Subhra Sankar Dhar, Wicher Bergsma and Angelos Dassios

Testing independence of covariates and errors in nonparametric regression

Article (Accepted version) (Refereed)

Original citation:
Sankar, Subhra and Bergsma, Wicher and Dassios, Angelos (2017) *Testing independence of covariates and errors in nonparametric regression.* Scandinavian Journal of Statistics. ISSN 0303-6898

DOI: 10.1111/sjos.12301

Testing independence of covariates and errors in nonparametric regression

Subhra Sankar Dhar[1], Wicher Bergsma[2], Angelos Dassios[2]

Department of Mathematics and Statistics, IIT Kanpur, India[1]

Department of Statistics, London School of Economics and Political Science, UK[2]

Abstract

Consider a nonparametric regression model $Y = m(X)+\epsilon$, where m is an unknown regression function, Y is a real-valued response variable, X is a real co-variate, and ϵ is the error term. In this article, we extend the usual tests for homoscedasticity by developing consistent tests for independence between X and ϵ. Further, we investigate the local power of the proposed tests using Le Cam's contiguous alternatives. An asymptotic power study under local alternatives along with extensive finite sample simulation study shows the performance of the new tests is competitive with existing ones. Furthermore, the practicality of the new tests is shown using two real data sets.

Kewwords: asymptotic power, contiguous alternatives, distance covariance, kendall's tau, nonparametric regression model, measure of association.

Running headline: Test of independence.

1 Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n independent replications of (X, Y), where Y is the response variable, and X is the covariate. We consider a nonparametric regression model $Y = m(X) + \epsilon$, where m is an unknown regression function, ϵ is the error random variable, and for simplicity, we assume that $E(\epsilon|X) = 0$. We would like to mention that the results of this article also hold for other identification restriction such as *median* $(\epsilon|X) = 0$ (see, e.g., Einmahl and Van Keilegom (2008a)). In this article, we develop tests to check whether the random variables ϵ and X are independent or not.

There has been an enormous amount of research done on tests for homoscedasticity in the last fifty years (e.g., see Goldfeld and Quandt (1965), Glejser (1969), Cook and Weisberg (1983) etc.). Due to the assumption of homoscedasticity, i.e., when the variance of the error random variable does not depend on the covariate, the variance of the response variable does not depend on the covariate. However, even for the homoscedastic model, the statistical inference about the unknown regression function may be misleading sometimes. For instance, in case of isotonic regression model, the confidence interval for the unknown regression function at a given point will be wrong even if the homoscedasticity assumption holds. In this example, in order to have trimmed mean isotonic regression estimator based correct confidence interval for the regression function at a given point, one needs to assume that the error and the covariate random variables are independent (see Remark 4 in Dhar (2016)).

In the last few years, several articles on testing independence of covariates and errors in different types of non-parametric regression models have appeared. Among these articles, Einmahl and Van Keilegom (2008a) proposed three tests based on Kolmogorov-Smirnov, Cramer-von-Mises and Anderson-Darling distances between the estimated joint distribution of (X, Y^*) and the product of the estimated marginal distributions of X and Y^*, where Y^* is the second order difference of Y. In another article, Einmahl and Van Keilegom (2008b) formulated the test statistics based on the same set of distances between the estimated joint distribution of (X, ϵ) and the product of the estimated marginal distributions of X and ϵ. The asymptotic distributions of the test statistics considered in Einmahl and Van Keilegom (2008a, 2008b) are developed using the empirical process theory. Within the same spirit of the distance based approach, Neumeyer (2009) considered a test based on a kernel estimator for the L_2-distance between the conditional distribution and the unconditional distribution of the covariates. A recently proposed test described in Hlavka, Huskova and Meintanis (2011) was based on the difference between the empirical characteristic function of the joint distribution of (ϵ, X) and the product of the empirical characteristic functions of the marginal distributions of ϵ and X.

Unlike the distance based tests described in the last paragraph, the tests we propose are based on association measures, namely, Kendall's τ (see Kendall (1938)), τ^*, which is a modification of Kendall's τ (see Bergsma and Dassios (2014)) and distance covariance (see Szekely, Rizzo and Bakirov (2007)). In fact, the methodology can be applied to more general association measure also. The construction of the test statistics are based on the second order differences of neighbouring triplets of responses, which is analogous to the approach of Einmahl and Van Keilegom (2008a). In other words, the test statistics essentially measure local association. In this article, we derive the asymptotic distribution of the test statistics based on U-statistic theory, and this approach allows us to investigate the power of our proposed tests under local alternatives, which has not been done before. In fact, since the kernel of the test statistics involved dependent observations, one needs to use the theory of degenerate U-statistic based on dependent random variables (see, e.g., Lee (1990)). The formulation of the test statistics and their asymptotic properties, which involved aforementioned issues, are thoroughly studied in the subsequent sections.

In connection with the discussion in the last paragraph, note here that one cannot observe the errors ϵ_i, $i = 1, \ldots, n$ since the data $(x_1, y_1), \ldots, (x_n, y_n)$ are based on the observations obtained from the joint distribution associated with (X, Y). In order to deal with this situation, one needs to use the Y values in such a way that the effect of m will be cancelled out under the smoothness assumption on m. Here based on the data on (X, Y), as it is mentioned earlier, the tests based on Kendall's τ, τ^* and distance covariance are considered. We investigate the asymptotic theory of the test statistics under null and contiguous alternatives, and also compare their performances for large samples as well as for small samples. We would here like to mention that as a tool to study the quality of a test, investigating the power property of a test under contiguous alternatives (see, e.g., Van Der Vaart (1998), Chapter 6) is well-established in the literature (see, e.g., Lehmann and Romano (2005)). Roughly speaking, contiguous alternatives are local alternatives asymptotically converging to the null hypothesis of interest. A precise definition of contiguity is given at the beginning of Section 3.

The rest of the article is organized as follows. In Section 2, we discuss the formulation of the test statistics and related issues. Section 3 investigates the asymptotic distribution of the test statistics under contiguous alternatives and compare the asymptotic power of the tests under those contiguous alternatives. Section 4 compares the performances of the tests for finite samples using simulations, and those tests are implemented on real data. Some concluding remarks are discussed in Section 5, and the more technical proofs are given in Appendix A. Appendix B reports the result of the asymptotic and the finite sample power studies in the tabular form.

2 Test statistics

We first recall the hypotheses again and formulate the hypothesis problem formally. For the nonparametric regression model $Y = m(X) + \epsilon$, we want to test $H_0 : \epsilon \perp\!\!\!\perp X (\Leftrightarrow F_{\epsilon,X} = G_\epsilon H_X)$ against the alternative $H_1 : \epsilon \not\perp\!\!\!\perp X (\Leftrightarrow F_{\epsilon,X} \neq G_\epsilon H_X)$, where $F_{\epsilon,X}$ is the joint distribution function of ϵ and X, and G_ϵ and H_X are the marginal distribution functions of ϵ and X, respectively. However, since one doesn't have the observations on the error ϵ as mentioned in the Introduction, one needs to construct the test statistics based on the data $(x_1, y_1), \ldots, (x_n, y_n)$ obtained from the joint distribution of (X, Y). Let $x_{(1)} \leq x_{(2)} \ldots \leq x_{(n)}$ be the order statistics of the observations $x_1, \ldots, x_n$, and $y_{(1)}, \ldots, y_{(n)}$ are the Y-values corresponding to the ordered X-values, and we now propose the test statistics $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ based on the observations $x_{(i)}$ and $y^*_{(i)} := y_{(i-1)} - 2y_{(i)} + y_{(i+1)}$, $i = 1, \ldots, n$; which are the following. In order to define all the quantities properly, we assume that $y_{(0)} = y_{(1)}$ and $y_{(n+1)} = y_{(n)}$, which results $y^*_{(1)} = -y_{(1)} + y_{(2)}$ and $y^*_{(n)} = y_{(n-1)} - y_{(n)}$. We then define

$$T_{n,1} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \mathrm{sign}\{(x_{(i)} - x_{(j)})(y^*_{(i)} - y^*_{(j)})\}, \tag{1}$$

$$T_{n,2} = \frac{1}{\binom{n}{4}} \sum_{1 \le i < j < k < l \le n} a(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)}) a(y^*_{(i)}, y^*_{(j)}, y^*_{(k)}, y^*_{(l)}) \tag{2}$$

and

$$T_{n,3} = \frac{1}{\binom{n}{4}} \sum_{1 \le i < j < k < l \le n} \frac{1}{4} h(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)}) h(y^*_{(i)}, y^*_{(j)}, y^*_{(k)}, y^*_{(l)}), \tag{3}$$

where $\text{sign}(x) = x/|x|$ when $x \neq 0$ and $= 0$ otherwise, $a(z_1, z_2, z_3, z_4) = \text{sign}(|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|)$, and $h(z_1, z_2, z_3, z_4) = (|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|)$. Note that formulas (1), (2) and (3) apply to data with and without ties on the x variable (see, e.g., Bergsma and Dassios (2014), Weihs, Drton and Leung (2016), Nandy, Weihs and Drton (2016)). Note further that $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are measures of independence, namely, Kendall's τ, a modified version of Kendall's τ (see Bergsma and Dassios (2014)) and the distance covariance (see Szekely et al. (2007)) based on the ordered X-values and the corresponding Y^*-values. For the unordered (X, Y), the population versions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are $T_1 = E\ \text{sign}\{(X_1 - X_2)(Y_1 - Y_3)\}$, $T_2 = Ea(X_1, X_2, X_3, X_4)a(Y_1, Y_2, Y_3, Y_4)$ and $T_3 = \frac{1}{4}Eh(X_1, X_2, X_3, X_4)h(Y_1, Y_2, Y_3, Y_4)$, respectively, where $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ and (X_4, Y_4) are independent replications of (X, Y). The formulations of the T_is, $i = 1, 2, 3$ imply that $T_i(X, Y) = 0$ for $i = 1, 2, 3$ *if* $X \perp\!\!\!\perp Y$. In fact, we should mention that $T_2(X, Y) = T_3(X, Y) = 0$ *if and only if* $X \perp\!\!\!\perp Y$ (see Bergsma and Dassios (2014)). Moreover, it is established that $T_{n,i} \xrightarrow{p} 0$ for $i = 1, 2, 3$ as $n \to \infty$ when $X \perp\!\!\!\perp \epsilon$ (see Proposition 1).

We should also have an explanation regarding the construction of y^*, i.e., the second order differences of y's. In view of the fact $X \perp\!\!\!\perp \epsilon \Rightarrow X \perp\!\!\!\perp g(\epsilon)$ for a proper function g, we consider the appropriate differences of $y_{(i)}$'s, which enables us to cancel out the effect of m's when m is a sufficiently smooth function. As a consequence of the smoothness of m, one can approximate $T_{n,i}(X_{(j)}, Y^*_{(j)})$ by $T_{n,i}(X_{(j)}, \epsilon^*_j)$ for sufficiently large n, where $\epsilon^*_j = \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1} := g(\epsilon_j)$, $i = 1, 2, 3$ and $j = 1, \ldots, n$. Instead of looking at $y^*_{(i)}$, one may also consider the usual difference $y_{(i)} - y_{(i-1)}$ (or the slope $\{y_{(i)} - y_{(i-1)}\}/\{x_{(i)} - x_{(i-1)}\}$), which will be eventually an approximation of $\epsilon_{(i)} - \epsilon_{(i-1)}$ (or the ratio between $\epsilon_{(i)} - \epsilon_{(i-1)}$ and $x_{(i)} - x_{(i-1)}$) having negligible third moment. It follows from the fact that the first order differences are symmetrically distributed, and consequently, the third moment of the difference between two nearly identical ϵ_is is close to zero. In other words, this test based on the first order differences will be essentially a similar test as the nonparametric test of homoscedasticity (see also the discussion in Einmahl and Van Keilegom (2008a)). For sake of completeness, we should provide an explanation why our proposed tests are more general over test of homoscedasticity. In literature, the test of homoscedasticity states that $H^*_0 : E(\epsilon^2_i | X = x) = \sigma^2$ for all x against H^*_1 : all possible alternatives. The construction of the hypothesis problem H^*_0 against H^*_1 indicates that it cannot detect the nature of the third moment of the conditional distribution of ϵ conditioning on $X = x$. As we mentioned earlier, the third moment of the difference between two nearly identical ϵ_is is close to zero, and for that reason, the tests based on $y_{(i)} - y_{(i-1)}$ have also not been considered. Now, in order to detect the ϵ_is varying third moment, we next consider a linear combination $ay_{(i-1)} + by_{(i)} + cy_{(i+1)}$, where $a + b + c = 0$ for which the absolute value of the third moment of the corresponding linear combinations of ϵ's is maximal when the second moment of the conditional distribution of ϵ is fixed. This fact leads to the construction of $y^*_{(i)}$ since $\epsilon_{(i-1)} - 2\epsilon_{(i)} + \epsilon_{(i+1)}$ has the maximum third moment among all possible choices of $a\epsilon_{(i-1)} + b\epsilon_{(i)} + c\epsilon_{(i+1)}$, $a, b, c \in \mathbb{R}$ when $a + b + c = 0$, and $a^2 + b^2 + c^2$ is a constant. With the same spirit, as we mentioned in the Introduction with an example related to skew normal distribution, our proposed test can detect dependence between the covariate X and the error random variable ϵ in the skewness of the conditional distribution even though when the conditional variance of ϵ conditioning on X is constant.

