Thailand Statistician
October 2025; 23(4): 798-815
http://statassoc.or.th
Contributed paper

An Extended Approach to Test of Independence between Error and Covariates under Nonparametric Regression Model

Sthitadhi Das, Soutik Halder and Saran Ishika Maiti *
Department of Statistics, Institute of Sciences, Visva Bharati, Santiniketan, India
*Corresponding author; e-mail: saran.ishika@gmail.com

Received: 6 November 2020
Revised: 20 March 2021
Accepted: 1 April 2021

Abstract

In 2014, Bergsma et al. (2014) proposed a generalized measure of association τ^* as an extension of widely used Kendall's τ. Later, in testing of independence between error and covariate, under nonparametric regression model $Y = m(X) + \epsilon$, with unknown regression function m and observation error ϵ, test statistic tailored on τ^* was suggested by Dhar et al. (2018). In this article, we develop a test, constructed on further extension of τ^*, considering the ordered X and the third order difference of Y with an motive to address the same issue of independence. We deduce the asymptotic distributions of test statistics using the theory of degenerate U-statistics. Moreover, we unravel the power of the proposed tests using Le Cam's concept of contiguous alternatives. A couple of simulated examples on normal and non normal distribution are furnished. Also, the performance of the test statistics is honed through a real data analysis.

Keywords: Kendall's τ, measures of association, asymptotic power, contiguous alternative, nonparametric regression model.

1. Introduction

Let us consider a set of n independent and identically distributed data points $(X_i, Y_i), i = 1(1)n$; Y being the response and X the covariate. We assume that X and Y are related via nonparametric regression model $Y = m(x) + \epsilon$, where m is an unknown regression function and ϵ is the error. To avoid the problem of identification, suppose that $E(\epsilon|X) = 0$. Objective of this article is to engineer a method of testing the independence of ϵ and X.

In last few years, a handful of articles tackling this issue under nonparametric regression model have been documented in literature. Most pioneering of them is by Einmahl et al. (2008) who 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 adopted as the second order difference of Y. Later, Neumeyer (2009) proposed a relevant test statistic based on a kernel estimator on L_2-distance between the conditional distribution and the unconditional distribution of the covariate X.

Apart from distance based tests, association-measures based tests too emerged in latest literature. The stepping stone in this regard is the popular Kendall's τ which measures the association between X and Y using the relative position label (rank) of the variables. Much later, Bergsma et al.

(2014) proposed a direct modification of τ, named τ^*, which is constructed on the four neighbouring triplicates of (X, Y).

$$\tau^* = \tau^*(X, Y) = Ea(X_1, X_2, X_3, X_4)a(Y_1, Y_2, Y_3, Y_4)$$

where $a(z_1, z_2, z_3, z_4) = sign(|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|)$. Further, clubbing the ideas from Einmahl et al. (2008) and Bergsma et al. (2014), Dhar et al. (2018) used τ^*, now constructed on the second order differences of neighbouring triplets of response, as a competent test statistic in order to checking the independence of ϵ and X.

Paving the same way, here we propose an association based test of independence constructed on ordered X and third order differences of response variable Y. As the kernel of test statics is woven by the dependent observations, theory of degenerate U-statistic for dependent random variables (see, Lee (1990)) is required to be brushed up in formulating the distributions of test statistics. Keeping in mind that one cannot observe the errors ϵ_i's directly, $i = 1, ..., n$ since the data (x_1, y_1),..., (x_n, y_n) are based on the observations obtained from the joint distribution associated with (X, Y), we need to choose Y values trickily so that the effect of m will be canceled out under the smoothness assumption of m. Further, to study the quality of the proposed test, power of the test is investigated under contiguous alternatives(see, Lehmann et al. (2005)).

The remainder of the article is outlined as follows. Section 2 discusses on the development of test statistics and some pertinent issues. Section 3 demonstrates the asymptotic distribution of the test statistics and asymptotic power under contiguous alternatives. A pair of finite sample simulated examples, designed on both normal and non normal alternative respectively, is placed in Section 4. Also, to incite readers' interest this very section includes a real data analysis performed by the proposed tests. Finally, Section 5 concludes this article with some general discussions indicating towards future explorations.

2. Developing the Test Statistics

For nonparametric regression model $Y = m(X) + \epsilon$, we would like to test

$$H_0 : X \perp\!\!\!\perp \epsilon (\Leftrightarrow F_{X,\epsilon} = G_X H_\epsilon) \ against \ H_1 : X \not\!\perp\!\!\!\perp \epsilon (\Leftrightarrow F_{X,\epsilon} \neq G_X H_\epsilon),$$

where $F_{X,\epsilon}$ being the joint distribution function of X and ϵ, H_ϵ and G_X the marginal distribution functions of ϵ and X respectively. Let $x_{(1)} \leq x_{(2)} ... \leq x_{(n)}$ be the order statistics, sorted through the increasing values of X, obtained from a random sample (x_1, y_1), (x_2, y_2),...,(x_n, y_n). Also, $y_{(1)}$,...,$y_{(n)}$ are the induced order statistics of Y, i.e., Y-values corresponding to the ordered X-values.

Under this setting, three test statistics $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are proposed on the observations $x_{(i)}$ and the third forward difference of $y_{(i)}$, say $y^*_{(i)}$. So, $y^*_{(i)} := y_{(i+1)} - 3y_{(i)} + 3y_{(i-1)} - y_{(i-2)}$, $i = 1, ..., n$. In order to define all the quantities properly, we assume that $y_{(-1)} = y_{(0)} = y_{(1)}$. Also, $y_{(n+1)} = y_{(n)}$, which results $y^*_{(1)} = y_{(2)}$, $y^*_{(2)} = 2y_{(1)} - 3y_{(2)} + y_{(3)}$ and $y^*_{(n)} = -y_{(n-2)} + 3y_{(n-1)} - 2y_{(n)}$. Then, we define

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

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

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

where $sign(x) = \frac{x}{|x|}$ when $x \neq 0$ and =0 otherwise, $a(z_1, z_2, z_3, z_4) = sign(|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|)$, 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 further that $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are three separate measures of association. All three $T_{n,1}, T_{n,3}$ and $T_{n,3}$ can handle occurrences of ties in explanatory variable (see Bergsma et al. (2014)). Tacitly, $T_{n,1}$ is the sample Kendall's τ while $T_{n,2}$ is our proposed statistic stemming from Bergsma's τ^* and $T_{n,3}$ is the statistic motivated from Dhar et al. (2018).

For the unordered (X, Y), the population versions of $T_{n,1}, T_{n,2}$ and $T_{n,3}$ are $\tau_1 = Esign\{(X_1 - X_2)(Y_1 - Y_3)\}$, $\tau_2 = Ea(x_1, x_2, x_3, x_4)a(y_1, y_2, y_3, y_4)$ and $\tau_3 = \frac{1}{4}Eh(x_1, x_2, x_3, x_4)h(y_1, y_2, y_3, y_4)$, 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). Also, the formulations of the T_i's are such that $T_i = 0$, $i = 1, 2, 3$ if $X \perp\!\!\!\perp Y$. Actually, $T_2(X, Y) = T_3(X, Y) = 0$ iff $X \perp\!\!\!\perp Y$ (see Bergsma et al. (2014)).

Next, let us delve in the rationale of constructing y^*, i.e., the third order differences of Y's. Since $X \perp\!\!\!\perp \epsilon \Rightarrow X \perp\!\!\!\perp g(\epsilon)$ for any proper function g, we may consider the appropriate differences of $y_{(i)}$'s, which accredits 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 $\epsilon^*_j = \epsilon_{j+1} - 3\epsilon_j + 3\epsilon_{j-1} - \epsilon_{j-2} := g(\epsilon_j)$, $i = 1, 2, 3$ and $j = 1, \cdots, n$. Instead of looking at $y^*_{(i)}$, the usual first difference $(y_{(i)} - y_{(i-1)})$ (or the slope $(y_{(i)} - y_{(i-1)})/(x_{(i)} - x_{(i-1)})$), can also serve as an approximation of $(\epsilon_{(i)} - \epsilon_{(i-1)})$ (or the slope $(\epsilon_{(i)} - \epsilon_{(i-1)})/(x_{(i)} - x_{(i-1)})$) having negligible third moment. But the third moment of the difference between two nearly identical ϵ_i's is close to zero as the first order differences are symmetrically distributed. So the test based on the first order differences will be an analogous one to any nonparametric test of homoscedasticity (see also the discussion in Einmahl et al. (2008)). Instead of first difference, choosing the third difference of $Y_{(i)}$ turns our test more general over the existent tests of homoscedasticity. To be more thorough, let us provide an explanation why it is so.

In literature, test of homoscedasticity states $H^*_0 : E(\epsilon^2_i | X = x) = \sigma^2$ for all x against H^*_1: all possible alternatives (see Goldfeld (1965) for basic study of test of homoscedasticity). But it does not ensure anything about the higher order moments. In our case, we want to detect not only the conditional error distributions with constant variance but also the varying higher order moments. Einmahl et al. (2008) already worked on this testing problem of homoscedasticity with varying third order moment using second difference of Y's. Motivated by the similar spirit, we tackle the same situation with third difference of $Y_{(i)}$'s. In order to detect ϵ_i's varying fourth moment we take a linear combination of the form $ay_{(i-2)} + by_{(i-1)} + cy_{(i)} + dy_{(i+1)}$; $a, b, c, d \in \mathbb{R}$. We would like to maximize the fourth order moment of this linear combination keeping variance fixed and third moment zero. Hence, the problem boils down to maximizing the forth order moment subject to three constraints— $a + b + c + d = 0$, $a^2 + b^2 + c^2 + d^2 = k$(a constant) and $a^3 + b^3 + c^3 + d^3 = 0$. For the second constraint, without loss of generality one might take $k = 1$. The constraint, evolved due to third moment (which might be either positive or negative), is also assumed as 0 for the sake of computational ease. The technique of discrete optimization leads us to obtain the values of a, b, c, d as $a = 1, b = -3, c = 3$ and $d = -1$. Thus the linear combination $y_{(i-2)} - 3y_{(i-1)} + 3y_{(i)} - y_{(i+1)}$ has the highest conditional fourth order moment among all possible choice of a, b, c and d.

