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 \(\tau_1\)-th quantile of \(Y_1\) conditional on \(Y_2\) and \(Q_{Y_1}(\tau_1)\) is the unconditional \(\tau_1\)-th quantile of \(Y_1\). The \(\tau_1\)-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 \(\tau_1\) 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 \(\tau_1\) 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 \(qcorr_{\tau_1}(Y_1 | Y_2) \overset{\text{def}}{=} \text{cov}(\{Y_1 \geq Q_{Y_1}(\tau_1)\}, Y_2)/(\{\tau_1(1 - \tau_1)\text{var}(Y_2)\}^{1/2})\), 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 (16JJD010002) 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

$$H_0: Q_{Y_1|Y_2=(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1) \text{ for } (\tau_1, \tau_2) \in I_1 \otimes I_2 \subseteq (0, 1) \otimes (0, 1),$$

(1.1) versus $H_1: \text{others}$.

We refer to $H_0$ in (1.1) as the interval quantile independence because both $I_1$ and $I_2$ can be intervals and we are concerned with quantile independence over two intervals. In particular, if $I_1$ is a singleton, say, $I_1 = \{\tau_1\}$, and $I_2 = (0, 1)$, $H_0$ 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 $\tau_1$. If $I_1 = I_2 = (0, 1)$, 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 $I_1 \subseteq (0, 1)$ and $I_2 = (0, 1)$. 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 $I_1 \otimes I_2$. 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 $I_1 \otimes I_2 \subseteq (0,1) \otimes (0,1)$. However, even if the interval quantile independence holds true for some $I_1 \otimes I_2 \subseteq (0,1) \otimes (0,1)$, $Y_1$ 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 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 $\tau_1 \in I_1$ and $I_2 = (0,1)$. 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,

$$\lim_{\tau_1 \to 0} \Pr \{ Y_2 \leq Q_{Y_2} (\tau_1) \mid Y_1 \leq Q_{Y_1} (\tau_1) \} \quad \text{and} \quad \lim_{\tau_1 \to 1} \Pr \{ Y_2 \geq Q_{Y_2} (\tau_1) \mid Y_1 \geq Q_{Y_1} (\tau_1) \}.$$ 

The lower and upper tail dependence of $(Y_1, Y_2)$ corresponds to the interval quantile dependence with $I_1 = I_2 = (0, \tau_1)$ for $\tau_1 \to 0$, and $I_1 = I_2 = (\tau_1, 1)$ 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 $I_1 \otimes I_2$ 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 $q(Y_1, Y_2; I_1, I_2)$, 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 $q(Y_1, Y_2; I_1, I_2) \geq 0$ 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 $q(Y_1, Y_2; I_1, I_2) = q(m_1(Y_1), m_2(Y_2); I_1, I_2)$ for monotonically increasing functions $m_1$ and $m_2$.

It can also be used to test quantile independence through setting $I_1 = \{ \tau_1 \}$ and $I_2 = (0,1)$ and statistical independence through setting $I_1 = I_2 = (0,1)$. We suggest a moment estimator to implement our proposed test. The resulting estimate, denoted by $\hat{q}(Y_1, Y_2; I_1, I_2)$, depends only on the ranks of the observations. We show that, in the general case of $q(Y_1, Y_2; I_1, I_2) > 0$, $n^{1/2} \{ \hat{q}(Y_1, Y_2; I_1, I_2) - q(Y_1, Y_2; I_1, I_2) \}$ is asymptotically normal, and in the particular case of $q(Y_1, Y_2; I_1, I_2) = 0$, $n \hat{q}(Y_1, Y_2; I_1, I_2)$ 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 “\(\Omega_k\)” the support of \(Y_k\), namely, \(\Omega_k \overset{\text{def}}{=} \{y_k : f_k(y_k) > 0\}\) where \(f_k\) stands for the marginal density function of \(Y_k\). Denote by \(Q_{Y_k}(\tau_k)\) the \(\tau_k\)-th quantile of \(Y_k\) and \(Q_{Y_1|Y_2}(\tau_1)\) the \(\tau_1\)-th quantile of \(Y_1\) conditional on \(Y_2\). In general, \(Q_{Y_1|Y_2}(\tau_1)\) varies with \((\tau_1, Y_2)\).

We assume throughout that \(Q_{Y_1|Y_2}(\tau_1)\) is uniquely defined as a function of \(\tau_1\) for each \(y_2 \in \Omega_2\). Let “\(\iff\)” stand for “is equivalent to”, “\(\overset{d}{\rightarrow}\)” stand for “converges in distribution”, “\(\overset{pr}{\rightarrow}\)” stand for “converges in probability” and “\(\overset{d}{=}\)” stand for “has the same distribution as”. Define \(F_k(y_k) \overset{\text{def}}{=} \text{pr}(Y_k \leq y_k)\) and \(F_{1,2}(y_1, y_2) \overset{\text{def}}{=} \text{pr}(Y_1 \leq y_1, Y_2 \leq y_2)\). 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 \(f_{1|2}(y_1 | y_2)\) 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, \(F_{n,k}(y_k) \overset{\text{def}}{=} n^{-1} \sum_{i=1}^n I(Y_{i,k} \leq y_k)\) and \(F_{n,1,2}(y_1, y_2) \overset{\text{def}}{=} n^{-1} \sum_{i=1}^n I(Y_{i,1} \leq y_1, Y_{i,2} \leq y_2)\), 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

$$Q_{Y_1|Y_2}(\tau_1) = Q_{Y_1}(\tau_1) \iff E\{I(Y_1 \leq Q_{Y_1}(\tau_1)) \mid Y_2\} = \tau_1$$

$$\iff \text{cov}\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq y_2)\} = 0, \text{ for all } y_2 \in \Omega_2.$$  

$$\iff \int_{\Omega_2} \frac{\text{cov}^2\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq y_2)\}}{\tau_1(1 - \tau_1)F_2(y_2)\{1 - F_2(y_2)\}} dy_2 = 0.$$  

$$\iff \int_0^1 \frac{\text{cov}^2\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq Q_{Y_2}(\tau_2))\}}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)} d\tau_2 = 0.$$  

