Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 1, 2024 (New Series), pp 193–217
http://www.ssca.org.in/journal

A Rank-based Test of Independence of Covariate and Error in Nonparametric Regression with Missing Completely at Random Response Situation

Sthitadhi Das and Saran Ishika Maiti
Department of Statistics
Visva Bharati, Santiniketan, India

Received: 17 January 2023; Revised: 18 July 2023; Accepted: 22 July 2023

Abstract

In the context of nonparametric regression, statistical relationship between the covariate and the random error is a matter of interest. For a traditional nonparametric regression model $Y = g(X) + \epsilon$ where Y is the response, X the covariate, ϵ the random error and $g(\cdot)$ a suitably chosen smooth function, null hypothesis may be framed as the independence of X and ϵ against all possible alternatives citing dependence between them. It may be of further concern, whether for an incomplete data set with several missing observations, such rank based testing of independence can be performed. For example, some observations on Y are unreported whereas the covariate X has complete data. On this structure of missingness completely at random (MCAR) situation, process of rank based testing on independence between X and ϵ may be thought of. This article delineates such testing techniques, based on Kendall's τ or Bergsma's (2014) τ^* and Blum *et al.* (1961) distance based test statistics, in order to develop consistent test procedures against a sequence of contiguous alternatives. The asymptotic powers of these test statistics are further studied through the finite sample simulation study, choosing different levels of missingness percentage. Finally, a real data analysis presents a comparative testimony of those proposed test statistics.

Key words: Asymptotic power; Contiguous alternative; Distance covariance; Kendall's τ; Missing completely at random; Nonparametric regression model; Local linear smoothing.

AMS Subject Classifications: 62G08, 62G30

1. Introduction

For a quite substantial period of time in statistics literature, missing data context continues to be a live topic. The impact of missing data on quantitative research can be serious, heading to biased estimates of parameters, loss of information, increased standard errors and debilitated the generalizability of findings. Usually, most statistical processes are designed for complete data. In the presence of missing values, failing to edit the incomplete

Corresponding Author: Saran Ishika Maiti
Email: saranishika.maiti@visva-bharati.ac.in

data into "complete" one can turn the data statistically unsuitable. Particularly, statistical inference process experiences a huge toll in presence of missingness. Thus, as a default approach, one may delete those missing observations before going to conduct the necessary analysis using statistical methods. Most inevitable drawback of such listwise deletion is that a large fraction of sample might get trimmed causing severe loss to statistical power. Some articles by Anderson(1957), Wilks(1932), Afifi and Elashoff(1966), Hartley and Hocking(1971) discussed the problem of listwise deletion where each value of data set is equally likely to be missing.

In regression set up, missing scenario mostly occurs in response variable Y where some of the observations in Y are not available. The chance mechanism of this missingness may be independent of X and Y or may depend fully on the covariate, X. The first case is termed as missing completely at random (MCAR) while the second type of missingness is missing at random (MAR) (Little and Rubin (2014)). Mathematically speaking, in regression set-up, missingness can be interpreted via a triplet (X_i, Y_i, δ_i) for a set of n observations on (X, Y). At a given point X_i, the response Y_i is either observed or missing. The indicator variable δ takes the value 1 or 0 according as the value of Y is reported or not. Clearly for MCAR, Prob$[\delta = 1/X, Y] = p$ (a constant) while for MAR $Prob[\delta = 1/X, Y] = P[\delta = 1/X] = p(X)$ (a function of X). We shall proceed with an MCAR data to test the association in the context of nonparametric regression further.

Suppose in nonparametric regression model $Y = g(X) + \epsilon$ with g being the unknown regression function and ϵ the error, missingness at random occurs in Y. Instead of complete deletion of those unavailable (X, Y) observations, imputation techniques may be used where substitutes for missing values are looked for. In contrast to imputing certain global estimates such as mean/median of available Y figures, it may be worthwhile to opt for some other imputation alternatives based on nonparametric regression estimation, like local linear smoothing, kernel density estimation *etc.* (Chung *et al.* (1993), Cheng (1994)), thereafter examining the impact of missingness on their performances. One may note that downside of imputation technique is to produce underestimates of standard errors, which leads in turn to inflated test statistics.