So, our proposed test can detect dependence between X and ϵ through the kurtosis of conditional distribution when the conditional variance of $(\epsilon | X = x)$ is fixed on all x .

3. Asymptotic Power Study Considering Contiguous Alternatives

Next, we are going to inspect the power of the proposed tests under contiguous alternatives. In order to do so, we determine the asymptotic distributions of the proposed test statistics $T_{n,1}, T_{n,2}$ and $T_{n,3}$ under the null hypothesis and contiguous alternatives. As a corner stone to study the asymptotic powers based under contiguous (or, local) alternatives (see in Hajek et al., 1999, p.249), the theory of contiguity needs a brief level of familiarity.

Suppose we have a sequence of measurable spaces $(\Omega_n, \mathcal{F}_n)$ where Ω_n is the sequence of sample spaces and $\mathcal{F}_n$ is the sequence of sigma fields generated by the classes of subsets of Ω_n under some desirable properties (see in Sokol et al., 2013, Chapter 1). Let P_n and Q_n be two probability measures defined on $(\Omega_n, \mathcal{F}_n)$. Q_n is said to be *contiguous* with respect to P_n if $P_n(A_n) \to 0 \Rightarrow Q_n(A_n) \to 0$ as $n \to \infty$, for every measurable sequence of events $A_n \in \mathcal{F}_n$. Symbolically, we denote $Q_n \triangleleft P_n$.

Le Cam proposed three lemmas which describe contiguity in terms of the asymptotic behaviour of the likelihood ratios between P_n and Q_n(e.g., see Hajek et al. (1999)). Le Cam's First Lemma is stated as follows.

Lemma 1 *Q_n is contiguous with respect to P_n if $\log \frac{Q_n}{P_n}$ asymptotically follows a Gaussian distribution with mean -$\frac{\sigma^2}{2}$ and variance σ^2 under P_n, where $\sigma > 0$ is a constant.*

Proof: See (Hajek et al., 1999, p.253), for the detailed proof.

Now for null hypothesis $H_0 : X \perp\!\!\!\perp \epsilon (\Leftrightarrow F_{X,\epsilon} = G_X H_\epsilon)$ under nonparametric model $Y = m(X) + \epsilon$ where $F_{\epsilon,x}$ is the joint distribution of (ϵ, X) and G_ϵ and H_X the marginal distributions of ϵ and X respectively, let us construct a sequence of alternatives

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

where $\gamma > 0$, $n = 1, 2, ...$, and K being a proper distribution function. It is to be noted that A_n is a sequence of sets which is changing over n along with its sigma field $\mathcal{F}_n$. So, the definition of contiguity can not directly affirm that $F_{n;X,\epsilon}$ is contiguous with respect to $F_{X,\epsilon}$.

Therefore, in Theorem 1 below, evolved from Le Cam's first lemma, we would establish that the sequence of alternatives H_n will be contiguous under some conditions. Let us assume the following conditions before stating the theorem.

Assumption 1
1. $f_{X,\epsilon}(x, e) > 0$ for all x and e, where $f_{X,\epsilon}$ is the joint probability density function of (X, ϵ).
2. $E_{Y \sim F_{X,\epsilon}} (\frac{k(Y)}{f_{X,\epsilon}(Y)} - 1)^2 < \infty$ where $k(.)$ is the proper probability density function.

Theorem 1 *Under Assumption 1, the sequence of alternatives H_n defined by (4) forms a contiguous sequence.*

Proof: Our objective is to show $F_{n;X,\epsilon} \triangleleft F_{X,\epsilon}$ where $\triangleleft$ denotes contiguity relation. Assume $f_{n;X,\epsilon}$ and $f_{X,\epsilon}$ being the joint probability density functions associated with $F_{n;X,\epsilon}$ and $F_{X,\epsilon}$ respectively. To establish contiguity between these two sequences, it is enough to show that L_n, the logarithm of the likelihood ratio of $f_{n;X,\epsilon}$ and $f_{X,\epsilon}$, is asymptotically normal with mean $-\frac{\sigma^2}{2}$ and variance σ^2, where σ^2 is a positive constant. Now moment generating function of $V_{n;X,\epsilon} = \log \frac{f_{n;X,\epsilon}}{f_{X,\epsilon}}$ can be proved to be approximately equal to $\exp((-\frac{1}{2}\frac{\gamma^2}{n}M_2)t + (\frac{\gamma^2}{n}M_2)\frac{t^2}{2})$ (detailed derivation would be found in Appendix II).

Clearly, the above is the moment generating function of a Gaussian distribution with mean $-\frac{\gamma^2}{2n}M_2$ and variance $\frac{\gamma^2}{n}M_2$. Thus, $\log \frac{f_{n;X,\epsilon}}{f_{X,\epsilon}} \overset{asym}{\sim} N(-\frac{\gamma^2}{2n}M_2, \frac{\gamma^2}{n}M_2)$.
Therefore, by Le Cam's first lemma, $F_{n;X,\epsilon} \triangleleft F_{X,\epsilon}$.
Note that for simplicity of writing, we drop the random variable Y from the expression of the condition in subsequent places.

It is worth mentioning that $E_{f_{X,\epsilon}}(log \frac{k}{f_{X,\epsilon}})$ captures dissimilarity between two densities $f_{\epsilon,k}$ and k. The larger values of $(1 - \frac{k}{f_{X,\epsilon}})$ indicate that k and $f_{X,\epsilon}$ are dissimilar, i.e., in either way round, X and ϵ are more statistically dependent through k. Summing up, Theorem 1 substantiates that the sequence of alternatives H_n will be contiguous with respect to H_0. In order to explore the asymptotic behaviors of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ (given in (1), (2) and (3), respectively) under H_0 next we present Proposition 1 with the following assumptions.

Assumption 2
1. $X_1, ..., X_n$ are i.i.d. random variables with common distribution function H_X.

2. $Y_1, ..., Y_n$ satisfy the model $Y_i = m(X_i) + \epsilon_i$, $i = 1, ..., n$, where the unknown function m has 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 $i = 1, ..., n$.

Proposition 1 *Under Assumption 2, we have* $T_{n,i} \xrightarrow{P} 0$ *for* $i = 1, 2, 3$ *as* $n \to \infty$ *when* H_0 *is true.*

To prove the Proposition 1, we refer two important results from the article *the spacings around the order statistics* (Nagaraja et al. (2015)). The following result unravels about the limiting joint distributions of uniform spacings from G_X.

Result 1 For any fixed j, $n(X_{(k+j+1)} - X_{(k+j)}) \stackrel{d}{=} n(U_{(k+j+1)} - U_{(k+j)}) \left[\frac{G^{-1}(U_{(k+j+1)}) - G^{-1}(U_{(k+j)})}{U_{(k+j+1)} - U_{(k+j)}}\right]$ where $\stackrel{d}{=}$ means distributionally identical.

In contrast to the last one, the next result demonstrates about the convergence of $\left[\frac{G^{-1}(U_{(k+j+1)}) - G^{-1}(U_{(k+j)})}{U_{(k+j+1)} - U_{(k+j)}}\right]$ to a non-zero constant in probability.

Result 2 If G' (first derivative of G) is positive, finite and continuous, then

$$\frac{G^{-1}(U_{(k+j+1)}) - G^{-1}(U_{(k+j)})}{U_{(k+j+1)} - U_{(k+j)}} \xrightarrow{a.s} \frac{1}{G'(x_p)} \equiv \frac{1}{g(x_p)}$$

Here $U_1, ..., U_n$ are i.i.d. U(0,1) random variables, and $U_{(1)}, ..., U_{(n)}$ are the corresponding order statistics. The factor $n(U_{(k+j+1)} - U_{(k+j)})$ converges to 0 almost surely due to (Smirnov , 1944, Theorem 3) , stating that a central order statistic, $X_{(k)}$, $1 \leq k \leq n$ (if $\frac{k}{n} \to p$ as $n \to \infty$, then $X_{(k)}$ is called a central order statistic) converges almost surely to x_p, i.e. $X_{(k)} \xrightarrow{a.s} x_p$ if x_p is an unique solution of $G(x) = p$, where $X_{(i)}$ is the i^{th} order statistic of the random variables X_1,...,X_n. Using these two results, the proof of Proposition 1 could be shaped up. The proof of this proposition is a lengthy one. (see Appendix II for the elaborate proof).

The main overtone of Proposition 1 is that it can be devised in checking whether the evidence obtained from the data favours H_0. Moreover, in order to carry out the tests based on $T_{n,i}$, $i = 1, 2, 3$, one needs to know approximated distributions of $T_{n,i}$'s. Clearly, $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are 2-dependent U-statistics. In addition to it, one needs to know the order of degeneracy of each $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ for deriving the asymptotic distributions. To incite readers' interest, let us recapitulate the definition of U-statistic and its order of degeneracy in a nutshell.