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 $E(\varepsilon \mid X) = 0 \iff E\{\varepsilon I(X \leq x)\} = 0$, for all $x$ lies in the support of $X$. The third equivalency is obvious because the integrand is nonnegative. Note that

$$\int_{\Omega_2} \frac{\text{cov}^2\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq y_2)\}}{\tau_1(1 - \tau_1)F_2(y_2)\{1 - F_2(y_2)\}} dy_2$$

$$= \int_0^1 \frac{\text{cov}^2\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq Q_{Y_2}(\tau_2))\}}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)f_2(Q_{Y_2}(\tau_2))} d\tau_2$$

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 $I_k$s be two subsets of $(0, 1)$, namely, $I_k \subseteq (0, 1)$, $I_k$ can be a singleton, say, $I_k = \{\tau_k\}$. We define the following index to measure and test $H_0$ in (1.1),

$$q(Y_1, Y_2; I_1, I_2) \overset{\text{def}}{=} \int_{I_1} \int_{I_2} \frac{\text{cov}^2\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq Q_{Y_2}(\tau_2))\}}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)} d\mu_1(\tau_1)d\mu_2(\tau_2),$$

where $\mu_1$s are two probability measures which can be different and depend on $I_k$. 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 $I_1 = \{\tau_1\}$ and $I_2 = (0, 1)$ in $q(Y_1, Y_2; I_1, I_2)$. 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 $q(Y_1, Y_2; I_1, I_2)$ at the population level.
Proposition 1. We assume that \( I_k = \{ \tau_k : d\mu_k(\tau_k)/d\tau_k > 0 \} \).

(i) If \( Q_{Y_1|Y_2 = Q_{Y_2}(\tau_2)}(\tau_1) \) is unique for \((\tau_1, \tau_2) \in I_1 \otimes I_2\), then \( q(Y_1, Y_2; I_1, I_2) = 0 \) if and only if \( q(Y_1, Y_2; I_1, (0, 1)) = 0 \) if and only if \( Y_1 \) and \( Y_2 \) are statistically independent.

(ii) If \( Q_{Y_1|Y_2 = Q_{Y_2}(\tau_2)}(\tau_1) = Q_{Y_1}(\tau_1) \) for \((\tau_1, \tau_2) \in I_1 \otimes I_2\), then \( q(Y_1, Y_2; I_1, (0, 1)) = 0 \) if and only if \( Y_1 \) and \( Y_2 \) are statistically independent.

(iii) If \( m_1 \) and \( m_2 \) are monotonically increasing functions, \( q(Y_1, Y_2; I_1, I_2) = q(m_1(Y_1), m_2(Y_2); I_1, I_2) \).

The first property in Proposition 1 indicates that, through using different \( I_k \)s, \( q(Y_1, Y_2; I_1, I_2) \) 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, \( q(Y_1, Y_2; I_1, I_2) \) can be used to test statistical independence and quantile independence by choosing \( I_k \)s properly. Note that \( I_k \)s are not tuning parameters. The role of \( I_k \)s in \( q(Y_1, Y_2; I_1, I_2) \) is in spirit the same as that of quantile levels in quantile regressions [14, 15]. How to choose \( I_k \)s depends on our purposes. Different users may specify different \( I_k \)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 \( I_1 = \{0.5\} \) and \( I_2 = (0, 1) \); if we aim to test whether the \( \tau_1 \)-th quantile function of \( Y_1 \) depends on \( Y_2 \) for \( \tau_1 \in (0.25, 0.75) \), we may specify \( I_1 = (0.25, 0.75) \) and \( I_2 = (0, 1) \); if we hope to test whether the first and the third quartiles of \( Y_1 \) depend on \( Y_2 \), we may specify \( I_1 = \{0.25, 0.75\} \) and \( I_2 = (0, 1) \); and if we aim to test whether \( Y_1 \) is independent of \( Y_2 \), we may choose \( I_1 = I_2 = (0, 1) \).

How to choose the probability measures \( \mu_k \)s depends on the intervals \( I_k \)s. We require throughout that \( I_k = \{ \tau_k : d\mu_k(\tau_k)/d\tau_k > 0 \} \). We specify \( \mu_k \) as a Lebesgue measure if \( I_k \) is an interval and a counting measure if \( I_k \) is a countable set. If \( I_k = (\tau_{k,1}, \tau_{k,2}) \) 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 \( \tau_{k,1} \leq \tau_k < \tau_{k,2} \), \( \mu_k(\tau_k) = 0 \) if \( \tau_k < \tau_{k,1} \) and \( \mu_k(\tau_k) = 1 \) if \( \tau_k \geq \tau_{k,2} \). In this case, \( d\mu_k(\tau_k)/d\tau_k = 1/(\tau_{k,2} - \tau_{k,1})I(\tau_{k,1} \leq \tau_k < \tau_{k,2}) \). If \( I_k = \{\tau_{k,1}, \tau_{k,2}\} \), we set the counting measure \( \mu_k(\tau_k) = 0 \) if \( \tau_k < \tau_{k,1} \), \( \mu_k(\tau_k) = 1/2 \) if \( \tau_{k,1} \leq \tau_k < \tau_{k,2} \), and \( \mu_k(\tau_k) = 1 \) if \( \tau_k \geq \tau_{k,2} \). In this case, \( d\mu_k(\tau_k)/d\tau_k = 1/2I(\tau_k \in I_k) \). In the particular case of \( I_1 = (0.25, 0.75) \) and \( I_2 = (0.25, 0.75) \), our proposed index reduces to the following simple form:

\[
q(Y_1, Y_2; I_1, I_2) = \sum_{\tau_1 \in I_1} \int_{I_2} \frac{\text{cov}^2\{I(Y_1 \leq Q_{Y_1}(\tau_1)), I(Y_2 \leq Q_{Y_2}(\tau_2))\}}{\tau_1(1-\tau_1)\tau_2(1-\tau_2)} d\tau_2.
\]

The third property in Proposition 1 shows that, \( q(Y_1, Y_2; I_1, I_2) \) 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

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

This invariant property play an important role here in that it allows us to assume subsequently that \(Y_k\)s have compact support because otherwise we replace \(Y_k\)s 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 \(\{(Y_{1,i}, Y_{2,i}), i = 1, \ldots, n\}\) 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

\[
q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = q(F_1(Y_1), F_2(Y_2); \mathcal{I}_1, \mathcal{I}_2)
\]

\[
= \sum_{j_1=1}^n \sum_{j_2=1}^n \int_{\mathcal{I}_1 \cap [(j_1-1)/n, j_1/n]} \int_{\mathcal{I}_2 \cap [(j_2-1)/n, j_2/n]} \frac{\text{cov}^2\{I(F_1(Y_1) \leq \tau_1), I(F_2(Y_2) \leq \tau_2)\}}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2),
\]