In nonparametric regression, a fundamental assumption is homoscedasticity, *i.e.* $E(\epsilon^2/X = x) = \sigma^2 > 0$. However even for homoscedastic model, inference based on unknown regression function $g(x)$ may be unconvincing, for instance in isotonic mean/median regression model, confidence interval for the regression function at a given point will be wrong even if the homoscedasticity holds. In such cases, it is safer to assume the independence between X and ϵ. This issue of checking the independence against all possible alternatives, has been addressed in the literature by Einmahl *et al.*(2008), Neumeyer(2009), Hlavka *et al.* (2011), Dhar *et al.*(2018). Most of the test statistics proposed are distance based except the rank based test statistic by Bergsma (2014), followed by Dhar *et al.* (2018), Das *et al.*(2022) where the test statistic is constructed on the sign function of second/third order differences of neighbouring quadruplet of responses.

The present article is evolved on the adoption of such rank based test statistic to investigate the independence of ϵ and X in nonprametric regression when the data has MCAR in Y. At the first stage, the missing places are imputed by the regression estimator through Nadaraya- Watson estimation and local linear smoothing technique respectively. Thereafter, filling those unregistered Y values we try to form rank based test statistic following the road-map by Bergsma (2014). We also investigate the asymptotic theory of those test statistics under null and contiguous alternative (Lehmann and Romano, 2005).

The rest of the article is organized as follows. Section 2 describes original regression model and the transformed imputed model. Section 3 provides the methodologies to estimate the regression function $g(.)$ using various estimation techniques. In section 4, test statistics are constructed based on the newly obtained bivariate observations X and Y. The asymptotic local powers of the test statistics under contiguous alternatives are computed in Section 5. Section 6 includes a real data study. A precise conclusion is presented in section 6. Appendix 1 contains derivation of technical details while appendix 2 contains numerical results of asymptotic power study.

2. Regression setting

Let the nonparametric regression model to be considered as $Y = g(X) + \epsilon$. Consider the following imcomplete data: (X_i, Y_i, δ_i), $i = 1, 2 \cdots, n$ where $\delta = 1$ if Y_i is observed otherwise $\delta_0 = 0$ if Y_i is missing. Also, $Prob(\delta = 1/X, Y) = Prob(\delta = 1/X) = p\,(0 < p < 1)$ where p being a fixed constant, *i.e.*, missingness is MCAR type. Let there be k bivariate observations assuming missingness on Y and the remaining $(n-k)$ pairs are complete. Suppose (X_i', Y_i') denote the i-th complete observation of (X, Y), $i = 1, 2, \cdots, (n-k)$. A nonparametric sub-model can be formulated on these **complete pairs** as

$$Y' = g_1(X') + \epsilon' \tag{1}$$

with the assumptions on error ϵ' similar to the assumptions, already drawn on error ϵ of the original model, as $E(\epsilon'|X' = x') = 0 \;\forall\; x'$ and $E(\epsilon'^2|X' = x') = \sigma^2(x')$ where $\sigma^2(x') > 0$. The regression function $g_1(\cdot)$ is the **first step regression function.** Its nonparametric estimator may be treated as a naive alternative against the estimator of $g(X)$ in the original model. After deducing the estimator of $g_1(\cdot)$ as $\hat{g}_1(\cdot)$, the missing observations on Y will be filled up by $\hat{g}_1(\cdot)$ at the values of the covariate X corresponding to the missing responses. These fillers are known as **imputed responses**. Thus, by imputing the missing values of Y, the complete data set (X^*, Y^*) of size n can be re-framed as follows.

$$Y_i^* = \begin{cases} Y_i' & \text{when}\, \delta = 1 \\ \hat{g}_1(X_i), & \text{when} \;\; \delta = 0\,;\; i = 1, 2, \cdots, n \end{cases}$$

Then, the following regression model is proposed on the hence completed bivariate data (X^*, Y^*).

$$Y^* = g_2(X^*) + \epsilon^* \tag{2}$$

where X^* being the covariate and ϵ^* being the error. Finally, $g_2(X^*)$ is estimated using the conventional methods like Nadaraya-Watson (NW) estimation and local linear smoothing method respectively.

3. Estimation of regression functions

3.1. Estimation using Nadaraya-Watson method

The first step regression function $g_1(.)$ in (1) can be estimated using Nadaraya Watson (NW) estimation process at $X' = x'$ as

$$\hat{g}_1(x') = \frac{\sum_{i=1}^{n} k\left(\frac{X_i' - x'}{h}\right) Y_i'}{\sum_{i=1}^{n} k\left(\frac{X_i' - x'}{h}\right)} \tag{3}$$

where $k(\cdot)$ is the kernel density function and h is the bandwidth satisfying $h \to 0$ with $nh \to \infty$ where $n \to \infty$. A variety of kernel functions are possible to be chosen but for practical and theoretical considerations we choose a very common one, Epanichnikov kernel $k(u)$, where $k(u) = .75(1-u^2).I(|u| \leq 1)$. This parabolic shape kernel enjoys some optimality properties.