Definition 1 For a given data set $\chi = \{X_1, ..., X_n\}$, $U_n = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < ... < i_m} k(x_{i_1}, ..., x_{i_n})$ is said to be a U-statistic of order m with kernel $k(\cdot)$ having the order of degeneracy l if

$$E_{X_{l+1},...,X_m} k(x_1, ..., x_l, X_{l+1}, ..., X_m) = 0, \tag{5}$$

for all $x_1, ..., x_l$.

Moreover, the collection of the random variables $\chi^* = \{X_1, ..., X_n\}$ will be called m-dependent random variables if X_b and X_a are independent for all $b - a > m(a, b = 1, ..., n; m \geq 1)$ and the corresponding U-statistic is said to be the m-dependent U-statistic. From an intuitive hunch, it is clear that $T_{n,1}$ has order of degeneracy 0. Also, the following lemma affirms that the asymptotic order of degeneracy for both $T_{n,2}$ and $T_{n,3}$ is 1.

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

Proof: Recalling the proof of Proposition 1, and arguing in a similar way, we have $E[T_{n,2}|X_{(i)} = x, Y^*_{(i)} = y] - \frac{1}{\binom{n}{4}} \sum_{1 \leq i < j < k < l \leq n} E[\{a(x, X_{(j)}, X_{(k)}, X_{(l)})\} E\{a(\epsilon, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)})\}] \longrightarrow 0$, where

$\epsilon^*_{(p)} = \epsilon_{(p-2)} - 3\epsilon_{(p-1)} + 3\epsilon_{(p)} - \epsilon_{(p+1)}$ and ϵ is a fixed value of $\epsilon^*_{(i)}$. Note that, $E[a(\epsilon, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)})] = 0 \; \forall \; \epsilon$ unless $j, k, l \in A^*$, where $A^* = \{(j,k,l) : \mathcal{P}\{j-2, j-1, j, j+1\} = \mathcal{P}\{k-2, k-1, k, k+1\} = \mathcal{P}\{l-2, l-1, l, l+1\}\}$, $\mathcal{P}$ being the class of all permutations. The construction of the set A^* justifies that the number of elements in A^* is finite and independent of n. Hence, $E[T_{n,2} | X_{(i)} = x, Y^*_{(i)} = y] \longrightarrow 0$ as $n \to \infty \; \forall \; x, y$. Since $T_{n,3}$ is based on sign function, reasoning in a similar way, one can show that $E[T_{n,3} | X_{(i)} = x, Y^*_{(i)} = y] \longrightarrow 0$ as $n \to \infty \; \forall \; x, y$.

Lemma 2 marks that $T_{n,2} = O_p(\frac{1}{n})$ and $T_{n,3} = O_p(\frac{1}{n})$, which can be established through the asymptotic theory of 2-dependent U-statistic, having order of degeneracy 1 whereas $T_{n,1} = O_p(1/\sqrt{n})$ since it has the order of degeneracy 0 (Lee (1990)). Next, in Theorem 2, 3 and 4 we would explore the asymptotic distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ under a sequence of contiguous alternative H_n.

Theorem 2 *If Assumptions 1 and 2 hold 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 μ_1 and variance σ^2, where*

$$\mu_1 = 2\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[2\int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{d}G_X(u)\,\mathrm{d}H^*(v) + 2\int_{x}^{\infty}\int_{y}^{\infty} \mathrm{d}G_X(u)\,\mathrm{d}H^*(v) - 1 \right] \mathrm{d}K(x,y)$$

$$\sigma_1^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[2\int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{d}G_X(u)\,\mathrm{d}H^*(v) + 2\int_{x}^{\infty}\int_{y}^{\infty} \mathrm{d}G_X(u)\,\mathrm{d}H^*(v) - 1 \right]^2 \mathrm{d}K(x,y)$$

Here, H^* is the distribution function of $(\epsilon_1 - 3\epsilon_2 + 3\epsilon_3 - \epsilon_4)$.

Proof: To prove this theorem, we take cue to *Le Cam's Third Lemma* stated as below.

Lemma 3 *Let $\{X_n\} \in \mathbb{R}^d$ be a sequence of random vectors. Then, the sequence of measures Q_n is contiguous with respect to the sequence of measures P_n if $(X_n, \log \frac{dQ_n}{dP_n})$ converges weakly under P_n, to a random vector in $\mathbb{R}^{d+1}$ associated with $(d+1)$-dimensional normal distribution with the location parameter μ_* and the dispersion parameter Σ_*, where*

$$\mu_* = \begin{bmatrix} \mu \\ -\frac{\sigma^2}{2} \end{bmatrix} \text{ and } \Sigma_* = \begin{bmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{bmatrix}$$

.

Then, $\{X_n\}$ converges weakly to a random vector in $\mathbb{R}^d$ which follows d-dimensional normal distribution with the location parameter $\mu + \tau$ and the dispersion parameter Σ under Q_n.

For the proof of this lemma, readers may check on (Van Der Vaart , 1998, page 90). Using this lemma, the proof of the asymptotic distribution of $T_{n,1}$ can be settled down as follows. Define

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

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

We have already mentioned that $T_{n,1}$ is a 2-dependent U statistic with the order of degeneracy 1. By Lemma 2 and by the expansion of L_n, the joint distribution of $(\sqrt{n}\{T_{n,1} - E(T_{n,1})\}, L_n/\sqrt{n})$ is asymptotically bivariate normal. The asymptotic covariance between $\sqrt{n}\{T_{n,1} - E(T_{n,1})\}$ and L_n is

$\frac{2\gamma}{\sqrt{n}} E_{f_{X,\epsilon}}[\sum_{i=1}^{n} E(sign\{(X - X_{(i)})(Y - Y^*_{(i)})\}|X, Y) \times \{\frac{k(z_i)}{f_{X,\epsilon}(z_i)} - 1\}]$
$= 2\gamma E_k E[sign\{(X - X_1)(Y - (\epsilon_1 - 3\epsilon_2 + 3\epsilon_3 - \epsilon_4))\}|X, Y]$
$(\because 2E_{f_{X,\epsilon}} E[sign\{(X - X_{(i)})(Y - Y^*_{(i)})\}|X, Y] = 0, \max_{i\in\{1,...,n\}} |x_{(i)} - x_{(i-1)}| = o_p(1).)$
$= 2\gamma \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [2 \int_{-\infty}^{x} \int_{-\infty}^{y} \mathrm{d}G_X(u)\, \mathrm{d}H^*(v) + 2 \int_{x}^{\infty} \int_{y}^{\infty} \mathrm{d}G_X(u)\, \mathrm{d}H^*(v) - 1]\, \mathrm{d}K(x, y).$

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

$$\mu_1 = 2\gamma \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [2 \int_{-\infty}^{x} \int_{-\infty}^{y} \mathrm{d}G_X(u)\, \mathrm{d}H^*(v) + 2 \int_{x}^{\infty} \int_{y}^{\infty} \mathrm{d}G_X(u)\, \mathrm{d}H^*(v) - 1]\, \mathrm{d}K(x, y)$$

and variance

$$\sigma_1^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [2 \int_{-\infty}^{x} \int_{-\infty}^{y} \mathrm{d}G_X(u)\, \mathrm{d}H^*(v) + 2 \int_{x}^{\infty} \int_{y}^{\infty} \mathrm{d}G_X(u)\, \mathrm{d}H^*(v) - 1]^2\, \mathrm{d}H_\epsilon(y)\, \mathrm{d}G_X(x).$$

Hence, the proof is complete.

The next theorem about the asymptotic distribution of $T_{n,2}$ under H_n can be designed similarly.

Theorem 3 *Under the Assumptions 1 and 2, for 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 the i.i.d. $N(0, 1)$ and λ_i are the eigenvalues associated with the function $l(x, y)$, defined as*
$l(x, y) = E[sign(|X_{(1)} - X_{(2)}| + |X_{(3)} - X_{(4)}| - |X_{(1)} - X_{(3)}| - |X_{(2)} - X_{(4)}|) \times sign(|Y^*_{(1)} - Y^*_{(2)}| + |Y^*_{(3)} - Y^*_{(4)}| - |Y^*_{(1)} - X^*_{(3)}| - |Y^*_{(2)} - Y^*_{(4)}|)|X_{(1)} = x, Y^*_{(1)} = y]$.

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,X_4; $Y^*_{(i)} = Y_{(i+1)} - 3Y_{(i)} + 3Y_{(i-1)} - Y_{(i-2)}$, $i = 3, 4, ..., (n - 1)$, where $Y_{(i)}$ is the Y-value corresponding to $X_{(i)}$, (x, y) is the realized value of $(X_{(1)}, Y_{(1)})$, and

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

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

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

for all (x, y). Proof of this theorem is due to (Gregory , 1977, Theorem 2.1). The detailed proof of this theorem is included in Appendix II.

Theorem 4 *Under Assumption 1 and Assumption 2 for 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 the i.i.d. $N(0, 1)$ and λ_i are the eigenvalues associated with the function $l^*(x, y)$, defined as*
$l^*(x, y) = E[sign(|X_{(1)} - X_{(2)}| + |X_{(3)} - X_{(4)}| - |X_{(1)} - X_{(3)}| - |X_{(2)} - X_{(4)}|) \times sign(|Y^*_{(1)} - Y^*_{(2)}| + |Y^*_{(3)} - Y^*_{(4)}| - |Y^*_{(1)} - X^*_{(3)}| - |Y^*_{(2)} - Y^*_{(4)}|)|X_{(1)} = x, Y^*_{(1)} = y]$.