3 Asymptotic power study: contiguous alternatives

Recall from the discussion in the Introduction that as a toolkit of comparing different tests, one can investigate the power of the tests under contiguous alternatives. In Section 3.1, we study the asymptotic distributions $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ under contiguous alternatives, which help us to derive the asymptotic local power of the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$. The implementation of those tests is also briefly described there. Further, in Section 3.2, the asymptotic power under contiguous alternatives of the proposed tests are investigated for various examples.

3.1 Main results

As we mentioned in Section 2 that $T_{n,i} \xrightarrow{p} 0$ for $i = 1, 2, 3$ as $n \to \infty$ under H_0, and since the tests based on $T_{n,2}$ and $T_{n,3}$ are consistent against any alternatives, it is of interest to see the asymptotic power of the tests based on $T_{n,i}$'s, $i = 1, 2, 3$ under contiguous or local alternatives (e.g., see Hajek, Sidak and Sen (1999), p. 249). Precisely, the sequence of probability measures Q_n is contiguous with respect to the sequence of probability measures P_n if $P_n(A_n) \to 0$ implies that $Q_n(A_n) \to 0$ for every sequence of measurable sets A_n, where $(\Omega_n, \mathcal{A}_n)$ is the sequence of measurable spaces, and P_n and Q_n are two probability measures defined on $(\Omega_n, \mathcal{A}_n)$. In order to characterise the contiguity in terms of the asymptotic behaviour of the likelihood ratios between P_n and Q_n, Le Cam proposed some results, which are popularly known as Le Cam's Lemmas (e.g., see Hajek, Sidak and Sen (1999)). A consequence of Le Cam's first lemma is that the sequence Q_n will be contiguous with respect to the sequence P_n if $\log \frac{Q_n}{P_n}$ asymptotically follows the Gaussian distribution with mean $= -\frac{\sigma^2}{2}$ and variance $= \sigma^2$ under P_n (e.g., see Hajek, Sidak and Sen (1999, p. 253, Corollary to Le Cam's first lemma)), where $\sigma > 0$ is a constant, and we use this fact to establish the contiguity in this article (see the proof of Theorem 1). Now, $H_0 : \epsilon \perp\!\!\!\perp X (\Leftrightarrow F_{\epsilon,X} = G_\epsilon H_X)$ for the model $Y = m(X) + \epsilon$, where $F_{\epsilon,X}$ is the joint distribution of (ϵ, X), and G_ϵ and H_X are the marginal distributions of ϵ and X, respectively. We then consider a sequence of alternatives

$$H_n : F_{n;\epsilon,X} = \left(1 - \frac{\gamma}{\sqrt{n}}\right) G_\epsilon H_X + \frac{\gamma}{\sqrt{n}} K, \tag{4}$$

where $\gamma > 0$, $n = 1, 2, \ldots$, and K is a proper distribution function. Here we should point out that A_n is a sequence of sets, which is changing over n along with its σ-field $\mathcal{A}_n$, and for that reason, it does *not follow directly* from the definition of contiguity that $F_{n;\epsilon,X}$ is contiguous with respect to $F_{\epsilon,X}$. In Theorem 1, based on Le Cam's first lemma, we establish that the alternatives H_n will be contiguous alternatives under some conditions. For sake of concise presentation, we assume the following conditions before stating Theorem 1.

Assumptions:

(A1) $f_{\epsilon,X}(e, x) > 0$ for all e and x, where $f_{\epsilon,X}$ is the joint probability density function of (ϵ, X).

(A2) $E_{Y \sim F_{\epsilon,X}} \left(\frac{k(Y)}{f_{\epsilon,X}(Y)} - 1\right)^2 < \infty$.

Theorem 1: *Under (A1) and (A2), the sequence of alternatives H_n defined by (4) forms a contiguous sequence.*

Here we would like to discuss a few issues related to the condition assumed in (A2), and for simplicity of writing, we drop the random variable Y from the expression of the condition in subsequent places. Note that $E_{f_{\epsilon,X}}\left(\log \frac{k}{f_{\epsilon,X}}\right) = E_{f_{\epsilon,X}}\left(1 - (1 - \log \frac{k}{f_{\epsilon,X}})\right) \approx -\frac{1}{2} E_{f_{\epsilon,X}}\left(1 - \frac{k}{f_{\epsilon,X}}\right)^2$ since $E_{f_{\epsilon,X}}\left(1 - \frac{k}{f_{\epsilon,X}}\right) = 0$, and hence, $E_{f_{\epsilon,X}}\left(\frac{k}{f_{\epsilon,X}} - 1\right)^2$ can be expressed as the first order approximation of an entropy $E_{f_{\epsilon,X}}\left(\log \frac{k}{f_{\epsilon,X}}\right)$, which measures dissimilarity between two densities $f_{\epsilon,X}$ and k. In other words, $E_{f_{\epsilon,X}}\left(\frac{k}{f_{\epsilon,X}} - 1\right)^2$ is also called the *mean square contingency* (see Renyi (1959, p. 446)) of $f_{\epsilon,X}$ and k. Further, note that if $k = f_{\epsilon,X}$, we have $\left(\frac{k}{f_{\epsilon,X}} - 1\right) = 0$, i.e., k and $f_{\epsilon,X}$ are similar. At the same time, the larger values of $\left(\frac{k}{f_{\epsilon,X}} - 1\right)$ indicate that k and $f_{\epsilon,X}$ are more dissimilar, i.e., in other words, ϵ and X are more statistically dependent through k. To summarize, Theorem 1 asserts that the sequence of alternatives H_n will be contiguous with respect to H_0 when the *mean square contingency* of $f_{\epsilon,X}$ and k is finite.

We establish the asymptotic behaviour of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ (given in (1), (2) and (3), respectively) under H_0 in Proposition 1. We assume the following conditions to establish Proposition 1.

Assumptions

(B1) $X_1, \ldots, X_n$ are i. i. d. random variables with common distribution function H_X.

(B2) $Y_1, \ldots, Y_n$ satisfy the model $Y_i = m(X_i) + \epsilon_i$, $i = 1, \ldots, n$, where the unknown function m possesses a bounded derivative, the random error ϵ_is are i. i. d. with bounded probability density function, and $E(\epsilon_i|X_i) = 0$ for all $i = 1, \ldots, n$.

Proposition 1: *Under (B1) and (B2), we have $T_{n,i} \xrightarrow{p} 0$ for $i = 1, 2, 3$ as $n \to \infty$ when H_0 is true.*

Note that the conditions assumed in (B1) and (B2) are realistic in nature. Assumption (B1) allows us to consider i. i. d. covariates following any type of distribution. The condition $E(\epsilon_i|X_i) = 0$ for all $i = 1, \ldots, n$ assumed in (B2) is used to avoid the problem of identifiability. Further, condition assumed on $m(.)$ holds when $m(.)$ is differentiable function and has uniformly bounded first derivative (see, e.g., Rudin (1976)).

The main implication of Proposition 1 is that one can use $T_{n,i}$, $i = 1, 2, 3$ to check whether the evidence obtained from the data favours H_0 or not. In addition, in order to carry out the tests based on $T_{n,i}$, $i = 1, 2, 3$, one needs to know the distributions (or an approximation of the distributions) of $T_{n,i}$'s. In this context, note that $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are 2-dependent U-statistics (e.g., see Lee (1990)) and to derive the asymptotic distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$, one needs to know the order of degeneracy of each $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$. For sake of completeness, the definition of U-statistic and its order of degeneracy are mentioned in the following. For a given data set $\mathcal{X} = \{x_1, \ldots, x_n\}$, $U_n = \frac{1}{\binom{n}{m}} \sum\limits_{1 \le i_1 < \ldots < i_m} k(x_{i_1}, \ldots, x_{i_m})$ is said to be a U-statistic of order m with kernel $k(.)$ having the order of degeneracy $= l$ if $E_{X_{l+1}, \ldots, X_m} k(x_1, \ldots, x_l, X_{l+1}, \ldots, X_m) = 0$ for all $x_1, \ldots, x_l$. Moreover, the collection of the random variables $\mathcal{X}^* = \{X_1, \ldots, X_n\}$ will be called m-dependent random variables if X_b and X_a are independent for all $b - a > m$ $(a, b = 1, \ldots, n;\ m \ge 1)$, and the corresponding U-statistic is said to be m-dependent U-statistic. It is easy to see that $T_{n,1}$ has the order of degeneracy 0 whereas it is established in Lemma 1 that the asymptotic order of degeneracy of $T_{n,2}$ and $T_{n,3}$ are 1.

Lemma 1: *Under (B1) and (B2), $E[T_{n,2}|X_{(i)} = x, Y^*_{(i)} = y] \to 0$ and $E[T_{n,3}|X_{(i)} = x, Y^*_{(i)} = y] \to 0$ for all $i = 1, \ldots, n$ as $n \to \infty$ under H_0, where $x, y \in \mathbb{R}$ are fixed constants.*

Lemma 1 implies that $T_{n,2} = O_p(\frac{1}{n})$ and $T_{n,3} = O_p(\frac{1}{n})$ as well, which can be established based on the asymptotic theory of 2-dependent U-statistic whereas $T_{n,1} = O_p(\frac{1}{\sqrt{n}})$ as it has the order of degeneracy $= 0$ (e.g., see Lee (1990)). We state the asymptotic distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ under H_n in Theorems 3.2, 3.3 and 3.4, respectively.

Theorem 2: *Suppose that (A1)-(A2) and (B1)-(B2) are true. Then, under H_n defined by (4), $\sqrt{n}\{T_{n,1}-$*

$E(T_{n,1})\}$ converges weakly to a Gaussian distribution with mean

$$\mu_1 = 2\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[2\int_{-\infty}^{x}\int_{-\infty}^{y} dH_X(u)dG^*(v) + 2\int_{x}^{\infty}\int_{y}^{\infty} dH_X(u)dG^*(v) - 1\right] dK(x,y)$$

and variance

$$\sigma_1^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[2\int_{-\infty}^{x}\int_{-\infty}^{y} dH_X(u)dG^*(v) + 2\int_{x}^{\infty}\int_{y}^{\infty} dH_X(u)dG^*(v) - 1\right]^2 dG_{\epsilon}(y)dH_X(x).$$

Here H_X is the marginal distribution function of X, and G^* is the distribution function of $\epsilon_1 - 2\epsilon_2 + \epsilon_3$.

In the proof of Theorem 2, Le Cam's third lemma has been used to get the asymptotic normality of $\sqrt{n}\{T_{n,1} - E(T_{n,1})\}$ under H_n, and in the course of using Le Cam's third lemma, we have used the fact that $\log L_n$ converges weakly to a random variable associated with a normal distribution having certain location and scale parameters (see the proofs of Theorems 3.1 and 3.2). It is here appropriate to mention that the asymptotic normality of $\log L_n$ is a sufficient condition but in general, it is not a necessary condition to establish the contiguity of Q_n with respect to P_n. Instead of Le Cam's third lemma, one can also follow Behnen and Neuhaus (1975)'s approach based on a specific truncation method for contiguity of Q_n with respect to P_n. Also, Behnen (1971) investigated the asymptotic relative efficiency of some tests for independence against general contiguous alternatives of positive quadrant dependence but none of Behnen (1971) and Behnen and Neuhaus (1975) considered Q_n as a mixture distribution as we consider here. Recently, Banerjee (2005) studied the behaviour of the likelihood ratio statistics for testing a finite dimensional parameter under local contiguous hypotheses. He perturbed the null hypothesized parameter to get the local (or contiguous) alternatives, which is different from the perturbance on the distribution function that has been considered by us.

