Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

MEASURING AND TESTING FOR INTERVAL QUANTILE DEPENDENCE*

By LIPING ZHU[†,‡], YAOWU ZHANG[§] AND KAI XU[§]

Renmin University of China[‡] *and Shanghai University of Finance and Economics*[§]

In this article we introduce the notion of interval quantile independence which generalizes the notions of statistical independence and quantile independence. We suggest an index to measure and test departure from interval quantile independence. The proposed index is invariant to monotone transformations, nonnegative and equals zero if and only if the interval quantile independence holds true. We suggest a moment estimate to implement the test. The resultant estimator is root-n-consistent if the index is positive and n-consistent otherwise, leading to a consistent test of interval quantile independence. The asymptotic distribution of the moment estimator is free of parent distribution, which facilitates to decide the critical values for tests of interval quantile independence. When our proposed index is used to perform feature screening for ultrahigh dimensional data, it has the desirable sure screening property.

1. Introduction. Suppose Y_1 and Y_2 are two univariate random variables, $Q_{Y_1|Y_2}(\tau_1)$ is the τ_1-th quantile of Y_1 conditional on Y_2 and $Q_{Y_1}(\tau_1)$ is the unconditional τ_1-th quantile of Y_1. The τ_1-th quantile of Y_1 is independent of Y_2 if $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$, and Y_1 is independent of Y_2 if $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$ for all $\tau_1 \in (0,1)$. In other words, the difference between $Q_{Y_1|Y_2}(\tau_1)$ and $Q_{Y_1}(\tau_1)$ characterizes the deviation from quantile independence at a single τ_1 and statistical independence for all $\tau_1 \in (0,1)$. Characterizing the difference between $Q_{Y_1|Y_2}(\tau_1)$ and $Q_{Y_1}(\tau_1)$ requires to estimate $Q_{Y_1|Y_2}(\tau_1)$ and $Q_{Y_1}(\tau_1)$ and test whether $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$ at a single τ_1 or for all $\tau_1 \in (0,1)$. Estimating the unconditional quantile $Q_{Y_1}(\tau_1)$ is straightforward. However, estimating the conditional quantile function $Q_{Y_1|Y_2}(\tau_1)$ is nontrivial and has received considerable attention in the past two decades, by assuming either $Q_{Y_1|Y_2}(\tau_1)$ is a linear [14, 15] or nonlinear function of Y_2 [8, 11]. In contrast to estimation, testing whether $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$ received little attention in the literature, partly because the covariance structures of the quantile estimates are complicated. [17] proposed a linear quantile correlation coefficient, defined as $\mathrm{qcorr}_{\tau_1}(Y_1 \mid Y_2) \stackrel{\mathrm{def}}{=} \mathrm{cov}\{I(Y_1 \geq Q_{Y_1}(\tau_1)), Y_2\}/\{\tau_1(1-\tau_1)\mathrm{var}(Y_2)\}^{1/2}$, to test whether $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$. However, the test based on quantile correla-

*The authors thank the Editor, the Associate Editor and the three anonymous reviewers for their constructive comments, which have led to a dramatic improvement of the earlier version of this article. The authors also thank Professor Ai Chunrong from University of Florida for sharing us the happiness-income dataset and Professor Hu Feifang from George Washington University for giving the name "Interval Quantile Independence".

[†]This work was supported by National Natural Science Foundation of China (11371236 and 11422107), MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (16JJD910002) and National Youth Top-notch Talent Support Program.

MSC 2010 subject classifications: Primary 62G10, 62H20; secondary 68Q32

Keywords and phrases: correlation, independence, quantile regression, rank test, sure screening property

tion is possibly inconsistent if $Q_{Y_1|Y_2}(\tau_1)$ is a nonlinear function of Y_2.

In this article we aim to measure and test the departure from

$$H_0: \quad Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1) \text{ for } (\tau_1,\tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2 \subseteq (0,1) \otimes (0,1), \tag{1.1}$$
$$\text{versus } H_1 : \text{others.}$$

We refer to H_0 in (1.1) as the *interval quantile independence* because both $\mathcal{I}_1$ and $\mathcal{I}_2$ can be intervals and we are concerned with quantile independence over two intervals. In particular, if $\mathcal{I}_1$ is a singleton, say, $\mathcal{I}_1 = \{\tau_1\}$, and $\mathcal{I}_2 = (0,1)$, H_0 in (1.1) boils down to the *quantile independence* H_0 : $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$ at a single τ_1. If $\mathcal{I}_1 = \mathcal{I}_2 = (0,1)$, then H_0 in (1.1) reduces to the *statistical independence* $H_0 : Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$ for all $\tau_1 \in (0,1)$. In this sense, the interval quantile independence defined in H_0 of (1.1) bridges the gap between quantile independence and statistical independence by choosing $\mathcal{I}_1 \subseteq (0,1)$ and $\mathcal{I}_2 = (0,1)$. The concept of interval quantile independence generalizes the notions of both quantile independence and statistical independence.

The interval quantile dependence allows practitioners to draw interpretable conclusions obtained through various quantile ranges $\mathcal{I}_1 \otimes \mathcal{I}_2$. In what follows we illustrate the usefulness of the interval quantile dependence through two motivating examples.

1. *Hypertension study*: It is common knowledge that hypertension is age related, possibly due to reduction in vascular compliance and stiffening of the arteries. However, aging effect on the systolic blood pressure is possibly different for young, middle-aged and old people. In other words, how the systolic blood pressure varies with age may vary at different stages. Measuring aging effect at different ages amounts to testing departure from interval quantile independence for different quantile ranges of ages. Our analysis indicates that the aging effect on the systolic blood pressure is much more significant for middle-aged people than for both young and old people.
2. *Happiness study*: It is generally believed that household income has a small and positive impact on happiness, which diminishes as income increases. In other words, money does buy happiness, but up to a certain point. Measuring the relationship between household income and happiness at different income levels amounts to testing departure from interval quantile independence for different quantile ranges of household incomes. Our analysis shows that a household need to make RMB 372121 yuan (around 53931 US$) a year if one lives in rural areas in China and RMB 462102 yuan (around 66971 US$) a year if

one lives in urban areas in China, but some extra income does not really translate into more happiness.

The interval quantile independence is different from statistical independence and quantile independence. In particular, if Y_1 is statistically independent of Y_2, H_0 in (1.1) is true for all $\mathcal{I}_1 \otimes \mathcal{I}_2 \subseteq (0,1) \otimes (0,1)$. However, even if the interval quantile independence holds true for some $\mathcal{I}_1 \otimes \mathcal{I}_2 \subseteq (0,1) \otimes (0,1)$, Y_1 is not necessarily independent of Y_2. Consequently, the independence tests, such as those based on distance correlation [23, 24], ranks of distances [12] and sign covariance related to Kendall's tau [3], may have an inflated test size when used to test (1.1). The quantile independence, $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$, may hold true when the interval quantile independence does not, even if $\tau_1 \in \mathcal{I}_1$ and $\mathcal{I}_2 = (0,1)$. Therefore, the quantile independence tests, such as those based on martingale difference correlation [22, 21] and [25], may lose power when used to test (1.1).

The interval quantile independence is related to but conceptually different from both the lower and the upper tail dependence [13], which are defined respectively as follows,

$$\lim_{\tau_1 \to 0} \mathrm{pr}\{Y_2 \le Q_{Y_2}(\tau_1) \mid Y_1 \le Q_{Y_1}(\tau_1)\} \text{ and } \lim_{\tau_1 \to 1} \mathrm{pr}\{Y_2 \ge Q_{Y_2}(\tau_1) \mid Y_1 \ge Q_{Y_1}(\tau_1)\}.$$

The lower and upper tail dependence of (Y_1, Y_2) corresponds to the interval quantile dependence with $\mathcal{I}_1 = \mathcal{I}_2 = (0, \tau_1)$ for $\tau_1 \to 0$, and $\mathcal{I}_1 = \mathcal{I}_2 = (\tau_1, 1)$ for $\tau_1 \to 1$, respectively. It describes the comovements in the tails of their distributions. In contrast, the interval quantile dependence allows for general intervals $\mathcal{I}_1 \otimes \mathcal{I}_2$ and does not concern necessarily the tail behaviors of the distributions of (Y_1, Y_2).

In this article we introduce an index, denoted by $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$, to test and measure the departure from the interval quantile independence defined in H_0 of (1.1). Our proposed index can be used to measure nonlinear quantile dependence. We will show that $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \ge 0$ with equality holding if and only if H_0 in (1.1) holds true. The proposed index is invariant to monotone transformations in the sense that $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(m_1(Y_1), m_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)$ for monotonically increasing functions m_1 and m_2. It can also be used to test quantile independence through setting $\mathcal{I}_1 = \{\tau_1\}$ and $\mathcal{I}_2 = (0,1)$ and statistical independence through setting $\mathcal{I}_1 = \mathcal{I}_2 = (0,1)$. We suggest a moment estimator to implement our proposed test. The resulting estimate, denoted by $\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$, depends only on the ranks of the observations. We show that, in the general case of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) > 0$, $n^{1/2}$ $\{\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) - q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)\}$ is asymptotically normal, and in the particular case of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = 0$, $n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ follows a non-normal limiting distribution. These asymptotic null distributions are free of parent distribution of (Y_1, Y_2), which facilitates the determination of critical values when the proposed index is used to test (1.1).

This paper is organized as follows. In Section 2 we introduce the concept of interval quantile independence and propose an index to measure the departure from interval quantile independence. The theoretical properties of our proposed index are also studied under both the population and sample levels. We also demonstrate the theoretical properties through numerical studies. In Section 3 we generalize the application of our proposed index to feature screening for ultrahigh dimensional data. We conclude this paper in Section 4. The proof of Proposition 1 is given in the Appendix and the proofs of Theorems 1-5 are given in the online supplement.

2. Interval Quantile Independence.

2.1. *Some notations.* The following notations will be used repetitively in subsequent exposition. Denote by "Ω_k" the support of Y_k, namely, $\Omega_k \stackrel{\text{def}}{=} \{y_k : f_k(y_k) > 0\}$ where f_k stands for the marginal density function of Y_k. Denote by $Q_{Y_k}(\tau_k)$ the τ_k-th quantile of Y_k and $Q_{Y_1|Y_2}(\tau_1)$ the τ_1-th quantile of Y_1 conditional on Y_2. In general, $Q_{Y_1|Y_2}(\tau_1)$ varies with (τ_1, Y_2). We assume throughout that $Q_{Y_1|Y_2}(\tau_1)$ is uniquely defined as a function of τ_1 for each $y_2 \in \Omega_2$. Let " $\Leftrightarrow$ " stand for "is equivalent to", " $\stackrel{d}{\longrightarrow}$ " stand for "converges in distribution", " $\stackrel{pr}{\longrightarrow}$ " stand for "converges in probability" and " $\stackrel{d}{=}$ " stand for "has the same distribution as". Define $F_k(y_k) \stackrel{\text{def}}{=} \text{pr}(Y_k \leq y_k)$ and $F_{1,2}(y_1, y_2) \stackrel{\text{def}}{=} \text{pr}(Y_1 \leq y_1, Y_2 \leq y_2)$. We further assume the joint distribution function $F_{1,2}(y_1, y_2)$ of (Y_1, Y_2) is continuous. Let $f_{1,2}(y_1, y_2)$ be the joint density function of (Y_1, Y_2) and $f_{1|2}(y_1 \mid y_2)$ be the conditional density of Y_1 given Y_2. Let $F_{n,k}$ and $F_{n,1,2}$ be the respective empirical versions of F_k and $F_{1,2}$ when a random sample of size n, denoted by $\{(Y_{i,1}, Y_{i,2}), i = 1, \ldots, n\}$, is available. To be precise, $F_{n,k}(y_k) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n I(Y_{i,k} \leq y_k)$ and $F_{n,1,2}(y_1, y_2) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n I(Y_{i,1} \leq y_1, Y_{i,2} \leq y_2)$, where $I(A)$ stands for an indicator function which equals one if the event A is true and zero otherwise.