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,X_4; $Y^*_{(i)} = Y_{(i+1)} - 3Y_{(i)} + 3Y_{(i-1)} - Y_{(i-2)}$,

$i = 3, 4, ..., (n-1)$, where $Y_{(i)}$ is the Y-value corresponding to $X_{(i)}$, (x, y) is the realized value of $(X_{(1)}, Y_{(1)})$, and

$$a_i^* = \gamma \int (\frac{k}{f} - 1) g_i^*(x) g_i^*(y) f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

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

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

for all (x, y). Proof of this theorem is shifted to Appendix II section.

Theorems 3 and 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 follows from the assertions in Theorems 3 and 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 beat the problem of infinitely many eigenvalues one can 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 kernel matrix.

4. Simulation Study and Real Data Analysis

In this section we investigate the performance of the tests based on $T_{n,1}, T_{n,2}$ and $T_{n,3}$ for finite sample simulation study and also for a real life data structure. In section 4.1, we execute a simulation study on a couple of distributions (normal and non-normal) to compare among the finite sample power of these three association based statistic. In contrast to section 4.1, section 4.2 deals with a real data analysis. The entire analysis is carried out through R programming. The related R codes are available on request with the corresponding author.

4.1. Finite sample simulation study

In this section, sample size is deliberately chosen large(say for example; 100/150/200) since we are interested in power based on asymptotic distribution (readers might be interested to study the power for small sample size but that would only lose the flavour of asymptotic theory).

The overall fixture of experiment is analogous to that described in Einmahl et al. (2008)'s, except the choice of regression function $m(x)$. Instead of the quadratic regression function, here we consider cubic equation just to have a parity with the test statistics which are constructed on third difference of Y^*. Also, along with the power of proposed test statistics $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$, the same due to Einmahl et al. (2008) tests are added herewith as a prospective tool of comparative study. Einmahl et al. (2008) considered three statistics rooted from Kolmogorov-Smirnov's statistic (KS), Cramer-von Mises statistic (CM) and Anderson-Darling statistic (AD). For the sake of brevity, we report the result of first two of them only here. Let us first define $F_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} I\{x_{(i)} \le x, y_{(i)} \le y\}$ where $I(.)$ being the indicator function. The forms of KS and CM statistics under Einmahl et al. (2008)'s paper are

$$\begin{aligned} T_{n,KS} &= \sqrt{n}\, sup_{x,y} |F_n(x, y) - \hat{F}_X(x) \hat{G}_Y(y)| \\ T_{n,CM} &= n \int \int (F_n(x, y) - \hat{F}_X(x) \hat{G}_Y(y))^2\, dF(\hat{x}) dG(\hat{y}) \end{aligned}$$

where $\hat{F}_X(x) = F_n(x, \infty)$ and $G(\hat{y}) = F_n(\infty, y)$.

The simulation study begins with the consideration that the covariate X follows $U(0,1)$ and $m(x) = x - .5x^2 - .9x^3$. The null model corresponds to a normal error term with zero mean and standard deviation equal to 0.1, i.e. $\epsilon \sim N(0, .1^2)$. Under alternative hypothesis we consider two cases of following conditional distributions of ϵ given the value of $X = x$.

$$H_{1,A} : (\epsilon/X = x) \sim N(0, \frac{1+ax}{100}),\ a > 0\ (a \text{ is variance-controlling parameter})$$
$$H_{1,B} : (\epsilon/X = x) \overset{\mathcal{D}}{=} \frac{W_x - r_x}{10\sqrt{2r_x}}, \text{ where } W_x \sim {\chi^2}_{r_x}, r_x = \frac{1}{bx}$$

Along with $T_{n,1}, T_{n,2}, T_{n,3}$ for every individual value of x and y, based on the proposed regression function, we compute $T_{n,KS}$ and $T_{n,CM}$ as well using numerical integration. The computations of power due to these five competitors, viz, $T_{n,1}, T_{n,2}$, $T_{n,3}$, $T_{n,KS}$ and $T_{n,CM}$ are executed by R3.5.3.

The stepwise algorithms of performing the asymptotic test under above-mentioned null and alternative hypothesis are mentioned below for better clarity.

1. Under H_0 we first generate a data $(x_1, y_1), \cdots, (x_n, y_n)$ from the regression model where x_i's are coming from $U(0,1)$ and errors ϵ's from $N(0, .1^2)$.
2. Considering the regression function as $m(x) = x - .5x^2 - .9x^3$ and ϵ on the model $y_i = m(x_i) + \epsilon_i$ next we generate y_is.
3. We estimate the error $\hat{\epsilon_i} = y_i - \hat{m}(x_i)$ where $\hat{m}$ is well-known Nadaraya-Watson estimator (Nadaraya (1964)), using Epanechnikov kernel.
4. To have estimated errors around 0, compute centered error,i.e.,$\hat{\epsilon}_{i,cent} = \hat{\epsilon_i} - \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon_i}$.
5. Let F_n^* be the empirical distribution function of centered errors. Using the empirical distribution function of centered $\hat{\epsilon}_{i,cent}$ we generate 200 many bootstrap sets of resample of $\hat{\epsilon}_{i,cent}$ with size $n = 100$. Here bootstrap sample (B) (viz.200) is taken as par the rule of thumb (Efron , 1993, p. 50) where it is suggested that a choice of bootstrap replications (B) $\geq$ 200 would give a reasonable value of standard error of an estimate.
6. So each bootstrap replication consists 100 $\hat{\epsilon}_{i,cent}$. Using original sample of covariates $x_1, x_2, \cdots, x_n$ and regression model we now obtain 200 sets of 100 bootstrap responses $y_i^* = \hat{m}(x) + \hat{\epsilon}_{i,cent}$.
7. Next from each bootstrap sample we calculate the test statistics using their functional expressions and $(1-\alpha)$-th quantile bootstrap distribution of the test statistic is considered as the estimated critical value.
8. To generate the bivariate data $\{(x_{(i)}, y^*_{(i)}) : i = 1, \ldots, n\}$ under the alternative hypothesis where $\epsilon \sim N(0, \frac{1+ax}{100})$, $a \in \mathbb{R}$, first we store the observations on X as already chosen from $U(0,1)$ (Step 1) under H_0. Next, to obtain $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ from $N(0, \frac{1+ax_i}{100})$ for fixed quantity a, we plug-in those X observations, thereby having $N(0, \frac{1+ax_1}{100})$, $N(0, \frac{1+ax_2}{100})$, …, $N(0, \frac{1+ax_n}{100})$ respectively. For instance, when $a = 1$, we generate $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ from $N(0, \frac{1+1.x}{100})$ upon the random sample $x_1, x_2, \ldots, x_n$ of X .
9. Subsequently, we also get $m(x_1)$, $m(x_2)$, …, $m(x_n)$ using the regression function $m(x) = x - .5x^2 - .9x^3$. Hence $y_1, y_2, \ldots, y_n$ can be generated by the relation $y = m(x) + \epsilon$.
10. Repeat the steps 4, 5, 6 and 7 with this new frameworks of X and Y.
11. Calculate the test statistics and other two competitors T_{KS} and T_{CM} for this sample under each set of 200 bootstrap replications.
12. Finally we need to count the number of times the calculated test statistics exceeding the critical point and then divide it by number of bootstrap replications (200). That will be the value of the bootstrap power for $a = 1$.
13. In a similar fashion, fixing $a = 2, 3, \ldots, 10$, one can evaluate the bootstrap powers using the aforementioned steps. Then for each a, one can obtain the corresponding bootstrap power

Exactly same procedure is carried out for the second example taking different values of b, $n = 200$ and $B = 200$, thereafter calculating out power for each bx. For both of the distribution power graphs are furnished here. Since the second example deals with Chi-square distribution–a positively skewed distribution, required sample size to authenticate the asymptotic power is higher than sample size (100) taken up in example one. The powers of five statistics for various values of a and b are reported in Table 2 and Table 3 in the Appendix I.

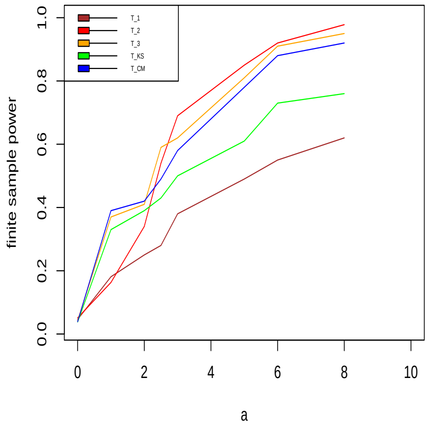

Figure 1 Power for Example 1 for varying values of a

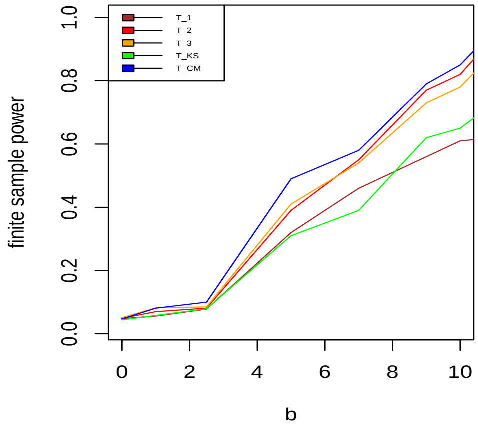

Figure 2 Power for Example 2 for varying values of b

Figure 1 delineates that for a cubic regression function the proposed statistic $T_{n,2}$ performs the best against different values of a. In fact for $n = 100$ and substantially large bootstrap sample (B=200), power of the test due to $T_{n,2}$ shoots almost close to 1 as a becomes larger. Also, $T_{n,CM}$ performs neck to neck with $T_{n,2}$. Noticeably, for smaller values of a all five statistics show up with poor power in detecting the underlying cubic relationship between y and x which is quite usual as in nonparametric regression the assumptions are far less restrictive and lower values of a indicate less dependence between ϵ and X. In fact, not the underlying distribution but it is a which actually probes the difference between null and alternative hypotheses. Loosely speaking, more the value of a clearer the discrimination between null and alternative hypothesis is. That's why for larger value of a, power increases. The same reasoning can be employed on the second example too. In contrast to the first example, based on normal distribution, the second one on Chi-square distribution (positively skewed) requires higher values of b to attain significant power. Figure 2 affirms that T_{CM} keeps hold of superiority in power with that of T_{n2} lagging just a little behind of it. Naturally, larger the values of the skewness parameter b closer will be the power to 1.

