neq j} (\nu_i - E\nu_i)(\nu_j - E\nu_j)\right]\\
&= \frac{1}{nh^2} Var(\nu_1).
\end{aligned}$$

Note that $\nu_i^2 = (\tau^2 - 2\tau I_i(\tau) + I_i(\tau))(\phi_\tau^{(1)} K_i)^2$, then

$$\begin{aligned}
E\nu_i^2 &= E\Big[(\tau^2 - 2\tau I_i(\tau) + I_i(\tau))(\phi_\tau^{(1)} K_i)^2\Big]\\
&= E\Big[\tau^2 - 2\tau I_i(\tau) + I_i(\tau)\Big] E K_i^2 \cdot (\phi_\tau^{(1)})^2\\
&= \tau(1 - \tau)(\phi_\tau^{(1)})^2 \int K^2\left(\frac{s_1 - x}{h}\right) f_1(s_1) ds_1\\
&= \tau(1 - \tau)(\phi_\tau^{(1)})^2 \int K^2(u_1) f_1(x + u_1 h) h du_1\\
&= (1 + o(1))\tau(1 - \tau)(\phi_\tau^{(1)})^2 h f_1(x) \int K^2(u) du.
\end{aligned}$$

Thus,

$$EQ_1^2 = (1 + o(1)) \frac{1}{nh} \frac{\tau(1 - \tau) \int K^2(u) du}{[f_1(g_1(x)|x)]^2 f_1(x)}. \tag{10}$$

Based on the central limit theorem, from (7), (9) and (10) we can obtain Theorem 4.1. □

Proof of Theorem 4.2. Set $\mathbf{u} = (\frac{X_i - x}{h}, \frac{Z_i - z}{h})^T$, $\chi = (x, z)^T$, $K_i = K(\mathbf{u})$, and $T_{ni} = (X_i - x, Z_i - z)\nabla_{g_{12}}(x, z)$, where $h = h_{12}$ is the pairwise bandwidth, and $\nabla_{g_{12}}(x, z) = (\frac{\partial g_{12}(x, z)}{\partial x}, \frac{\partial g_{12}(x, z)}{\partial z})^T$ is the gradient vector. Denote

$H_{g_{12}}(x,z)$ as the Hessian matrix with second partial derivatives of function $g_{12}(x,z)$ as elements. Lemma 2 and (6) entail that

$$\begin{aligned}
&\hat{g}_{12}(x,z) - g_{12}(x,z)\\
&\quad= \frac{1}{nh^2}\phi_\tau^{(1,2)} \sum_{i=1}^n \psi_\tau(Y_i^{(1,2)*})K_i + o_p(\frac{1}{\sqrt{nh^2}})\\
&\quad= \frac{1}{nh^2}\phi_\tau^{(1,2)} \sum_{i=1}^n \left[\psi_\tau(Y_i^{(1,2)*})K_i - E\psi_\tau(Y_i^{(1,2)*})K_i\right]\\
&\qquad + \frac{1}{nh^2}\phi_\tau^{(1,2)} \sum_{i=1}^n E\psi_\tau(Y_i^{(1,2)*})K_i + o_p(\frac{1}{\sqrt{nh^2}})\\
&\quad= Q_1 + Q_2 + o_p(\frac{1}{\sqrt{nh^2}}),
\end{aligned} \tag{11}$$

where $Y_i^{(1,2)*} = Y_i^{(1,2)} - g_{12}(x,z) - \frac{\partial g_{12}(x,z)}{\partial x}(X_i - x) - \frac{\partial g_{12}(x,z)}{\partial z}(Z_i - z) = Y_i^{(1,2)} - g_{12}(x,z) - T_{ni}$ and $\phi_\tau^{(1,2)} = \phi_\tau^{(1,2)}(x,z) = (f_{12}(g_{12}(x,z)|(x,z))f_{12}(x,z))^{-1}$. From (6), note that $\tau = F_{12}(g_{12}(X_i,Z_i)|(X_i,Z_i))$ and when $K_i > 0$, there exists $0 < \theta < 1$ such that

$$\begin{aligned}
\Delta_i(x,z) &= g_{12}(X_i,Z_i) - g_{12}(x,z) - T_{ni}\\
&= \frac{1}{2}h^2\mathbf{u}^T H_{g_{12}}(x,z)\mathbf{u}.
\end{aligned} \tag{12}$$