2.2. *The rationale.* We start with the test for quantile independence. Suppose for now we aim to test H_0: $Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$ for a single $\tau_1 \in (0, 1)$, versus H_1: otherwise. It follows from the uniqueness of $Q_{Y_1|Y_2}(\tau)$ for each $y_2 \in \Omega_2$ that

$$
\begin{aligned}
Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1) \quad &\Leftrightarrow \quad E\{I(Y_1 \le Q_{Y_1}(\tau_1)) \mid Y_2\} = \tau_1 \\
&\Leftrightarrow \quad \mathrm{cov}\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le y_2)\} = 0, \text{ for all } y_2 \in \Omega_2. \\
&\Leftrightarrow \quad \int_{\Omega_2} \frac{\mathrm{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le y_2)\}}{\tau_1(1-\tau_1)F_2(y_2)\{1-F_2(y_2)\}} dy_2 = 0. \\
&\Leftrightarrow \quad \int_0^1 \frac{\mathrm{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le Q_{Y_2}(\tau_2))\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\tau_2 = 0.
\end{aligned}
$$

The first equivalency follows from the definition and the uniqueness of $Q_{Y_1|Y_2}(\tau_1)$. The second follows from the fact that $E(\varepsilon \mid X) = 0 \Leftrightarrow E\{\varepsilon I(X \le x)\} = 0$, for all x lies in the support of X. The third equivalency is obvious because the integrand is nonnegative. Note that

$$
\begin{aligned}
&\int_{\Omega_2} \frac{\mathrm{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le y_2)\}}{\tau_1(1-\tau_1)F_2(y_2)\{1-F_2(y_2)\}} dy_2 \\
= &\int_0^1 \frac{\mathrm{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le Q_{Y_2}(\tau_2))\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)f_2(Q_{Y_2}(\tau_2))} d\tau_2
\end{aligned}
$$

and $f_2(y_2) > 0$ for all $y_2 \in \Omega_2$. This immediately entails the last equivalency. The denominators in the last two equivalencies are used to rescale the integrand to be not greater than one.

The above discussion motivates us to define the following index to measure and test the interval quantile independence between Y_1 and Y_2. Specifically, we let $\mathcal{I}_k$s be two subsets of $(0,1)$, namely, $\mathcal{I}_k \subseteq (0,1)$, $\mathcal{I}_k$ can be a singleton, say, $\mathcal{I}_k = \{\tau_k\}$. We define the following index to measure and test H_0 in (1.1),

$$
\begin{aligned}
(2.1) \qquad & q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \\
\stackrel{\text{def}}{=} \; & \int_{\mathcal{I}_1} \int_{\mathcal{I}_2} \frac{\mathrm{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le Q_{Y_2}(\tau_2))\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2),
\end{aligned}
$$

where μ_ks are two probability measures which can be different and depend on $\mathcal{I}_k$. The denominator of the integrand in (2.1) is used for normalization. We define $0/0 = 0$ to avoid possible confusion in calculation. Our proposed index is related to the martingale difference correlation [21, 22] if we set $\mathcal{I}_1 = \{\tau_1\}$ and $\mathcal{I}_2 = (0,1)$ in $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$. In this sense, the martingale difference correlation is a special case of our proposed index. There are however two distinctions. The martingale difference correlation is based on characteristic function while our proposed index is in spirit based on distribution function, and the martingale difference correlation allows for multivariate Y_1 and Y_2 and our proposed index requires that both Y_1 and Y_2 be univariate.

We first present some properties of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ at the population level.

PROPOSITION 1. *We assume that* $\mathcal{I}_k = \{\tau_k : d\mu_k(\tau_k)/d\tau_k > 0\}$.
(i) If $Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1)$ *is unique for* $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$*, then* $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = 0 \Leftrightarrow Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1)$*, for* $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$.
(ii) If $Q_{Y_1|Y_2}(\tau_1)$ *is unique for* $\tau_1 \in \mathcal{I}_1$ *given* Y_2*, then* $q(Y_1, Y_2; \mathcal{I}_1, (0,1)) = 0 \Leftrightarrow Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1)$*, for all* $\tau_1 \in \mathcal{I}_1$*;* $q(Y_1, Y_2; (0,1), (0,1)) = 0$ *if and only if* Y_1 *and* Y_2 *are statistically independent.*
(iii) If m_1 *and* m_2 *are monotonically increasing functions,* $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(m_1(Y_1), m_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)$.

The first property in Proposition 1 indicates that, through using different $\mathcal{I}_k$s, $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ can be used to quantify nonlinear dependence between a certain range of Y_1 and a certain range of Y_2. The second property states that, $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ can be used to test statistical independence and quantile independence by choosing $\mathcal{I}_k$s properly. Note that $\mathcal{I}_k$s are not tuning parameters. The role of $\mathcal{I}_k$s in $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ is in spirit the same as that of quantile levels in quantile regressions [14, 15]. How to choose $\mathcal{I}_k$s depends on our purposes. Different users may specify different $\mathcal{I}_k$s for different purposes, leading to different conclusions. For instance, if we hope to test whether the median function of Y_1 depends on Y_2, we may set $\mathcal{I}_1 = \{0.5\}$ and $\mathcal{I}_2 = (0,1)$; if we aim to test whether the τ_1-th quantile function of Y_1 depends on Y_2 for $\tau_1 \in (0.25, 0.75)$, we may specify $\mathcal{I}_1 = (0.25, 0.75)$ and $\mathcal{I}_2 = (0,1)$; if we hope to test whether the first and the third quartiles of Y_1 depend on Y_2, we may specify $\mathcal{I}_1 = \{0.25, 0.75\}$ and $\mathcal{I}_2 = (0,1)$; and if we aim to test whether Y_1 is independent of Y_2, we may choose $\mathcal{I}_1 = \mathcal{I}_2 = (0,1)$.

How to choose the probability measures μ_ks depends on the intervals $\mathcal{I}_k$s. We require throughout that $\mathcal{I}_k = \{\tau_k : d\mu_k(\tau_k)/d\tau_k > 0\}$. We specify μ_k as a Lebesgue measure if $\mathcal{I}_k$ is an interval and a counting measure if $\mathcal{I}_k$ is a countable set. If $\mathcal{I}_k = (\tau_{k,1}, \tau_{k,2})$ for $\tau_{k,1} < \tau_{k,2}$, we can set the Lebesgue measure $\mu_k(\tau_k) = (\tau_k - \tau_{k,1})/(\tau_{k,2} - \tau_{k,1})$ if $\tau_{k,1} \le \tau_k < \tau_{k,2}$, $\mu_k(\tau_k) = 0$ if $\tau_k < \tau_{k,1}$ and $\mu_k(\tau_k) = 1$ if $\tau_k \ge \tau_{k,2}$. In this case, $d\mu_k(\tau_k)/d\tau_k = 1/(\tau_{k,2} - \tau_{k,1}) I(\tau_{k,1} \le \tau_k < \tau_{k,2})$. If $\mathcal{I}_k = \{\tau_{k,1}, \tau_{k,2}\}$, we set the counting measure $\mu_k(\tau_k) = 0$ if $\tau_k < \tau_{k,1}$, $\mu_k(\tau_k) = 1/2$ if $\tau_{k,1} \le \tau_k < \tau_{k,2}$, and $\mu_k(\tau_k) = 1$ if $\tau_k \ge \tau_{k,2}$. In this case, $d\mu_k(\tau_k)/d\tau_k = 1/2 I(\tau_k \in \mathcal{I}_k)$. In the particular case of $\mathcal{I}_1 = \{0.25, 0.75\}$ and $\mathcal{I}_2 = (0.25, 0.75)$, our proposed index reduces to the following simple form:

$$q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = \sum_{\tau_1 \in \mathcal{I}_1} \int_{\mathcal{I}_2} \frac{\mathrm{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau_1)), I(Y_2 \le Q_{Y_2}(\tau_2))\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\tau_2.$$

The third property in Proposition 1 shows that, $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ is invariant when monotonically increasing transformations are used. This property is not shared with Pearson correlation, distance correlation or quantile correlation [17, 23, 24]. Because the cumulative distribution functions $F_k(y_k)$s are strictly increasing, we can simply choose $m_k(y_k) = F_k(y_k)$ in Theorem 1, then our proposed index has an equivalent form of

$$q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = \int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \frac{\text{cov}^2\{I(F_1(Y_1) \le \tau_1), I(F_2(Y_2) \le \tau_2)\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1)d\mu_2(\tau_2).$$

This invariant property play an important role here in that it allows us to assume subsequently that Y_ks have compact support because otherwise we replace Y_ks with their respective monotonically increasing transformations $F_k(Y_k)$s. In what follows we shall work with the above form of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ in that it facilitates estimation of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$.

2.3. *Asymptotic properties.* Next we study estimation of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$. Suppose the observations $\{(Y_{i,1}, Y_{i,2}), i = 1, \ldots, n\}$ are independent and identically distributed. Estimating $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ is nontrivial, because integral approximation is typically not straightforward. In the present context, we make use of the fact that the empirical distributions are step functions to simplify estimation. We first note that

$$\begin{aligned} & q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(F_1(Y_1), F_2(Y_2); \mathcal{I}_1, \mathcal{I}_2) \\ = & \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} \int_{\mathcal{I}_1\cap[(j_1-1)/n, j_1/n)}\int_{\mathcal{I}_2\cap[(j_2-1)/n, j_2/n)} \\ & \frac{\text{cov}^2\{I(F_1(Y_1) \le \tau_1), I(F_2(Y_2) \le \tau_2)\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1)d\mu_2(\tau_2), \end{aligned}$$

which motivates us to estimate $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ through

$$\begin{aligned} \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \stackrel{\text{def}}{=} & \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} \int_{\mathcal{I}_1\cap[(j_1-1)/n, j_1/n)}\int_{\mathcal{I}_2\cap[(j_2-1)/n, j_2/n)} \\ & \frac{\widehat{\text{cov}}^2\{I(F_{n,1}(Y_1) \le \tau_1), I(F_{n,2}(Y_2) \le \tau_2)\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1)d\mu_2(\tau_2), \end{aligned}$$

where $\mathcal{I}_k \cap [(j_k-1)/n, j_k/n)$ stands for the intersection of $\mathcal{I}_k$ and $[(j_k-1)/n, j_k/n)$, and

$$\begin{aligned} & \widehat{\text{cov}}\{I(F_{n,1}(Y_1) \le \tau_1), I(F_{n,2}(Y_2) \le \tau_2)\} \\ \stackrel{\text{def}}{=} & n^{-1}\sum_{i=1}^{n} I(F_{n,1}(Y_{i,1}) \le \tau_1) I(F_{n,2}(Y_{i,2}) \le \tau_2) \\ - & n^{-2}\sum_{i=1}^{n} I(F_{n,1}(Y_{i,1}) \le \tau_1)\sum_{i=1}^{n} I(F_{n,2}(Y_{i,2}) \le \tau_2). \end{aligned}$$