4.2. Real data analysis

In real data analysis part we consider a popular data set, Airfoil Self Noise data which is available in the University of California, Irvine at UCI Machine Learning Repository Data Set page (http://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise). This data consists of 1503 instances, each with 6 attributes obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. The bunch of attributes consist of five independent variables,viz, frequency (in hertzs), angle of attack (in degrees), chord length (in meters), free-stream velocity (in meters per second), suction side displacement thickness (in meters) and one dependent variable – scaled sound pressure level (in decibels). In our study we adopt only one regressor, viz. frequency(X) and scaled sound pressure level as regressand (Y). Scatter plot (Figure 3) of frequency vs. scaled sound pressure level shows that data is almost free of outliers.

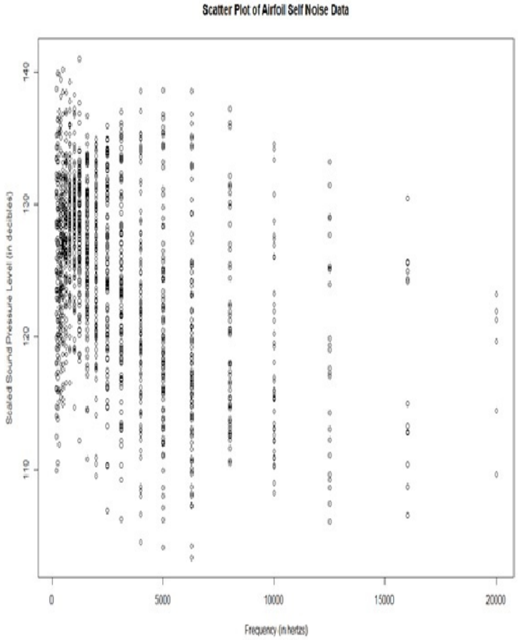

Figure 3 Scatter Plots on frequency vs scaled sound pressure

Under a set-up of nonparametric regression $Y = m(X) + \epsilon$ we would like to test if X is independent of ϵ. We first compute the value of $T_{n,i}$, $i = 1, 2, 3$ on the given data. Those $T_{n,i}$ are taken as the critical points to compute p-value. Next, we generate 200 bootstrap resamples from the data and in each resample we find an estimate of $T_{n,i}$. Let us name the estimate for kth sample as $T^k_{n,i}$. Then the estimate of p-value is $\frac{number\ of\ T^k_{n,i} > T_{n,i}}{500}$. Since $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are U statistics, the critical value obtained by the bootstrap method is asymptotically valid for the critical value done by asymptotic distribution of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ (Bickel , 1981, Sec. 3).

We report the p-values due to the statistics $T_{n,i}$ constructed on taking second difference of Y as well as the third difference of Y for a comparative perspective.

Table 1 Table for P-values in Airfoil self noise data set

Test Stats	P-value for statistic on third order difference	P-value for statistic on second order difference
T_{n1}	.420	.396
T_{n2}	.498	.521
T_{n3}	.567	.611

So having the knowledge on independence between covariate and errors under nonparametric set up enables us for further statistical exposition. In this real data analysis, due to large sample size, one can directly execute the tests based on the asymptotic distributions $T_{n,1}$,$T_{n,2}$ and $T_{n,3}$.

5. Conclusion

This article delivers a method of testing the independence between regressor and the error under nonparametric regression model $Y = m(X) + \epsilon$. The proposed test statistics are based on difference of ordered covariate and third difference on induced ordered response. Among these three, first two are the functions of the rank or the positions of the observations whereas third one is distance based. Quite intuitively, the first two statistics will be more robust against the presence of extreme observations while the third one succumbs to the presence of outliers. To fathom on robustness of these statistics one can derive the influence curve or break down points of $T_{n,1}, T_{n,2}$ and $T_{n,3}$ against outliers.

An intrinsic idea can be inculcated on higher order differences of $Y_{(i)}$ in test statistics construction. 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 expectation of the different power of the order statistics. The optimal choice of the order of the moment and the investigation on the new versions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ could be a potential concern for future research. Moreover, the proposed test might also be auditioned as a tool of measuring independence in some hard core real data analysis too. Besides, in case of more than one regressor, choice of association based test statistic regarding test of independence might be of prospective interest for future study.

Acknowledgments

Both of the authors are indebted to anonymous referee and the Editor for pointing out few prime issues, corrections to which lead to much improvement of the draft. The corresponding author is financially supported by Core Research Project Grant (EMR/2017/005421) from Science and Engineering Research Board, Department of Science and Technology, Govt. of India.

References

Bergsma W, Dassios A. A consistent test of independence based on a sign covariance related to Kendall's tau. Bernoulli. 2014; 20: 1006-1028.

Bickel PJ, Freedman DA. Some asymptotic theory for the bootstrap. Ann Stat. 1981; 9: 1196-1217.

Bradley RC. Basic Properties of strong mixing conditions. A survey and some open questions. Probab Surv. 2005; 2: 107-144.

Dhar SS, Dassios A. and Bergsma W. Testing Independence of Covariates and Errors in Nonparametric regression. Scand J Stat. 2018; 45: 421-443.

Efron B., Tibshirani RJ. An introduction to the bootstrap (1st Edition). Chapman & Hall/ CRC; 1993.

Einmahl JHJ, Van Keilegom I. Tests for independence in nonparametric regression. Stat Sin. 2008; 18: 601-615.

Goldfeld, SM and Quandt RE. Some tests for homoscedasticity. J Am Stat. 1965; 60: 539-547.

Gregory GG. Large sample theory for U statistics and test of fit. Ann Stat. 1977; 5(1), 110-123.

Hajek J, Sidak Z. and Sen PK. Theory of Rank Tests. Academic Press; 1999.

Hosking, JRM. L-moments: analysis and estimation of distributions using linear combinations of order statistics. J Royal Stat Soc B. 1990; 52: 105-124.

Lee AJ. U-Statistics: Theory and Practice. 1990: Marcel Dekker: New York and Basel.

Lehmann EL, Romano JP. Testing Statistical Hypotheses. 2005; Springer text books in Statistics, Springer.

Nadaraya EA. On non-parametric estimates of density functions and regression curves. Stat Sin. 1964; 15: 635-644.

Nagaraja HN, Bharath K. and Zhang F. Spacings around an order statistic. Ann Inst Stat Math. 2015; 67: 515-540.

Neumeyer, N. Testing independence in nonparametric regression. J Multivar Anal. 2009; 100 : 1551-1566.

Smirnov NV. Approximate laws of distribution of random variables from empirical data. 1944; Uspekhi Mat. Nauk; 10: 179206.
Sokol A and Røn-Nielson A. Lecture notes-Advanced Probability. 2013; Available from:http://web.math.ku.dk/noter/filer/vidsand12.pdf.
Van Der Vaart AW. Asymptotic Statistics. Cambridge University Press; 1998.

Appendix I

Table 2 Finite sample power of different tests over different values of a at 5% level of significance for Example 1

n	a	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	T_{KS}	T_{CM}
100	0	.046	.051	.043	.038	.039
	1	.181	.163	.37	.33	.39
	2	.25	.34	.41	.39	.42
	2.5	.28	.54	.59	.43	.49
	3	.38	.69	.62	.50	.58
	5	.49	.85	.81	.61	.78
	6	.55	.92	.91	.73	.88
	8	.62	.98	.95	.76	.92

Table 3 Finite sample power of different tests over different values of b at 5% level of significance for Example 1

n	b	$T_{n,1}$	$T_{n,2}$	$T_{n,3}$	T_{KS}	T_{CM}
200	0	.048	.049	.051	.044	.0.047
	1	.056	.08	.082	.058	.39
	2.5	.08	.081	.085	.079	.42
	5	.32	.39	.41	.31	.49
	7	.46	.55	.54	.39	.58
	9	.6	.77	.73	.62	.78
	10	.62	.82	.78	.65	.88
	11	.64	.94	.89	.73	.96

Appendix II

Proof of Theorem 1: Let's define

$$V_{n;X,\epsilon} = \log \frac{f_{n;X,\epsilon}}{f_{X,\epsilon}}$$

We have to prove that $V_{n;X,\epsilon} \overset{a}{\sim} N(.,.)$ with some parameters. Let the moment generating function of $V_{n;X,\epsilon}$ is $M_{V_{n;X,\epsilon}}(t)$. Now,

$M_{V_{n;X,\epsilon}}(t) = E[\exp(\log(\frac{f_{n;X,\epsilon}}{f_{X,\epsilon}})t)]$

$=E[\frac{f_{n;X,\epsilon}}{f_{X,\epsilon}}]^t$

$=E[\frac{(1-\frac{\gamma}{\sqrt{n}})f_{X,\epsilon}+\frac{\gamma}{\sqrt{n}}k}{f_{X,\epsilon}}]^t$

$=1 - \frac{t\gamma}{\sqrt{n}}E(1-\frac{k}{f_{X,\epsilon}}) + \frac{t(t-1)}{2}\frac{\gamma^2}{n}E(1-\frac{k}{f_{X,\epsilon}})^2 + o(\frac{1}{n})$

$=1 + \frac{t(t-1)}{2}\frac{\gamma^2}{n}M_2 + o(\frac{1}{n})$ where $M_2 = E(1-\frac{k}{f_{X,\epsilon}})^2$ being the mean square contingency of $f_{\epsilon,k}$ and k.