Using (12), it follows that there exists some $0 < \xi < 1$ such that

$$\begin{aligned}
Q_2 &= \frac{1}{h^2}\phi_\tau^{(1,2)} E(\tau - I(Y_i^{(1,2)*} < 0))K_i\\
&= \frac{1}{h^2}\phi_\tau^{(1,2)} E\Big\{\big[F_{12}(g_{12}(X_i,Z_i)|(X_i,Z_i))\\
&\qquad - F_{12}(g_{12}(x,z) + T_{ni}|(X_i,Z_i))\big]K_i\Big\}\\
&= \frac{1}{h^2}\phi_\tau^{(1,2)} E\Big[f_{12}(g_{12}(x,z) + T_{ni} + \xi\Delta_i(x,z)|(X_i,Z_i))\\
&\qquad \times \frac{1}{2}h^2\mathbf{u}^T H_{g_{12}}(x,z)\mathbf{u}K_i\Big]\\
&= (1+o(1))\frac{1}{2h^2}\phi_\tau^{(1,2)} f_{12}(g_{12}(x,z)|(x,z)\int\Big[h^2\mathbf{u}^T\\
&\qquad \times H_{g_{12}}(x,z)\mathbf{u}K\left(\frac{s_1-x}{h},\frac{s_2-z}{h}\right)f_{12}(s_1,s_2)\Big]ds_1ds_2\\
&= (1+o(1))\frac{1}{2h^2}\phi_\tau^{(1,2)} f_{12}(g_{12}(x,z)|(x,z))\\
&\qquad \times \int\big[h^2\mathbf{u}^T H_{g_{12}}(x,z)\mathbf{u}K(u_1,u_2)\\
&\qquad \times f_{12}(x+u_1h, z+u_2h)h^2\big]du_1du_2\\
&= (1+o(1))h^2\left[\frac{1}{2}\int \mathbf{u}^T H_{g_{12}}(x,z)\mathbf{u}K(\mathbf{u})d\mathbf{u}\right].
\end{aligned} \tag{13}$$

Set $\nu_i = \phi_\tau^{(1,2)}\psi_\tau(Y_i^{(1,2)*})K_i$ and $I_i(\tau) = I(Y_i^{(1,2)*}(\tau) < 0)$, then

$$\begin{aligned}
EQ_1^2 &= \frac{1}{(nh^2)^2}E\left[\sum_{i=1}^n(\nu_i - E\nu_i)^2 + \sum_{i\neq j}(\nu_i - E\nu_i)(\nu_j - E\nu_j)\right]\\
&= \frac{1}{nh^4}Var(\nu_1).
\end{aligned}$$

Note that $\nu_i^2 = [(\tau^2 - 2\tau I_i(\tau) + I_i(\tau))(\phi_\tau^{(1,2)})^2K_i^2]$, then

$$\begin{aligned}
E\nu_i^2 &= E[(\tau^2 - 2\tau I_i(\tau) + I_i(\tau))(\phi_\tau^{(1,2)})^2K_i^2]\\
&= E[\tau^2 - 2\tau I_i(\tau) + I_i(\tau)]EK_i^2(\phi_\tau^{(1,2)})^2\\
&= \tau(1-\tau)(\phi_\tau^{(1,2)})^2\int K^2\left(\frac{s_1-x}{h},\frac{s_2-z}{h}\right)\\
&\qquad \times f_{12}(s_1,s_2)ds_1ds_2\\
&= \tau(1-\tau)(\phi_\tau^{(1,2)})^2\int K^2(u_1,u_2)\\
&\qquad \times f_{12}(x+u_1h, z+u_2h)h^2du_1du_2\\
&= (1+o(1))\tau(1-\tau)(\phi_\tau^{(1,2)})^2h^2f_{12}(x,z)\int K^2(u)du.
\end{aligned}$$