The second stage estimator of the regression function $g_2(X^*)$ in (4) is also deduced in a similar manner.

$$\hat{g}_2(x^*) = \frac{\sum_{i=1}^{n} k\left(\frac{X_i^* - x^*}{h}\right) Y_i^*}{\sum_{i=1}^{n} k\left(\frac{X_i^* - x^*}{h}\right)} \tag{4}$$

Further, proposition of some test statistics are made.

3.2. Estimation using local linear smoothing (LLS)

In addressing the same issue, another alternative approach against NW estimation can be the technique of local linear smoothing (Chu *et al.*, 1995). This method begins with the minimization of the local weighted least squares based on all bivariate observations, *i.e.* minimization of the following expression.

$$\sum_{i=1}^{n} [Y_i - r_0 - r_1(x - X_i)]^2 \, k\left(\frac{x - X_i}{h}\right) \delta_i \tag{5}$$

As per the notation stated in section 2, specifically for non missing pairs of observations (X', Y') the above expression of minimization can be re-framed as minimization of

$$\sum_{i=1}^{n-k} [Y_i' - r_0 - r_1(x' - X_i')]^2 \, k\left(\frac{x' - X_i'}{h}\right) \tag{6}$$

The minimization yields the solutions of the constants r_0 and r_1. (5) gives

$$\hat{r}_0 = \frac{\sum_{i=1}^{n} (M_2 - (x - X_i)M_1)\; k\left(\frac{x - X_i}{h}\right) \delta_i \, Y_i}{\sum_{i=1}^{n} [M_2 - (x - X_i)M_1]\; k\left(\frac{x - X_i}{h}\right) \delta_i} \tag{7}$$

where $M_l = \sum_{i=1}^{n}(x - X_i)^l \; k\left(\frac{x - X_i}{h}\right)\delta_i,\; l = 1, 2$. Clearly, for non-missing pairs of observations (X', Y') (6) would be reshaped as

$$\hat{r}_0 = \frac{\sum_{i=1}^{n-k}\left[M_2' - (x' - X_i')M_1'\right]\, k\left(\frac{x' - X_i'}{h}\right) Y_i}{\sum_{i=1}^{n-k}\left[M_2' - (x' - X_i')M_1'\right]\, k\left(\frac{x' - X_i'}{h}\right)} \tag{8}$$

where $M_l' = \sum_{i=1}^{n-k}(x' - X_i')^l \; k\left(\frac{x' - X_i'}{h}\right),\; l = 1, 2$. The least square estimate $\hat{r}_1$ of r_1 can be deduced in a similar way from (5) or (6) which is simply

$$\hat{r}_1 = \frac{\sum_{i=1}^{n}(x' - X_i')\, k\left(\frac{x - X_i}{h}\right)\delta_i\, Y_i - \hat{r}_0 M_1'}{M_2'}.$$

Next, by the first order Taylor's expansion, $g(X_i)$ can be expanded in the neighbourhood of x as

$$g(X_i) = g(x) - (x - X_i)g^{(1)}(x) \tag{9}$$

where $g^{(1)}(x)$ is the first order derivative of $g(x)$. Hence the response Y_i can be approximated as $\{g(x) - (x - X_i)g^{(1)}(x) + \epsilon_i\}$, $i = 1, \ldots, n$. Synonymously, under non missing set up Y_i' may be approximated as $\{g(x') - (x' - X_i')g^{(1)}(x') + \epsilon_i'\}$, $i = 1, \ldots, n$. Then substituting Y_i' in (3), we obtain

$$\begin{aligned}
\hat{g}_1(x') &= \frac{\sum_{i=1}^{n} k\left(\frac{X_i' - x'}{h}\right)\{\hat{r}_0 + \hat{r}_1\,(x' - X_i')\}}{\sum_{i=1}^{n} k\left(\frac{X_i' - x'}{h}\right)} \\
&= \hat{r}_0 - h\hat{r}_1 \frac{\sum_{i=1}^{n}\left(\frac{X_i' - x'}{h}\right) k\left(\frac{X_i' - x'}{h}\right)}{\sum_{i=1}^{n} k\left(\frac{X_i' - x'}{h}\right)}
\end{aligned}$$

which approaches to $\hat{r}_0$ mentioned in (7) for relatively small bandwidth h such that $h \to 0$. Noticeably, the estimator $\hat{r}_1$ is not of use when $h \to 0$. Denote $\beta_i' = M_2' - (x' - X_i')M_1'$ $\forall$ $i = 1, \ldots, n$. Then the estimate of $g_1(x)$ will be