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

\[
\hat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \overset{\text{def}}{=} \sum_{j_1=1}^n \sum_{j_2=1}^n \int_{\mathcal{I}_1 \cap [(j_1-1)/n, j_1/n]} \int_{\mathcal{I}_2 \cap [(j_2-1)/n, j_2/n]} \frac{\text{cov}^2\{I(F_{n,1}(Y_1) \leq \tau_1), I(F_{n,2}(Y_2) \leq \tau_2)\}}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2),
\]

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

\[
\text{cov}\{I(F_{n,1}(Y_1) \leq \tau_1), I(F_{n,2}(Y_2) \leq \tau_2)\} \overset{\text{def}}{=} n^{-1} \sum_{i=1}^n I(F_{n,1}(Y_{1,i}) \leq \tau_1) I(F_{n,2}(Y_{2,i}) \leq \tau_2)
\]

\[
- n^{-2} \sum_{i=1}^n I(F_{n,1}(Y_{1,i}) \leq \tau_1) \sum_{i=1}^n I(F_{n,2}(Y_{2,i}) \leq \tau_2).
\]
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,

$$
\tilde{q}(Y_1, Y_2; (0, 1), (0, 1)) \overset{def}{=} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \left[ \tilde{\operatorname{cov}}^2 \{ I(F_{n,1}(Y_1) \leq j_1/n), I(F_{n,2}(Y_2) \leq j_2/n) \} \right]
$$

$$
= \frac{1}{\tau_1(1 - \tau_1)} \int_{I_1 \cap [(j_1 - 1)/n, j_1/n)} d\mu_1(\tau_1)
$$

$$
+ \frac{1}{\tau_2(1 - \tau_2)} \int_{I_2 \cap [(j_2 - 1)/n, j_2/n)} d\mu_2(\tau_2)
$$

and $\tilde{q}(Y_1, Y_2; \tau_1, (0, 1)) \overset{def}{=} \sum_{j_2=1}^{n} \left[ \tilde{\operatorname{cov}}^2 \{ I(F_{n,1}(Y_1) \leq \tau_1), I(F_{n,2}(Y_2) \leq j_2/n) \} \right]

$$

$$
= \frac{1}{\tau_1(1 - \tau_1)} \int_{I_1 \cap [(j_1 - 1)/n, j_1/n)} d\mu_1(\tau_1)
$$

$$
+ \frac{1}{\tau_2(1 - \tau_2)} \int_{I_2 \cap [(j_2 - 1)/n, j_2/n)} d\mu_2(\tau_2)
$$

Note that the integrals

$$
\int_{I_1 \cap [(j_k - 1)/n, j_k/n)} \frac{1}{\tau_k(1 - \tau_k)} d\mu_k(\tau_k), k = 1, 2,
$$

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

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

If $\mu_k$ is a counting measure, say, $\mu_k(\tau) = I(\tau \geq \tau_k, 0)$, then, for $\tau_k, 0 \in [a, b]$,

$$
\int_{a}^{b} \frac{1}{\tau_k(1 - \tau_k)} d\mu_k(\tau_k) = \frac{1}{\tau_k, 0(1 - \tau_k, 0)}.
$$

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

Theorem 1 states that $\tilde{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ 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 $\tau_k$ 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

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

where $\sigma^2 \overset{def}{=} 4 \operatorname{var}(Z)$ and $Z$ is defined in (S1.5).
2. If $H_0$ in (1.1) is true, then $q(Y_1, Y_2; I_1, I_2) = 0$, and

$$n \tilde{q}(Y_1, Y_2; I_1, I_2) \overset{d}{\to} \int_{I_1} \int_{I_2} \frac{B^2(\tau_1, \tau_2)}{\tau_1(1 - \tau_1) \tau_2(1 - \tau_2)} d\mu_1(\tau_1)d\mu_2(\tau_2)$$

$$= \sum_{j=1}^{\infty} \lambda_j \chi_j^2(1).$$

where $B(\tau_1, \tau_2)$ is a separable Gaussian process depending on $(\tau_1, \tau_2)$ for $(\tau_1, \tau_2) \in I_1 \otimes I_2 \subseteq (0, 1) \otimes (0, 1)$, $E\{B(\tau_1, \tau_2)\} = 0$ 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 $\lambda_j$'s are eigenvalues defined in (S1.7) which depend on $(I_1, I_2)$ 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 $Q_{Y_k}(\tau_k)$ converges in probability to $Q_{Y_k}(\tau_k)$ uniformly for $\tau_k \in I_k \subseteq (0, 1)$. 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 $Y_k$'s 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 $q(Y_1, Y_2; I_1, I_2) = q(F_1(Y_1), F_2(Y_2); I_1, I_2)$. We estimate $q(Y_1, Y_2; I_1, I_2)$ using the ranks of $Y_k$'s only. The empirical distribution functions $F_{n,k}(Y_k)$'s are monotonically increasing. It follows immediately that $\tilde{q}(Y_1, Y_2; I_1, I_2) = \tilde{q}(F_{n,1}(Y_1), F_{n,2}(Y_2); I_1, I_2)$. In general, if $m_k(Y_k)$'s are monotonically increasing, we can replace $Y_k$'s 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 $q(Y_1, Y_2; I_1, I_2) = q(m_1(Y_1), m_2(Y_2); I_1, I_2)$ and $\tilde{q}(Y_1, Y_2; I_1, I_2) = \tilde{q}(m_1(Y_1), m_2(Y_2); I_1, I_2)$. In addition, we require that the boundedness assumption hold uniformly for $\tau_k \in I_k$ 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 $n \tilde{q}(Y_1, Y_2; I_1, I_2) > c_\alpha$, where the critical value at the significance level $\alpha$, $c_\alpha$, is defined as the upper $\alpha$ quantile of the asymptotic null distribution of $n \tilde{q}(Y_1, Y_2; I_1, I_2)$ under $H_0$. Theorem 1 ensures that using $q(Y_1, Y_2; I_1, I_2)$ to test (1.1) is consistent, and the power is approximately

$$\beta_n \overset{d}{=} \text{pr} \{n \tilde{q}(Y_1, Y_2; I_1, I_2) > c_\alpha \mid q(Y_1, Y_2; I_1, I_2) > 0\}$$

$$\approx 1 - \Phi\left\{\frac{c_\alpha - n q(Y_1, Y_2; I_1, I_2)}{\sqrt{n^{1/2} \sigma}}\right\},$$
where $\Phi$ is the cumulative distribution function of $N(0, 1)$. Apparently, $\beta_n \to 1$ as $n \to \infty$, indicating that our test for (1.1) is consistent.