Thus,

$$EQ_1^2 = (1+o(1))\frac{1}{nh^2}\frac{\tau(1-\tau)\int K^2(u)du}{[f_{12}(g_{12}(x,z)|(x,z))]^2f_{12}(x,z)}. \tag{14}$$

Based on the central limit theorem, from (11), (13) and (14) we prove Theorem 4.2. □

Proof of Theorem 4.3. Under the null hypothesis H_0, data $\{(\mathbf{X}_i^T, Y_i)\}_{i=1}^n$ satisfies $Y_i = \sum_{\alpha=1}^d g_\alpha(X_\alpha) + \epsilon_i$. Let $\theta_\tau(\mathbf{X}_i) = C_\tau + \sum_{\alpha=1}^d g_\alpha(X_{\alpha i})$ be the quantile function under H_0, $\tilde{\theta}_\tau(\mathbf{X}_i) = \tilde{C}_\tau + \sum_{\alpha=1}^d \tilde{g}_\alpha(X_{\alpha i})$ be the estimate of $\theta_\tau(\mathbf{X}_i)$ under H_0, and $\hat{\theta}_\tau(\mathbf{X}_i) = \hat{C}_\tau + \sum_{\alpha=1}^d \hat{g}_\alpha(X_{\alpha i}) + \sum_{1\le\alpha<j\le d}\hat{g}_{\alpha j}(X_{\alpha i}, X_{ji})$ be the estimate of $\theta_\tau(\mathbf{X}_i)$ under the alternative hypothesis H_1. Denote $\chi_n = (\mathbf{X}_1,\ldots,\mathbf{X}_n)$, $\psi_\tau(u) = \tau - I(u<0)$ and $\rho_\tau(u) = u\psi_\tau(u)$. Then

$$\begin{aligned}
ARS_0(\tau) &= \frac{1}{n}\sum_{i=1}^n \rho_\tau\Big(Y_i - \tilde{\theta}_\tau(\mathbf{X}_i)\Big),\\
ARS_1(\tau) &= \frac{1}{n}\sum_{i=1}^n \rho_\tau\Big(Y_i - \hat{\theta}_\tau(\mathbf{X}_i)\Big).
\end{aligned}$$

Apply Knight equation in Knight (1998) as follows

$$\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \Big(I(u\le s) - I(u\le 0)\Big)ds,$$

then we obtain

$$\begin{aligned}
&ARS_0(\tau)\\
&= \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(\epsilon_{i\tau}^* - (\tilde{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i))\right)
\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\Big(\rho_\tau(\epsilon^*_{i\tau}) - \big(\tilde{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\big)\psi_\tau(\epsilon^*_{i\tau}) + \int_0^{\tilde{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} [I(\epsilon^*_{i\tau} \le t) - I(\epsilon^*_{i\tau} \le 0)]\, dt\Big)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(\rho_\tau(\epsilon^*_{i\tau}) - \big(\tilde{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\big)\psi_\tau(\epsilon^*_{i\tau}) + I_0\right),$$

and

$$ARS_1(\tau) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(\epsilon^*_{i\tau} - (\hat{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i))\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\Big(\rho_\tau(\epsilon^*_{i\tau}) - \big(\hat{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\big)\psi_\tau(\epsilon^*_{i\tau}) + \int_0^{\hat{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} [I(\epsilon^*_{i\tau} \le t) - I(\epsilon^*_{i\tau} \le 0)]\, dt\Big)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(\rho_\tau(\epsilon^*_{i\tau}) - \big(\hat{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\big)\psi_\tau(\epsilon^*_{i\tau}) + I_1\right),$$

where $\epsilon^*_{i\tau} = \epsilon_{i\tau} - C_\tau$ and

$$I_0 = \int_0^{\tilde{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} [I(\epsilon^*_{i\tau} \le t) - I(\epsilon^*_{i\tau} \le 0)]\, dt,$$
$$I_1 = \int_0^{\hat{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} [I(\epsilon^*_{i\tau} \le t) - I(\epsilon^*_{i\tau} \le 0)]\, dt.$$