$$\hat{g}_1(x') = \frac{\sum_{i=1}^{n}\beta_i' Y_i'}{\sum_{i=1}^{n}\beta_i'} \tag{10}$$

or a slightly modified estimator $\hat{g}_1(x') = \dfrac{\sum_{i=1}^{n} \beta_i' Y_i'}{\sum_{i=1}^{n} \beta_i' + n^{-2}}$ where n^{-2} is added to the denominator to avoid the situation of $\sum_{i=1}^{n} \beta_i' \approx 0$. This $\hat{g}_1(x')$ is called *simplified local linear smoother* (SLLS) of $g_1(x')$.

As we mentioned in the introduction, deletion of incomplete pairs may cause loss of information in data analysis. Hence the technique of refilling the missing observations or imputation would be thought of. $\hat{g}_1(x')$ can be treated as the imputed estimator for those k missing responses at the values of corresponding X. Subsequently, the estimator $\hat{g}_2(\cdot)$ is to be derived on the basis of complete bivariate observations (X, Y), denoted as (X^*, Y^*) after the imputation process.

Thus, in this concocted data $X^* = X$ and $Y_i^* = \delta_i\, Y_i' + (1 - \delta_i)\, \hat{g}_1(X_i')$.

Minimizing $\sum_{i=1}^{n} [Y_i^* - s_0 - s_1(x^* - X_i^*)]^2\, k\left(\dfrac{x^* - X_i^*}{h}\right)$ with respect to the linear constants s_0 and s_1 following the same arguments already proposed in (5) and (6),

$$\hat{s}_0 = \frac{\sum_{i=1}^{n} (M_2^* - (x^* - X_i^*)M_1^*)\ k\left(\dfrac{x^* - X_i^*}{h}\right) \delta_i\, Y_i^*}{\sum_{i=1}^{n} (M_2^* - (x^* - X_i^*)M_1^*)\ k\left(\dfrac{x^* - X_i^*}{h}\right)} \tag{11}$$

where

$$M_l^* = \sum_{i=1}^{n} (x^* - X_i^*)^l\ k\left(\frac{x^* - X_i^*}{h}\right),\ l = 1, 2.$$

and $\hat{s}_1$ be the solution of s_1.

Ultimately, using the same logic as projected in (10), the final estimator $\hat{g}_2(\cdot)$ at $X^* = x^*$ is derived as

$$\hat{g}_2(x^*) = \frac{\sum_{i=1}^{n} \beta_i^* Y_i^*}{\sum_{i=1}^{n} \beta_i^*} \tag{12}$$

where $\beta_i^* = M_2^* - (x^* - X_i^*)M_1^*\ \forall\ i = 1, \ldots, n$. Alternatively, (11) can be written as $\hat{g}_2(x^*) = \dfrac{\sum_{i=1}^{n} \beta_i^* Y_i^*}{\sum_{i=1}^{n} \beta_i^* + n^{-2}}$ in order to avoid the possibility of the inflation of $\hat{g}_2(x^*)$. This estimator $\hat{g}_2(\cdot)$ is called the *imputed local linear smoother* (**ILLS**) of $g(x)$.

4. Relevant test statistics

In order to test $H_0 : X^* \perp\!\!\!\perp \epsilon^*$ ($\perp\!\!\!\perp$ means independence) we consider a sequence of contiguous alternatives, say H_n, that converges to H_0 as $n \to \infty$. In this case, the sequence of contiguous alternative H_n, indicating to the dependence between X^* and ϵ^*, has the following expression

$$H_n: \quad F_{n;X^*,\epsilon^*}(x^*, e^*) = (1 - \frac{\gamma}{\sqrt{n}})\, G_{X^*}(x^*)\, H_{\epsilon^*}(e^*) + \frac{\gamma}{\sqrt{n}}\, K_{X^*,\epsilon^*}(x^*, e^*) \tag{13}$$