Because $F_{n,k}(Y_k)$ is a step function, the numerator of the integrand remains unchanged for $\tau_k \in \mathcal{I}_k \cap [(j_k-1)/n, j_k/n)$. Consequently, the integral approximation is straightforward. In particular,

$$\begin{aligned}\widehat{q}(Y_1, Y_2; (0,1), (0,1)) \quad &\stackrel{\text{def}}{=} \quad \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \Bigg[\widehat{\text{cov}}^2 \{ I(F_{n,1}(Y_1) \le j_1/n), I(F_{n,2}(Y_2) \le j_2/n) \} \\ &\int_{\mathcal{I}_1 \cap [(j_1-1)/n, j_1/n)} \frac{1}{\tau_1(1-\tau_1)} d\mu_1(\tau_1) \\ &\int_{\mathcal{I}_2 \cap [(j_2-1)/n, j_2/n)} \frac{1}{\tau_2(1-\tau_2)} d\mu_2(\tau_2) \Bigg],\end{aligned}$$

$$\begin{aligned}\text{and } \widehat{q}(Y_1, Y_2; \tau_1, (0,1)) \quad &\stackrel{\text{def}}{=} \quad \sum_{j_2=1}^{n} \Bigg[\widehat{\text{cov}}^2 \{ I(F_{n,1}(Y_1) \le \tau_1), I(F_{n,2}(Y_2) \le j_2/n) \} \\ &\frac{1}{\tau_1(1-\tau_1)} \int_{\mathcal{I}_2 \cap [(j_2-1)/n, j_2/n)} \frac{1}{\tau_2(1-\tau_2)} d\mu_2(\tau_2) \Bigg].\end{aligned}$$

Note that the integrals

$$\int_{\mathcal{I}_k \cap [(j_k-1)/n, j_k/n)} \frac{1}{\tau_k(1-\tau_k)} d\mu_k(\tau_k), k = 1, 2,$$

have closed forms for μ_k being either a counting or a Lebesgue measure. For instance, if μ_k is a Lebesgue measure, say, $\mu_k(\tau_k) = \tau_k$,

$$\int_a^b \frac{1}{\tau_k(1-\tau_k)} d\mu_k(\tau_k) \quad = \quad \{\log(b) - \log(1-b)\} - \{\log(a) - \log(1-a)\}.$$

If μ_k is a counting measure, say, $\mu_k(\tau) = I(\tau \ge \tau_{k,0})$, then, for $\tau_{k,0} \in [a, b]$,

$$\int_a^b \frac{1}{\tau_k(1-\tau_k)} d\mu_k(\tau_k) \quad = \quad \frac{1}{\tau_{k,0}(1-\tau_{k,0})}.$$

To avoid potential ambiguity in practice, we define $\log(0) = 0$ and $1/0 = 0$.

Theorem 1 states that $\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ converges in distribution.

Theorem 1. *Assume that the density of Y_k, $f_k\{Q_{Y_k}(\tau_k)\}$, and its first derivative with respect to τ_k are bounded away from zero and infinity on $\mathcal{I}_k \subseteq (0,1)$.*

1. If H_0 in (1.1) is false, then $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) > 0$, and

$$n^{1/2} \{ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) - q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \} \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 \stackrel{def}{=} 4 \text{ var}(Z)$ and Z is defined in (S1.5).

2. *If H_0 in (1.1) is true, then $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = 0$, and*

$$n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \xrightarrow{d} \int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \frac{B^2(\tau_1,\tau_2)}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2)$$
$$\stackrel{d}{=} \sum_{j=1}^{\infty} \lambda_j \chi_j^2(1).$$

where $B(\tau_1, \tau_2)$ is a separable Gaussian process depending on (τ_1, τ_2) for $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2 \subseteq (0,1) \otimes (0,1)$, $E\{B(\tau_1,\tau_2)\} = 0$ and

$$E\{B(\tau_1,\tau_2)B(\tau_1',\tau_2')\} = \{\min(\tau_1,\tau_1') - \tau_1\tau_1'\}\{\min(\tau_2,\tau_2') - \tau_2\tau_2'\}.$$

The loadings λ_js are eigenvalues defined in (S1.7) which depend on $(\mathcal{I}_1, \mathcal{I}_2)$ rather than the joint distribution of (Y_1, Y_2), and $\chi_j^2(1)$s are independent chi-square random variables with one degree of freedom.

We remark on the boundedness assumption on $f_k\{Q_{Y_k}(\tau_k)\}$. We assume such conditions to ensure that $\widehat{Q}_{Y_k}(\tau_k)$ converges in probability to $Q_{Y_k}(\tau_k)$ uniformly for $\tau_k \in \mathcal{I}_k \subseteq (0,1)$. Similar conditions are also used in the literature. See, for example, condition (C1) in [26] and condition (F) in [16]. The boundedness assumption is satisfied if both Y_ks have compact support. If Y_k does not have a compact support, we can simply replace Y_k with $F_k(Y_k)$ which apparently has a compact support. The invariant property in Proposition 1 ensures that $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(F_1(Y_1), F_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)$. We estimate $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ using the ranks of Y_ks only. The empirical distribution functions $F_{n,k}(Y_k)$s are monotonically increasing. It follows immediately that $\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = \widehat{q}(F_{n,1}(Y_1), F_{n,2}(Y_2); \mathcal{I}_1, \mathcal{I}_2)$. In general, if $m_k(Y_k)$s are monotonically increasing, we can replace Y_ks with $m_k(Y_k)$s as long as $m_k(Y_k)$s have compact support. The invariant property in Proposition 1 ensures that $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(m_1(Y_1), m_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)$ and $\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = \widehat{q}(m_1(Y_1), m_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)$. In addition, we require that the boundedness assumption hold uniformly for $\tau_k \in \mathcal{I}_k$ only. Therefore, the condition on f_k is regarded as reasonable and acceptable in the present context.

Given a random sample of size n from a bivariate population, our test for (1.1) can be carried out as follows: we reject H_0 if $n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) > c_\alpha$, where the critical value at the significance level α, c_α, is defined as the upper α quantile of the asymptotic null distribution of $n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ under H_0. Theorem 1 ensures that using $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ to test (1.1) is consistent, and the power is approximately

$$\beta_n \stackrel{\text{def}}{=} \text{pr}\{n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) > c_\alpha \mid q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) > 0\}$$
$$\approx 1 - \Phi\big[\{c_\alpha - n\ q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)\}/(n^{1/2}\sigma)\big],$$

where Φ is the cumulative distribution function of $\mathcal{N}(0,1)$. Apparently, $\beta_n \to 1$ as $n \to \infty$, indicating that our test for (1.1) is consistent.

How to decide a critical value c_α is nontrivial. Suppose $\mathcal{I}_1 = \{\tau_1\}$ and $\mathcal{I}_2 = (0,1)$, $\mu_1(\tau) = I(\tau \geq \tau_1)$ and $\mu_2(\tau) = \tau$. Accordingly, $d\mu_1(\tau)/d\tau = I(\tau = \tau_1)$ and $d\mu_2(\tau) = d\tau$. Following [1] and [4], we can show that

$$\int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \frac{B^2(\tau_1,\tau_2)}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2) \overset{d}{=} \sum_{j=1}^{\infty} \frac{\chi_j^2(1)}{j(j+1)}.$$

The limiting distribution can be approximated with

$$\sum_{j=1}^{N} \frac{\chi_j^2(1)}{j(j+1)},$$

for a sufficiently large N. Suppose $\mathcal{I}_1 = \mathcal{I}_2 = (0,1)$, $\mu_1(\tau) = \mu_2(\tau) = \tau$. Accordingly, $d\mu_k(\tau) = d\tau$. Following [1] and [4], we can also show that

$$\int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \frac{B^2(\tau_1,\tau_2)}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2) \overset{d}{=} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{\chi_{ij}^2(1)}{i(i+1)j(j+1)},$$

where $\chi_{ij}^2(1)$s are independent chi-square random variables with one degree of freedom. This limit distribution can also be approximated with

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\chi_{ij}^2(1)}{i(i+1)j(j+1)}. \tag{2.2}$$

Because the asymptotic distributions are approximately tractable, the critical value c_α can be easily decided under these two situations. We use a toy example to demonstrate how accurate these approximates are. We choose $N = 10, 20, 50$ and 100 in Figure 1 (A), from which it can be clearly seen that as long as $N \geq 20$, such approximations are very accurate.

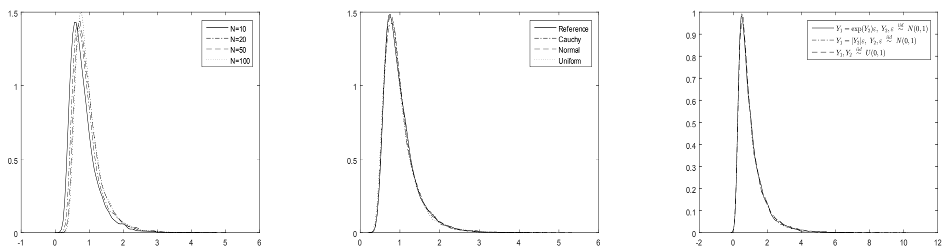
In general, we suggest a simulation-based procedure to decide c_α. Theorem 1 states that, under H_0 in (1.1), the asymptotic distribution of $n\,\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ does not depend on the joint distribution of (Y_1, Y_2). This inspires us to randomly generate new samples from uniform distribution to approximate the asymptotic null distribution. To be precise, we generate $Y_{i,k}^\star$ independently from uniform distribution, $i = 1, \ldots, n, k = 1, 2$, and re-estimate $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ based on $\{(Y_{i,1}^\star, Y_{i,2}^\star), i = 1, \ldots, n\}$. We repeat this procedure for B times and set c_α to be the upper α quantile of the estimates of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ obtained from the randomly generated samples. Because all $Y^\star_{i,k}$s are independent, it is natural to anticipate that this procedure provides a reasonable approximation of the asymptotic null distribution of $n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ for a sufficiently large B. Throughout we use this method to decide c_α in the test for (1.1).

Theorem 2 establishes the consistency of this simulation-based procedure.

Theorem 2. *Under the conditions of Theorem 1, it follows that*

$$n\ \widehat{q}(Y_1^\star, Y_2^\star; \mathcal{I}_1, \mathcal{I}_2) \quad \xrightarrow{d} \quad \int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \frac{B^2(\tau_1,\tau_2)}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2)$$

where $B(\tau_1, \tau_2)$ is defined in Theorem 1.



(A) (B) (C)

FIG 1. *(A): We approximate the limiting distribution of $n\ \widehat{q}(Y_1, Y_2; (0,1), (0,1))$ with the first $N \times N$ terms when Y_1 and Y_2 are statistically independent, $N = 10$, 20, 50 and 100. (B): The density functions of $n\ \widehat{q}(Y_1, Y_2; (0,1), (0,1))$ when Y_1 and Y_2 are drawn independently from cauchy distribution, standard normal distribution and uniform distribution. The reference density function is the limiting distribution of $n\ \widehat{q}(Y_1, Y_2; (0,1), (0,1))$ when Y_1 and Y_2 are statistically independent. (C): The density functions of $n\ \widehat{q}(Y_1, Y_2; \{0.5\}, (0,1))$ for two simulated models: $Y_1 = \exp(Y_2)\varepsilon$ and $Y_1 = |Y_2|\varepsilon$, where Y_1 and ε are independent and standard normal. We also present the density function of $n\ \widehat{q}(Y_1, Y_2; \{0.5\}, (0,1))$ when Y_1 and Y_2 are independent and uniformly distributed.*

To illustrate the appealing distribution-free property of our proposed test, we generate Y_1 and Y_2 independently from Cauchy, standard normal and uniform distribution, and draw the density functions of the test statistic $n\ \widehat{q}(Y_1, Y_2; (0,1), (0,1))$ in Figure 1(B). A reference density function is also given, which is obtained through choosing $N = 100$ in the right hand side of (2.2). It can be clearly seen that all four curves match perfectly, indicating that our proposed test is indeed distribution-free.

We consider three additional toy examples. In the first example, $Y_1 = \exp(Y_2)\varepsilon$; in the second example, $Y_1 = |Y_2|\varepsilon$. In both examples, we draw ε and Y_2 independently from standard normal distribution. In the third example, we generate Y_1 and Y_2 independently from uniform distribution. In all examples, $q(Y_1, Y_2, \{0.5\}, (0,1)) = 0$. The sample size $n = 100$. We repeat this procedure 1000 times and plot the density functions of $n\,\widehat{q}(Y_1, Y_2; \{0.5\}, (0,1))$ in Figure 1(C). Again we can see that these three density functions are almost identical, indicating that the simulation-based procedure is consistent.