Also, we would like to point out that $G^*(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{1 - G_{\epsilon}\left(\frac{u+v-y}{2}\right)\right\} dG_{\epsilon}(u)dG_{\epsilon}(v)$, which follows from the arguments based on convolutions. Here G_{ϵ} is the marginal distribution function of ϵ. Further, it directly follows from Theorem 2 that $\sqrt{n}\{T_{n,1} - E(T_{n,1})\}$ converges weakly to normal distribution with zero mean and variance $= \sigma_1^2$ under H_0 since the sequence of alternatives H_n coincide with the null hypothesis H_0 when $\gamma = 0$. In order to carry out the test having the level of significance $= \alpha$, one needs to find out the α% critical value (denote it as $c_{1,\alpha}$), i.e., $(1-\alpha)$-th quantile of the normal distribution having zero mean and variance $= \sigma_1^2$, and the asymptotic power of the test under H_n will be $P[X > c_{1,\alpha}]$, where X is the random variable associated with the normal distribution having mean $= \mu_1$ and variance $= \sigma_1^2$.

Theorem 3: *Suppose that (A1)-(A2) and (B1)-(B2) are true. Then, under H_n defined by (4), $n\{T_{n,2} -$*

$E(T_{n,2})\}$ *converges weakly to* $\sum_{i=1}^{\infty} \lambda_i\{(Z_i + a_i)^2 - 1\}$, *where* Z_i*'s are i.i.d.* $N(0,1)$ *random variables, and* λ_i*'s are the eigenvalues associated with* $l(x,y) = E\{\text{sign}(|X_{(1)} - X_{(2)}| + |X_{(3)} - X_{(4)}| - |X_{(1)} - X_{(3)}| - |X_{(2)} - X_{(4)}|) \times \text{ sign } (|Y^*_{(1)} - Y^*_{(2)}| + |Y^*_{(3)} - Y^*_{(4)}| - |Y_{(1)} - Y^*_{(3)}| - |Y^*_{(2)} - Y^*_{(4)}|)|X_{(1)} = x, Y^*_{(1)} = y\}$. *Here* (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) *and* (X_4, Y_4) *are i.i.d. bivariate random vectors,* $X_{(i)}$ *is the i-th order statistic of the random variables* X_1, X_2, X_3 *and* X_4, $Y^*_{(i)} = Y_{(i-1)} - 2Y_{(i)} + Y_{(i+1)}$, *where* $Y_{(i)}$ *is the* Y *value corresponding to* $X_{(i)}$, (x, y) *is the realized value of* $(X_{(1)}, Y_{(1)})$, *and*

$$a_i = \gamma \int \left\{ \frac{k}{f} - 1 \right\} g_i(x) g_i(y) f_{X,Y} dxdy.$$

Here $g_i(x)$ *and* $g_i(y)$ *are such that*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} l(x,y) \prod_{i=2}^{4} g_k(X_i) g_k(Y_i) d\left(\prod_{i=2}^{4} F_{X_i,Y_i} \right) = \lambda_k g_k(x) g_k(y)$$

for all (x, y).

Theorem 4: *Suppose that (A1)-(A2) and (B1)-(B2) are true. Then, under* H_n *defined by (4),* $n\{T_{n,3} - E(T_{n,3})\}$ *converges weakly to* $\sum_{i=1}^{\infty} \lambda^*_i\{(Z^*_i + a^*_i)^2 - 1\}$, *where* Z^*_i*'s are i.i.d.* $N(0,1)$ *random variables, and* λ^*_i*'s are the eigenvalues associated with* $l^*(x,y) = E\{(|X_{(1)} - X_{(2)}| + |X_{(3)} - X_{(4)}| - |X_{(1)} - X_{(3)}| - |X_{(2)} - X_{(4)}|) \times (|Y^*_{(1)} - Y^*_{(2)}| + |Y^*_{(3)} - Y^*_{(4)}| - |Y^*_{(1)} - Y^*_{(3)}| - |Y^*_{(2)} - Y^*_{(4)}|)|X_{(1)} = x, Y^*_{(1)} = y\}$. *Here* (X_1, Y_1), (X_2, Y_2), (X_3, Y_3) *and* (X_4, Y_4) *are i.i.d. bivariate random vectors,* $X_{(i)}$ *is the i-th order statistic of the random variables* X_1, X_2, X_3 *and* X_4, $Y^*_{(i)} = Y_{(i-1)} - 2Y_{(i)} + Y_{(i+1)}$, *where* $Y_{(i)}$ *is the* Y *value corresponding to* $X_{(i)}$, (x, y) *is the realized value of* $(X_{(1)}, Y_{(1)})$ *and*

$$a^*_i = \gamma \int \left\{ \frac{k}{f} - 1 \right\} g^*_i(x) g^*_i(y) f_{X,Y} dxdy.$$

Here $g^*_i(x)$ *and* $g^*_i(y)$ *are such that*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} l(x,y) \prod_{i=2}^{4} g^*_k(X_i) g^*_k(Y_i) d\left(\prod_{i=2}^{4} F_{X_i,Y_i} \right) = \lambda_k g^*_k(x) g^*_k(y)$$

for all (x, y).

Theorems 3.3 and 3.4 enable us to compute the asymptotic power of the tests based on $T_{n,2}$ and $T_{n,3}$ under H_n. In this context, note that under H_0, $n\{T_{n,2} - E(T_{n,2})\}$ and $n\{T_{n,3} - E(T_{n,3})\}$ converge weakly to $\sum_{i=1}^{\infty} \lambda_i\{Z_i^2 - 1\}$ and $\sum_{i=1}^{\infty} \lambda^*_i\{Z_i^{*2} - 1\}$, respectively, which follow from the assertions in Theorems 3.3 and 3.4

since both $a_i = a_i^* = 0$ if $\gamma = 0$ (i.e., when H_0 is true). The corresponding asymptotic critical values (denote those are as $c_{2,\alpha}$ and $c_{3,\alpha}$, respectively) can be obtained from $(1-\alpha)$-th quantile of the distributions described at the beginning of this paragraph. However, it is difficult to derive the explicit expression of the quantiles of the distribution since the distribution involves the infinite sum of the weighted chi-squared distribution, where weights are the eigenvalues of the kernels associated with $T_{n,2}$ (or $T_{n,3}$). In order to overcome the problem related to infinitely many eigenvalues and the infinite sum, we approximate the kernel function at $n_1 \times n_1$ many marginal quantile points and compute the eigenvalues of $n_1 \times n_1$ finite-dimensional matrix associated with the kernel function. The (i, j)-th element of the matrix is the $(i/n_1, j/n_1)$-th marginal quantile (e.g., see Babu and Rao (1988)) of the joint distribution associated with the bivariate random vector (X, Y), where $i = 1, \dots, n_1$ and $j = 1, \dots, n_1$. Then, we generate a sample with size 1000 from that approximated finite sum of the wighted chi-squared distribution, and $(1 - \alpha)$-th quantile of that sample is considered as the approximated value of the asymptotic critical value at $\alpha\%$ level of significance. In order to compute the power, we similarly approximate the infinite sum of the weighted chi-squared distributions described in Theorems 3.3 and 3.4 by an appropriate finite sum of the chi-squared distributions and simulate sample with size 1000 from the approximated distributions, and finally, the proportion of the observations in the sample larger than the approximated critical value is considered as the estimated value of the asymptotic power. In the asymptotic power studies of different tests, we consider $n_1 = 10$ unless mentioned otherwise.

3.2 Examples

We consider three examples, where we investigated the asymptotic power of the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$. In all of the examples, we take the null model $f_{\epsilon,X}(\epsilon, x) = \frac{1}{2\pi} e^{-\frac{\epsilon^2 + x^2}{2}}$, where ϵ, $x \in \mathbb{R}$, and $f_{\epsilon,X}$ is the density function of $F_{\epsilon,X}$. As the contiguous alternative model is of the form

$$\left(1 - \frac{\gamma}{\sqrt{n}}\right) f_{\epsilon,X} + \frac{\gamma}{\sqrt{n}} k, \tag{5}$$

where $f_{\epsilon,X}$ is same as considered in the null model for all examples but the choices of k are different over different examples, which are the following.

Example 1: The joint density $k_{\epsilon,X}(e, x)$ is such that $(\epsilon | X = x) \stackrel{\mathcal{D}}{=} N_x^*$, where $N_x^* \sim N(0, \frac{1+5x}{100})$, and the marginal density function of X is $N(0, 1)$.

Example 2: The joint density $k_{\epsilon,X}(e, x)$ is such that $(\epsilon | X = x) \stackrel{\mathcal{D}}{=} C_x^*$, where C_x^* is a random variable associated with Cauchy distribution with the location parameter $= 0$ and the scale parameter $= x^2$. The marginal density function of X is $N(0, 1)$.

It is an appropriate place to mention that the integrals involved in μ_1 and σ_1^2 described in the statement of Theorem 2 are calculated theoretically. Note that in the above examples, the joint density function of X and ϵ is the standard bivariate normal density under H_0, and hence, H_X and G_ϵ are the cumulative distribution functions of the standard normal distribution because under H_0, $F_{\epsilon,X} = H_X G_\epsilon$. To compute G^*, we use the relation $G^*(y) = \int\limits_0^\infty \int\limits_0^\infty \{1 - G_\epsilon(\frac{u+v-y}{2})\} dG_\epsilon(u) dG_\epsilon(v)$ as we mentioned earlier. Since both in μ_1 and σ_1^2, the integrands have tractable expressions in terms of the standard normal density functions, we are able to compute μ_1 and σ_1^2 theoretically without using Monte Carlo Methods. For Examples 1 and 2, the results are reported in Tables 1 and 2 (see Appendix B), and also summarized in Figure 1.

It follows from the figures in Table 1 that the tests based on $T_{n,1}$ (given in (1)), $T_{n,2}$ (given in (2)) and $T_{n,3}$ (given in (3)) are comparable as expected since k and $f_{\epsilon,X}$ are related by only scale transformation. Hence, there is neither any effect of outliers nor any significant effect of the alternative distribution. It is amply indicated by the figures in Table 2 that the tests based on $T_{n,1}$ and $T_{n,2}$ are more powerful than then test based on $T_{n,3}$ under contiguous alternatives. It is here appropriate to point out that $T_{n,1}$ and $T_{n,2}$ are rank based test statistic whereas $T_{n,3}$ is a moment based test statistic. For that reason, the better robustness properties of $T_{n,1}$ and $T_{n,2}$ compared to $T_{n,3}$ have significant effect on the results as the conditional density associated with k has heavier tail than that of $f_{\epsilon,X}$.

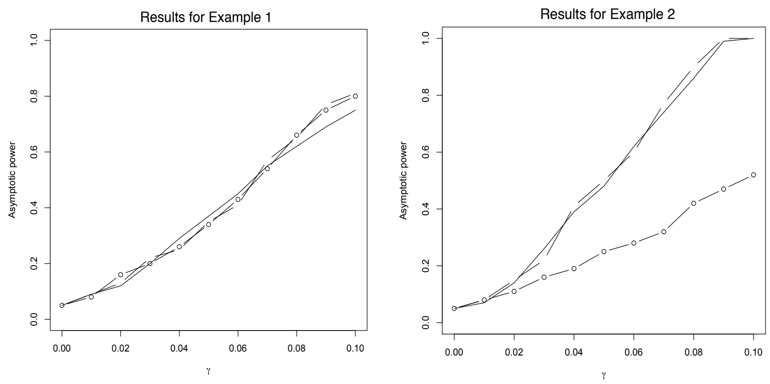

Figure 1: The asymptotic power of the test based on $T_{n,1}$ (solid curve $-$), the test based on $T_{n,2}$ (lined curve $- \ -$) and the test based on $T_{n,3}$ (dotted lined curve $-o-$) for different values of γ when the contiguous alternative model is of the form (5).

4 Finite sample study and real data analysis

In Section 3, we investigated the performance of the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ when the sample size n tends to infinity. It is now of interest to see the performance of the tests when the sample sizes are finite, and for that reason, in Section 4.1, we carry out some simulation studies to compare the finite sample power of these three tests along with the tests considered by Einmahl and Van Keilegom (2008a), Neumeyer (2009), Hlavka et al. (2011) and Breusch-Pagan test (see Breusch and Pagan (1979)) specially for checking heteroskedasticity. In addition, Section 4.2 shows the applicability of our proposed tests to real data.