where $F_{n;X^*,\epsilon^*}(\cdot,\cdot)$ denote the joint CDF of (X^*, ϵ^*) under H_n while, $H_{\epsilon^*}(\cdot)$ and $G_{X^*}(\cdot)$ are the marginal CDFs of ϵ^* and X^* respectively and $K_{X^*,\epsilon^*}(\cdot,\cdot)$ is the proper joint distribution function of (X^*, ϵ^*). $\gamma > 0$ is the mixing constant for $F_0(\cdot,\cdot)$ and $K_{X^*,\epsilon^*}(\cdot,\cdot)$ where $F_0(x^*, e^*) = G_{X^*}(x^*)\, H_{\epsilon^*}(e^*)$ is the joint CDF of (X^*, ϵ^*) under H_0. First we generate a bivariate sample $\{(x_1^*, e_1^*), \ldots, (x_n^*, e_n^*)\}$ of size n from $F_0(x^*, e^*)$ under H_0. Then, using the regression model $Y^* = g(X^*) + \epsilon^*$ we obtain the bivariate observations $(x_1^*, y_1^*), \ldots, (x_n^*, y_n^*)$ can be from the joint distribution function of (X^*, Y^*). Taking the ordered observations on X^* as $x_{(1)}^*, \ldots, x_{(n)}^*$ and then the corresponding Y^*-values as $y_{(1)}^*, \ldots, y_{(n)}^*$ ($y_{(i)}^*$'s termed as *induced ordered statistics*), we achieve the ordered set $\{(x_{(1)}^*, y_{(1)}^*), \ldots, (x_{(n)}^*, y_{(n)}^*)\}$. The related errors are $\epsilon_{(1)}^*, \ldots, \epsilon_{(n)}^*$ which are also viewed as the induced ordered values of $\epsilon_1^*, \ldots, \epsilon_n^*$. The second order differences of these induced ordered observations $y_{(i)}^*$'s, $i = 1, \ldots, n$ are defined as $y_{(i)}^{*(2)} := y_{(i+1)}^* - 2y_{(i)}^* + y_{(i-1)}^*$ with the marginal considerations as $y_{(0)}^* = y_{(1)}^*$, $y_{(n+1)}^* = y_{(n)}^*$, resulting two threshold figures as $y_{(1)}^{*(2)} = y_{(2)}^* - y_{(1)}^*$ and $y_{(n)}^{*(2)} = y_{(n-1)}^* - y_{(n)}^*$. Based on the these bivariate observation $(x_{(i)}^*, y_{(i)}^{*(2)})$s for $i = 1, \ldots, n$, the following test statistics (Dhar *et al.* (2018)) are proposed as

$$T_{n,1} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} sign\{(x_{(i)}^* - x_{(j)}^*)(y_{(i)}^{*(2)} - y_{(j)}^{*(2)})\} \tag{14}$$

$$T_{n,2} = \frac{1}{\binom{n}{4}} \sum_{1 \le i < j \le n} a(x_{(i)}^*, x_{(j)}^*, x_{(k)}^*, x_{(l)}^*)\, a(y_{(i)}^{*(2)}, y_{(j)}^{*(2)}, y_{(k)}^{*(2)}, y_{(l)}^{*(2)}) \tag{15}$$

$$T_{n,3} = \frac{1}{\binom{n}{4}} \sum_{1 \le i < j \le n} \frac{1}{4} h(x_{(i)}^*, x_{(j)}^*, x_{(k)}^*, x_{(l)}^*)\, h(y_{(i)}^{*(2)}, y_{(j)}^{*(2)}, y_{(k)}^{*(2)}, y_{(l)}^{*(2)}) \tag{16}$$

where $sign(t) = \frac{t}{|t|}$ if $t \neq 0$ or 0 otherwise, $h(p, q, r, s) = \{|p - q| + |r - s| - |p - r| - |q - s|\}$; $p, q, r, s \in \mathbb{R}$ and $a(p, q, r, s) = sign\{|p-q| + |r-s| - |p-r| - |q-s|\}$. (14) is the sample version of Kendall's tau statistics between X^* and $Y^{*(2)}$ while (15) is the sample statistic in favour to τ^* which is an extended version of Kendall's tau by Bergsma *et al.* (2014). In contrast, (16) is the sample counterpart of the distance based measure D introduced by Blum-Kiefer-Rosenblatt (1961).

For the sake of readers' interest the population versions of the aforementioned test statistics for unordered observations on (X^*, Y^*), $i = 1, 2, 3$ are too presented herewith.

$$T_1 = E[sign(X_1^* - X_2^*)(Y_1^* - Y_3^*)]$$

$$T_2 = E[a(X_1^*, X_2^*, X_3^*, X_4^*)a(Y_1^*, Y_2^*, Y_3^*, Y_4^*)]$$

$$T_3 = E[\frac{1}{4}h(X_1^*, X_2^*, X_3^*, X_4^*)h(Y_1^*, Y_2^*, Y_3^*, Y_4^*)].$$