=exp(log(1 + $\frac{t(t-1)}{2}\frac{\gamma^2}{n}M_2 + o(\frac{1}{n})$)) ignoring the terms involving higher order to n^{-1}
$\approx \exp(\log(1 + \frac{t(t-1)}{2}\frac{\gamma^2}{n}M_2))$
=exp($\frac{t(t-1)}{2}\frac{\gamma^2}{n}M_2 - \frac{1}{2}(\frac{t(t-1)}{2}\frac{\gamma^2}{n}M_2)^2 + ...$)
=exp($\frac{t(t-1)}{2}\frac{\gamma^2}{n}M_2 + o(\frac{1}{n})$)
$\approx \exp((-\frac{1}{2}\frac{\gamma^2}{n}M_2)t + (\frac{\gamma^2}{n}M_2)\frac{t^2}{2})$ ($\because E(1 - \frac{k}{f_{X,\epsilon}}) = 0$ and $E(1 - \frac{k}{f_{X,\epsilon}})^2 < \infty$). Note that for simplicity of writing, we drop the random variable Y from the expression of the condition in subsequent places.

Proof of Proposition 1:

Let us start with $T_{n,1}$. We would like to show that as $n \to \infty$, $T_{n,1}$ approaches towards 0. Lets define A_{ij} for all i, j as $A_{ij} = sign(X_{(i)} - X_{(j)})(Y^*_{(i)} - Y^*_{(j)})$.
Then

$$T_{n,1} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} A_{ij}.$$

Now, all possible values taken by A_{ij} are $-1, 0, +1$. Suppose a_{ij} be the value of A_{ij}. Then, a_{ij}'s are interpreted as

$$a_{ij} = \begin{cases} 1, & \text{if } (X_{(i)} - X_{(j)})(Y^*_{(i)} - Y^*_{(j)}) > 0 \\ -1, & \text{if } (X_{(i)} - X_{(j)})(Y^*_{(i)} - Y^*_{(j)}) < 0 \\ 0, & \text{if } (X_{(i)} - X_{(j)})(Y^*_{(i)} - Y^*_{(j)}) = 0. \end{cases}$$

Define two probabilities as follows.
$p_c = P[(X_{(i)} - X_{(j)})(Y^*_{(i)} - Y^*_{(j)}) > 0]$ and $p_d = P[(X_{(i)} - X_{(j)})(Y^*_{(i)} - Y^*_{(j)}) < 0]$.
Then $H_0 : p_c = p_d \Leftrightarrow \tau' = 0$, where τ' is *Kendall's Tau* based on the bivariate random sample from the joint distribution of X and Y, defined as $\tau^{'} = p_c - p_d$. The marginal probability distribution of A_{ij} is given by

$$f_{A_{ij}}(a_{ij}) = \begin{cases} p_c, & \text{if } a_{ij} = 1 \\ p_d, & \text{if } a_{ij} = -1 \\ (1 - p_c - p_d), & \text{if } a_{ij} = 0 \end{cases}$$

$\therefore E(A_{ij}) = 1 \times p_c + (-1) \times p_d + 0 \times (1 - p_c - p_d) = p_c - p_d = \tau', \forall i < j$.
So, $\Rightarrow E(T_{n,1}) = \frac{1}{\binom{n}{2}} \times \binom{n}{2} \times \tau' = \tau'$. Under H_0, $\tau' = 0$, which implies that $E(T_{n,1}) = 0$, and $E(A^2_{ij}) = p_c + p_d \Rightarrow Var(A_{ij}) = p_c + p_d = 2p_c$. Next, the joint distribution of $(A_{ij}, A_{ik}), j \ne k$ is given by

$$f_{A_{ij},A_{ik}}(a_{ij}, a_{ik}) = \begin{cases} p_{cc}, & \text{if } a_{ij} = a_{ik} = 1 \\ p_{dd}, & \text{if } a_{ij} = a_{ik} = -1 \\ p_{cd}, & \text{if } a_{ij} = 1, a_{ik} = -1 \text{ or } a_{ij} = -1, a_{ik} = 1 \\ (1 - p_{cc} - p_{dd} - 2p_{cd}), & \text{if } a_{ij} = 0, a_{ik} = -1, 0, 1 \text{ or } a_{ik} = 0, a_{ij} = -1, 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

for all $i < j$, $i < k$, $j \ne k$, $i = 1, ..., n$ and some quantities p_{cc}, p_{dd} and p_{cd} such that $0 \le p_{cc}, p_{dd}, p_{cd} \le 1$.

$$\begin{aligned} E(A_{ij}A_{ik}) &= \sum_{\substack{i=1 \\ 1 \le i < j \ne k \le n}}^{n} \sum_{j,k=1}^{n} a_{ij}a_{ik}P(A_{ij} = a_{ij}, A_{ik} = a_{ik}) \\ &= p_{cc} + p_{dd} - 2p_{cd} \\ &\Rightarrow cov(A_{ij}, A_{ik}) = E(A_{ij}A_{ik}) - E(A_{ij})E(A_{ik}) = p_{cc} + p_{dd} - 2p_{cd} \text{ under } H_0. \end{aligned}$$

Using the marginal distribution of A_{ij} and the joint distribution of (A_{ij}, A_{ik})

$$\begin{aligned}
E(A_{ij}A_{ik}) &= \frac{1}{\binom{n}{2}^2}[\sum_{\substack{i=1\\1\le i<j\le n}}^{n}\sum_{j=1}^{n} Var(A_{ij}) + \sum_{\substack{i=1\\i\neq h}}^{n}\sum_{j=1}^{n}\sum_{\substack{h=1\\j\neq k}}^{n}\sum_{k=1}^{n} cov(A_{ij}, A_{hk})] \\
&= \frac{1}{\binom{n}{2}^2}[\binom{n}{2}Var(A_{ij}) + \sum_{i=1}^{n-1}\sum_{\substack{j=i+1\\j\neq k}}^{n}\sum_{k=i+1}^{n} cov(A_{ij}, A_{ik}) \\
&\quad + \sum_{j=2}^{n}\sum_{\substack{i=1\\i\neq k}}^{j-1}\sum_{k=1}^{j-1} cov(A_{ij}, A_{kj}) + \sum_{j=2}^{n}\sum_{\substack{i=1\\i\neq k}}^{j-1}\sum_{k=j+1}^{n} cov(A_{ij}, A_{jk}) \\
&\quad + \sum_{i=2}^{n-1}\sum_{\substack{j=i+1\\j\neq k}}^{n}\sum_{k=1}^{n} cov(A_{ij}, A_{ki})]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\binom{n}{2}^2}[\binom{n}{2}Var(A_{ij}) + 2\binom{n}{3}cov(A_{ij}, A_{ik}) + 2\binom{n}{3}cov(A_{ij}, A_{ik}) \\
&\quad + \binom{n}{3}cov(A_{ij}, A_{ik}) + \binom{n}{3}cov(A_{ij}, A_{ik})] \quad [since\ A'_{ij}s\ are\ identical, \\
&\quad cov(A_{ij}, A_{ik}) = cov(A_{ij}, A_{kj}) = cov(A_{ij}, A_{jk}) = cov(A_{ij}, A_{ki})] \\
&= \frac{Var(A_{ij})}{\binom{n}{2}} + 6 \times \frac{\binom{n}{3}}{\binom{n}{2}^2} \times cov(A_{ij}, A_{ik}) \\
&= \frac{2p_c}{\binom{n}{2}} + 4 \times \frac{n(n-1)(n-2)}{n^2(n-1)^2} \times (p_{cc} + p_{dd} - 2p_{cd}) \\
&= \frac{4p_c}{n(n-1)} + 4 \times \frac{(n-2)}{n(n-1)} \times (p_{cc} + p_{dd} - 2p_{cd}) \longrightarrow 0\ as\ n \to \infty
\end{aligned}$$

Therefore, $T_{n,1} \xrightarrow{P} 0$. Next we prove $T_{n,2} \xrightarrow{P} 0$ as $n \to \infty$. Note that, $Y^*_{(p)}$ is the third order difference for induced order variable $Y_{(p)}$.

$$\begin{aligned}
Y^*_{(p)} &= Y_{(p-2)} - 3Y_{(p-1)} + 3Y_{(p)} - Y_{(p+1)} \\
&= (Y_{(p-2)} - Y_{(p-1)}) - 2(Y_{(p-1)} - Y_{(p)}) + (Y_{(p)} - Y_{(p+1)}),\ p = i, j, k, l. \qquad (6)
\end{aligned}$$

We assumed that m is a smooth function, hence $m^{(r)}(x)$ exists for all $r = 1, 2, ...,$ and $\sup_x m^{(r)}(x) < \infty$. Also, by mean value theorem (since $m(.)$ is differentiable)

$$\begin{aligned}
Y_{(p-2)} - Y_{(p-1)} &= m(X_{(p-2)}) + \epsilon_{(p-2)} - m(X_{(p-1)}) + \epsilon_{(p-1)} \\
&= m'(\xi_{p-2,p-1})(X_{(p-2)} - X_{(p-1)}) + (\epsilon_{(p-2)} - \epsilon_{(p-1)}), \\
&\quad \text{where } \xi_{p-2,p-1} \text{ is a point between } X_{(p-2)} \text{ and } X_{(p-1)}. \\
&= m'(\xi_{p-2,p-1})o_p(1) + (\epsilon_{(p-2)} - \epsilon_{(p-1)})
\end{aligned}$$

Similarly,

$Y_{(p-1)} - Y_{(p)} = m'(\xi_{p-1,p})o_p(1) + (\epsilon_{(p-1)} - \epsilon_{(p)})$, where $\xi_{p-1,p}$is a point between $X_{(p-1)}$ and $X_{(p)}$.