The asymptotic normality of $\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ stated in Theorem 1 allows us to construct confidence intervals for $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ when it is nonzero, if we can have a consistent estimate of σ^2. In what follows we discuss how to estimate σ^2 consistently. We estimate $Q_{Y_k}(\tau_k)$ with $\widehat{Q}_{Y_k}(\tau_k) = \inf\{x : F_{n,k}(x) \geq \tau_k\}$, and estimate the conditional distribution of $(Y_k \mid Y_l)$, denoted by $F_{k|l}(y_k \mid y_l)$, with the following Nadaraya-Watson kernel estimate,

$$\widehat{F}_{k|l}(y_k \mid y_l) \stackrel{\text{def}}{=} \sum_{i=1}^{n} I(Y_{i,k} \leq y_k) K_{h_l}(Y_{i,l} - y_l) \Big/ \sum_{i=1}^{n} K_{h_l}(Y_{i,l} - y_l),$$

where $K_{h_l}(\cdot) = K(\cdot/h_l)/h_l$, K is a second-order kernel function and h_l is the associated bandwidth, $l = 1, 2$. Throughout our numerical studies we simply use $h_l = 1.06 n^{-1/5} \widehat{\text{std}}(Y_l)$, where $\widehat{\text{std}}(Y_l)$ is a robust estimate of the standard deviation of Y_l. For notational clarity we write $q_k = Q_{Y_k}(\tau_k)$ and $\widehat{q}_k = \widehat{Q}_{Y_k}(\tau_k)$. Define $\widehat{\Delta}(\widehat{q}_1, \widehat{q}_2) = F_{n,1,2}(\widehat{q}_1, \widehat{q}_2) - \tau_1\tau_2$. We replace all unknowns in Z_i defined in (S1.5) with their respective estimates. This gives

$$\widehat{Z}_i \stackrel{\text{def}}{=} \int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \widehat{T}_i(\tau_1, \tau_2) \Big/ \{\tau_1(1-\tau_1)\tau_2(1-\tau_2)\}\, d\mu_1(\tau_1) d\mu_2(\tau_2), \tag{2.3}$$

where

$$\begin{aligned} \widehat{T}_i(\tau_1, \tau_2) \stackrel{\text{def}}{=} \widehat{\Delta}(\widehat{q}_1, \widehat{q}_2)\Big[&\{I(F_{n,1}(Y_{i,1}) \leq \tau_1, F_{n,2}(Y_{i,2}) \leq \tau_2) - F_{n,1,2}(\widehat{q}_1, \widehat{q}_2) \\ &- \tau_1 I(F_{n,2}(Y_{i,2}) \leq \tau_2) - \tau_2 I(F_{n,1}(Y_{i,1}) \leq \tau_1) + 2\tau_1\tau_2\} \\ &+ \{\widehat{F}_{2|1}(\widehat{q}_2 \mid \widehat{q}_1) - \tau_2\}\{\tau_1 - I(F_{n,1}(Y_{i,1}) \leq \tau_1)\} \\ &+ \{\widehat{F}_{1|2}(\widehat{q}_1 \mid \widehat{q}_2) - \tau_1\}\{\tau_2 - I(F_{n,2}(Y_{i,2}) \leq \tau_2)\}\Big]. \end{aligned} \tag{2.4}$$

By noting that $F_{n,k}$, $F_{n,1,2}$ and $\widehat{F}_{k|l}$ are all step functions, we evaluate the integrals using the same ideas as we used to estimate $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$. The estimator of σ^2 is given by

$$\widehat{\sigma}^2 \stackrel{\text{def}}{=} 4n^{-1} \sum_{i=1}^{n} \widehat{Z}_i^2. \tag{2.5}$$

The following theorem establishes the consistency of $\widehat{\sigma}^2$.

Theorem 3. *In addition to the conditions in Theorem 1, we assume that the first derivative of f_k, denoted by f_k', the density function of $(Y_k \mid Y_l)$, denoted by $f_{k|l}$, and the first derivative of $F_{k|l}(y_k \mid y_l)$ with respect to y_l, denoted by $F'_{k|l}(y_k \mid y_l)$, $k \neq l$, are all Lipschitz continuous uniformly, i.e., there exists a positive constant C such that*

$$\sup_{y_l \in \Omega_l} |f_l'(y_l + u) - f_l'(y_l)| \leq C|u|, \text{ and}$$

$$\sup_{(y_k, y_l) \in \Omega_k \otimes \Omega_l} |f_{k|l}(y_k + u \mid y_l) - f_{k|l}(y_k \mid y_l)| \leq C|u|, \text{ and}$$

$$\sup_{(y_k, y_l) \in \Omega_k \otimes \Omega_l} |F'_{k|l}(y_k \mid y_l + u) - F'_{k|l}(y_k \mid y_l)| \leq C|u|.$$

In addition, we assume that the kernel K is a probability density function, K is symmetric and Lipschitz continuous, and has a compact support. We further assume that the bandwidth h_l satisfies $nh_l^4 \to \infty$ and $nh_l^8 \to 0$ as $n \to \infty$, for $l = 1, 2$. Then $\widehat{\sigma}^2 \xrightarrow{pr} \sigma^2$ as $n \to \infty$.

Theorem 2 ensures that, the asymptotic null distribution can be well approximated through our proposed simulation-based method. When the null hypothesis H_0 in (1.1) is rejected, the asymptotic normality presented in Theorem 1, together with Theorem 3, allows us to construct a reasonable confidence interval for nonzero $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$. Some alternative methods, such as the pairwise bootstrap, may also be used to construct confidence intervals. However, theoretical justification for the validity of the pairwise bootstrap appears not straightforward. We also remark here that, it is highly nontrivial, yet theoretically challenging [2, 6], to devise an adaptive method that can be used to construct confidence intervals for general $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$. The theoretical challenge lies in the nonstandard asymptotics, where the limiting distribution of $\widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ is discontinuous on the boundary of the parameter space. Such discontinuity and nonstandard asymptotics pose huge challenges for us to design a uniform method to construct confidence intervals for general $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$. [2] gave several examples that the usual bootstrap does not work when the null hypothesis is on the boundary of the parameter space. This type of non-regularity occurs in many other settings as well, such as change-point detection [5] and post-selection inference [20].

Next we consider local alternatives of the following form:

$$F_{1,2}\{Q_{Y_1}(\tau_1), Q_{Y_2}(\tau_2)\} - \tau_1\tau_2 = n^{-1/2} h(\tau_1, \tau_2), \tag{2.6}$$

for all $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$, where $h(\cdot)$ satisfies

$$\sup_{(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2} h^2(\tau_1, \tau_2) > 0.$$

Taking derivative on both sides of (2.6) with respect to τ_2, we obtain that $F_{1|2}\{Q_{Y_1}(\tau_1) \mid Q_{Y_2}(\tau_2)\} - \tau_1 = n^{-1/2}\partial h(\tau_1, \tau_2)/\partial \tau_2$. This indicates that

$$Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}\{\tau_1 + n^{-1/2}\partial h(\tau_1, \tau_2)/\partial \tau_2\} = Q_{Y_1}(\tau_1).$$

It follows from Taylor expansion that $Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}\{\tau_1+n^{-1/2}\partial h(\tau_1, \tau_2)/\partial \tau_2\} = Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1) + n^{-1/2}\{\partial h(\tau_1, \tau_2)/\partial \tau_2\}/f_{1|2}\{Q_{Y_1|Y_2=Q_{Y_2}}(\tau_1) \mid Q_{Y_2}(\tau_2)\} + o(n^{-1/2})$. Therefore, the local alternative (2.6) implies that

$$Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1) + \frac{n^{-1/2}\partial h(\tau_1, \tau_2)/\partial \tau_2}{f_{1|2}\{Q_{Y_1|Y_2=Q_{Y_2}}(\tau_1) \mid Q_{Y_2}(\tau_2)\}} + o(n^{-1/2}),$$

for $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$, which seems to match the hypotheses in (1.1) more naturally than (2.6). However, we consider the local alternative of the form (2.6) for technical reasons. Theorem 4 indicates that the test for (1.1) using $n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ has nontrivial power under the local alternative (2.6).

Theorem 4. *Suppose that $\partial h(\tau_1, \tau_2)/\partial \tau_1$ and $\partial h(\tau_1, \tau_2)/\partial \tau_2$ are bounded uniformly on $\mathcal{I}_1 \otimes \mathcal{I}_2$. Under the conditions of Theorem 1 we have*

$$n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \quad \xrightarrow{d} \quad \int_{\mathcal{I}_1}\int_{\mathcal{I}_2} \frac{\{B(\tau_1, \tau_2) + h(\tau_1, \tau_2)\}^2}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2)$$

where $B(\tau_1, \tau_2)$ is defined in Theorem 1.

2.4. *Numerical studies.* In this section we investigate the finite sample behavior of our proposed test for (1.1).

Example 1: A Simulation Study We consider three simulated models:

$$Y_2 = A\{Y_1^2 I\,(Y_1 > 0) + \widetilde{Y}_1^2 I\,(Y_1 \le 0)\} + \varepsilon; \tag{2.7}$$

$$Y_2 = \exp\left(AY_1^2\right)\varepsilon; \tag{2.8}$$

$$Y_2 = AY_1^2 + \varepsilon; \tag{2.9}$$

where ε, Y_1 and $\widetilde{Y}_1$ are drawn independently from standard Cauchy distribution. We set $A = 0$, 0.5, 1.0, 1.5 and 2. When $A = 0$, Y_1 and Y_2 are independent in all three models. When $A \ne 0$, $Q_{Y_1|Y_2}(\tau_1)$ depends on Y_2 for $\tau_1 \in \mathcal{I}_1 = (0.5, 1)$ in model (2.7), for $\tau_1 \in \mathcal{I}_1 = (0, 0.5) \cup (0.5, 1)$ in model (2.8) and for $\tau_1 \in (0, 1)$ in model (2.9). In other words, $q(Y_1, Y_2; \mathcal{I}_1, (0, 1))$ attains its maximum when $\mathcal{I}_1 \supseteq (0.5, 1)$ in model (2.7), $\mathcal{I}_1 \supseteq (0, 0.5) \cup (0.5, 1)$ in model (2.8) and $\mathcal{I}_1 = (0, 1)$ in model (2.9) for any nonzero A.

We compare our proposed test for the interval quantile independence (1.1) with the Kendall's rank test ("Kendall's tau(Y_1, Y_2)", [1]), the rank-based distance correlation test ("dcorr$\{F_1(Y_1), F_2(Y_2)\}$", [24]), the linear quantile correlation test [17] and the martingale difference correlation test for quantile dependence [22] at three different quantile levels ("$\text{qcorr}_{\tau_1}(Y_1 \mid Y_2)$" and "$\text{MDC}_{\tau_1}(Y_1 \mid Y_2)$" for $\tau_1 = 0.50, 0.75$ and 0.90). To implement our proposed test using the statistic $\{n\ \widehat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)\}$, we vary $\mathcal{I}_1 = \{0.50\}$, $\{0.75\}$, $\{0.90\}$, $(0, 0.25)$, $(0, 0.5)$, $(0, 0.75)$, $(0.25, 0.75)$, $(0.5, 1)$, $(0.75, 1)$, $(0, 0.5) \cup (0.5, 1)$ and $(0, 1)$, and fix $\mathcal{I}_2 = (0, 1)$. We set the sample size $n = 50$ and the significance level $\alpha = 0.05$, and repeat each scenario 1000 times. We report both the sizes and the powers of the aforementioned tests in Table 1.