4.1 Finite sample simulation study

We here study the finite sample power and the level of our tests using simulations and compared them with the tests described in Einmahl and Van Keilegom (2008a), Neumeyer (2009), Hlavka et al. (2011) and Breusch-Pagan test (see Breusch and Pagan (1979)). The Breusch-Pagan test checks whether the conditional variance of the residuals obtained from the regression are independent of the conditioning variable (i.e, the covariates) or not. In other words, it tests the presence of heteroskedasticity. The test statistics of Einmahl and Van Keilegom (2008a)'s tests are denoted by $T_{n,KS}$, $T_{n,CM}$ and $T_{n,AD}$ as they were defined in that article. For sake of completeness, the expressions of $T_{n,KS}$, $T_{n,CM}$ and $T_{n,AD}$ are provided in the following. Let us first define $F_n(x,y) = \frac{1}{n}\sum\limits_{i=1}^{n} 1_{\{x_{(i)}\leq x,\ y^*_{(i)}\leq y\}}$, where $1_A = 1$ if A is true, and $= 0$, otherwise. The forms of the test statistics are

$$T_{n,KS} = \sqrt{n}\sup_{x\in\mathbb{R},y\in\mathbb{R}} |F_n(x,y) - \hat{F}_X(x)\hat{G}(y)|, \tag{6}$$

$$T_{n,CM} = n\int\int (F_n(x,y) - \hat{F}_X(x)\hat{G}(y))^2 d\hat{F}_X(x)d\hat{G}(y) \tag{7}$$

and

$$T_{n,AD} = n\int\int \frac{(F_n(x,y) - \hat{F}_X(x)\hat{G}(y))^2}{\hat{F}_X(x)\hat{G}(y)(1-\hat{F}_{X-}(x))(1-\hat{G}_{-}(y))} d\hat{F}_X(x)d\hat{G}(y). \tag{8}$$

Here, $\hat{F}_X(x) = F_n(x,\infty)$ and $\hat{G}(y) = F_n(\infty, y)$, and F_- denotes the left-continuous version of for any distribution function F. The test statistic considered in Neumeyer (2009) is denoted by

$$T_{n,Neu} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{1}{h_n}K\left(\frac{\hat{\epsilon}_i - \hat{\epsilon}_j}{h_n}\right)\int (I\{X_i\leq x\} - F_{X,n}(x))(I\{X_j\leq x\} - F_{X,n}(x))w(x)dx, \tag{9}$$

where $\hat{\epsilon}_i = y_i - \hat{m}(x_i)$, $\hat{\epsilon}_j = y_j - \hat{m}(x_j)$, $F_{X,n}$ is the empirical distribution of the covariates $X_1,\ldots,X_n$, w is a weight function, K and h_n denote the kernel and the bandwidth, respectively. The test statistic described in Hlavka et al. (2011) as

$$T_{n,W} = n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |D_n(t_1, t_2)|^2 W(t_1, t_2) dt_1 dt_2, \tag{10}$$

where $D_n(t_1, t_2) = \hat{\phi}(t_1, t_2) - \hat{\phi}_X(t_1)\hat{\phi}_{\hat{e}}(t_2)$, and $W(t_1, t_2)$ denotes a suitable weight function. Here $\hat{\phi}(t_1, t_2) = \frac{1}{n}\sum_{j=1}^{n} e^{it_1 X_j + it_2 \hat{e}_j}$, $\hat{\phi}_X(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j}$, and $\hat{\phi}_{\hat{e}}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{it\hat{e}_j}$, and $\hat{e}_j = y_j - \hat{m}(X_j)$.

In this simulation study, our overall setting is similar to that in Einmahl and Van Keilegom (2008a) and Neumeyer (2009). As they considered, suppose that the covariate random variable X is associated with the uniform distribution on $(0, 1)$, $m(x) = x - 0.5x^2$, and the simulations are carried out for the sample sizes $n = 100$ and $= 1000$ with the level of significance $= 5\%$. Under H_0 (i.e., null hypothesis), the error random variable ϵ follows normal distribution with zero mean and standard deviation equal to 0.1. Also, we consider first three alternative distributions as studied in Einmahl and Van Keilegom (2008a) and the fourth one is related to skew normal distribution. For fourth one, we compare the performances of our tests with Breusch-Pagan test.

Example 3: $(\epsilon|X = x) \sim N(0, \frac{1+ax}{100})$. Here a controls the variance.

Example 4: $(\epsilon|X = x) \stackrel{\mathscr{D}}{=} \frac{W_x - r_x}{10\sqrt{2r_x}}$, where $W_x \sim \chi^2_{r_x}$, $r_x = \frac{1}{bx}$. Here $b > 0$ controls the skewness.

Example 5: $(\epsilon|X = x) \stackrel{\mathscr{D}}{=} \frac{1}{10}\sqrt{1 - (cx)^{1/4}} T_x$, where $T_x \sim t_{2/(cx)^{1/4}}$. Here $c \in (0, 1]$ controls the kurtosis.

Example 6: $(\epsilon|X = x)$ follows a skew normal distribution (see Azzalini (1985)) with the location parameter $= \xi(x)$, the scale parameter $= \omega(x)$ and the shape parameter $= \alpha(x)$, where $\xi(x) = x$, $\omega(x) = \sqrt{x^2 + 3}$, and $\alpha(x) = \frac{3\pi}{\sqrt{2(x^2+3) - 9\pi^2}}$.

To implement the proposed tests, i.e., the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$, one needs to compute the $\alpha\%$ (here $\alpha = 5$) critical value of the test. For computing the critical value, we adopt the Bootstrap procedure, which is described as follows (see also Neumeyer (2009)). We first generate a data $(x_1, y_1), \ldots, (x_n, y_n)$ from the regression model described in the paragraph before Example 3, and based on this data, we estimate the error $\hat{\epsilon}_i = y_i - \hat{m}(x_i)$, where $\hat{m}$ is well-known Nadaraya-Watson estimator (see Nadaraya (1964) and Watson (1964)) using Epanechnikov kernel (see, e.g., Silverman (1986)). We now define the centered error $\epsilon_{cen,i} = \hat{\epsilon}_i - \frac{1}{n}\sum_{i}^{n} \hat{\epsilon}_i$, where $i = 1, \ldots, n$. Let F_n^* be the empirical distribution function of $\epsilon_{cen,i}$ $(i = 1, \ldots, n)$, and the bootstrap sample $\epsilon_1^*, \ldots, \epsilon_n^*$ are generated from the empirical distribution function of centered error, i.e., F_n^*. Based on the original sample of the covariate $(x_1, \ldots, x_n)$ and the Bootstrap sample of the errors $(\epsilon_1^*, \ldots, \epsilon_n^*)$, we now have the bootstrap responses $y_i^* = \hat{m}(x_i) + \epsilon_i^*$. Finally, the Bootstrap sample was $(y_1^*, x_1), \ldots, (y_n^*, x_n)$. Following the same procedure, we generate $B = 500$ many bootstrap resamples, and $(1 - \alpha)$-th quantile of the Bootstrap distribution of the test statistic is considered as the estimated critical value.

The computations of $T_{n,1}$, $T_{n,2}$, $T_{n,3}$, $T_{n,KS}$, $T_{n,CM}$, $T_{n,AD}$ and $T_{n,Neu}$ are done in R and Matlab codes, which are available to the first author. For the test based on $T_{n,Neu}$, we report the same results as provided in Neumeyer (2009) when $n = 100$. In this context, we would here like to mention that Neumeyer (2009) reported the results of her proposed test for three choices of estimated bandwidth (denoted as h_n^* in that article); however, for sake of concise presentation, we report the maximum value of the rejection probability (i.e., estimated power) among three choices of values. In case of the test described in Hlavka et al. (2011), two versions of $W(t_1, t_2)$, namely, $W_1(t_1, t_2)$ and $W_2(t_1, t_2)$, are chosen as they considered in their article. For detailed description of $W_1(t_1, t_2)$ and $W_2(t_1, t_2)$, we refer the readers to Section 4.1.1 in Hlavka et al. (2011). The computation of $T_{n,W}$ is carried out by an R code available in `http://www.karlin.mff.cuni.cz/~hlavka/stat.html`. All results are summarized in Figure 2 and in Tables 3, 4, 5 and 6 in Appendix B.

The figures in Table 3 indicates that the tests based on $T_{n,1}$ and $T_{n,2}$ performs well for Example 3 and comparable with the performances of the tests based on $T_{n,Neu}$ (given in (9)), T_{n,W_1} and T_{n,W_2} (given in (10)). In that example, the test based on $T_{n,3}$ is under achieved particularly for large values of a. For Example 4, the performances of the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are comparable (see Table 4). However, for Example 5, the tests based on $T_{n,1}$ and $T_{n,2}$ and other tests as well outperform the test based on $T_{n,3}$ (see Table 5). It is expected since $T_{n,3}$ is moment based estimator, and consequently, it is not a robust estimator against the outliers generated from the heavy-tailed distribution. The reported values in Table 6 (see Appendix B) clearly indicates that a test of homoscedasticity, namely, Breusch-Pagan test is outperformed by our proposed tests since it fails to detect the varying conditional skewness of the distribution with constant conditional variance.

4.2 Real data analysis

As we have observed in the asymptotic and the finite sample power study, the test based on $T_{n,3}$ fails to perform well in the presence of the influential observations, we wanted to see how different tests perform on real data having (and not having) outliers. To investigate the performance of the tests, we consider here two real data, which are the following with detailed descriptions.

We first consider a real data set **Airfoil Self Noise Data**, and this data set is available in `https://archive.ics.uci.edu/ml/machine-learning-databases/00291/`. It consists of six independent variables, namely, frequency (in Hertzs), angle of attack (in degrees), chord length (in meters), free-stream velocity (in

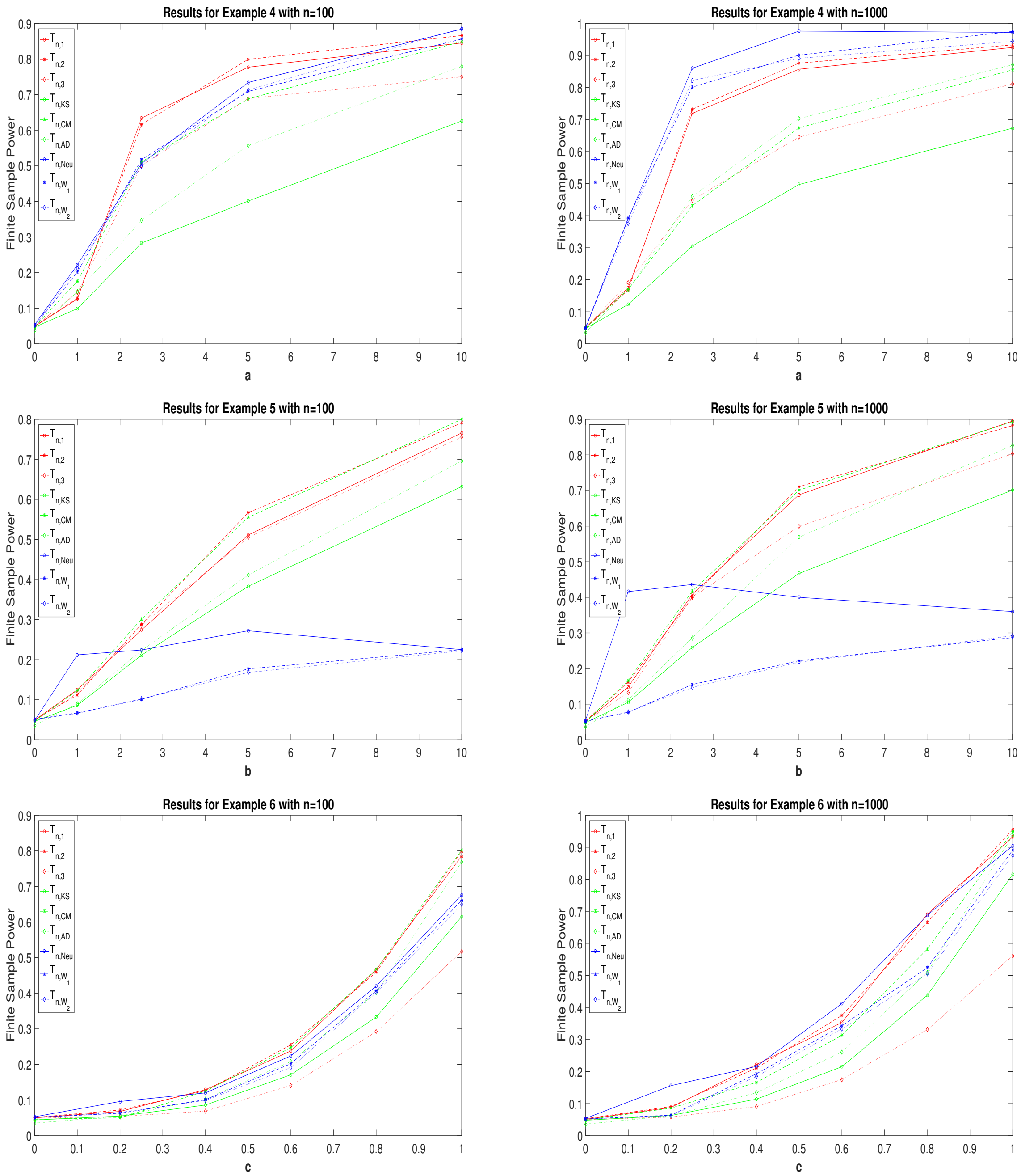

Figure 2: The finite sample power of different tests with 5% level of significance for Examples 4, 5 and 6. In first row, a is the parameter described in Example 3. In second row, b is the parameter described in Example 4, and in third row, c is the parameter described in Example 5.