Then, equation (3.2) boils down to
$Y^*_{(p)} = (m'(\xi_{p-2,p-1}) - 2m'(\xi_{p-1,p}) + m'(\xi_{p,p+1}))o_p(1) + \epsilon^*_{(p)}$
$\Rightarrow Y^*_{(p)} - \epsilon^*_{(p)} \xrightarrow{P} 0$ for all $p = i, j, k, l$. So

$$T_{n,2} \xrightarrow{P} \frac{1}{\binom{n}{4}} \sum_{1\le i<j<k<l\le n} a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) a(\epsilon^*_{(i)}, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)}).$$

Again, $E[\frac{1}{\binom{n}{4}} \sum\limits_{1\le i<j<k<l\le n} a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) a(\epsilon^*_{(i)}, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)})]$
$=\frac{1}{\binom{n}{4}} \sum\limits_{1\le i<j<k<l\le n} E[a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) a(\epsilon^*_{(i)}, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)})]$
$=\frac{1}{\binom{n}{4}} \sum\limits_{1\le i<j<k<l\le n} E[a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)})] \times E[a(\epsilon^*_{(i)}, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)})]$ (since X is independent of ϵ under H_0). Note that, $\epsilon_{(1)}, ..., \epsilon_{(n)}$ are i.i.d. random variables from G_ϵ. Hence,

$$E[a(\epsilon^*_{(i)}, \epsilon^*_{(j)}, \epsilon^*_{(k)}, \epsilon^*_{(l)})] = 0, \text{ unless } (i, j, k, l) \in \mathcal{A},$$

where $\mathcal{A} = \{(i, j, k, l) : \mathcal{P}(i-2, i-1, i, i+1) = \mathcal{P}(j-2, j-1, j, j+1) = \mathcal{P}(k-2, k-1, k, k+1) = \mathcal{P}(l-2, l-1, l, l+1)\}$, $\mathcal{P}$ denoting the class of all permutations. It follows from the construction of the set $\mathcal{A}$ that the number of elements in it is finite and independent of n, and consequently $\frac{card(\mathcal{A})}{n} \longrightarrow 0$ as $n \to \infty$, where $card(\mathcal{A})$ denotes the number of elements in $\mathcal{A}$. Therefore we conclude that,
$E[\frac{1}{\binom{n}{4}} \sum\limits_{1\le i<j<k<l\le n} a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) a(Y^*_{(i)}, Y^*_{(j)}, Y^*_{(k)}, Y^*_{(l)})] \longrightarrow 0$, as $n \to \infty$. Arguing in similar way we conclude that,
$E[\frac{1}{\binom{n}{4}} \sum\limits_{1\le i<j<k<l\le n} a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) a(Y^*_{(i)}, Y^*_{(j)}, Y^*_{(k)}, Y^*_{(l)})]^2 \longrightarrow 0$, as $n \to \infty$, i.e., variance converges to 0 as n increases.
Therefore, $\frac{1}{\binom{n}{4}} \sum\limits_{1\le i<j<k<l\le n} a(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) a(Y^*_{(i)}, Y^*_{(j)}, Y^*_{(k)}, Y^*_{(l)}) \xrightarrow{P} 0 \Rightarrow T_{n,2} \xrightarrow{P} 0.$

Further, to prove $T_{n,3} \xrightarrow{P} 0$, we use Result 1, Result 2 and Lemma 1. From Result 2 and Lemma 1, and using Slutsky's Theorem, we obtain

$$\begin{aligned}\frac{G^{-1}(U_{(k+j+1)}) - G^{-1}(U_{(k+j)})}{U_{(k+j+1)} - U_{(k+j)}} &\xrightarrow{P} \frac{1}{G'(x_p)} \equiv \frac{1}{g(x_p)} \text{ and} \\ (n(U_{(k+j+1)} - U_{(k+j)})) &\xrightarrow{d} Z, \text{ where } Z \sim exp(1) \\ \Rightarrow n(X_{(k+j+1)} - X_{(k+j)}) &\xrightarrow{d} \frac{1}{g(x_p)} Z. \end{aligned} \tag{7}$$

Thus, for $1 \le i < j < k < l \le n$,

$$\begin{aligned} & h(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) \times h(Y^*_{(i)}, Y^*_{(j)}, Y^*_{(k)}, Y^*_{(l)}) \\ = & (|X_{(i)} - X_{(j)}| + |X_{(k)} - X_{(l)}| - |X_{(i)} - X_{(k)}| - |X_{(j)} - X_{(l)}|) \\ & \times(|Y^*_{(i)} - Y^*_{(j)}| + |Y^*_{(k)} - Y^*_{(l)}| - |Y^*_{(i)} - Y^*_{(k)}| - |Y^*_{(j)} - Y^*_{(l)}|) \\ = & (|X_{(i)} - X_{(j)}| \times |Y^*_{(i)} - Y^*_{(j)}| + |X_{(i)} - X_{(j)}| \times |Y^*_{(k)} - Y^*_{(l)}| \\ - & |X_{(i)} - X_{(j)}| \times |Y^*_{(i)} - Y^*_{(k)}| - |X_{(i)} - X_{(j)}| \times |Y^*_{(j)} - Y^*_{(l)}|) \\ & +(|X_{(k)} - X_{(l)}| \times |Y^*_{(i)} - Y^*_{(j)}| + |X_{(k)} - X_{(l)}| \times |Y^*_{(k)} - Y^*_{(l)}| \\ & -|X_{(k)} - X_{(l)}| \times |Y^*_{(i)} - Y^*_{(k)}| - |X_{(k)} - X_{(l)}| \times |Y^*_{(j)} - Y^*_{(l)}|) \\ & -(|X_{(i)} - X_{(k)}| \times |Y^*_{(i)} - Y^*_{(j)}| + |X_{(i)} - X_{(k)}| \times |Y^*_{(k)} - Y^*_{(l)}| \\ & -|X_{(i)} - X_{(k)}| \times |Y^*_{(i)} - Y^*_{(k)}| - |X_{(i)} - X_{(k)}| \times |Y^*_{(j)} - Y^*_{(l)}|) \\ & -(|X_{(j)} - X_{(l)}| \times |Y^*_{(i)} - Y^*_{(j)}| + |X_{(j)} - X_{(l)}| \times |Y^*_{(k)} - Y^*_{(l)}| \\ & -|X_{(j)} - X_{(l)}| \times |Y^*_{(i)} - Y^*_{(k)}| - |X_{(j)} - X_{(l)}| \times |Y^*_{(j)} - Y^*_{(l)}|). \end{aligned}$$

Tacitly to say,

$$
\begin{aligned}
|X_{(i)} - X_{(j)}| &= |X_{(i)} - X_{(i+1)} + X_{(i+1)} - X_{(i+2)} + ... + X_{(j+1)} - X_{(j)}| \\
&\leq |X_{(i)} - X_{(i+1)}| + |X_{(i+1)} - X_{(i+2)}| + ... + |X_{(j-1)} - X_{(j)}|
\end{aligned}
$$

and

$$
\begin{aligned}
|Y^*_{(i)} - Y^*_{(j)}| &= |(Y_{(i-2)} - 3Y_{(i-1)} + 3Y_{(i)} - Y_{(i+1)}) - (Y_{(j-2)} - 3Y_{(j-1)} + 3Y_{(j)} - Y_{(j+1)})| \\
&= |(Y_{(i-2)} - Y_{(i-1)}) - 2(Y_{(i-1)} - Y_{(i)}) + (Y_{(i)} - Y_{(i+1)}) \\
&\quad -(Y_{(j-2)} - Y_{(j-1)}) + 2(Y_{(j-1)} - Y_{(j)}) - (Y_{(j)} - Y_{(j+1)})| \\
&\leq |Y_{(i-2)} - Y_{(i-1)}| + 2|Y_{(i-1)} - Y_{(i)}| + |Y_{(i)} - Y_{(i+1)}| \\
&\quad +|Y_{(j-2)} - Y_{(j-1)}| + 2|Y_{(j-1)} - Y_{(j)}| + |Y_{(j)} - Y_{(j+1)}| \\
&\leq |m(X_{(i-2)}) - m(X_{(i-1)})| + |\epsilon_{(i-2)} - \epsilon_{(i-1)}| + 2|m(X_{(i-1)}) - m(X_{(i)})| \\
&\quad +2|\epsilon_{(i-1)} - \epsilon_{(i)}| + |m(X_{(i)}) - m(X_{(i+1)})| + |\epsilon_{(i)} - \epsilon_{(i+1)}| \\
&\quad +|m(X_{(j-2)}) - m(X_{(j-1)})| + |\epsilon_{(j-2)} - \epsilon_{(j-1)}| + 2|m(X_{(j-1)}) - m(X_{(j)})| \\
&\quad +2|\epsilon_{(j-1)} - \epsilon_{(j)}| + |m(X_{(j)}) - m(X_{(j+1)})| + |\epsilon_{(j)} - \epsilon_{(j+1)}|
\end{aligned}
$$