Therefore,

$$\lambda(\tau) = ARS_0(\tau) - ARS_1(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\big(\hat{\theta}_\tau(\mathbf{X}_i) - \tilde{\theta}_\tau(\mathbf{X}_i)\big)\psi_\tau(\epsilon^*_{i\tau}) + I_0 - I_1\right). \tag{15}$$

Set $Z_{i,\tau} = \big(\hat{\theta}_\tau(\mathbf{X}_i) - \tilde{\theta}_\tau(\mathbf{X}_i)\big)\psi_\tau(\epsilon^*_{i\tau})$. Since $E\psi_\tau(\epsilon^*_{i\tau}) = \tau - E[I(\epsilon_{i\tau} < C_\tau)] = 0$, we have

$$E(Z_{i,\tau}|\mathbf{X}_i) = \big(\hat{\theta}_\tau(\mathbf{X}_i) - \tilde{\theta}_\tau(\mathbf{X}_i)\big)E\psi_\tau(\epsilon^*_{i\tau}) = 0. \tag{16}$$

$$E(I_0|\mathbf{X}_i) = E\Big\{\int_0^{\tilde{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} \big[I(\epsilon^*_{i\tau} \le t) - I(\epsilon^*_{i\tau} \le 0)\big]dt\Big\}$$
$$= \int_0^{\tilde{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} E\big[I(\epsilon^*_{i\tau} \le t) - I(\epsilon^*_{i\tau} \le 0)\big]dt$$
$$= \int_0^{\tilde{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} \big[F(C_\tau + t|\mathbf{X}_i) - F(C_\tau|\mathbf{X}_i)\big]dt$$
$$= \int_0^{\tilde{\theta}_\tau(\mathbf{X}_i)-\theta_\tau(\mathbf{X}_i)} f(C_\tau|\mathbf{X}_i)t\,dt + o(1)$$
$$= \frac{1}{2}f(C_\tau|\mathbf{X}_i)\big(\tilde{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\big)^2 + o(1),$$

and similarly

$$E(I_1|\mathbf{X}_i) = \frac{1}{2}f(C_\tau|\mathbf{X}_i)\big(\hat{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i)\big)^2 + o(1),$$

therefore,

$$E(\lambda(\tau)|\chi_n) = \frac{1}{n}\sum_{i=1}^{n}E[Z_{i,\tau} + (I_0 - I_1)|\mathbf{X}_i] = \frac{1}{2n}\sum_{i=1}^{n}f(C_\tau|\mathbf{X}_i)\Big[(\tilde{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i))^2 - (\hat{\theta}_\tau(\mathbf{X}_i) - \theta_\tau(\mathbf{X}_i))^2\Big] + o(1). \tag{17}$$

We next consider the variance of $\lambda(\tau)$ given χ_n. Under Theorems 4.1–4.2 and Assumption 4,

$$Var(Z_{i,\tau}|\mathbf{X}_i) = E(Z^2_{i,\tau}|\mathbf{X}_i) = \tau(1-\tau)\left(\hat{\theta}_\tau(\mathbf{X}_i) - \tilde{\theta}_\tau(\mathbf{X}_i)\right)^2 = O_p(\frac{1}{nh^2}),$$

and $E[|Z_{i,\tau}|^3|\mathbf{X}_i] = O_p((nh^2)^{-3/2})$, then

$$\frac{\sum_{i=1}^n E[|Z_{i,\tau}|^3|\mathbf{X}_i]}{\big(\sum_{i=1}^n Var(Z_{i,\tau}|\mathbf{X}_i)\big)^{3/2}} = O_p(\frac{1}{\sqrt{n}}) \to 0 \tag{18}$$

in probability as $n \to \infty$, and

$$Var(\lambda(\tau)|\chi_n) = \frac{\tau(1-\tau)}{n^2}\sum_{i=1}^{n}\left(\hat{\theta}_\tau(\mathbf{X}_i) - \tilde{\theta}_\tau(\mathbf{X}_i)\right)^2. \tag{19}$$