It can be seen from Table 1 that the empirical sizes of almost all tests are pretty close to the significance level α. The power performance is however quite different. In particular, the Kendall's rank test fails to detect the heterogeneity effect in model (2.8) and the "symmetric pattern" in the sense that $E(Y_1 \mid Y_2) = E(Y_1)$ in models (2.7) and (2.9). The rank-based distance correlation test and $\widehat{q}(Y_1, Y_2; (0, 1), (0, 1))$ have comparable power performance in model (2.9). In models (2.7) and (2.8), $\widehat{q}(Y_1, Y_2; (0, 1), (0, 1))$ is significantly superior to the rank-based distance correlation test. However, the independence tests cannot tell which quantile levels of Y_1 depend on Y_2.

We compare the interval quantile independence test $(q(Y_1, Y_2; \{\tau_1\}, (0, 1)))$ with the martingale difference correlation test $(\text{MDC}_{\tau_1}(Y_1 \mid Y_2))$ and the linear quantile correlation test $(\text{qcorr}_{\tau_1}(Y_1 \mid Y_2))$ in testing the quantile independence at a single quantile level $\tau_1 = 0.75$ and 0.90. All these three quantile independence tests have different power performance at different quantile levels. Given each quantile level τ_1, in all three models, our proposed test is the most powerful, followed by the martingale difference correlation test. The quantile correlation test has the smallest power in that it is designed to detect linear quantile dependence.

Recall that in model (2.8) $Q_{Y_1|Y_2}(0.50) = 0$ for all $\tau_2 \in \mathcal{I}_2 = (0, 1)$. This partly explains why the powers of the linear quantile correlation test $\text{qcorr}_{0.50}(Y_1 \mid Y_2)$, the martingale difference correlation test $\text{MDC}_{0.50}(Y_1 \mid Y_2)$ and our proposed test $\widehat{q}(Y_1, Y_2; \{0.50\}, (0, 1))$ are close to the significance level α. In this heterogeneous model, our proposed tests $\widehat{q}(Y_1, Y_2; (0, 1), (0, 1))$ and $\widehat{q}(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0, 1))$ have comparable power performance.

Next we demonstrate how our proposed index tells at which quantile levels $Q_{Y_1|Y_2}(\tau_1)$ depends on Y_2. In model (2.7) with $A = 2$, the power of $\widehat{q}(Y_1, Y_2; (0, 0.5), (0, 1))$ is 0.056 whereas that of $\widehat{q}(Y_1, Y_2; (0.5, 1), (0, 1))$ is 0.910. This indicates that $Q_{Y_1|Y_2}(\tau_1)$ depends on Y_2 for $\tau_1 \in (0.5, 1)$ and yet are independent of Y_2 for $\tau_1 \in (0, 0.5)$. In model (2.8) with $A = 2$, the power of $\widehat{q}(Y_1, Y_2; \{0.50\}, (0, 1))$ is 0.074 whereas that of $\widehat{q}(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0, 1))$ is as large as 0.951. This again indicates that $Q_{Y_1|Y_2}(\tau_1)$ depends upon Y_2 for $\tau_1 \in (0, 0.5) \cup (0.5, 1)$ and yet are independent of Y_2 for $\tau_1 \in \{0.5\}$. None of the competitors can convey such messages.

The above discussion also motivates us to expect a test which is consistent with respect to a large class of alternatives will have a lower power with regard to a sub-class of alternatives than a test which has optimum properties with respect to this particular sub-class. This consideration suggests the problem of selecting from a given class of tests a test which is most powerful with respect to certain alternatives.

Example 2: The Hypertension Study High blood pressure is perhaps one of the most common medical problems in China. Long-term, prolonged high blood pressure put added stress on the heart and arteries and is thus a main risk factor for cardiovascular, cerebrovascular and renal diseases. It is estimated that one-third of adults in China have hypertension. However, many who have it are unaware that they have it. Therefore, it is important to disseminate the awareness of taking precautions against hypertension.

It is common knowledge that hypertension is age related, possibly due to reduction in vascular compliance and stiffening of the arteries. Aging effect on the systolic blood pressure is however possibly different for young, middle-aged and old people. The Chinese government conducted a hypertension study in Inner Mongolian Autonomous Region in 2002. In this study, both the systolic blood pressure (Y_1) and age (Y_2) of 1051 subjects are recorded simultaneously. The goal of our study is to quantify the aging effect on the systolic blood pressure at different stages, which can be achieved through studying how $q(Y_1, Y_2; (0, 1), \mathcal{I}_2)$ varies with $\mathcal{I}_2$. We choose $\mathcal{I}_2 = (0, 0.1)$, $(0.1, 0.2)$, …, $(0.9, 1)$. In this study, $Q_{Y_2}(\tau_2) = 20$, 33, 36, 39, 42, 45, 49, 52, 56, 64 and 83 for $\tau_2 = 0$, 0.1, …, 1.0, respectively. Table 2 shows that $q(Y_1, Y_2; (0, 1), \mathcal{I}_2)$ is concentrated at $\mathcal{I}_2 = (0.2, 0.9)$, which corresponds to age ranging from 36 to 64, and decreases significantly on either side. This apparently indicates that the aging effect is much more significant for middle-aged people than for both young and old people.

We use the simulation-based approach introduced in Section 2.3 to test whether $\widehat{q}(Y_1, Y_2; (0, 1), \mathcal{I}_2)$s are significantly different from zero. All resulting p-values are less than 10^{-3}, which strikes the chord with common knowledge that hypertersion is age related. Table 2 charts the estimates of σ^2 given by (2.5). It can be seen from Table 2 that $\widehat{\sigma}^2$ for $\mathcal{I}_2 = (0.2, 0.9)$ is also comparatively larger than that for $\mathcal{I}_2 = (0, 0.1)$ or $\mathcal{I}_2 = (0.9, 1)$, indicating that the aging effect for middle-aged people is also more diversified than for both young and old people.

Example 3: The Happiness Study Can money buy us happiness? Would more money really make us happier? These interesting questions fascinate

TABLE 1

The empirical size and power of the Kendall's rank test, the rank-based distance correlation test, the linear quantile correlation test ($qcorr_{\tau_1}(Y_1 \mid Y_2)$), the martingale difference correlation test for quantile dependence ($MDC_{\tau_1}(Y_1 \mid Y_2)$) and our proposed test $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ for models (2.7)-(2.9) at the significance level 0.05.

model	method	$A = 0$	$A = 0.5$	$A = 1$	$A = 1.5$	$A = 2$
(2.7)	Kendall tau(Y_1, Y_2)	0.044	0.231	0.307	0.375	0.395
	dcorr$\{F_1(Y_1), F_2(Y_2)\}$	0.050	0.332	0.451	0.536	0.573
	qcorr$_{0.50}(Y_1 \mid Y_2)$	0.025	0.005	0.004	0.009	0.007
	qcorr$_{0.75}(Y_1 \mid Y_2)$	0.047	0.183	0.197	0.197	0.203
	qcorr$_{0.90}(Y_1 \mid Y_2)$	0.089	0.520	0.530	0.543	0.541
	MDC$_{0.50}(Y_1 \mid Y_2)$	0.050	0.054	0.064	0.051	0.061
	MDC$_{0.75}(Y_1 \mid Y_2)$	0.071	0.332	0.340	0.357	0.379
	MDC$_{0.90}(Y_1 \mid Y_2)$	0.051	0.578	0.591	0.610	0.614
	$q(Y_1, Y_2; \{0.50\}, (0,1))$	0.042	0.068	0.073	0.067	0.060
	$q(Y_1, Y_2; \{0.75\}, (0,1))$	0.054	0.514	0.652	0.731	0.785
	$q(Y_1, Y_2; \{0.90\}, (0,1))$	0.063	0.882	0.933	0.960	0.973
	$q(Y_1, Y_2; (0, 0.25), (0,1))$	0.040	0.059	0.050	0.049	0.049
	$q(Y_1, Y_2; (0, 0.5), (0,1))$	0.042	0.055	0.055	0.061	0.056
	$q(Y_1, Y_2; (0, 0.75), (0,1))$	0.051	0.148	0.200	0.226	0.241
	$q(Y_1, Y_2; (0.25, 0.75), (0,1))$	0.053	0.576	0.730	0.797	0.852
	$q(Y_1, Y_2; (0.5, 1), (0,1))$	0.056	0.730	0.821	0.880	0.910
	$q(Y_1, Y_2; (0.75, 1), (0,1))$	0.061	0.894	0.949	0.970	0.976
	$q(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0,1))$	0.041	0.668	0.754	0.785	0.801
	$q(Y_1, Y_2; (0, 1), (0,1))$	0.048	0.517	0.631	0.733	0.775
(2.8)	Kendall tau(Y_1, Y_2)	0.054	0.173	0.228	0.198	0.206
	dcorr$\{F_1(Y_1), F_2(Y_2)\}$	0.062	0.245	0.337	0.358	0.393
	qcorr$_{0.50}(Y_1 \mid Y_2)$	0.027	0.016	0.027	0.025	0.017
	qcorr$_{0.75}(Y_1 \mid Y_2)$	0.040	0.131	0.130	0.132	0.118
	qcorr$_{0.90}(Y_1 \mid Y_2)$	0.078	0.414	0.390	0.411	0.399
	MDC$_{0.50}(Y_1 \mid Y_2)$	0.063	0.073	0.080	0.076	0.067
	MDC$_{0.75}(Y_1 \mid Y_2)$	0.047	0.476	0.492	0.498	0.510
	MDC$_{0.90}(Y_1 \mid Y_2)$	0.044	0.742	0.699	0.720	0.727
	$q(Y_1, Y_2; \{0.50\}, (0,1))$	0.048	0.062	0.067	0.072	0.074
	$q(Y_1, Y_2; \{0.75\}, (0,1))$	0.058	0.308	0.439	0.442	0.491
	$q(Y_1, Y_2; \{0.90\}, (0,1))$	0.045	0.748	0.827	0.830	0.817
	$q(Y_1, Y_2; (0, 0.25), (0,1))$	0.054	0.800	0.862	0.881	0.893
	$q(Y_1, Y_2; (0, 0.5), (0,1))$	0.058	0.573	0.666	0.674	0.708
	$q(Y_1, Y_2; (0, 0.75), (0,1))$	0.054	0.486	0.595	0.646	0.708
	$q(Y_1, Y_2; (0.25, 0.75), (0,1))$	0.052	0.436	0.554	0.603	0.648
	$q(Y_1, Y_2; (0.5, 1), (0,1))$	0.050	0.521	0.645	0.660	0.662
	$q(Y_1, Y_2; (0.75, 1), (0,1))$	0.049	0.806	0.878	0.888	0.891
	$q(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0,1))$	0.048	0.880	0.928	0.930	0.951
	$q(Y_1, Y_2; (0, 1), (0,1))$	0.051	0.777	0.919	0.957	0.981
(2.9)	Kendall tau(Y_1, Y_2)	0.054	0.128	0.160	0.178	0.174
	dcorr$\{F_1(Y_1), F_2(Y_2)\}$	0.057	0.810	0.962	0.986	0.999
	qcorr$_{0.50}(Y_1 \mid Y_2)$	0.025	0.007	0.012	0.010	0.011
	qcorr$_{0.75}(Y_1 \mid Y_2)$	0.049	0.205	0.211	0.196	0.189
	qcorr$_{0.90}(Y_1 \mid Y_2)$	0.087	0.524	0.535	0.549	0.525
	MDC$_{0.50}(Y_1 \mid Y_2)$	0.049	0.059	0.084	0.076	0.080
	MDC$_{0.75}(Y_1 \mid Y_2)$	0.064	0.337	0.367	0.349	0.344
	MDC$_{0.90}(Y_1 \mid Y_2)$	0.054	0.576	0.604	0.612	0.601
	$q(Y_1, Y_2; \{0.50\}, (0,1))$	0.057	0.088	0.119	0.125	0.141
	$q(Y_1, Y_2; \{0.75\}, (0,1))$	0.045	0.477	0.641	0.713	0.752
	$q(Y_1, Y_2; \{0.90\}, (0,1))$	0.058	0.871	0.941	0.961	0.960
	$q(Y_1, Y_2; (0, 0.25), (0,1))$	0.036	0.878	0.925	0.948	0.956
	$q(Y_1, Y_2; (0, 0.5), (0,1))$	0.047	0.764	0.844	0.890	0.911
	$q(Y_1, Y_2; (0, 0.75), (0,1))$	0.056	0.856	0.974	0.991	0.997
	$q(Y_1, Y_2; (0.25, 0.75), (0,1))$	0.057	0.813	0.960	0.987	0.994
	$q(Y_1, Y_2; (0.5, 1), (0,1))$	0.052	0.712	0.827	0.890	0.902
	$q(Y_1, Y_2; (0.75, 1), (0,1))$	0.050	0.874	0.947	0.958	0.970
	$q(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0,1))$	0.046	0.936	0.961	0.966	0.964
	$q(Y_1, Y_2; (0, 1), (0,1))$	0.053	0.980	1.000	1.000	1.000