meters per second) and suction side displacement thickness (in meters) and one dependent variable (i.e., Y), namely, scaled sound pressure level (in decibels). The size of the data set is 1503. In this study, we consider only one independent variable (i.e., X), namely, frequency (in Hertzs). In order to check whether X and the errors ϵ in the model (i.e., the model described in the beginning of Section 1) are independent or not, we carry out bootstrap tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ and compute the p-values of the corresponding tests. We first compute the value of $T_{n,i}$, $i = 1, 2, 3$ (denote it as $t_{0,i}$) from the given data and to estimate $P_{H_0}[T_{n,i} > t_{0,i}]$, (i.e., p-value) $i = 1, 2, 3$, we generate j many bootstrap resamples from the given data as we described in the paragraph after Example 6. Let $T^k_{n,i}$ denote the estimate of $T_{n,i}$ for k-th resample ($k = 1, \ldots, j$), and the p-value of the i-th test is defined as $\frac{\sum_{k=1}^{j} 1_{\{T^k_{n,i} > t_{0,i}\}}}{j}$. In this numerical study, we choose $j = 500$, and the p-values of the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are 0.57, 0.49 and 0.53, respectively for this data set. These p-values indicate that one can use the model $Y = m(X) + \epsilon$ for this data set, where X and ϵ are independent random variables, and having the knowledge of the independence between the covariates (i.e., X) and the errors (i.e., ϵ) helps us to use different statistical procedures in a straightforward way.

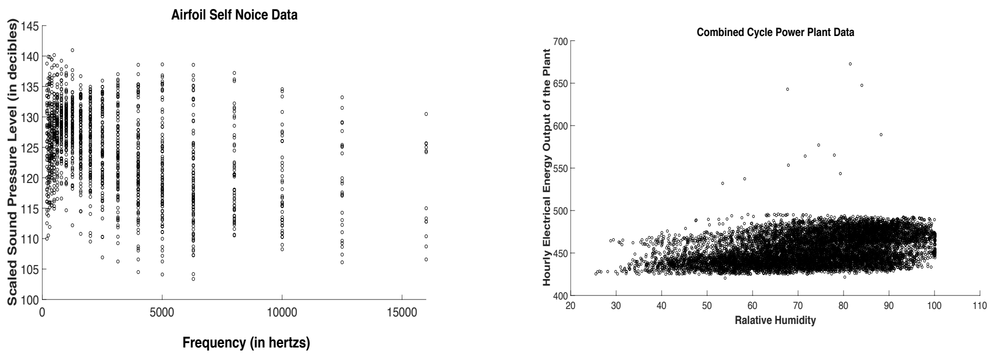

Figure 3: The scatter plots for Airfoil self noise data and Combined cycle power plant data.

We next consider a real data set **Combined Cycle Power Plant Data**, which is available in `https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant`. This data set contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. This data set consists of four independent variables, which are hourly average ambient variables, namely, temperature, ambient pressure, relative humidity and exhaust vacuum and one dependent variable, namely, net hourly electrical energy output of the plant. In our study, we consider relative humidity as the independent variable (X). Following the same procedure described for Airfoil Self Noise Data, we have computed the p-values of the tests based on $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$, and those p-values are 0.35, 0.40 and 0.14, respectively. The large p-values associated with the tests based on $T_{n,1}$ and $T_{n,2}$ favours the null hypothesis whereas the small value associated with the test based on $T_{n,3}$ supports the assertion in the alternative hypothesis. Since this data set contains some outliers (see the second diagram from right in Figure 3), the small p-value associated with the test based on $T_{n,3}$ indicates that it is not robust against the outliers.

5 Concluding remarks

In this article, we develop three tests to check whether the regressor and the errors are independent or not for the non-parametric regression model $Y = m(X) + \epsilon$, where Y, $m(.)$, X and ϵ are same as defined at the beginning of introduction. Among these three tests studied here, it is observed that the tests based on $T_{n,1}$ and $T_{n,2}$ are more robust against the outliers than the test based on $T_{n,3}$. The reason behind this fact is that $T_{n,1}$ and $T_{n,2}$ are based on the rank or the positions of the observations whereas $T_{n,3}$ is based on the absolute values of the observations or their powers. Besides, one can also formulate Kolmogorov-Smirnov or Cramer-von Mises type test statistics (e.g., see Einmahl and Van Keilegom (2008a)) based on $x_{(i)}$ and $y^*_{(i)}$, and the performances of those tests were investigated in finite sample study. Those tests along with the tests based on $T_{n,Neu}$ and $T_{n,W}$ also perform well when the data are generated from heavy tailed distribution since those tests are also based on the rank of the observations.

We should also point out that one can also consider higher order differences of $y_{(i-1)}$. In this article, we considered the second order differences of $y_{(i-1)}$, i.e., $y_{(i+1)} - 2y_{(i)} + y_{(i+1)}$ as we wanted to maximize the third moment of $\epsilon_{(i+1)} - 2\epsilon_{(i)} + \epsilon_{(i-1)}$ among all possible choices of $a\epsilon_{(i+1)} + b\epsilon_{(i)} + c\epsilon_{(i-1)}$, where $a, b, c \in \mathbb{R}$ along with the conditions (i) $a^2 + b^2 + c^2$ is a constant and (ii) $a + b + c = 0$, i.e., in other words, this construction gave us *more general* tests than usual test of homoscedasticity. In fact, any higher order odd moments can be maximized by the L-moments (see, e.g., Hosking (1990), which follows from the combinatorial arguments and the form of the expectation of the different power of the order statistics. The optimal choice of the order of the moment and the investigation of the new versions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ based on the other order of moment could be a subject of future research.

In addition, as we mentioned in the Introduction and investigated in the finite sample simulation study, our proposed tests can detect different kind of dependence structure between the covariates and the error random variable whereas a test of homoscedasticity fails to detect it when the conditional variance is constant.

For further illustration of this fact, we consider the following example, where the model is $Y = X^2 + \epsilon$. Let $(\epsilon|X = x)$ follow skew normal distribution (see Azzalini (1985)) with the location parameter $= \xi(x)$, the scale parameter $= \omega(x)$ and the shape parameter $= \alpha(x)$, where $\xi(x) = x$, $\omega(x) = \sqrt{x^2+3}$, and $\alpha(x) = \frac{3\pi}{\sqrt{2(x^2+3)-9\pi^2}}$, and X has the standard normal distribution. Here $\epsilon \in \mathbb{R}$ and $x \in \mathbb{R}$. We then generate a sample (x_1, y_1), ..., (x_{100}, y_{100}) based on the distributions of (ϵ, X), and for this sample, we carry out the test based on $T_{n,2}$ and a well-known homoscedasticity test, namely, Breusch-Pagan test (see Breusch and Pagan (1979)). To test $H_0 : \epsilon \perp\!\!\!\perp X$ against $H_1 : \epsilon \not\perp\!\!\!\perp X$, we have obtained a high p-value $= 0.657$ (i.e., favours H_0) for Breusch-Pagan test whereas a small p-value $= 0.041$ (i.e., H_0 is rejected) is obtained for the test based on $T_{n,2}$. In other words, it clearly indicates that Breusch-Pagan test fails to detect the dependence structure between ϵ and X since $Var(\epsilon|X)$ is constant $(= 3)$ here whereas the test based on $T_{n,2}$ is able to detect that dependence structure as the skewness of the conditional distribution of $\epsilon|X = x$ depends on x.

Furthermore, we would like to discuss another issue related to τ and τ^*. As we indicated in Section 2 that for two random variables X and Y with distributions F and G, respectively, $\tau = 0$ does *not* necessarily imply that $X \perp\!\!\!\perp Y$ whereas $\tau^* = 0 \Leftrightarrow X \perp\!\!\!\perp Y$. For example, let $C(x, y)$ denote the bivariate distributions on the unit cube with uniform marginals, and a direct algebra implies that $\tau = 4EC(X, Y) - 1$, and consequently, we have $\tau = 0$ if and only if $EC(X, Y) = 1/4$. In particular, note that if $X \perp\!\!\!\perp Y$, so are $U \perp\!\!\!\perp V$ (here $U = F(X)$ and $V = G(Y)$), then $C(u, v) = uv$, and hence, $EC(U, V) = \int_0^1 \int_0^1 uv\ dudv = 1/4$, which implies that $\tau = 4EC(U, V) - 1 = 0$. Next, let us consider another form of $C(u, v) = uv + \alpha u(u-1)(2u-1)v(v-1)(2v-1)$, where U and V follows uniform distribution on $[0, 1]$, and $\alpha \in [-1, 2]$. Note here that for any $\alpha \in [-1, 2]$, $\tau = 0$; however, $U \perp\!\!\!\perp V$ only when $\alpha = 0$. In other words, if we consider $\alpha = 1.5$, we have $\tau = 0$, but U and V are *not* independent. Motivated by this example, in the model $Y = m(X) + \epsilon$ considered in this article, suppose that the joint distribution of X and ϵ is $L_{\epsilon,X}(e, x) = ex + \alpha e(e-1)(2e-1)x(x-1)(2x-1)$ (consider $\alpha = 0$ under H_0 and $\alpha = 1.5$ under H_1), where ϵ and X both follow uniform distribution on $[0, 1]$. Considering $m(X) = X^2$, we then compute the p-value the tests based on both $T_{n,1}$ and $T_{n,2}$. For this example, the test based on $T_{n,1}$ gives us a high p-value $= 0.71$ whereas p-value of the test based on $T_{n,2}$ is 0.08, which is quite small. This investigation clearly indicates that the test based on $T_{n,1}$ fails since even under alternative $\tau = 0$.

The issue of robustness is also a potential research interest since $T_{n,1}$ and $T_{n,2}$ are rank based test statistics whereas $T_{n,3}$ is a moment based test statistic. To measure the robustness, one can compare the population versions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ in terms of maximum bias (e.g., see Huber (1981)). Roughly speaking, for a functional $T(F)$, where F is a distribution function and T is the population version of the estimator T_n, the maximum bias measures the effect on $T(F)$ of an arbitrary large observation with a specified mass. In fact, it is shown in Dhar, Bergsma and Dassios (2016) that a version of maximum bias of $T_{n,1}$ and $T_{n,2}$ based on (x_i, y_i) is bounded whereas that of $T_{n,3}$ based on (x_i, y_i) is unbounded, and it is expected that those results can be extended for $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ based on $(x_{(i)}, y^*_{(i)})$ also. Besides, one can derive the influence curve or the breakdown points of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ to measure the robustness of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ against the outliers. On computational issue, we would like to point out that the direct computation of $T_{n,2}$ based on (x_i, y_i) requires $O(n^4)$ operations whereas Weihs et al. (2016) showed that it can be reduced to $O(n^2 \log n)$, and Heller and Heller (2016) further reduced it to $O(n^2)$ operations.

In real data analysis, we carried out bootstrap tests and made the decision based on the p-values obtained from the bootstrap tests. Since $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are U-statistics, in view of Bickel and Freedman (1981, Section 3, p. 1203), the bootstrap version of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ is asymptotically (i.e., number of replications $j \to \infty$) valid in the sense of approximating the the original distributions of those test statistics after appropriate normalization, and hence, the critical value obtained by bootstrap method is asymptotically valid for the critical value that can be obtained by asymptotic distribution of the test statistics. Further, note that since the sample sizes of those real data set are large (1503 and 9568, respectively), one can also directly carry out the tests based on the asymptotic distributions described in Theorems 3.2, 3.3 and 3.4 as the asymptotic distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ under H_0 follow from the assertions of those theorems when $\gamma = 0$. However, one needs to estimate the unknown distribution function and the associated parameters from the given data to implement the tests based on the asymptotic distribution. Besides, to carry out such tests, one may also adopt some other methodologies proposed in Pfister, Buhlmann, Scholkopf and Peters (2017) and Leung and Drton (2017).