How to decide a critical value $c_\alpha$ 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, $d\mu_1(\tau)/d\tau = \mathcal{I}(\tau = \tau_1)$ and $d\mu_2(\tau) = d\tau$. Following [1] and [4], we can show that

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

The limiting distribution can be approximated with

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

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

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

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

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\chi^2_{ij}(1)}{i(i+1)j(j+1)}.$$ 

Because the asymptotic distributions are approximately tractable, the critical value $c_\alpha$ 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 $c_\alpha$. Theorem 1 states that, under $H_0$ in (1.1), the asymptotic distribution of $n \tilde{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ 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 $Y_{i,k}$ 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 $\{Y^*_{i,1}, Y^*_{i,2}\}, i = 1, \ldots, n$. We repeat this procedure for $B$ times and set $c_\alpha$ to be the upper $\alpha$ quantile of the estimates of $q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)$ obtained from the randomly generated samples. Because
all $Y_{i,k}^*$s are independent, it is natural to anticipate that this procedure
provides a reasonable approximation of the asymptotic null distribution of
$n \tilde{q}(Y_1, Y_2; I_1, I_2)$ for a sufficiently large $B$. Throughout we use this method
to decide $c_\alpha$ in the test for (1.1).

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

\textbf{Theorem 2.} Under the conditions of Theorem 1, it follows that
\[ n \tilde{q}(Y_1^*, Y_2^*; I_1, I_2) \xrightarrow{d} \int_{I_1} \int_{I_2} \frac{B^2(\tau_1, \tau_2)}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)} d\mu_1(\tau_1)d\mu_2(\tau_2) \]
where $B(\tau_1, \tau_2)$ is defined in Theorem 1.

![Fig 1](image)

(A) (B) (C)

\textbf{Fig 1.} (A): We approximate the limiting distribution of $n \tilde{q}(Y_1, Y_2; (0, 1), (0, 1))$ 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 $n \tilde{q}(Y_1, Y_2; (0, 1), (0, 1))$ 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
$n \tilde{q}(Y_1, Y_2; (0, 1), (0, 1))$ when $Y_1$ and $Y_2$ are statistically independent. (C): The den-
sity functions of $n \tilde{q}(Y_1, Y_2; \{0.5\}, (0, 1))$ for two simulated models: $Y_1 = \exp(Y_2)\varepsilon$ and
$Y_1 = Y_2|\varepsilon$, where $Y_1$ and $\varepsilon$ are independent and standard normal. We also present the
density function of $n \tilde{q}(Y_1, Y_2; \{0.5\}, (0, 1))$ 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
$n \tilde{q}(Y_1, Y_2; (0, 1), (0, 1))$ 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 $\varepsilon$
and $Y_2$ independently from standard normal distribution. In the third ex-
ample, 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 \(n \, \hat{q}(Y_1, Y_2; \{0.5\}, (0, 1))\) 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 \(\hat{q}(Y_1, Y_2; I_1, I_2)\) stated in Theorem 1 allows us to construct confidence intervals for \(q(Y_1, Y_2; I_1, I_2)\) when it is nonzero, if we can have a consistent estimate of \(\sigma^2\). In what follows we discuss how to estimate \(\sigma^2\) consistently. We estimate \(\hat{Q}Y_k(\tau_k)\) with \(\hat{Q}Y_k(\tau_k) = \inf\{x : F_{n,k}(x) \geq \tau_k\}\), and estimate the conditional distribution of \((Y_k \mid Y_l)\), denoted by \(F_{k\mid l}(y_k \mid y_l)\), with the following Nadaraya-Watson kernel estimate,

\[
\hat{F}_{k\mid l}(y_k \mid y_l) \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} I(Y_{i,k} \leq y_k) K_{h_l}(Y_{i,l} - y_l) \sum_{i=1}^{n} K_{h_l}(Y_{i,l} - y_l),
\]

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 \(h_l = 1.06n^{-1/5}\text{std}(Y_l)\), where \(\text{std}(Y_l)\) is a robust estimate of the standard deviation of \(Y_l\). For notational clarity we write \(q_k = QY_k(\tau_k)\) and \(\hat{q}_k = \hat{Q}Y_k(\tau_k)\). Define \(\hat{\Delta}(\hat{q}_1, \hat{q}_2) = F_{n,1,2}(\hat{q}_1, \hat{q}_2) - \tau_1\tau_2\). We replace all unknowns in \(Z_i\) defined in (S1.5) with their respective estimates. This gives

\[
(2.3) \quad \hat{Z}_i \overset{\text{def}}{=} \int_{I_1} \int_{I_2} \hat{T}_i(\tau_1, \tau_2) \mid \{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)\} \text{d}\mu_1(\tau_1)\text{d}\mu_2(\tau_2),
\]

where

\[
\hat{T}_i(\tau_1, \tau_2) \overset{\text{def}}{=} \hat{\Delta}(\hat{q}_1, \hat{q}_2) \left\{I(F_{n,1}(Y_{i,1}) \leq \tau_1, F_{n,2}(Y_{i,2}) \leq \tau_2) - F_{n,1,2}(\hat{q}_1, \hat{q}_2) - \tau_1 I(F_{n,2}(Y_{i,2}) \leq \tau_2) - \tau_2 I(F_{n,1}(Y_{i,1}) \leq \tau_1) + 2\tau_1\tau_2\right\}
\]

\[
+ \left\{\hat{F}_{2\mid 1}(\hat{q}_2 \mid \hat{q}_1) - \tau_2\right\} \{\tau_1 - I(F_{n,1}(Y_{i,1}) \leq \tau_1)\}
\]

\[
+ \left\{\hat{F}_{1\mid 2}(\hat{q}_1 \mid \hat{q}_2) - \tau_1\right\} \{\tau_2 - I(F_{n,2}(Y_{i,2}) \leq \tau_2)\}\right\}.
\]