To check $H_0 : X^* \perp\!\!\!\perp \epsilon^*$ is analogous of checking $H_0 : X^* \perp\!\!\!\perp f(\epsilon^*)$ for any proper function $f(\cdot)$. Let us assume the form of the function as $f(\epsilon^*) = \epsilon^{*(2)} = \epsilon_{(i+1)} - 2\epsilon_{(i)} + \epsilon_{(i-1)}$, $i = 1, \ldots, n$, the second order difference of ϵ^*. Thus modified H_0 is $H_0 : X^* \perp\!\!\!\perp \epsilon^{*(2)}$. Since ϵ_i's are unobservable, so is $\epsilon^{*(2)}$. Thus instead of $\epsilon^{*(2)}$ we may judiciously approximate it by $Y^{*(2)}$ provided the function $g(.)$ is sufficiently smooth. Thus H_0 can further be modified to $H_0 : X^* \perp\!\!\!\perp Y^{*(2)}$. Evidently, independence of X^* and ϵ^* implies and implied by $T_k = 0$ for $k = 1, 2, 3$. So, $H_0 : X^* \perp\!\!\!\perp Y^{*(2)}$ implies $T_k = 0$, $k = 1, 2, 3$ and vice versa. Therefore their sample representatives, *viz.*, $T_{n,k}$ for $k = 1, 2, 3$ would be regarded as the desired test statistics to carry out the test of independence.

To kick-start the test process it is reasonable to approximate $T_{n,k}((x^*_{(1)}, e^{*(2)}_{(1)}), \ldots, (x^*_{(4)}, e^{*(2)}_{(4)}))$ by $T_{n,k}((x^*_{(1)}, y^{*(2)}_{(1)}), \ldots, (x^*_{(4)}, y^{*(2)}_{(4)}))$ for $k = 1, 2, 3$, as due to smoothness of $g(.)$, $y^{*(2)}$ would enable to sweep out the effect of g for large n. In fact, any function sorting out the effect of $g(\cdot)$ can be chosen instead of $y^{*(2)}_{(i)}$. For instance, the test statistic based on first order differences of Y^* may be applicable also for testing homoscedasticity of errors against all possible alternatives, which coincides with any traditional nonparametric test of homoscedasticity [see the discussion in Einmahl *et al.*, 2008]. Under H_0 the critical regions can be determined by the test statistics $T_{n,i}$'s $(i = 1, 2, 3)$ as $\omega_{n,i} : T_{n,i} > c_{\alpha,i}$, $i = 1, 2, 3$, where $\alpha \in (0, 1)$ is the level of significance satisfying $P_{H_0}[T_{n,i} > c_{\alpha,i}] = \alpha$ and $c_{\alpha,i}$ is the α-th critical point of the limiting distribution of $T_{n,i}$ under H_0. To study the statistical powers of all $T_{n,i}$'s under H_n for different values of γ, we have to ascertain their limiting distributions.

5. Study on asymptotic powers of the test statistics

It can be shown that the proposed test statistics $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are all degenerate U-statistics. In order to study their asymptotic powers we would use various asymptotic properties such as consistency, efficiency, limiting law related to degenerate U statistic. Hence, the order of degeneracy of $T_{n,i}$ for each $i = 1, 2, 3$ is derived hereafter so that their asymptotic distributions under H_0 and H_n can be established.

5.1. Contiguity

For two arbitrary sequences of probability measures, say P_n and Q_n, the definition of contiguity of P_n and Q_n on the sequence of measurable spaces $(\chi_n, \mathcal{A}_n)$ is stated from Le Cam (1960a).

Definition 1: For an arbitrary sequence of events $A_n \in \mathcal{A}_n$, if $P_n(A_n) \longrightarrow 0 \implies Q_n(A_n) \longrightarrow 0$ for sufficiently large sample size n, then Q_n is concluded as contiguous with respect to P_n. It is symbolically expressed as $P_n \triangleleft Q_n$.

To detect whether $P_n \triangleleft Q_n$ holds, the theory of local asymptotic normality (LAN) needs to be expounded. Le Cam's first lemma describes the asymptotic Gaussian nature of the quantity $\log \frac{dQ_n}{dP_n}$ under the probability measure P_n (p.253, Hajek *et al.*, 1999)

Lemma 1: Let $l_n = \frac{dQ_n}{dP_n}$ be a sequence of likelihood ratios corresponding to P_n and Q_n.

Define G_n to be the sequence of distribution functions of l_n. Furthermore, G_n converges to another distribution function G such that

$$\int_0^\infty v\,dG(v) = 1.$$

Then, $P_n \triangleleft Q_n$.

Corollary 1 below delves out an useful consequence of Lemma 1 .

Corollary 0.1: $\log l_n \overset{P_n}{\rightsquigarrow} N(-\frac{1}{2}\theta, \theta)$ implies that Q_n is contiguous with respect to P_n.