Acknowledgement: The authors are thankful to the editor Prof. Peter Dalgaard, an anonymous associate editor and two anonymous reviewers for several constructive suggestions and comments, which improved the article significantly. The work was carried out when the first author was visiting in the Department of Statistics at the London School of Economics and Political Science, and he would like to thank them for providing local hospitality and travel grant. The first author is also thankful to Mr. Prashant Jha for helping him to draw the figures and to Ms. Harsha Mulchandani for editing the first draft of the article.

References

[1] Arcones, M. A. (1995) On the central limit theorem for U-statistics under absolute regularity. *Statistics & Probability Letters*, **2**4, 245–249.

[2] Akritas, M. G. and Van Keilegom, I. (2001) Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, **2**8, 549–568.

[3] Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

[4] Babu, J. and Rao, C. R. (1988) Joint asymptotic distribution of marginal quantile functions in samples from a multivariate population. *Journal of Multivariate Analysis*, **2**7, 15–23.

[5] Banerjee, M. (2005) Likelihood ratio tests under local alternatives in regular semiparametric models. *Statistica Sinica*, **1**5, 635–644.

[6] Bergsma, W. and Dassios, A. (2014) A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, **2**0, 1006–1028.

[7] Behnen, K. (1971) Asymptotic optimality and ARE of certain rank-order tests under contiguity. *The Annals of Mathematical Statistics*, **4**2, 325–329.

[8] Behnen, K. and Neuhaus, G. (1975) A central limit theorem under contiguous alternatives. *The Annals of Statistics*, **3**, 1349–1353.

[9] Bickel, P. J. and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.

[10] Billingsley, P. (1995) *Probability and Measure.* Third Edition.

[11] Bradley, R. C. (2005) Basic Properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, **2**, 107–144.

[12] Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **4**7, 1287–1294.

[13] Cook, R. D. and Weisberg, S. (1983) Diagnostics for heteroscedasticity in regression. *Biometrika*, **7**0, 1–10.

[14] Dhar, S. S. (2016) Trimmed mean isotonic regression. *Scandinavian Journal of Statistics*, **4**3, 202–212.

[15] Dhar, S. S., Dassios, A. and Bergsma, W. (2016) A study of the power and robustness of a new test for independence against contiguous alternatives. *Electronic Journal of Statistics*, **1**0, 330–351.

[16] Einmahl, J. H. J. and Van Keilegom, I. (2008a) Tests for independence in nonparametric regression. *Statistica Sinica*, **1**8, 601–615.

[17] Einmahl, J. H. J. and Van Keilegom, I. (2008b) Specification tests in nonparametric regression. *Journal of Econometrics*, **1**43, 88–102.

[18] Glejser, H. (1969) A new test for heteroscedasticity. *Journal of the American Statistical Association*, **6**4, 316–323.

[19] Goldfeld, S. M. and Quandt, R. E. (1965) Some tests for homoscedasticity. *Journal of the American Statistical Association*, **6**0, 539–547.

[20] Hajek, J., Sidak, Z. and Sen, P. K. (1999) *Theory of Rank Tests.* Academic Press.

[21] Heller, Y. and Heller, R. (2016) Computing the Bergsma Dassios sign-covariance. https://arxiv.org/abs/1605.08732

[22] Hlavka, Z., Huskova, M. and Meintanis, S. G. (2011) Tests for independence in non-parametric heteroscedastic regression models. *Journal of Multivariate Analysis*, **1**02, 816–827.

[23] Hosking, J. R. M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society; Series B*, **5**2, 105–124.

[24] Huber, P. J. (1981) *Robust Statistics.* New York: Wiley.

[25] Kendall, M. G. (1938) A new measure of rank correlation. *Biometrika*, **3**0, 81–93.

[26] Lee, A. J. (1990) *U-Statistics: Theory and Practice.* Marcel Dekker: New York and Basel.

[27] Leucht, A. and Neumann, M. H. (2013) Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Annals of Institute of Statistical Mathematics*, **6**5, 349–386.

[28] Leung, D. and Drton, M. (2017) Testing independence in high dimensions with sums of squares of rank correlations. *The Annals of Statistics.* To appear.

[29] Lehmann, E. L. and Romano, J. P. (2005) *Testing Statistical Hypotheses.* New Delhi: Springer.

[30] Mijatovic, A. and Vysotsky, A. (2015) On the weak limit law of the maximal uniform k-spacing. http://arxiv.org/abs/1508.00944

[31] Nadaraya, E. A. (1964) On non-parametric estimates of density functions and regression curves. *Theory of Probability and Its Applications*, **1**0, 186–190.

[32] Nandy, P., Weihs, L. and Drton, M. (2016) Large-sample theory for the Bergsma-Dassios sign covariance. *Electronic Journal of Statistics*, **1**0, 2287–2311.

[33] Neumeyer, N. (2009) Testing independence in nonparametric regression. *Journal of Multivariate Analysis*, **1**00, 1551–1566.

[34] Pfister, N., Buhlmann, P., Scholkopf, B. and Peters, J. (2017) Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B.* To appear.

[35] Renyi, A. (1959) On measures of dependence. *Acta Mathematica Hungarica*, **1**0, 441–451.

[36] Rudin, W. (1976) *Principles of Mathematical Analysis.* Mcgraw-hill International Editions. Mathematics Series.

[37] Shorack, G. R. and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York

[38] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, vol. 26. CRC press.

[39] Szekely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, **3**5, 2769–2794.

[40] Van Der Vaart, A. W. (1998) *Asymptotic Statistics.* Cambridge University Press.

[41] Watson, G. S. (1964) Smooth regression analysis. *Sankhya*, **2**6, 359–372.

[42] Weihs, L., Drton, M. and Leung, D. (2016) Efficient computation of the Bergsma-Dassios sign covariance. *Computational Statistics*, **3**1, 315–328.

Subhra Sankar Dhar
Department of Mathematics and Statistics, IIT Kanpur, Kanpur 208016, Uttar Pradesh, India.
E-mail: subhra@iitk.ac.in

6 Appendix A: Proofs

Proof of Theorem 1: In order to establish the contiguity of the sequence $F_{n;\epsilon,X}$ relative to $F_{\epsilon,X}$, it is enough to show that L_n, the logarithm of the likelihood ratio, is asymptotically normal with mean $-\frac{1}{2}\sigma^2$ and variance σ^2 (see Hajek, Sidak and Sen (1999, p. 254, Corollary to Le Cam's first Lemma) and the first paragraph in Section 3), where σ is a positive constant. For notational convenience, we denote $\mathbf{Z} = (\epsilon, X)$, and $f_{n;\epsilon,X}$ and $f_{\epsilon,X}$ are the density functions associated with $F_{n;\epsilon,X}$ and $F_{\epsilon,X}$, respectively. Now, we have

$$
\begin{aligned}
L_n &= \sum_{i=1}^{n} \log \frac{f_{n;\epsilon,X}(\mathbf{z}_i)}{f_{\epsilon,X}(\mathbf{z}_i)} = \sum_{i=1}^{n} \log \frac{(1-\gamma/\sqrt{n})f_{\epsilon,X}(\mathbf{z}_i) + \gamma/\sqrt{n}k(\mathbf{z}_i)}{f_{\epsilon,X}(\mathbf{z}_i)} \\
&= \sum_{i=1}^{n} \log \left[1 + \gamma/\sqrt{n}\left\{\frac{k(\mathbf{z}_i)}{f_{\epsilon,X}(\mathbf{z}_i)} - 1\right\}\right] \\
&= \frac{\gamma}{\sqrt{n}} \sum_{i=1}^{n} \{m(\mathbf{z}_i) - 1\} - \frac{\gamma^2}{2n} \sum_{i=1}^{n} \{m(\mathbf{z}_i) - 1\}^2 / \left\{1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i) - 1\}}{\sqrt{n}}\right\}^2,
\end{aligned}
$$

where $m(\mathbf{z}_i) = \frac{k(\mathbf{z}_i)}{f_{\epsilon,X}(\mathbf{z}_i)}$, and $a_{in} \in (0,1)$ with probability 1.

Now, we define $W_n = \sum_{i=1}^{n} \frac{\gamma}{\sqrt{n}}\{m(\mathbf{z}_i) - 1\} - \frac{\gamma^2}{2} E_{f_{\epsilon,X}}\{m(\mathbf{z}_1) - 1\}^2$. Note that by straightforward application of C.L.T., it follows that W_n is asymptotically normal with mean $-\frac{\gamma^2}{2} E_{f_{\epsilon,X}}\{m(\mathbf{z}_1) - 1\}^2$ and variance $\gamma^2 E_{f_{\epsilon,X}}\{m(\mathbf{z}_1) - 1\}^2$ since $E_{f_{\epsilon,X}}\left\{\frac{k(\mathbf{z})}{f_{\epsilon,X}(\mathbf{z})} - 1\right\}^2 < \infty$. So, in order to prove contiguity of the sequence of densities associated with H_n, it is enough to show that $|L_n - W_n| \xrightarrow{p} 0$ as $n \to \infty$.

For convenience of writing, we denote $\sigma^2 = E_{f_{\epsilon,X}}\{m(\mathbf{z}_1) - 1\}^2$, $\sigma_{1l}^2 = E_{f_{\epsilon,X}}\{m(\mathbf{z}_1) - 1\}^2 1_{\{m(\mathbf{z}_1) \le l\}}$ and $\sigma_{2l}^2 = E_{f_{\epsilon,X}}\{m(\mathbf{z}_1) - l\}^2 1_{\{m(\mathbf{z}_1) > l\}}$, where $l > 0$ is a constant. So, we have

$$
|L_n - W_n| \le \left|T_{1n} - \frac{\gamma^2\sigma_{1l}^2}{2}\right| + \left|T_{2n} - \frac{\gamma^2\sigma_{2l}^2}{2}\right|,
$$

where T_{1n} and T_{2n} are given by

$$
T_{1n} = \sum_{i=1}^{n} \frac{\gamma^2\{m(\mathbf{z}_i) - 1\}^2}{2n} / \left[1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i) - 1\}}{\sqrt{n}}\right]^2 1_{\{m(\mathbf{z}_i) \le l\}}
$$

and

$$
T_{2n} = \sum_{i=1}^{n} \frac{\gamma^2\{m(\mathbf{z}_i) - 1\}^2}{2n} / \left[1 + \frac{a_{in}\gamma\{m(\mathbf{z}_i) - 1\}}{\sqrt{n}}\right]^2 1_{\{m(\mathbf{z}_i) > l\}}.
$$

Now, for a fixed $\epsilon > 0$, we choose l_0 sufficiently large such that $\gamma^2\sigma_{2l_0}^2 < \epsilon$ and $l_0 > 1$. Then,

$$
\begin{aligned}
& P\left[\left|T_{2n} - \frac{\gamma^2\sigma_{2l_0}^2}{2}\right| > \epsilon/2\right] \\
=\; & P\left[T_{2n} > \frac{\gamma^2\sigma_{2l_0}^2}{2} + \frac{\epsilon}{2}\right] + P\left[T_{2n} < \frac{\gamma^2\sigma_{2l_0}^2}{2} - \frac{\epsilon}{2}\right] \\
=\; & P\left[T_{2n} > \frac{\gamma^2\sigma_{2l_0}^2}{2} + \frac{\epsilon}{2}\right] + 0 \text{ (since } \tfrac{\gamma^2\sigma_{2l_0}^2}{2} - \tfrac{\epsilon}{2} < 0) \\
\le\; & P\left[\sum_{i=1}^{n} \frac{\gamma^2\{m(\mathbf{z}_i) - 1\}^2}{2n} 1_{\{m(\mathbf{z}_i) > l_0\}} > \frac{\gamma^2\sigma_{2l_0}^2}{2} + \frac{\epsilon}{2}\right] \to 0 \text{ as } n \to \infty.
\end{aligned}
$$

The last implication follows from the fact that $\sum_{i=1}^{n}\frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2}{2n}1_{\{m(\mathbf{z}_i)>l_0\}}\xrightarrow{p}\frac{\gamma^2\sigma^2_{2l_0}}{2}$.