By noting that \(F_{n,k}, F_{n,1,2}\) and \(\hat{F}_{k\mid l}\) are all step functions, we evaluate the integrals using the same ideas as we used to estimate \(q(Y_1, Y_2; I_1, I_2)\). The estimator of \(\sigma^2\) is given by

\[
(2.5) \quad \hat{\sigma}^2 \overset{\text{def}}{=} 4n^{-1} \sum_{i=1}^{n} \hat{Z}_i^2.
\]

The following theorem establishes the consistency of \(\hat{\sigma}^2\).
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

\[
\sup_{y_l \in \Omega_l} |f_k'(y_l + u) - f_k'(y_l)| \leq C|u|, \quad \text{and}
\]

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

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

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 \( \hat{\sigma}^2 \xrightarrow{p} \sigma^2 \) 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 \( q(Y_1, Y_2; I_2, I_2) \). 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 \( q(Y_1, Y_2; I_1, I_2) \). The theoretical challenge lies in the nonstandard asymptotics, where the limiting distribution of \( \hat{q}(Y_1, Y_2; I_1, I_2) \) 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 \( q(Y_1, Y_2; I_1, I_2) \). [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:

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

for all \((\tau_1, \tau_2) \in I_1 \otimes I_2\), where \( h(\cdot) \) satisfies

\[
\sup_{(\tau_1, \tau_2) \in I_1 \otimes I_2} h^2(\tau_1, \tau_2) > 0.
\]
Taking derivative on both sides of (2.6) with respect to \( \tau_2 \), 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

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

It follows from Taylor expansion that

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

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

\[
n \hat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2)\] 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

\[
n \hat{q}(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) \overset{d}{=} \int_{\mathcal{I}_1} \int_{\mathcal{I}_2} \frac{\{B(\tau_1, \tau_2) + h(\tau_1, \tau_2)\}^2}{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)} d\mu_1(\tau_1) d\mu_2(\tau_2)
\]

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:

\[
Y_1 = A\{Y_1^2 I(1 > 0) + \hat{Y}_1^2 I(Y_1 \leq 0)\} + \varepsilon; \tag{2.7}
\]

\[
Y_2 = \exp(AY_1^2) \varepsilon; \tag{2.8}
\]

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

where \( \varepsilon, Y_1 \) and \( \hat{Y}_1 \) 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 \( A \neq 0 \), \( Q_{Y_1|Y_2}(\tau_1) \) 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 (“$qcorr_{\tau_1}(Y_1 | Y_2)$” and “$\text{MDC}_{\tau_1}(Y_1 | Y_2)$” for $\tau_1 = 0.50, 0.75$ and 0.90). To implement our proposed test using the statistic \(n \tilde{q}(Y_1, Y_2; I_1, I_2)\), we vary \(I_1 = \{0.50\}, \{0.75\}, \{0.90\}, (0, 0.25), (0, 0.5), (0, 0.75), (0.25, 0.75), (0.5, 1), (0.75, 1), (0, 0.5) \cup (0.5, 1) \) and \(I_2 = (0, 1)\). 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 \(\alpha\). 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 \(E(Y_1 | Y_2) = E(Y_1)\) in models (2.7) and (2.9). The rank-based distance correlation test and \(\tilde{q}(Y_1, Y_2; (0, 1), (0, 1))\) have comparable power performance in model (2.9). In models (2.7) and (2.8), \(\tilde{q}(Y_1, Y_2; (0, 1), (0, 1))\) 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 \((\text{MDC}_{\tau_1}(Y_1 | Y_2))\) and the linear quantile correlation test \((qcorr_{\tau_1}(Y_1 | Y_2))\) 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 \(\tau_1\), 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 \(\tau_2 \in I_2 = (0, 1)\). This partly explains why the powers of the linear quantile correlation test \(qcorr_{0,50}(Y_1 | Y_2)\), the martingale difference correlation test \(\text{MDC}_{0,50}(Y_1 | Y_2)\) and our proposed test \(\tilde{q}(Y_1, Y_2; \{0.50\}, (0, 1))\) are close to the significance level \(\alpha\). In this heterogeneous model, our proposed tests \(\tilde{q}(Y_1, Y_2; (0, 1), (0, 1))\) and \(\tilde{q}(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0, 1))\) 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 \(\tilde{q}(Y_1, Y_2; (0, 0.5), (0, 1))\) is 0.056 whereas that of \(\tilde{q}(Y_1, Y_2; (0.5, 1), (0, 1))\) 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 \(\tilde{q}(Y_1, Y_2; \{0.50\}, (0, 1))\) is 0.074 whereas that of \(\tilde{q}(Y_1, Y_2; (0, 0.5) \cup (0.5, 1), (0, 1))\) is as large as 0.951. This again indicates that \(Q_{Y_1|Y_2}(\tau_1)\) de-
pends 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 $q(Y_1, Y_2; (0, 1), I_2)$ varies with $I_2$. We choose $I_2 = (0, 0.1), (0.1, 0.2), \ldots, (0.9, 1)$. In this study, $Q_{Y_2}(\tau_2) = 20, 33, 36, 39, 42, 45, 49, 52, 56, 64$ and $83$ for $\tau_2 = 0, 0.1, \ldots, 1.0$, respectively. Table 2 shows that $q(Y_1, Y_2; (0, 1), I_2)$ is concentrated at $I_2 = (0.2, 0.9)$, 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 $\bar{q}(Y_1, Y_2; (0, 1), I_2)$s are significantly different from zero. All resulting p-values are less than $10^{-3}$, which strikes the chord with common knowledge that hypertension is age related. Table 2 charts the estimates of $\sigma^2$ given by (2.5). It can be seen from Table 2 that $\bar{\sigma}^2$ for $I_2 = (0.2, 0.9)$ is also comparatively larger than that for $I_2 = (0, 0.1)$ or $I_2 = (0.9, 1)$, 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
| (2.7) | | | |
|---|---|---|---|---|---|
| Kendall tau(Y_1, Y_2) | 0.054 | 0.179 | 0.228 | 0.198 | 0.296 |
| dcorr\{F_1(Y_1), F_2(Y_2)\} | 0.062 | 0.245 | 0.337 | 0.358 | 0.393 |
| qcorr_{0.50}(Y_1 | Y_2) | 0.027 | 0.016 | 0.027 | 0.025 | 0.017 |
| qcorr_{0.75}(Y_1 | Y_2) | 0.040 | 0.131 | 0.130 | 0.132 | 0.118 |
| qcorr_{0.90}(Y_1 | Y_2) | 0.078 | 0.414 | 0.390 | 0.411 | 0.399 |
| MDC_{0.50}(Y_1 | Y_2) | 0.063 | 0.073 | 0.080 | 0.076 | 0.067 |
| MDC_{0.75}(Y_1 | Y_2) | 0.047 | 0.476 | 0.492 | 0.498 | 0.510 |
| MDC_{0.90}(Y_1 | Y_2) | 0.044 | 0.742 | 0.699 | 0.720 | 0.727 |
| q(Y_1, Y_2; \{0.50\}, (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; \{0.251\}, (0, 1)) | 0.054 | 0.800 | 0.862 | 0.881 | 0.893 |
| q(Y_1, Y_2; \{0.5\}, (0, 1)) | 0.058 | 0.573 | 0.666 | 0.674 | 0.708 |
| q(Y_1, Y_2; \{0.751\}, (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; \{0.5, 0.5\} | \{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 |