The proof of Corollary 1 can be derived using Lemma 1 (for details see Van Der Vaart (2002)). To derive the asymptotic distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ using Le Cam's first lemma under contiguous alternatives H_n we assume

Assumption 1: $f_{X^*,\epsilon^*}(x^*, e^*) > 0$ for all x^* and e^*, where f_{X^*,ϵ^*} is the joint PDF of (X^*, ϵ^*).

Assumption 2: $E_{F_{X^*,\epsilon^*}}(\frac{k_{X^*,\epsilon^*}(x^*,e^*)}{f_{X^*,\epsilon^*}(x^*,e^*)} - 1)^2 < \infty$ where $k_{X^*,\epsilon^*}(\cdot,\cdot)$ is the joint proper PDF of (X^*, ϵ^*).

Theorem 1: Under Assumption 1 and Assumption 2, H_n is a sequence of contiguous alternatives.

The formal proof of Theorem 1 is provided in Appendix 1. Next, we explore out the limiting laws of an U-statistic with certain order of degeneracy so that limiting distributions of $T_{n,i}$'s under both hypotheses can be intuited further.

Definition 2: (U statistic) Suppose $\psi(z_1, \ldots, z_m)$ be a real-valued measurable function. Based on a sample $\{Z_1, \ldots, Z_n\}$ from $F_Z(\cdot) \in \mathcal{F}$, $m \leq n$, a U-statistic with kernel ψ is defined as

$$U_n \equiv U_n(\psi) \quad = \quad \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \ldots < i_m \leq n} \psi(Z_{i_1}, \ldots, Z_{i_m}). \tag{17}$$

U_n is an unbiased estimator of population parameter θ. Also, U_n attains the minimum variance among all other unbiased estimators of θ.
Let us define a sequence of functions related to ψ. For $c = 0, 1, \cdots, m$, let

$\psi_c(z_1, \ldots, z_c) = E[\psi(z_1, \ldots, z_c, Z_{c+1}, \ldots, Z_m)]$ where $X_{c+1}, \cdots, X_n$ are i.i.d. Clearly, $E\psi_c(z_1, \ldots, z_c) = \theta$.

Denote, $\psi_c^*(z_1, \ldots, z_c) = \psi_c(z_1, \ldots, z_c) - E[\psi_c(z_1, \ldots, z_c)]$ and $\xi_c = var[\psi_c^*(z_1, \ldots, z_c)]$, $0 \leq c \leq m$.

Under this notation, the degeneracy of U statistic of order m is defined as follows.

Definition 3: (Order of degeneracy) The order of degeneracy of a U statistic is p if $\xi_0 = \ldots = \xi_p = 0$ and $\xi_{p+1} > 0$.

Here p is the order of degeneracy for the associated kernel $\psi(.)$ and the corresponding U-statistic U_n as well. Some useful theorems, provided by Lee (1990), are pertinent in the context of variance of U_n.

Theorem 2: (i) $\psi_c(z_1, \ldots, z_c) = E[\psi_d(z_1, \ldots, z_c, Z_{c+1}, \ldots, Z_d)]$ for $1 \leq c < d \leq m$.

(ii) $E[\psi_c(Z_1, \ldots, Z_c)] = E[\psi(Z_1, \ldots, Z_m)]$.

Theorem 3: $\xi_c = cov(\psi(N_1), \psi(N_2))$ with N_1, N_2 being the subsets of $\mathcal{C}_{m,n}$, $c = 1, \ldots, m$ each with m number of elements.

Theorem 4: The variance of U_n based on kernel ψ of degree m is

$$Var(U_n) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c} \binom{n-m}{m-c} \xi_c \tag{18}$$

The asymptotic distribution of $\sqrt{n}(U_n - \theta)$ for large n is normal with mean 0 and variance $m^2\,\xi_1$ (Serfling, 1980) . Unfortunately, in degenerate situation the asymptotic distribution of U_n is no longer normally distributed. Also, it can be explained that $\sqrt{n}(U_n - \theta)$ does not converge to a random variable with degenerate distribution function. If the kernel ψ possesses order of degeneracy p, then the asymptotic distribution of $n^{\frac{d+1}{2}}(U_n - \theta)$ converges to a nonnormal distribution as n increases. The following theorem from Serfling (1980) unveils on the pattern of distribution when $p = 1$(*i.e.* order of degeneracy 1).