Now, we fix $0<\eta<1$ on the event $\{m(\mathbf{z}_i)\leq l_0\}$, and hence, we have $1+\frac{a_{in}\gamma\{m(\mathbf{z}_i)-1\}}{\sqrt{n}}\leq 1+\frac{\gamma(l_0-1)}{\sqrt{n}}<1+\eta$ for all $n\geq N$ (say). Also, since $m(\mathbf{z})\geq 0$, $1+\frac{a_{in}\gamma\{m(\mathbf{z}_i)-1\}}{\sqrt{n}}\geq 1-\frac{a_{in}\gamma}{\sqrt{n}}>1-\eta$ for all $n\geq N$. Next, we define $V_{1n}=\sum_{i=1}^{n}\frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2 1_{\{m(\mathbf{z}_i)\leq l_0\}}}{2n(1+\eta)^2}$ and $V_{2n}=\sum_{i=1}^{n}\frac{\gamma^2\{m(\mathbf{z}_i)-1\}^2 1_{\{m(\mathbf{z}_i)\leq l_0\}}}{2n(1-\eta)^2}$. The aforementioned facts imply that $T_{1n}\in(V_{1n},V_{2n})$ for all $n\geq N$, $V_{1n}\xrightarrow{p}\frac{\gamma^2\sigma^2_{1l_0}}{2(1+\eta)^2}$ and $V_{2n}\xrightarrow{p}\frac{\gamma^2\sigma^2_{1l_0}}{2(1-\eta)^2}$. Hence, we have

$$\begin{aligned}
& P\left[\left|T_{1n}-\frac{\gamma^2\sigma^2_{1l_0}}{2}\right|>\epsilon/2\right]\\
\leq\ & P\left[\left|V_{1n}-\frac{\gamma^2\sigma^2_{1l_0}}{2(1+\eta)^2}\right|>\epsilon/2-\frac{\gamma^2\sigma^2_{1l_0}}{2}+\frac{\gamma^2\sigma^2_{1l_0}}{2(1+\eta)^2}\right]\\
+\ & P\left[\left|V_{2n}-\frac{\gamma^2\sigma^2_{1l_0}}{2(1-\eta)^2}\right|>\epsilon/2+\frac{\gamma^2\sigma^2_{1l_0}}{2}+\frac{\gamma^2\sigma^2_{1l_0}}{2(1-\eta)^2}\right].
\end{aligned}$$

Now, we choose $\eta>0$ so small such that $\epsilon/2-\frac{\gamma^2\sigma^2_{1l_0}}{2}+\frac{\gamma^2\sigma^2_{1l_0}}{2(1+\eta)^2}>0$ and $\epsilon/2+\frac{\gamma^2\sigma^2_{ll_0}}{2}+\frac{\gamma^2\sigma^2_{1l_0}}{2(1-\eta)^2}>0$. Thus, we have $T_{1n}-\frac{\gamma^2\sigma^2_{1l_0}}{2}\xrightarrow{p}0$, and consequently, $L_n-W_n\xrightarrow{p}0$, which ensures the contiguity of the sequence of densities associated with H_n. This completes the proof. □

Proof of Proposition 1: Here we prove the result for $T_{n,2}$, and the proofs for $T_{n,1}$ and $T_{n,3}$ is similar to that of $T_{n,2}$. Note that $T_{n,2}=\frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n}a(x_{(i)},x_{(j)},x_{(k)},x_{(l)})a(y^*_{(i)},y^*_{(j)},y^*_{(k)},$ $y^*_{(l)})=\frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n}a(x_{(i)},x_{(j)},x_{(k)},x_{(l)})a(y_{(i-1)}-2y_{(i)}+y_{(i+1)},y_{(j-1)}-2y_{(j)}+y_{(j+1)},y_{(k-1)}-2y_{(k)}+y_{(k+1)},y_{(l-1)}-2y_{(l)}+y_{(l+1)})$, and under H_0, $X_1,\ldots,X_n$ and $\epsilon_1,\ldots,\epsilon_n$ are two independent i.i.d. samples. Further, let $R_1,\ldots,R_n$ be the ranks of $X_1,\ldots,X_n$, respectively, and also observe that $X_1,\ldots,X_n$ and $\epsilon_{R_1},\ldots,\epsilon_{R_n}$ are two independent i.i.d. samples. We now consider $(X_1,\epsilon_{R_1}),\ldots,(X_n,\epsilon_{R_n})$, which are i.i.d. random vectors having independent components, and it is easy to see that ϵ_{R_i} has the distribution G_ϵ (i.e., the distribution function of ϵ) in view of the fact that R_i's ($i=1,\ldots,n$) are distributed uniformly over $\{1,\ldots,n\}$, and hence, we can redefine the model as $Y_i=m(X_i)+\epsilon_{R_i}\Leftrightarrow Y_{(i)}=m(X_{(i)})+\epsilon_i$ using (B1). Under this model, we have

$$\begin{aligned}
T_{n,2} &= \frac{1}{\binom{n}{4}}\sum_{1\leq i<j<k<l\leq n}\{a(x_{(i)},x_{(j)},x_{(k)},x_{(l)})a(m(x_{(i-1)})-2m(x_{(i)})+m(x_{(i+1)})+\epsilon_{i-1}-2\epsilon_i\\
&+ \epsilon_{i+1},m(x_{(j-1)})-2m(x_{(j)})+m(x_{(j+1)})+\epsilon_{j-1}-2\epsilon_j+\epsilon_{j+1},m(x_{(k-1)})-2m(x_{(k)})\\
&+ m(x_{(k+1)})+\epsilon_{k-1}-2\epsilon_k+\epsilon_{k+1},m(x_{(l-1)})-2m(x_{(l)})+m(x_{(l+1)})+\epsilon_{l-1}-2\epsilon_l+\epsilon_{l+1})\}
\end{aligned}$$

Note that $m(x_{(j-1)})-2m(x_{(j)})+m(x_{(j+1)})=\{m(x_{(j-1)})-m(x_{(j)})\}+\{m(x_{(j+1)})-m(x_{(j)})\}=m'(\xi_{i-1})(x_{(i)}-x_{(i-1)})+m'(\xi_i)(x_{(i+1)}-x_{(i)})=\{m'(\xi_{i-1})+m'(\xi_i)\}O_p\left(\frac{\log n}{n}\right)$. The last inequality follows from the fact that $m(.)$ is differentiable (using (B2)), and $\max_{i\in\{1,\ldots,n\}}|x_{(i)}-x_{(i-1)}|=O_p\left(\frac{\log n}{n}\right)$ when the support of the probability density function of X is bounded (e.g., see Shorack and Wellner (2009, p.731, Section 4) and Mijatovic and Vladislav (2015); consider $k=1$ in Theorem 1). Note that since $T_{n,2}$ is invariant under monotone transformation, without loss of generality, one may consider x_is are in $[0,1]$ even when the the support of the probability density function of X is unbounded (e.g., by integral transformation $X\rightarrow F_X$, where F_X is the distribution function of X).

Hence, we now have

$$\begin{aligned}
T_{n,2} &= \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} \{a(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)}) a(m'(\xi_{i-1})(x_{(i)} - x_{(i-1)}) + m'(\xi_i)(x_{(i+1)} - x_{(i)}) \\
&+ \epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}, m'(\xi_{j-1})(x_{(j-1)} - x_{(j)}) + m'(\xi_j)(x_{(j)} - x_{(j+1)}) + \epsilon_{j-1} - 2\epsilon_j \\
&+ \epsilon_{j+1}, m'(\xi_{k-1})(x_{(k)} - x_{(k-1)}) + m'(\xi_k)(x_{(k+1)} - x_{(k)}) + \epsilon_{k-1} - 2\epsilon_k + \epsilon_{k+1}, \\
&\quad m'(\xi_{l-1})(x_{(l)} - x_{(l-1)}) + m'(\xi_l)(x_{(l+1)} - x_{(l)}) + \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\} \\
&= \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} \{a(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)}) a(\{m'(\xi_{i-1}) + m'(\xi_i)\} O_p\left(\frac{\log n}{n}\right) \\
&+ \epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}, \{m(\xi_{j-1}) + m(\xi_j)\} O_p\left(\frac{\log n}{n}\right) + \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \\
&\quad \{m'(\xi_{k-1}) + m'(\xi_k)\} O_p\left(\frac{\log n}{n}\right) + \epsilon_{k-1} - 2\epsilon_k + \epsilon_{k+1}, \{m'(\xi_{l-1}) + m'(\xi_l)\} \\
&\quad O_p\left(\frac{\log n}{n}\right) + \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\}.
\end{aligned}$$

Hence, $T_{n,2} - \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} \{a(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)}) a(\epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}, \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \epsilon_{k-1} - 2\epsilon_k + \epsilon_{k+1}, \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\} \xrightarrow{p} 0$ as $n \to \infty$ in view of the fact that $\sup_{x\in\mathbb{R}} |m'(x)| < \infty$. Since the product of two sign functions are bounded by 1, using dominated convergence theorem (e.g., see Billingsley (1995)), we have $E(T_{n,2}) - \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} E[\{a(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)}) a(\epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}, \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \epsilon_{k-1} - 2\epsilon_k + \epsilon_{k+1}, \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\}] \to 0$ as $n \to \infty$. Further, since $(x_{(1)}, \ldots, x_{(n)})$ and $(\epsilon_1, \ldots, \epsilon_n)$ are two independent sequence of random variables, $E(T_{n,2}) - \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} E[\{a(x_{(i)}, x_{(j)}, x_{(k)}, x_{(l)})\}] E[a(\epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}, \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \epsilon_{k-1} - 2\epsilon_k + \epsilon_{R_{k+1}}, \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\}] \to 0$ as $n \to \infty$. Note that as $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. random variables, $E[a(\epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}, \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \epsilon_{k-1} - 2\epsilon_k + \epsilon_{R_{k+1}}, \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\}] = 0$ unless $i, j, k, l \in A$, where $A = \{(i, j, k, l) : \mathcal{P}\{i-1, i, i+1\} = \mathcal{P}\{j-1, j, j+1\} = \mathcal{P}\{k-1, k, k+1\} = \mathcal{P}\{l-1, l, l+1\}\}$. Here $\mathcal{P}$ denote the class of all permutations. It follows from the construction of the set A that the number of elements in A is finite and independent of n, and consequently, $\frac{Card(A)}{n} \to 0$ as $n \to \infty$, where $Card(A)$ denotes the number of elements in A. Hence, $E(T_{n,2}) \to 0$ as $n \to \infty$. Arguing in a similar way as in the proof of $E(T_{n,2}) \to 0$ as $n \to \infty$, we have $E(T_{n,2}^2) \to 0$ as $n \to \infty$. These two facts imply that $T_{n,2} \xrightarrow{p} 0$ as $n \to \infty$. Since both $T_{n,1}$ and $T_{n,3}$ are based on sign functions, arguing in a similar way as in the case of $T_{n,2}$, one can establish that $T_{n,1} \xrightarrow{p} 0$ and $T_{n,3} \xrightarrow{p} 0$ as $n \to \infty$. □

Proof of Lemma 1: Here also, we shall provide the proof for $T_{n,2}$, and the similar arguments lead to the result for $T_{n,3}$. Arguing in a similar way as in the proof of Proposition 1, using (B1)-(B2), we have $E\left[T_{n,2} | X_{(i)} = x, Y^*_{(i)} = y\right] - \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} E[\{a(x, x_{(j)}, x_{(k)}, x_{(l)})$
$E(\epsilon, \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \epsilon_{k-1} - 2\epsilon_k + \epsilon_{k+1}, \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\}] \to 0$, where ϵ is a fixed value of $\epsilon_{i-1} - 2\epsilon_i + \epsilon_{i+1}$. Now, note that $E[a(\epsilon, \epsilon_{j-1} - 2\epsilon_j + \epsilon_{j+1}, \epsilon_{k-1} - 2\epsilon_k + \epsilon_{R_{k+1}}, \epsilon_{l-1} - 2\epsilon_l + \epsilon_{l+1})\}] = 0$ for all ϵ unless $j, k, l \in A^*$, where $A^* = \{(j, k, l) : \mathcal{P}\{j-1, j, j+1\} = \mathcal{P}\{k-1, k, k+1\}\} = \mathcal{P}\{l-1, l, l+1\}\}$. Here $\mathcal{P}$ is the class of all permutations. It follows from the construction of the set A^* that the number of elements in A is finite and independent of n, and hence, $E[T_{n,2} | x_{(i)} = x, y^*_{(i)} = y] \to 0$ as $n \to \infty$ for all x, y. Since $T_{n,3}$ is also based on sign function, arguing in a similar way, one can show that $E\left[T_{n,3} | X_{(i)} = x, Y^*_{(i)} = y\right] \to 0$

as $n \to \infty$ for all x, y. □

To prove Theorem 2, we should state the following lemma proposed by Le Cam.