m being a smooth function. Using Mean value theorem, we get $m(X_{(p)}) - m(X_{(p+1)}) = m'(\xi_{p,p+1})|X_{(p)} - X_{(p+1)}|, p = i-2, i-1, i, j-2, j-1, j$, where $\xi_{p,p+1}$ is a point between $X_{(p)}$ and $X_{(p+1)}$. Therefore,

$$
\begin{aligned}
|Y^*_{(i)} - Y^*_{(j)}| &\leq |m'(\xi_{i-2,i-1})||X_{(i-2)} - X_{(i-1)}| + |\epsilon_{(i-2)} - \epsilon_{(i-1)}| \\
&\quad +2|m'(\xi_{i-1,i})||X_{(i-1)} - X_{(i)}| + 2|\epsilon_{(i-1)} - \epsilon_{(i)}| \\
&\quad +|m'(\xi_{i,i+1})||X_{(i)} - X_{(i+1)}| + |\epsilon_{(i)} - \epsilon_{(i+1)}| \\
&\quad +|m'(\xi_{j-2,j-1})||X_{(j-2)} - X_{(j-1)}| + |\epsilon_{(j-2)} - \epsilon_{(j-1)}| \\
&\quad +2|m'(\xi_{j-1,j})||X_{(j-1)} - X_{(j)}| + 2|\epsilon_{(j-1)} - \epsilon_{(j)}| \\
&\quad +|m'(\xi_{j,j+1})||X_{(j)} - X_{(j+1)}| + |\epsilon_{(j)} - \epsilon_{(j+1)}|
\end{aligned}
$$

implying

$$
\begin{aligned}
E|Y^*_{(i)} - Y^*_{(j)}| &\leq |m'(\xi_{i-2,i-1})E|X_{(i-2)} - X_{(i-1)}| + E|\epsilon_{(i-2)} - \epsilon_{(i-1)}| \\
&\quad +2|m'(\xi_{i-1,i})|E|X_{(i-1)} - X_{(i)}| + 2E|\epsilon_{(i-1)} - \epsilon_{(i)}| \\
&\quad +|m'(\xi_{i,i+1})|E|X_{(i)} - X_{(i+1)}| + E|\epsilon_{(i)} - \epsilon_{(i+1)}| \\
&\quad +|m'(\xi_{j-2,j-1})|E|X_{(j-2)} - X_{(j-1)}| + E|\epsilon_{(j-2)} - \epsilon_{(j-1)}| \\
&\quad +2|m'(\xi_{j-1,j})|E|X_{(j-1)} - X_{(j)}| + 2E|\epsilon_{(j-1)} - \epsilon_{(j)}| \\
&\quad +|m'(\xi_{j,j+1})|E|X_{(j)} - X_{(j+1)}| + E|\epsilon_{(j)} - \epsilon_{(j+1)}|.
\end{aligned}
$$

We derive the expectation to find the asymptotic mean and variance of $T_{n,3}$. Let's start with $E|X_{(i)} - X_{(j)}||Y^*_{(i)} - Y^*_{(j)}|$. The rest of the terms will be derived in similar fashion as we handle the calculation of $E|X_{(i)} - X_{(j)}||Y^*_{(i)} - Y^*_{(j)}|$. Before derivation, we separately calculate the expectations of the terms in $h(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) \times h(Y^*_{(i)}, Y^*_{(j)}, Y^*_{(k)}, Y^*_{(l)})$ followed by $E(T_{n,3})$ and $Var(T_{n,3})$. There would be four types of terms depending on i, j in $E|X_{(i)} - X_{(j)}||Y^*_{(i)} - Y^*_{(j)}|$. Using equation (7),

1. $E|X_{(a+1)} - X_{(a)}| \; \forall a$
   $=E(X_{(a+1)} - X_{(a)})$
   $=\frac{1}{n}E(n(X_{(a+1)} - X_{(a)}))$
   $\because n(X_{(a+1)} - X_{(a)}) \xrightarrow{d} \frac{1}{g(x_p)}Z$, where $\frac{a}{n} \rightarrow p$ as $n \rightarrow \infty$, $Z \sim Exp(1)$.

$\therefore E(n(X_{(a+1)} - X_{(a)})) \longrightarrow E[\frac{1}{g(x_p)}Z] = \frac{1}{g(x_p)}E(Z) = \frac{1}{g(x_p)}$.
Then, $E[|X_{(a+1)} - X_{(a)}|] = \frac{1}{ng(x_p)} \longrightarrow 0$ as $n \to \infty$.

2. $E(X_{(a+1)} - X_{(a)})^2 \ \forall a$
$=\frac{1}{n^2}E(n(X_{(a+1)} - X_{(a)}))^2$
$\longrightarrow \frac{1}{n^2} \times \frac{1}{(g(x_p))^2} \times E(Z^2) = \frac{2}{n^2(g(x_p))^2} \longrightarrow 0$ as $n \to \infty$

3. $E(X_{(a+1)} - X_{(a)})(X_{(b+1)} - X_{(b)})$, where $a < b$
$=\frac{1}{n^2}E[n(X_{(a+1)} - X_{(a)}) \times n(X_{(b+1)} - X_{(b)})]$
$=\frac{1}{n^2}E[n(X_{(a+1)} - X_{(a)})] \times E[n(X_{(b+1)} - X_{(b)})]$ (using Lemma 1)
$\longrightarrow E(\frac{1}{g(x_p)}Z_a)E(\frac{1}{g(x_q)}Z_b)$, where $\frac{a}{n} \to p$, $\frac{b}{n} \to q$ as $n \to \infty$
$=\frac{1}{n^2 g(x_p)g(x_q)} \times E(Z_a) \times E(Z_b)$, where $Z_a, Z_b \sim Exp(1)$
$=\frac{1}{n^2 g(x_p)g(x_q)} \longrightarrow 0$ as $n \to \infty$.

4. $E_{X,\epsilon}(X_{(a+1)} - X_{(a)})|\epsilon_{(t+1)} - \epsilon_{(t)}| \ \forall t$
$=E_X(X_{(a+1)} - X_{(a)}) \times E_\epsilon|\epsilon_{(t+1)} - \epsilon_{(t)}|$ (since under null hypothesis, $X \perp\!\!\!\perp \epsilon$)
Now, $E_X(X_{(a+1)} - X_{(a)}) \longrightarrow 0$ as $n \to \infty$, and, $E_\epsilon|\epsilon_{(t+1)} - \epsilon_{(t)}| < \infty$.
$\therefore E_{X,\epsilon}(X_{(a+1)} - X_{(a)})|\epsilon_{(t+1)} - \epsilon_{(t)}| \longrightarrow 0$ as $n \to \infty$.

All the terms computed above converge to 0. As a whole it implies that $E(h(X_{(i)}, X_{(j)}, X_{(k)}, X_{(l)}) \times h(Y^*_{(i)}, Y^*_{(j)}, Y^*_{(k)}, Y^*_{(l)}))$ converge to 0. Hence, $E(T_{n,3}) \to 0$ as $n \to \infty$. In similar manner, we can prove that $E(T^2_{n,3}) \to 0$ as $n \to \infty$. Thus $T_{n,3} \xrightarrow{P} 0$.

Proof of Theorem 3

By applying Lemma 2 and Lemma 3, $n(T_{n,2} - E(T_{n,2}))$ converges 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 the iid $N(0, 1)$ random variables. Further, the sequence of densities (call it by q_n)associated with H_n is dominated by the density p_0 associated with H_0, along with Radon-Nikodym derivative $\dfrac{dq_n}{dp_0} = 1 + n^{-\frac{1}{2}}h_n$, where $h_n = \gamma(\frac{k}{f} - 1) \in L_2(p_0)$ (L^2 being the second order normed space) since $E_f(\frac{k}{f} - 1)^2 < \infty$, which is already taken care in the theorem. Hence, by Gregory (1977), $n(T_{n,2} - E(T_{n,2}))$ 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's are as defined before. Thus the proof.

Proof of Theorem 4

By Lemma 2, we conclude that $n(T_{n,3} - E(T_{n,3}))$ converges weakly to $\sum_{i=1}^{\infty} \lambda_i^*(Z_i^{*^2} - 1)$ under H_0, where λ_i^*'s are the eigenvalues corresponding to $l^*(x, y)$, and Z_i^*'s are the iid $N(0, 1)$ random variables, which is evident from Lemma 1. Arguing similarly as in the proof of the asymptotic distribution of $T_{n,2}$, we conclude that $n\{T_{n,3} - E(T_{n,3})\}$ 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^*'s are as defined before. This completes the proof.