| (2.8) | | | |
|---|---|---|---|---|---|
| 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 | Y_2) | 0.025 | 0.007 | 0.012 | 0.010 | 0.011 |
| qcorr_{0.75}(Y_1 | Y_2) | 0.049 | 0.205 | 0.211 | 0.196 | 0.189 |
| qcorr_{0.90}(Y_1 | Y_2) | 0.087 | 0.524 | 0.535 | 0.549 | 0.525 |
| MDC_{0.50}(Y_1 | Y_2) | 0.049 | 0.059 | 0.084 | 0.076 | 0.080 |
| MDC_{0.75}(Y_1 | Y_2) | 0.064 | 0.337 | 0.367 | 0.349 | 0.364 |
| MDC_{0.90}(Y_1 | 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; \{0.251\}, (0, 1)) | 0.036 | 0.878 | 0.925 | 0.948 | 0.956 |
| q(Y_1, Y_2; \{0.5\}, (0, 1)) | 0.047 | 0.764 | 0.844 | 0.890 | 0.911 |
| q(Y_1, Y_2; \{0.751\}, (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 | 0.952 |
| 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; \{0.5, 0.5\} | \{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 |
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 (Y₁) and the household income (Y₂) 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 \( q(Y_1, Y_2; (0, 1), I_2) \) to quantify the relations between \( Y_1 \) and \( Y_2 \), for \( I_2 = (0, 1), (0, 1), (0, 2, 1), (0.3, 1), (0.4, 1), (0.5, 1), (0.6, 1), (0.7, 1), (0.8, 1), (0.9, 1), (0.95, 1), (0.99, 1) \) and \( (0.995, 1) \). In this study, \( Q_{Y_2}(\tau_2) = 901, 9520, 16000, 19791, 26803, 30000, 36987, 45880, 56818, 68492, 108153, 247983, 372121, 1000000 \) 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 \( \tau_2 = 0, 0.1, \ldots, 0.9, 0.95, 0.99, 0.995 \) and 1. The circles and stars in Figure 2(B) exhibits the patterns of \( q(Y_1, Y_2; (0, 1), I_2) \) varying with \( I_2 \). 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 \( q(Y_1, Y_2; (0, 1), I_2) \) is zero for each \( I_2 \). 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
Fig. 2. The Happiness Study. Panel (A): The density functions of household income for both urban and rural households. (B): The circles “•” and stars “∗” exhibit how $q(Y_1, Y_2; (0, 1), I_2)$ varies with $I_2$ for both urban and rural households. From left to right in (B), $I_2 = (0, 1), (0.1, 1), (0.2, 1), \ldots, (0.9, 1), (0.95, 1), (0.99, 1)$ and $(0.995, 1)$.

Table 3

The p-values of the interval quantile independence tests using $\hat{q}(Y_1, Y_2; (0, 1), I_2)$ for different $I_2$.  

<table>
<thead>
<tr>
<th>$I_2$</th>
<th>(0, 1)</th>
<th>(0.1, 1)</th>
<th>(0.2, 1)</th>
<th>(0.3, 1)</th>
<th>(0.4, 1)</th>
<th>(0.5, 1)</th>
<th>(0.6, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rural</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>urban</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
</tr>
<tr>
<td>$I_2$</td>
<td>(0.7, 1)</td>
<td>(0.8, 1)</td>
<td>(0.9, 1)</td>
<td>(0.95, 1)</td>
<td>(0.99, 1)</td>
<td>(0.995, 1)</td>
<td></td>
</tr>
<tr>
<td>rural</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$0.025$</td>
<td>$0.156$</td>
</tr>
<tr>
<td>urban</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$&lt; 0.001$</td>
<td>$0.018$</td>
<td>$0.057$</td>
<td></td>
</tr>
</tbody>
</table>

$x \overset{\text{def}}{=} (X_1, \ldots, X_p)^T$ 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 $\{(x_i, Y_i), i = 1, \ldots, n\}$, we aim to identify which covariates are predictive for some quantile levels of the response variable $Y$. Denote $A$ the indices of the important covariates, namely, $A \overset{\text{def}}{=} \{k : \text{The } \tau_1\text{-th quantile of } Y \text{ conditional on } x = (X_1, \ldots, X_p)^T \text{ depends on } \tau_2\text{-th quantile level of } X_k\}$, for $(\tau_1, \tau_2) \in I_1 \otimes I_2 \subseteq (0, 1) \otimes (0, 1)$.