Lemma: Le Cam's third lemma: *Let $\{X_n\} \in \mathbb{R}^d$ be a sequence of random vectors, and the sequence of measures Q_n is contiguous with respect to the sequence of another probability measures P_n. If $(X_n, \log \frac{dQ_n}{dP_n})$ converges weakly to a random vector in $\mathbb{R}^{d+1}$ associated with $(d+1)$-dimensional normal distribution with the location parameter $= \begin{pmatrix} \mu \\ -\frac{\sigma^2}{2} \end{pmatrix}$ and the scatter parameter $= \begin{pmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{pmatrix}$ under P_n, then $\{X_n\}$ converges weakly to a random vector in $\mathbb{R}^d$ associated with d-dimensional normal distribution with the location parameter $= \mu + \tau$ and the scatter parameter $= \Sigma$ under Q_n.*

Proof: See page 90 of Van der Vaart (1998). □

Proof of Theorem 2: Since $T_{n,1}$ is a non-degenerate U-statistics (e.g., see Lee (1990), p. 14–15) as mentioned in the discussion after Lemma 1 and in view of expansion of L_n in the proof of Proposition 1, the joint distribution of $\sqrt{n}\{(T_{n,1} - E(T_{n,1})), L_n/\sqrt{n}\}$ is asymptotically bivariate normal distribution. Note also that the asymptotic covariance between $\sqrt{n}(T_{n,1} - E(T_{n,1}))$ and L_n is

$$\begin{aligned}
& \frac{2\gamma}{n} E_{f_{\epsilon,X}}\left[\sum_{i=1}^{n} E[\text{sign}\{(X - X_{(i)})(Y - Y^*_{(i)})\}|X,Y] \times \left\{\frac{k(\mathbf{z}_i)}{f_{\epsilon,X}(\mathbf{z}_i)} - 1\right\}\right] \\
= \;& 2\gamma E_k E\left[\text{sign}\{(X - X)(Y - (\epsilon_1 - 2\epsilon_2 + \epsilon_3))\}|X,Y\right] \\
& (\text{since } 2E_{f_{\epsilon,X}} E[\text{sign}\{(X - X_{(i)})(Y - Y^*_{(i)})\}|X,Y] = 0,\ Y_{(i)} = m(X_{(i)}) + \epsilon_i, \\
& \sup_x |m'(x)| < \infty \text{ and } \max_{i \in \{1,\ldots,n\}} |x_{(i)} - x_{(i-1)}| = o_P(1)) \\
= \;& 2\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[2\int_{-\infty}^{x}\int_{-\infty}^{y} dH_X dG^* + 2\int_{x}^{\infty}\int_{y}^{\infty} dH_X dG^* - 1\right] dK,
\end{aligned}$$

where H_X is the distribution function of X, and G^* is the distribution function of $\epsilon_1 - 2\epsilon_2 + \epsilon_3$.

Now, by a straightforward application of Le Cam's third lemma (see also Hajek, Sidak and Sen (1999, p. 257)) and the asymptotic distribution of a non-degenerate 2-dependent U-statistic (see Lee (1990), Arcones (1995) and Bradley (2005) for relationship between β-mixing and m-dependent random variables), one can establish that under contiguous alternatives H_n (see Theorem 1), $\sqrt{n}(T_{n,1} - E(T_{n,1}))$ converges weakly to a Gaussian distribution with mean

$$\mu_1 = 2\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[2\int_{-\infty}^{x}\int_{-\infty}^{y} dH_X dG^* + 2\int_{x}^{\infty}\int_{y}^{\infty} dH_X dG^* - 1\right] dK$$

and variance

$$\sigma_1^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[2\int_{-\infty}^{x}\int_{-\infty}^{y} dH_X dG^* + 2\int_{x}^{\infty}\int_{y}^{\infty} dH_X dG^* - 1\right]^2 dG_\epsilon dH_X.$$

Hence, the proof is complete. □

Proof of Theorem 3: It follows from Lemma 1 that the asymptotic degeneracy of $T_{n,2}$ is 1, which implies that $n(T_{n,2} - E(T_{n,2}))$ will converge weakly to $\sum_{i=1}^{\infty} \lambda_i\{Z_i^2 - 1\}$ under H_0, where λ_i's are the eigenvalues associated with $l(x,y)$, and Z_i's are i.i.d. $N(0,1)$ random variables (e.g., see Leucht and Neumann (2013)).

Further, note that the sequence of densities (denote it as q_n) associated with H_n is dominated by the density (denote it as p_0) associated with H_0 with Radon-Nikodym derivative $\frac{dq_n}{dp_0} = 1 + n^{-\frac{1}{2}} h_n$, where $h_n = \gamma \left(\frac{k}{f} - 1 \right) \in L_2(p_0)$ since $E_f \left(\frac{k}{f} - 1 \right)^2 < \infty$, which is assumed in the theorem. Hence, q_n and p_0 satisfy the assumptions stated in Theorem 2.1 in Gregory (1977), which concludes that $n(T_{n,2} - E(T_{n,2}))$ converges weakly to $\sum_{i=1}^{\infty} \lambda_i \{(Z_i + a_i)^2 - 1\}$ under H_n, where λ_i, Z_i and a_i are as defined in the statement of the theorem. This completes the proof. $\square$

Proof of Theorem 4: Lemma 1 also asserts that the asymptotic degeneracy of $T_{n,3}$ is 1, which implies that $n(T_{n,3} - E(T_{n,3}))$ will converge weakly to $\sum_{i=1}^{\infty} \lambda_i^* \{Z_i^{*2} - 1\}$ under H_0, where λ_i^*'s are the eigenvalues associated with $l^*(x, y)$, and Z_i^*'s are i.i.d. $N(0, 1)$ random variables (e.g., see Leucht and Neumann (2013)).

Arguing in the same way as in the proof of Theorem 3, $n(T_{n,3} - E(T_{n,3}))$ converges weakly to $\sum_{i=1}^{\infty} \lambda_i^* \{(Z_i^* + a_i^*)^2 - 1\}$ under H_n, where λ_i^*, Z_i^* and a_i^* are defined as in the statement of the theorem. This completes the proof. $\square$

7 Appendix B: Asymptotic power study and finite sample simulation study

γ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
Test based on $T_{n,1}$	0.05	0.09	0.12	0.20	0.29	0.37	0.45	0.55	0.62	0.69	0.75
Test based on $T_{n,2}$	0.05	0.09	0.13	0.22	0.25	0.35	0.41	0.57	0.65	0.77	0.81
Test based on $T_{n,3}$	0.05	0.08	0.16	0.20	0.26	0.34	0.43	0.54	0.66	0.75	0.80

Table 1: The results for **Example 1**: The asymptotic power of different tests for different values of γ. For different values of γ, the value within each cell of the second, the third and the fourth rows denote the asymptotic power of the corresponding tests at 5% level of significance under contiguous alternatives.

γ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
Test based on $T_{n,1}$	0.05	0.07	0.14	0.26	0.39	0.48	0.62	0.74	0.86	0.99	1
Test based on $T_{n,2}$	0.05	0.08	0.15	0.22	0.41	0.50	0.60	0.77	0.90	1	1
Test based on $T_{n,3}$	0.05	0.08	0.11	0.16	0.19	0.25	0.28	0.32	0.42	0.47	0.52

Table 2: The results for **Example 2**: The asymptotic power of different tests for different values of γ. For different values of γ, the value within each cell of the second, the third and the fourth rows denote the asymptotic power of the corresponding tests at 5% level of significance under contiguous alternatives.

Tests ($n = 100$)	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,KS}$	$T_{n,CM}$	$T_{n,AD}$	$T_{n,Neu}$	T_{n,W_1}	T_{n,W_2}
$a = 0$	0.049	0.050	0.049	0.047	0.048	0.038	0.054	0.050	0.052
$a = 1$	0.128	0.129	0.147	0.099	0.176	0.146	0.222	0.202	0.213
$a = 2.5$	0.637	0.610	0.509	0.283	0.511	0.347	0.506	0.517	0.499
$a = 5$	0.780	0.801	0.693	0.401	0.688	0.557	0.734	0.709	0.713
$a = 10$	0.849	0.869	0.756	0.626	0.849	0.779	0.884	0.856	0.886
Tests ($n = 1000$)	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,KS}$	$T_{n,CM}$	$T_{n,AD}$	$T_{n,Neu}$	T_{n,W_1}	T_{n,W_2}
$a = 0$	0.051	0.048	0.050	0.049	0.049	0.037	0.049	0.051	0.052
$a = 1$	0.181	0.159	0.195	0.126	0.180	0.168	0.395	0.395	0.375
$a = 2.5$	0.717	0.747	0.462	0.315	0.441	0.462	0.862	0.811	0.822
$a = 5$	0.856	0.870	0.655	0.487	0.664	0.713	0.976	0.911	0.899
$a = 10$	0.937	0.949	0.827	0.663	0.859	0.881	0.973	0.979	0.946

Table 3: The results for **Example 3**. The finite sample power of the different tests for different values of n at 5% level of significance. The number of repetitions is 1000.

Tests ($n = 100$)	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,KS}$	$T_{n,CM}$	$T_{n,AD}$	$T_{n,Neu}$	T_{n,W_1}	T_{n,W_2}
$b = 0$	0.048	0.050	0.048	0.046	0.047	0.036	0.048	0.051	0.050
$b = 1$	0.127	0.117	0.116	0.086	0.123	0.090	0.212	0.067	0.066
$b = 2.5$	0.276	0.282	0.286	0.211	0.301	0.223	0.224	0.101	0.102
$b = 5$	0.518	0.569	0.511	0.383	0.555	0.411	0.272	0.177	0.168
$b = 10$	0.769	0.799	0.761	0.632	0.800	0.696	0.225	0.225	0.222
Tests ($n = 1000$)	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,KS}$	$T_{n,CM}$	$T_{n,AD}$	$T_{n,Neu}$	T_{n,W_1}	T_{n,W_2}
$b = 0$	0.052	0.052	0.051	0.048	0.049	0.036	0.054	0.052	0.051
$b = 1$	0.157	0.175	0.137	0.105	0.166	0.112	0.417	0.079	0.078
$b = 2.5$	0.418	0.411	0.406	0.259	0.427	0.286	0.446	0.165	0.157
$b = 5$	0.690	0.716	0.592	0.467	0.711	0.569	0.400	0.232	0.228
$b = 10$	0.899	0.895	0.806	0.701	0.899	0.826	0.362	0.287	0.298

Table 4: The results for **Example 4**. The finite sample power of the different tests for different values of n at 5% level of significance. The number of repetitions is 1000.

Tests ($n = 100$)	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,KS}$	$T_{n,CM}$	$T_{n,AD}$	$T_{n,Neu}$	T_{n,W_1}	T_{n,W_2}
$c = 0$	0.050	0.051	0.049	0.044	0.046	0.035	0.053	0.051	0.052
$c = 0.2$	0.069	0.075	0.056	0.055	0.050	0.055	0.096	0.063	0.065
$c = 0.4$	0.131	0.128	0.070	0.086	0.125	0.103	0.120	0.101	0.099
$c = 0.6$	0.236	0.257	0.144	0.171	0.247	0.208	0.224	0.202	0.191
$c = 0.8$	0.461	0.461	0.296	0.333	0.467	0.401	0.420	0.407	0.402
$c = 1$	0780	0.800	0.522	0.615	0.802	0.769	0.676	0.661	0.649
Tests ($n = 1000$)	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	$T_{n,KS}$	$T_{n,CM}$	$T_{n,AD}$	$T_{n,Neu}$	T_{n,W_1}	T_{n,W_2}
$c = 0$	0.052	0.053	0.050	0.048	0.049	0.036	0.054	0.052	0.052
$c = 0.2$	0.091	0.096	0.061	0.063	0.086	0.062	0.166	0.064	0.062
$c = 0.4$	0.229	0.218	0.097	0.114	0.166	0.134	0.226	0.199	0.189
$c = 0.6$	0.357	0.377	0.166	0.215	0.313	0.261	0.422	0.349	0.343
$c = 0.8$	0.696	0.675	0.338	0.438	0.582	0.509	0.698	0.524	0.515
$c = 1$	0.938	0.960	0.577	0.815	0.949	0.937	0.904	0.891	0.885

Table 5: The results for **Example 5**. The finite sample power of the different tests for different values of n at 5% level of significance. The number of repetitions is 1000.

Tests	$T_{n,1}$ based test	$T_{n,2}$ based test	$T_{n,3}$ based test	Breusch-Pagan test
$n = 100$	0.715	0.841	0.577	0.052
$n = 1000$	0.811	0.925	0.608	0.066

Table 6: The results for **Example 6**. The finite sample power of the different tests for different values of n at 5% level of significance. The number of repetitions is 1000.