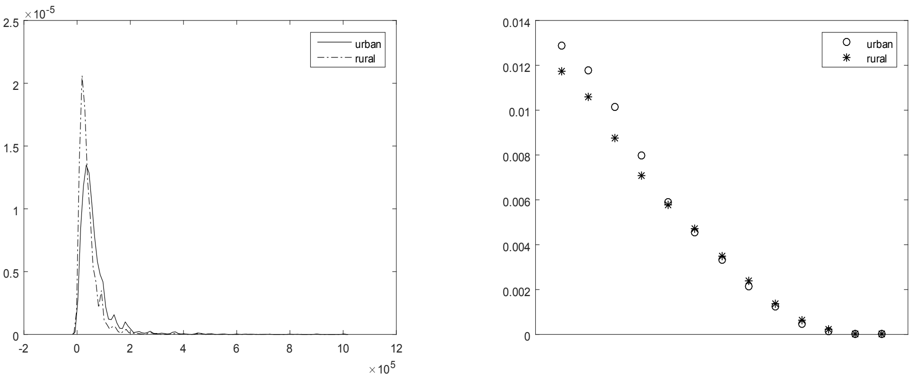
Table 2
The Hypertension Study: $\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)$ and its standard deviation, denoted by $\widehat{\sigma}\{\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)\}$, for different $\mathcal{I}_2$s. All numbers below are multiplied by 1000.

$\mathcal{I}_2$	$(0, 0.1)$	$(0.1, 0.2)$	$(0.2, 0.3)$	$(0.3, 0.4)$	$(0.4, 0.5)$
$\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)$	0.9592	1.8988	4.4976	5.2453	5.4556
$\widehat{\sigma}\{\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)\}$	0.1618	0.3108	0.4466	0.4839	0.5105
$\mathcal{I}_2$	$(0.5, 0.6)$	$(0.6, 0.7)$	$(0.7, 0.8)$	$(0.8, 0.9)$	$(0.9, 1)$
$\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)$	5.2591	5.3519	5.0317	4.7689	1.8907
$\widehat{\sigma}\{\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)\}$	0.5141	0.5275	0.5240	0.4842	0.2649

and divide both psychologists and econometricians. South West University of Finance and Economics conducted a large scale household finance survey in China. In this study, a total of 10332 households, 4321 from rural and 6011 from urban areas, are visited. For each household, both the self-reported levels of wellbeing (Y_1) and the household income (Y_2) are recorded. The densities of the household income are given in Figure 2 (A). Both indicate that the household income are highly skewed. Again we use $q(Y_1, Y_2; (0,1), \mathcal{I}_2)$ to quantify the relations between Y_1 and Y_2, for $\mathcal{I}_2 = (0,1)$, $(0.1,1)$, $(0.2,1)$, $(0.3,1)$, $(0.4,1)$, $(0.5,1)$, $(0.6,1)$, $(0.7,1)$, $(0.8,1)$, $(0.9,1)$, $(0.95,1)$, $(0.99,1)$ and $(0.995,1)$. In this study, $Q_{Y_2}(\tau_2) =$ 901, 9520, 16000, 19791, 26803, 30000, 36987, 45880, 56818, 83492, 108153, 247983, 372121, 1000000 in rural areas and $Q_{Y_2}(\tau_2)$ =400, 18046, 26870, 31425, 40000, 47483, 55856, 70000, 89897, 120392, 180000, 368240, 462102, 1000000 in urban areas, for τ_2 =0, 0.1, …, 0.9, 0.95, 0.99, 0.995 and 1. The circles and stars in Figure 2(B) exhibits the patterns of $q(Y_1, Y_2; (0,1), \mathcal{I}_2)$ varying with $\mathcal{I}_2$. Both indicate that the relations between personal sense of happiness and household incomes becomes weaker and weaker as the household income increases.

We use the simulation-based approach to test whether $q(Y_1, Y_2; (0,1), \mathcal{I}_2)$ is zero for each $\mathcal{I}_2$. The p-values are reported in Table 3. It can be clearly seen that the self-reported levels of wellbeing increased with annual household income up to RMB 372121 yuan (roughly 53931 US$) in rural areas and RMB 462102 yuan (roughly 66971 US$) in urban areas. But after that, increasing amounts of money had no further effect on happiness. In other words, once an individual can afford to satisfy their most basic needs, having more money no longer translates into more happiness. To put it in a nutshell, money does make us happier, but only up to a certain point.

3. Application to Feature Screening. In this section we generalize the application of our proposed index to feature screening in ultrahigh dimensional regressions. Suppose Y is a univariate response variable and



(A) (B)

FIG 2. *The Happiness Study. Panel (A): The density functions of household income for both urban and rural households. (B): The circles "o" and stars " ∗ " exhibit how $q(Y_1, Y_2; (0,1), \mathcal{I}_2)$ varies with $\mathcal{I}_2$ for both urban and rural households. From left to right in (B), $\mathcal{I}_2 = (0,1)$, $(0.1,1)$, $(0.2,1)$, ..., $(0.9,1)$, $(0.95,1)$, $(0.99,1)$ and $(0.995,1)$,*

TABLE 3
The p-values of the interval quantile independence tests using $n\,\widehat{q}(Y_1, Y_2; (0,1), \mathcal{I}_2)$ for different $\mathcal{I}_2$.

$\mathcal{I}_2$	$(0,1)$	$(0.1,1)$	$(0.2,1)$	$(0.3,1)$	$(0.4,1)$	$(0.5,1)$	$(0.6,1)$
rural	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
urban	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$\mathcal{I}_2$	$(0.7,1)$	$(0.8,1)$	$(0.9,1)$	$(0.95,1)$	$(0.99,1)$	$(0.995,1)$	
rural	< 0.001	< 0.001	< 0.001	< 0.001	0.025	0.156	
urban	< 0.001	< 0.001	< 0.001	< 0.001	0.018	0.057	

$\mathbf{x} \overset{\text{def}}{=} (X_1, \ldots, X_p)^{\mathrm{T}}$ is an ultrahigh dimensional covariate vector. We assume that the covariate dimension p is much larger than the sample size n. With a sample of size n, denoted by $\{(\mathbf{x}_i, Y_i), i = 1, \ldots, n\}$, we aim to identify which covariates are predictive for some quantile levels of the response variable Y. Denote $\mathcal{A}$ the indices of the important covariates, namely, $\mathcal{A} \overset{\text{def}}{=} \{k :$ The τ_1-th quantile of Y conditional on $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}}$ depends on the τ_2-th quantile level of X_k, for $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2 \subseteq (0,1) \otimes (0,1)\}$.

We propose the following screening procedure to remove as many unimportant covariates as possible. We calculate $\widehat{q}(Y, X_k; \mathcal{I}_1, \mathcal{I}_2)$ for each covariate and rank their relative importance in a descending order. It is natural to anticipate that $\widehat{q}(Y, X_k; \mathcal{I}_1, \mathcal{I}_2)$ for the important covariates is larger than that for unimportant covariates. This motivates us to retain the covariates indexed by

$$\widehat{\mathcal{A}} \overset{\text{def}}{=} \{k : \widehat{q}(Y, X_k; \mathcal{I}_1, \mathcal{I}_2) \geq c_1 n^{t-1/2}\}$$

for some $0 < t \leq 1/2$ and $c_1 > 0$. Theorem 5 ensures that $\mathcal{A} \subseteq \widehat{\mathcal{A}}$ with an overwhelming probability if the number s of elements in $\mathcal{A}$ satisfies $ns \exp(-c_2 n^{2t}) \to 0$ as $n \to \infty$ and

$$q(Y, X_k; \mathcal{I}_1, \mathcal{I}_2) \geq 2c_1 n^{t-1/2} \quad \text{for all } k \in \mathcal{A}, \tag{3.1}$$

where c_1 and c_2 will be defined shortly.

Theorem 5. *Assume the conditions in Theorem 1 hold. For any* $0 < t \leq 1/2$*, there exist positive constants* c_1 *and* c_2 *such that, as* $n \to \infty$*,*

$$pr\left\{ \mid \widehat{q}(Y, X_k; \mathcal{I}_1, \mathcal{I}_2) - q(Y, X_k; \mathcal{I}_1, \mathcal{I}_2) \mid > c_1 n^{t-1/2} \right\} = O\left\{ n \exp(-c_2 n^{2t}) \right\}.$$

If we further assume (3.1) holds, then

$$pr(\mathcal{A} \subseteq \widehat{\mathcal{A}}) \geq 1 - O\left\{ sn \exp(-c_2 n^{2t}) \right\}.$$

Assumption (3.1) allows that the marginal signal strength of the important covariates, which is quantified by $\widehat{q}(Y, X_k; \mathcal{I}_1, \mathcal{I}_2)$, shrinks to zero at a certain rate. It also requires that those signals be strong enough to be detectable. This is a key assumption to ensure our proposed screening procedure to have the desirable sure screening property. Similar conditions are widely assumed in the screening literature to ensure corresponding screening approaches to work properly. See, for example, condition 3 in [9], condition E in [10], condition C in [7], condition (C1) in [27], and condition (C2) in [19].

Example 4: A Simulation Study We use a simulated example to illustrate the finite-sample performance of this screening procedure. Consider

$$Y_i = 5X_{i,1} + X_{i,2}^2 + 2X_{i,3}X_{i,4} + \exp(X_{i,5})\varepsilon_i, \tag{3.2}$$

where $\mathbf{x}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,p})^{\mathrm{T}}$ is generated from a mixture of multivariate normal population with mean zero and covariance matrix $\mathbf{\Sigma} = \left(0.9^{|k-k'|}\right)_{p \times p}$ with probability 0.9 and standard Cauchy distribution with probability 0.1, and ε_i is drawn from (i) standard normal and (ii) standard Cauchy distribution. In this example, the active covariate set $\mathcal{A} = \{1, 2, 3, 4, 5\}$. We set $n = 200$ and $p = 5000$ in our simulations.