We propose the following screening procedure to remove as many unimportant covariates as possible. We calculate $\hat{q}(Y, X_k; I_1, I_2)$ for each covariate and rank their relative importance in a descending order. It is natural to anticipate that $\hat{q}(Y, X_k; I_1, I_2)$ for the important covariates is larger than that for unimportant covariates. This motivates us to retain the covariates
indexed by
\[ \hat{A} \overset{\text{def}}{=} \{ k : \tilde{q}(Y, X_k; I_1, I_2) \geq c_1 n t^{-1/2} \} \]
for some \( 0 < t \leq 1/2 \) and \( c_1 > 0 \). Theorem 5 ensures that \( A \subseteq \hat{A} \) with
an overwhelming probability if the number \( s \) of elements in \( A \) satisfies
\[ ns \exp(-c_2 n^{2t}) \rightarrow 0 \text{ as } n \rightarrow \infty \] and
\[ (3.1) \quad q(Y, X_k; I_1, I_2) \geq 2c_1 n t^{-1/2} \quad \text{for all } k \in A, \]
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 \( n \rightarrow \infty \),
\[ \Pr\{ | \tilde{q}(Y, X_k; I_1, I_2) - q(Y, X_k; I_1, I_2) | > c_1 n t^{-1/2} \} = O\{ n \exp(-c_2 n^{2t}) \}. \]
If we further assume (3.1) holds, then
\[ \Pr(A \subseteq \hat{A}) \geq 1 - O\{ sn \exp(-c_2 n^{2t}) \}. \]

Assumption (3.1) allows that the marginal signal strength of the important
covariates, which is quantified by \( q(Y, X_k; I_1, I_2) \), shrinks to zero at a certain
rate. It also requires that these 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
\[ (3.2) \quad Y_i = 5X_{i,1} + X_{i,2}^2 + 2X_{i,3}X_{i,4} + \exp(X_{i,5}) \varepsilon_i, \]
where \( X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,p})^T \) is generated from a mixture of multivariate
normal population with mean zero and covariance matrix \( \Sigma = (0.9^{|k-k'|})_{p \times p} \)
with probability 0.9 and standard Cauchy distribution with probability 0.1,
and \( \varepsilon_i \) is drawn from (i) standard normal and (ii) standard Cauchy distribu-
tion. In this example, the active covariate set \( A = \{ 1, 2, 3, 4, 5 \} \). We set
\( n = 200 \) and \( p = 5000 \) in our simulations.

We consider four choices for \((I_1, I_2)\) in \( q(Y, X_k; I_1, I_2) \) to perform screen-
ing: (i) \( I_1 = \{ 0.50 \}, I_2 = (0, 1) \); (ii) \( I_1 = \{ 0.75 \}, I_2 = (0, 1) \); (iii) \( I_1 = I_2 = (0.05, 0.95) \) and (iv) \( I_1 = I_2 = (0, 1) \). 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$_{\tau_{1/4}}$-SIS].

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

1. The minimal model size $S$ which is required to ensure inclusion of all truly important covariates. The closer $S$ 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 $S$ for each screening procedure out of 1000 replications.

2. The selection probability $P_A$ that all five important covariates are ranked at the top $[n/\log n]$ and $2[n/\log n]$ positions. The closer $P_A$ is to one, the better performance the corresponding screening procedure has. We report this empirical selection probability $P_A$ for each screening procedure out of 1000 replications.

3. The selection probability $P_S$ 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 $P_S$ will be close to one for this covariate. We report this empirical selection probability $P_S$ 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 $S$ 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 [19] 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 $S$.

<table>
<thead>
<tr>
<th>Error</th>
<th>Method</th>
<th>min</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauchy</td>
<td>$q(Y, X_1; (0.5), (0, 1))$</td>
<td>5</td>
<td>6</td>
<td>11</td>
<td>52</td>
<td>546</td>
<td>1241</td>
<td>2450</td>
</tr>
<tr>
<td></td>
<td>$q(Y, X_1; (0.75), (0, 1))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>22</td>
<td>122</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>$q(Y, X_1; (0.05, 0.95), (0.05, 0.95))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>26</td>
<td>90</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>$q(Y, X_1; (0, 1), (0, 1))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>22</td>
<td>79</td>
<td>181</td>
</tr>
<tr>
<td>SIS</td>
<td></td>
<td>5</td>
<td>2168</td>
<td>3578</td>
<td>4423</td>
<td>4891</td>
<td>4980</td>
<td>5000</td>
</tr>
<tr>
<td>Kendall’s tau</td>
<td></td>
<td>5</td>
<td>6</td>
<td>11</td>
<td>57</td>
<td>899</td>
<td>2865</td>
<td>4626</td>
</tr>
<tr>
<td>SIRS</td>
<td></td>
<td>5</td>
<td>305</td>
<td>1028</td>
<td>2334</td>
<td>4390</td>
<td>4971</td>
<td>5000</td>
</tr>
<tr>
<td>DC-SIS</td>
<td></td>
<td>5</td>
<td>173</td>
<td>1340</td>
<td>3672</td>
<td>4812</td>
<td>4955</td>
<td>4987</td>
</tr>
<tr>
<td>MDC_{0.5}-SIS</td>
<td></td>
<td>5</td>
<td>7</td>
<td>32</td>
<td>252</td>
<td>2020</td>
<td>3663</td>
<td>4145</td>
</tr>
<tr>
<td>MDC_{0.75}-SIS</td>
<td></td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>28</td>
<td>860</td>
<td>4097</td>
<td>4955</td>
</tr>
<tr>
<td>Q_a_{0.5}-SIS</td>
<td></td>
<td>5</td>
<td>305</td>
<td>437</td>
<td>952</td>
<td>4602</td>
<td>4943</td>
<td>4998</td>
</tr>
<tr>
<td>Q_a_{0.75}-SIS</td>
<td></td>
<td>5</td>
<td>345</td>
<td>506</td>
<td>1073</td>
<td>4602</td>
<td>4943</td>
<td>4997</td>
</tr>
<tr>
<td>Normal</td>
<td>$q(Y, X_1; (0.5), (0, 1))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>113</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>$q(Y, X_1; (0.75), (0, 1))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>$q(Y, X_1; (0.05, 0.95), (0.05, 0.95))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>21</td>
<td>77</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>$q(Y, X_1; (0, 1), (0, 1))$</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>19</td>
<td>58</td>
<td>119</td>
</tr>
<tr>
<td>SIS</td>
<td></td>
<td>5</td>
<td>1989</td>
<td>3575</td>
<td>4327</td>
<td>4868</td>
<td>4967</td>
<td>4976</td>
</tr>
<tr>
<td>Kendall’s tau</td>
<td></td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>55</td>
<td>530</td>
<td>4188</td>
<td>4626</td>
</tr>
<tr>
<td>SIRS</td>
<td></td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>38</td>
<td>88</td>
<td>5000</td>
</tr>
<tr>
<td>DC-SIS</td>
<td></td>
<td>5</td>
<td>274</td>
<td>1038</td>
<td>2245</td>
<td>4333</td>
<td>4958</td>
<td>4986</td>
</tr>
<tr>
<td>MDC_{0.5}-SIS</td>
<td></td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>59</td>
<td>831</td>
<td>2989</td>
</tr>
<tr>
<td>MDC_{0.75}-SIS</td>
<td></td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>294</td>
<td>2370</td>
</tr>
<tr>
<td>Q_a_{0.5}-SIS</td>
<td></td>
<td>5</td>
<td>142</td>
<td>227</td>
<td>425</td>
<td>4487</td>
<td>4897</td>
<td>4995</td>
</tr>
<tr>
<td>Q_a_{0.75}-SIS</td>
<td></td>
<td>5</td>
<td>170</td>
<td>294</td>
<td>530</td>
<td>4491</td>
<td>4867</td>
<td>4995</td>
</tr>
</tbody>
</table>