According to Lyapunov CLT theorem, from (17), (18) and (19) we obtain Theorem 4.3. □

Received 30 January 2015

REFERENCES

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19(2)**, 760–777. MR1105843

De Gooijer, J. G. and Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *J. Amer. Statist. Assoc.* **98(461)**, 135–146. MR1965680

Derbort, S., Dette, H. and Munk, A. (2002). A test for additivity in nonparametric regression. *Ann. Inst. Statist. Math.* **54(1)**, 60–82. MR1893542

Dette, H. and Scheder, R.(2011). Estimation of additive quantile regression. *Ann. Inst. Stat. Math.* **63**, 245–265. MR2780485

Eubank, R. L. and Hart, J. D. (1995). Testing for additivity in nonparametric regression. *Ann. Statist.* **23(6)**, 1896–1920. MR1389857

Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications.* Chapman and Hall, London. MR1383587

Fan, J. and Jiang, J. (2005). Nonparametric inferences for additive models. *J. Amer. Statist. Assoc.* **100**, 890–907. MR2201017

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153–193. MR1833962

Goh, C. (2005). *Nonparametric inferences on conditional quantile processes.* University of Toronto, Department of Economics Working Paper No. 277.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–1947. MR1245774

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models.* Chapman and Hall, London. MR1082147

Knight, K. (1998). Limiting distribution for L1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–770. MR1626024

Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100(472)**, 1238–1249. MR2236438

Koenker, R. and Xiao, Z. J. (2002). Inference on the quantile regression process. *Econometrica* **4**, 1583–1612. MR1929979

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50. MR0474644

Linton, O. and Nielson, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82(1)**, 93–100. MR1332841

Lu, Z., Hui, Y. V. and Zhao, Q. (2000). Local linear quantile regression under dependence: Bahadur representation and application. *Technical report*, Department of Management Sciences, City University of Hong Kong.

Mammen, E., Linton, O. and Nielson, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443–1490. MR1742496

Marron, J. S. and Nolan, D. (1988). Canonical kernels for density estimation. *Statistics and Probability Letters* **7(3)**, 195–199. MR0980921

Müller, H. G. and Yao, F. (2008). Functional additive models. *J. Amer. Statist. Assoc.* **103(484)**, 1534–1544. MR2504202

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization.* John Wiley & Sons, New York, Chichester. MR1191168

Sperlich, S., Tojsteim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* **18**, 197–251. MR1891823

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348–1360. MR0594650

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040–1053. MR0673642

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13(2)**, 689–705. MR0790566

Stone, C. J. (1986). The dimensional reduction principle for generalized additive models. *Ann. Statist.* **14**, 590–606. MR0840516

Tian, M. Z. (2006). A quantile regression analysis of family background factor effects on mathematical achievement. *Journal of Data Science* **4**, 461–478.

Yu, K. and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics* **31**, 333–346. MR2087829

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228–237. MR1614628

Qianqian Zhu
Center for Applied Statistics
School of Statistics
Renmin University of China
Beijing 100872
China

Department of Statistics and Actuarial Science
The University of Hong Kong
Hong Kong
E-mail address: zhunanapig@126.com

Yanan Hu
School of Statistics
Renmin University of China
Beijing 100872
China
E-mail address: yananhu@139.com

Maozai Tian
Center for Applied Statistics
School of Statistics
Renmin University of China
Beijing 100872
China

School of Statistics
Lanzhou University of Finance and Economics
Lanzhou, 730101
Gansu
China

School of Statistics and Information
Xinjiang University of Finance and Economics
Urumchi 830012
Xinjiang
China E-mail address: mztian@ruc.edu.cn
url: http://stat.ruc.edu.cn/teacher_more.php?id=54&cid=25