We consider four choices for $(\mathcal{I}_1, \mathcal{I}_2)$ in $q(Y, X_k; \mathcal{I}_1, \mathcal{I}_2)$ to perform screening: (i) $\mathcal{I}_1 = \{0.50\}, \mathcal{I}_2 = (0, 1)$; (ii) $\mathcal{I}_1 = \{0.75\}, \mathcal{I}_2 = (0, 1)$; (iii) $\mathcal{I}_1 = \mathcal{I}_2 = (0.05, 0.95)$ and (iv) $\mathcal{I}_1 = \mathcal{I}_2 = (0, 1)$. The third choice excludes 10%

data points in both X_k and Y, for $k = 1, \ldots, p$, because with probability 0.1, the observations of X_k may contain some extreme values. We compare our screening procedure with the following four competitors: the Pearson correlation based sure independence screening [9, SIS], the Kendall's rank correlation based sure independence screening [18, Kendall's tau], the distance correlation based sure independence screening [19, DC-SIS], the sure independent ranking and screening procedure [27, SIRS], MDC based quantile sure independence screening [22, MDC$_{\tau_1}$-SIS] and the quantile-adaptive sure independence screening [11, Qa$_{\tau_1}$-SIS].

We evaluate the performance of independence screening procedures using the following three criteria [19, 27].

1. The *minimal model size* $\mathcal{S}$ which is required to ensure inclusion of all truly important covariates. The closer $\mathcal{S}$ is to the number of truly important covariates in model (3.2), the better performance the corresponding screening procedure has. We report the minimum, the first quartile, the median, the third quartile, the 95-th percentile, the 99-th percentile and the maximum number of $\mathcal{S}$ for each screening procedure out of 1000 replications.
2. The *selection probability* $\mathcal{P}_{\mathcal{A}}$ that all five important covariates are ranked at the top $[n/\log n]$ and $2[n/\log n]$ positions. The closer $\mathcal{P}_{\mathcal{A}}$ is to one, the better performance the corresponding screening procedure has. We report this empirical selection probability $\mathcal{P}_{\mathcal{A}}$ for each screening procedure out of 1000 replications.
3. The *selection probability* $\mathcal{P}_{\mathcal{S}}$ that each individual important covariate is ranked at the top $[n/\log n]$ and $2[n/\log n]$ positions. If a screening procedure is able to identify X_k as an important covariate, it is reasonable to expect that $\mathcal{P}_{\mathcal{S}}$ will be close to one for this covariate. We report this empirical selection probability $\mathcal{P}_{\mathcal{S}}$ for each screening procedure and each important covariate out of 1000 replications.

It can be seen from Tables 4-5 that our proposed screening proposals perform the best throughout. In particular, the medians of $\mathcal{S}$ for both $q(Y_1, Y_2; (0.05, 0.95), (0.05, 0.95)$ and $q(Y_1, Y_2; (0, 1), (0, 1)$ equal exactly the number of truly important covariates and their inter-quartiles are at most 2. Table 5 also indicate that our proposal can detect all the truly important covariates with an overwhelming probability. Due to the presence of extreme values in the covariates, SIS [9], DC-SIS [19] and the quantile-adaptive sure independence screening procedure [11] (Qa$_{0.5}$-SIS and Qa$_{0.75}$-SIS) fail in this example. The SIRS [27] procedure fails when the error term follows Cauchy distribution. The Kendall's tau [18] and MDC$_{\tau_1}$-SIS [22] work satisfactorily

TABLE 4

The minimum, the first quartile, the median, the third quartile, the 95-th percentile, the 99-th percentile and the maximum of $\mathcal{S}$.

Error	Method	min	25%	50%	75%	95%	99%	max
Cauchy	$q(Y, X_k; \{0.5\}, (0,1))$	5	6	11	52	546	1241	2450
	$q(Y, X_k; \{0.75\}, (0,1))$	5	5	5	7	22	122	380
	$q(Y, X_k; (0.05, 0.95), (0.05, 0.95))$	5	5	5	7	26	90	241
	$q(Y, X_k; (0,1), (0,1))$	5	5	5	7	22	79	181
	SIS	5	2168	3578	4423	4891	4980	5000
	Kendall's tau	5	6	11	57	899	2865	4626
	SIRS	5	305	1028	2334	4390	4971	5000
	DC-SIS	5	173	1340	3672	4812	4955	4987
	$\text{MDC}_{0.5}$-SIS	5	7	32	252	2020	3663	4145
	$\text{MDC}_{0.75}$-SIS	5	5	7	28	860	4097	4955
	$\text{Qa}_{0.5}$-SIS	5	305	437	952	4602	4943	4998
	$\text{Qa}_{0.75}$-SIS	5	345	506	1073	4602	4943	4997
Normal	$q(Y, X_k; \{0.5\}, (0,1))$	5	5	5	6	10	113	360
	$q(Y, X_k; \{0.75\}, (0,1))$	5	5	5	5	6	8	12
	$q(Y, X_k; (0.05, 0.95), (0.05, 0.95))$	5	5	5	7	21	77	97
	$q(Y, X_k; (0,1), (0,1))$	5	5	5	6	19	58	119
	SIS	5	1989	3575	4327	4868	4967	4976
	Kendall's tau	5	6	13	55	530	4188	4626
	SIRS	5	5	6	8	38	88	5000
	DC-SIS	5	274	1038	2245	4333	4958	4986
	$\text{MDC}_{0.5}$-SIS	5	5	6	13	59	831	2989
	$\text{MDC}_{0.75}$-SIS	5	5	5	6	13	294	2370
	$\text{Qa}_{0.5}$-SIS	5	142	227	425	4487	4897	4995
	$\text{Qa}_{0.75}$-SIS	5	170	294	530	4491	4867	4995

in terms of the median values of $\mathcal{S}$. However, they are substantially inferior to our proposal in terms of inter-quartiles of $\mathcal{S}$.

Table 5 charts the empirical probabilities $\mathcal{P}_{\mathcal{A}}$ and $\mathcal{P}_{\mathcal{S}}$ that the important covariates are retained after screening for a given model size. The SIS, DC-SIS and Qa_{τ_1}-SIS are very inefficient in detecting either of the first four important covariates X_ks, for $k = 1, \ldots, 4$. The Kendall's tau and the MDC_{τ_1}-SIS are much better, but worse than our proposed screening method.

4. Concluding Remarks. In this article we introduce the concept of interval quantile independence, which generalizes the notions of both statistical independence and quantile independence. We also suggest an index in (2.1) to measure and test the departure from interval quantile independence. The proposed test based on (2.1) is consistent, unbiased and powerful. By contrast, the independence tests, such as those based on distance correlation, ranks of distances and sign covariance related to Kendall's tau, may have an inflated test size when used to test the interval quantile independence. The quantile independence tests, such as those based on linear quantile correlation and martingale difference correlation, may lose power when there exists

TABLE 5
The empirical probabilities $\mathcal{P}_S$ and $\mathcal{P}_A$.

Error	Model Size	Method	$\mathcal{P}_S$					$\mathcal{P}_A$
			X_1	X_2	X_3	X_4	X_5	
Cauchy	$[n/\log n]$	$q(Y, X_k; \{0.5\}, (0,1))$	1.000	1.000	0.970	0.886	0.741	0.709
		$q(Y, X_k; \{0.75\}, (0,1))$	1.000	1.000	0.996	0.988	0.974	0.967
		$q(Y, X_k; (0.05, 0.95), (0.05, 0.95))$	1.000	1.000	0.997	0.991	0.978	0.970
		$q(Y, X_k; (0,1), (0,1))$	1.000	1.000	0.997	0.990	0.983	0.974
		SIS	0.129	0.286	0.063	0.049	0.468	0.022
		Kendall's tau	1.000	0.998	0.958	0.852	0.761	0.700
		SIRS	0.912	0.567	0.513	0.467	0.500	0.082
		DC-SIS	0.294	0.433	0.144	0.118	0.542	0.056
		$\text{MDC}_{0.5}$-SIS	0.975	0.958	0.894	0.822	0.674	0.523
		$\text{MDC}_{0.75}$-SIS	0.973	0.974	0.944	0.930	0.904	0.765
		$\text{Qa}_{0.5}$-SIS	0.093	0.231	0.039	0.054	0.565	0.023
		$\text{Qa}_{0.75}$-SIS	0.070	0.118	0.037	0.058	0.492	0.023
	$2[n/\log n]$	$q(Y, X_k; \{0.5\}, (0,1))$	1.000	1.000	0.979	0.924	0.823	0.798
		$q(Y, X_k; \{0.75\}, (0,1))$	1.000	1.000	0.997	0.996	0.987	0.982
		$q(Y, X_k; (0.05, 0.95), (0.05, 0.95))$	1.000	1.000	0.999	0.995	0.989	0.986
		$q(Y, X_k; (0,1), (0,1))$	1.000	1.000	0.999	0.995	0.993	0.989
		SIS	0.197	0.339	0.084	0.072	0.525	0.022
		Kendall's tau	1.000	0.999	0.974	0.896	0.819	0.782
		SIRS	0.931	0.609	0.562	0.526	0.574	0.114
		DC-SIS	0.392	0.502	0.247	0.211	0.680	0.124
		$\text{MDC}_{0.5}$-SIS	0.982	0.966	0.915	0.863	0.754	0.610
		$\text{MDC}_{0.75}$-SIS	0.978	0.988	0.960	0.951	0.925	0.817
		$\text{Qa}_{0.5}$-SIS	0.185	0.387	0.076	0.090	0.644	0.025
		$\text{Qa}_{0.75}$-SIS	0.138	0.295	0.064	0.083	0.611	0.024
Normal	$[n/\log n]$	$q(Y, X_k; \{0.5\}, (0,1))$	1.000	1.000	1.000	0.990	0.900	0.895
		$q(Y, X_k; \{0.75\}, (0,1))$	1.000	1.000	1.000	1.000	1.000	1.000
		$q(Y, X_k; (0.05, 0.95), (0.05, 0.95))$	1.000	1.000	1.000	0.995	0.980	0.980
		$q(Y, X_k; (0,1), (0,1))$	1.000	1.000	1.000	0.995	0.990	0.985
		SIS	0.145	0.345	0.045	0.060	0.440	0.025
		Kendall's tau	1.000	1.000	0.970	0.885	0.760	0.690
		SIRS	0.940	0.605	0.520	0.470	0.555	0.085
		DC-SIS	0.335	0.490	0.120	0.130	0.500	0.065
		$\text{MDC}_{0.5}$-SIS	0.975	0.965	0.940	0.930	0.825	0.710
		$\text{MDC}_{0.75}$-SIS	0.980	0.975	0.975	0.955	0.950	0.845
		$\text{Qa}_{0.5}$-SIS	0.179	0.372	0.091	0.085	0.554	0.039
		$\text{Qa}_{0.75}$-SIS	0.151	0.229	0.083	0.091	0.491	0.042
	$2[n/\log n]$	$q(Y, X_k; \{0.5\}, (0,1))$	1.000	1.000	1.000	0.990	0.940	0.935
		$q(Y, X_k; \{0.75\}, (0,1))$	1.000	1.000	1.000	1.000	1.000	1.000
		$q(Y, X_k; (0.05, 0.95), (0.05, 0.95))$	1.000	1.000	1.000	0.995	0.995	0.990
		$q(Y, X_k; (0,1), (0,1))$	1.000	1.000	1.000	0.995	1.000	0.995
		SIS	0.230	0.400	0.070	0.070	0.495	0.025
		Kendall's tau	1.000	1.000	0.975	0.900	0.825	0.785
		SIRS	0.955	0.640	0.575	0.540	0.640	0.135
		DC-SIS	0.390	0.560	0.255	0.240	0.675	0.155
		$\text{MDC}_{0.5}$-SIS	0.985	0.970	0.955	0.945	0.865	0.770
		$\text{MDC}_{0.75}$-SIS	0.985	0.985	0.980	0.980	0.960	0.900
		$\text{Qa}_{0.5}$-SIS	0.343	0.540	0.192	0.171	0.667	0.067
		$\text{Qa}_{0.75}$-SIS	0.255	0.457	0.150	0.147	0.617	0.060

interval quantile dependence. We further utilize the proposed interval quantile index as a marginal utility to perform feature screening for ultrahigh dimensional data. This screening procedure is model-free, conceptually simple, convenient to implement with no tuning parameters or nonparametric model fitting involved. The desirable sure screening property is also established. We demonstrate the effectiveness of our proposed screening procedure in comparison with existing methods.