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

Table 5 charts the empirical probabilities $P_A$ and $P_S$ that the important covariates are retained after screening for a given model size. The SIS, DC-SIS and $Q_a_{\tau_1}$-SIS are very inefficient in detecting either of the first four important covariates $X_k$, for $k = 1, \ldots, 4$. The Kendall’s tau and the MDC$_{\tau_1}$-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
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

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

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 $Q_{Y_1 \mid Y_2 = Q_{Y_2}(\tau)}(\tau) = Q_{Y_1}(\tau)$, 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 $Q_{Y_1 \mid Y_2 = Q_{Y_2}(\tau)}(\tau) = Q_{Y_1}(\tau)$, 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, $y_1 = (Y_{11}, \ldots, Y_{1p})^T$ and $y_2 = (Y_{21}, \ldots, Y_{2q})^T$, is of interest at quantile levels $\mathcal{I}_1 = \mathcal{I}_{11} \times \mathcal{I}_{12} \cdots \times \mathcal{I}_{1p} \subseteq (0,1) \times (0,1) \cdots \times (0,1)$ and $\mathcal{I}_2 = \mathcal{I}_{21} \times \mathcal{I}_{22} \cdots \times \mathcal{I}_{2q} \subseteq (0,1) \times (0,1) \cdots \times (0,1)$, we can define

$$q_{\text{sum}}(y_1, y_2; \mathcal{I}_1, \mathcal{I}_2) \overset{\text{def}}{=} \int_{\mathcal{I}_1} \int_{\mathcal{I}_2} \frac{\text{cov}^2 \left\{ \prod_{k=1}^p I(Y_{1k} \leq Q_{Y_{1k}}(\tau_{1k})), \prod_{l=1}^q I(Y_{2l} \leq Q_{Y_{2l}}(\tau_{2l})) \right\}}{\prod_{k=1}^p \tau_{1k}(1-\tau_{1k}) \prod_{l=1}^q \tau_{2l}(1-\tau_{2l})} d\mu_1(\tau_1) d\mu_2(\tau_2),$$

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
\[
q_{\text{max}}(y_1, y_2; \mathcal{I}_1, \mathcal{I}_2) \overset{\text{def}}{=} \max_{1 \leq k \leq p} \max_{1 \leq l \leq q} \int_{\mathcal{I}_{1k}} \int_{\mathcal{I}_{2l}} \frac{\text{cov}^2\{I(Y_{1k} \leq Q_{Y_{1k}}(\tau_{1k})), I(Y_{2l} \leq Q_{Y_{2l}}(\tau_{2l}))\}}{\prod_{k=1}^{p} \tau_{1k}(1 - \tau_{1k}) \prod_{l=1}^{q} \tau_{2l}(1 - \tau_{2l})} d\mu_{1k}(\tau_{1k}) d\mu_{2l}(\tau_{2l}).
\]

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 \(\mathbf{y}_1\) and \(\mathbf{y}_2\)) in the presence of a high dimensional covariate vector \(\mathbf{x}\). This is an interesting and yet very challenging issue, which warrants thorough investigation.

\textbf{APPENDIX A: PROOF OF PROPOSITION 1}

(i) We first notice that \(q(Y_1, Y_2; \mathcal{I}_1, \mathcal{I}_2) = 0 \iff \text{pr}\{Y_1 \leq Q_{Y_1}(\tau_1), Y_2 \leq Q_{Y_2}(\tau_2)\} = \tau_1 \tau_2\) for \((\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2\). Taking derivative with respective to \(\tau_2\) 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 \(\text{pr}\{Y_1 \leq Q_{Y_1}(\tau_1) \mid Y_2 = Q_{Y_2}(\tau_2)\} = \tau_1\). 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 \(\text{pr}\{Y_1 \leq Q_{Y_1}(\tau_1) \mid Y_2 = Q_{Y_2}(\tau_2)\} = \tau_1\) for \((\tau_1, \tau_2) \in \mathcal{I}_1 \otimes \mathcal{I}_2\). Consequently,
\[
\text{pr}\{Y_1 \leq Q_{Y_1}(\tau_1), Y_2 \leq Q_{Y_2}(\tau_2)\} = E[\text{pr}\{Y_1 \leq Q_{Y_1}(\tau_1) \mid Y_2\} I(Y_2 \leq Q_{Y_2}(\tau_2))] = \tau_1 E\{I(Y_2 \leq Q_{Y_2}(\tau_2))\} = \tau_1 \tau_2.
\]

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
\[
q(Y_1, Y_2; \tau_1, (0, 1)) = 0 \iff E\{I(Y_1 \leq Q_{Y_1}(\tau_1)) \mid Y_2\} = \tau_1.
\]

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
\[
q(Y_1, Y_2; (0, 1), (0, 1)) = 0.
\]

\(\iff\) \(\text{cov}\{I(Y_1 \leq y_1), I(Y_2 \leq y_2)\} = 0\), \(\forall y_k \in \{Q_{Y_k}(\tau_k) : \tau_k \in (0, 1)\}, k = 1, 2\).

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 $q(Y_1, Y_2; I_1, I_2) = q(m_1(Y_1), m_2(Y_2); I_1, I_2)$. □

REFERENCES