There is another closely relevant measure which can also be used to quantify the degree of quantile dependence. It is defined as

$$q_{\text{common}}(Y_1, Y_2; \mathcal{I}) \stackrel{\text{def}}{=} \int_{\mathcal{I}} \frac{\text{cov}^2\{I(Y_1 \le Q_{Y_1}(\tau)), I(Y_2 \le Q_{Y_2}(\tau))\}}{\tau^2(1-\tau)^2} d\mu(\tau).$$

This metric is related to the tail dependence [13] if we set $\mathcal{I} = (0, \tau)$ or $\mathcal{I} = (1-\tau, 1)$ for $\tau \to 0$. One can show that $Q_{Y_1|Y_2=Q_{Y_2}(\tau)}(\tau) = Q_{Y_1}(\tau)$, for $\tau \in \mathcal{I}$, implies $q(Y_1, Y_2; \mathcal{I}) = 0$; and $q(Y_1, Y_2; \mathcal{I}) = 1$ if $Y_2 = m_1(Y_1)$ for some strictly monotone function m_1. However, $q(Y_1, Y_2; \mathcal{I}) = 0$ does not imply $Q_{Y_1|Y_2=Q_{Y_2}(\tau)}(\tau) = Q_{Y_1}(\tau)$, for $\tau \in \mathcal{I}$. In other words, $q_{\text{common}}(Y_1, Y_2; \mathcal{I})$ is possibly inconsistent in testing the interval quantile independence (1.1). Thus, we advocate using our proposed index defined in (2.1).

The distance correlation can be used to characterize statistical independence between two random vectors in arbitrary dimensions. The martingale difference correlation can be used to quantify quantile dependence of a univariate variable on another random vector, and the mean dependence of a random vector on another. It is thus natural to ask whether and how we can generalize the proposed index to the cases with random vectors. This is however not straightforward because defining quantiles for random vectors is essentially different from that for a univariate random variable. If the componentwise quantile dependence between two random vectors, $\mathbf{y}_1 = (Y_{11}, \ldots, Y_{1p})^{\text{T}}$ and $\mathbf{y}_2 = (Y_{21}, \ldots, Y_{2q})^{\text{T}}$, is of interest at quantile levels $\mathcal{I}_1 = \mathcal{I}_{11} \otimes \mathcal{I}_{12} \cdots \otimes \mathcal{I}_{1p} \subseteq (0,1) \otimes (0,1) \cdots \otimes (0,1)$ and $\mathcal{I}_2 = \mathcal{I}_{21} \otimes \mathcal{I}_{22} \cdots \otimes \mathcal{I}_{2q} \subseteq (0,1) \otimes (0,1) \cdots \otimes (0,1)$, we can define

$$\begin{aligned} & q_{\text{sum}}(\mathbf{y}_1, \mathbf{y}_2; \mathcal{I}_1, \mathcal{I}_2) \\ \stackrel{\text{def}}{=} & \int_{\mathcal{I}_1} \int_{\mathcal{I}_2} \frac{\text{cov}^2\left\{\prod_{k=1}^{p} I(Y_{1k} \le Q_{Y_{1k}}(\tau_{1k})), \prod_{l=1}^{q} I(Y_{2l} \le Q_{Y_{2l}}(\tau_{2l}))\right\}}{\prod_{k=1}^{p} \tau_{1k}(1-\tau_{1k}) \prod_{l=1}^{q} \tau_{2l}(1-\tau_{2l})} d\mu_1(\tau_1) d\mu_2(\tau_2), \end{aligned}$$

where $d\mu_1(\tau_1) = \prod_{k=1}^{p} d\mu_{1k}(\tau_{1k})$ and $d\mu_2(\tau_2) = \prod_{l=1}^{q} d\mu_{2l}(\tau_{2l})$. Similarly, we define

$$q_{\max}(\mathbf{y}_1, \mathbf{y}_2; \mathcal{I}_1, \mathcal{I}_2) \overset{\text{def}}{=} \max_{1\le k\le p} \max_{1\le l\le q} \int_{\mathcal{I}_{1k}} \int_{\mathcal{I}_{2l}} \frac{\operatorname{cov}^2\{I(Y_{1k} \le Q_{Y_{1k}}(\tau_{1k})), I(Y_{2l} \le Q_{Y_{2l}}(\tau_{2l}))\}}{\prod_{k=1}^{p} \tau_{1k}(1-\tau_{1k}) \prod_{l=1}^{q} \tau_{2l}(1-\tau_{2l})} d\mu_{1k}(\tau_{1k}) d\mu_{2l}(\tau_{2l}).$$

One may also wonder how to quantify the interval quantile dependence between two univariate random variables Y_1 and Y_2 (or two multivariate random vectors $\mathbf{y}_1$ an $\mathbf{y}_2$) in the presence of a high dimensional covariate vector $\mathbf{x}$. This is an interesting and yet very challenging issue, which warrants thorough investigation.

APPENDIX A: PROOF OF PROPOSITION 1

(i) We first notice that $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = 0 \Leftrightarrow \operatorname{pr}\{Y_1 \le Q_{Y_1}(\tau_1), Y_2 \le Q_{Y_2}(\tau_2)\} = \tau_1\tau_2$, for $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$. Taking derivative with respective to τ_2 on both sides of the above equation and using the fact that $f_2\{Q_{Y_2}(\tau_2)\}d\{Q_{Y_2}(\tau_2)\}/d\tau_2 = 1$, we obtain that $\operatorname{pr}\{Y_1 \le Q_{Y_1}(\tau_1) \mid Y_2 = Q_{Y_2}(\tau_2)\} = \tau_1$. Therefore, a direct consequence of the uniqueness of $Q_{Y_1|Y_2}(\tau_1)$ is that $Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1)$, for $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$, which completes the " $\Rightarrow$ " part.

Now we turn to the " $\Leftarrow$ " part. That $Q_{Y_1|Y_2=Q_{Y_2}(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1)$ yields immediately that $\operatorname{pr}\{Y_1 \le Q_{Y_1}(\tau_1) \mid Y_2 = Q_{Y_2}(\tau_2)\} = \tau_1$ for $(\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2$. Consequently,

$$\begin{aligned} \operatorname{pr}\{Y_1 \le Q_{Y_1}(\tau_1), Y_2 \le Q_{Y_2}(\tau_2)\} &= E\left[\operatorname{pr}\{Y_1 \le Q_{Y_1}(\tau_1) \mid Y_2\} I(Y_2 \le Q_{Y_2}(\tau_2))\right] \\ &= \tau_1 E\left\{I(Y_2 \le Q_{Y_2}(\tau_2))\right\} = \tau_1\tau_2. \end{aligned}$$

This completes the proof of the " $\Leftarrow$ " part.

(ii) To prove the first part, it suffices to prove the special case $\mathcal{I}_1 = \{\tau_1\}$ because the integrand is nonnegative. We note that

$$q(Y_1, Y_2; \tau_1, (0,1)) = 0 \Leftrightarrow E\{I(Y_1 \le Q_{Y_1}(\tau_1)) \mid Y_2\} = \tau_1.$$

Therefore, the first part is proven through the uniqueness of $Q_{Y_2|Y_1}(\tau_1)$.

We are in the position to prove the second part. Note that

$$\begin{aligned} & q(Y_1, Y_2; (0,1), (0,1)) = 0. \\ \Leftrightarrow \quad & \operatorname{cov}\{I(Y_1 \le y_1), I(Y_2 \le y_2)\} = 0, \forall y_k \in \{Q_{Y_k}(\tau_k) : \tau_k \in (0,1)\}, k = 1, 2. \end{aligned}$$

Therefore, $q(Y_1, Y_2; (0,1), (0,1)) = 0$ is tantamount to that the intergrand is zero. The second equivalency follows from the arbitratiness of $\tau_k \in (0,1)$. The right hand of the above display entails that Y_1 and Y_2 are independent.

(iii) Using the fact both m_1 and m_2 are strictly increasing functions, we have that $I(m_k(Y_k) \leq Q_{m_k(Y_k)}(\tau_k)) = I(m_k(Y_k) \leq m_k(Q_{Y_k}(\tau_k))) = I(Y_k \leq Q_{Y_k}(\tau_k))$, which yields that $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(m_1(Y_1), m_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)$. $\square$

REFERENCES

[1] Anderson, X. and Darling, X. (1952). Asymptotic theory of certain "Goodness-of-Fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics* **23** 193–212.

[2] Andrews,D. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrika* **68** 399–405.

[3] Bergsma, W. and Dassios, A. (2014) A consistent test of independence based on a sign covariance related to Kendalls tau. *Bernoulli* **20** 1006–1028.

[4] Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961) Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics* **32** 485–498.

[5] Csörgö, M. and Horváth, L. (1997) *Limit Theorems in Change-Point Analysis* John Wiley & Sons Inc

[6] Davies,R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254.

[7] Fan, J., Feng, Y. and Song, R. (2011) Nonparametric independence screening in sparse ultra-high dimensional additive models, *Journal of the American Statistical Association,* **106**, 544–557.

[8] Fan, J., Hu, T.-C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, **21** 433–446

[9] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society, Series B*, **70** 849–911.

[10] Fan, J. and Song, R. (2010) Sure independence screening in generalized linear models with np-dimensionality, *The Annals of Statistics,* **38**, 3567C-3604.

[11] He, X., Wang, L., and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, **41** 342–369.

[12] Heller, R., Heller, Y. and Corfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika*, **100** 503–510.

[13] Joe, H. (1997) *Multivariate Models and Dependence Concepts.* New York: Chapman.

[14] Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, Cambridge.

[15] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46** 33C-49.

[16] Koenker, R. and Portnoy, S. (1987) L-estimators for linear models. *Journal of the American Statistical Association* **82** 851–857.

[17] Li, G., Li, Y. and Tsai, C. L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association* **110** 246–261.

[18] Li, G., Peng, H., Zhang, J. and Zhu, L. (2012). Robust rank correlation based screening. *The Annals of Statistics*, **40** 1846–1877.

[19] Li, R., Zhong, W. and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, *107* 1129–1139.

[20] McKeague, I. W. and Qian, M. (2015). An adaptive resampling test for test detecting the presence of significant predictors. *Journal of the American Statistical Association*, *110* 1422–1433.

[21] Park, T., Shao, X. and Yao, S. (2015). Partial martingale difference correlation. *Electronic Journal of Statistics*, *9* 1492–1517.

[22] Shao, X. and Zhang, J. (2014) Martingale difference correlation and its use in high dimensional variable screening. *Journal of the American Statistical Association* **109** 1302–1318.

[23] Szekely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35** 2769–2794.

[24] Szekely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics.* **3** 1236–1265.

[25] Wu, Y. and Yin, G. (2014). Conditional quantile screening in ultrahigh-dimensional heterogeneous data. *Biometrika* **102** 65–76.

[26] Zheng, Q., Peng, L., and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics*, **43** 2225–2258.

[27] Zhu, L. P., Li, L., Li, R., and Zhu, L. X. (2011). Model-free feature screening for ultrahigh dimensional data. *Journal of the American Statistical Association*, **106** 1464–1475.

Center for Applied Statistics
Institute of Statistics and Big Data
Renmin University of China
59 Zhongguancun Avenue, Haidian District
Beijing 100872, P. R. China.
E-mail: zhu.liping@ruc.edu.cn

School of Information Management and Engineering
Shanghai University of Finance and Economics
777 Guoding Road
Shanghai 200433, P. R. China.
E-mail: zhangyaowucp@163.com

School of Statistics and Management
Shanghai University of Finance and Economics
777 Guoding Road
Shanghai 200433, P. R. China.
E-mail: tjxxukai@163.com