Theorem 5: Let $\tilde{\psi}_2(z_1, z_2) = E[\psi(Z_1, Z_2, Z_3, \ldots, Z_m)|Z_1 = z_1, Z_2 = z_2]$, and $\xi_2 = Var[\tilde{\psi}_2(z_1, z_2)]$. If $\xi_1 = 0 < \xi_2$ and $E[\psi^2(Z_1, \ldots, Z_m)] < \infty$, then for some real constants λ_1, λ_2, ... and *iid* $N(0, 1)$ random variables Γ_1, Γ_2, ...,

$$n(U_n - \theta) \xrightarrow{L} Y \tag{19}$$

where $Y \sim \binom{m}{2} \sum_{i=1}^{\infty} \lambda_i(\Gamma_i^2 - 1)$, $m \geq 2$.

The asymptotic non-Gaussian distribution of degenerate U-statistic may also be explicated through obtaining the variance of a symmetric and positive definite quadratic kernel $W(Z_1, Z_2)$ with order of degeneracy 1 where Z_1, Z_2 are i.i.d. random variables. The kernel $W(Z_1, Z_2)$ can be expanded as

$$W(z_1, z_2) = \sum_{k=1}^{\infty} \lambda_k \phi_k(z_1) \phi_k(z_2)$$

where λ_k's are the eigenvalues with corresponding eigenfunctions $\phi_k(z)$'s satisfying

$$\int_{-\infty}^{\infty} W(z, Z_2) \phi_k(Z_2) d\, Z_2 = \lambda_k \phi_k(z).$$

In contiguous set up, the distribution of degenerate U statistic can be deduced (Gregory, 1977). Let $Q_{n,1}$ be the sequence of probability measures with $Q_n = Q_{n,1} \times \ldots \times Q_{n,1}$ (n times). P_0 is the probability measure under H_0 with $P_n = P_0 \times \ldots \times P_0$ (n times). Further suppose, Q_n is contiguous with respect to P_n. Then, the following theorem asserts the limiting distribution of an U-statistic T_n under the probability measure Q_n.

Theorem 6: (Gregory, 1977) Suppose the Radon-Nikodym derivative $dQ_{n,1}/dP_0 = 1 + n^{-\frac{1}{2}}\, h_n$ holds for some sequence $\{h_n\}$ in $L_2(\chi, \mathcal{A})$ that converges to $h \in L_2$. Then, for an U-statistic T_n with order of degeneracy 1,

$$\lim_{n\to\infty} Q_{n,1}\{T_n \leq x\} = P\left(\sum_{k=1}^{\infty} \lambda_k\{(\Gamma_k + a_k)^2 - 1\} \leq x\right) \tag{20}$$

where $a_k = \int h\, \phi_k\, dP_0$ and $\Gamma_1,\, \Gamma_2,\, \ldots$ are *iid* $N(0,1)$ random variables.

The asymptotic distributions for $T_{n,2}$ and $T_{n,3}$ under H_0 and H_n are easily obtainable using Theorem 6.

Generally speaking, let us define an operator E on $L_2(\chi, \mathcal{A})$ for $\tilde{\psi}_2(z_1, z_2)$ associated with the kernel ψ as

$$E\, g(z) = \int_{-\infty}^{\infty} \tilde{\psi}_2(z, y) g(y) d(F(y)),\;\; z \in \mathbb{R},\;\; g \in L_2 \tag{21}$$

and corresponding to E the eigenvalues $\lambda_1,\, \lambda_2,\, \ldots$ satisfy $E\, g = \lambda\, g$. Hence one can conclude that $\tilde{\psi}_2(z_1, z_2) = \sum_{k=1}^{\infty} \lambda_k g_k(z_1) g_k(z_2)$ with being orthonormal sequence g_k's satisfying $E[g_k(Z_1)g_l(Z_2)] = 1$ if $k = l$ and 0 if $k \neq l$. Here g_k's are the eigenfunctions corresponding to λ_k's of the transformation

$$E[\tilde{\psi}_2(z, Z_1) g_k(Z_1)] = \lambda_k\, g_k(z) \tag{22}$$

and in L_2,

$$\sum_{k=1}^{n} \lambda_k g_k(Z_1) g_k(Z_2) \xrightarrow{q.m.} \tilde{\psi}_2(Z_1, Z_2). \tag{23}$$

5.2. Limiting distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$

These test statistics are constructed by the spacings function formed from the distribution function of X^* *i.e.* $G_{X^*}(\cdot)$. Regarding consistency of the test statistics under H_0, we prefer to mention below an important result related to the expectation of an ordered uniform spacing due to Bairamov *et al.* (2010).

Result 1: For $r \geq 1$ and $n \to \infty$,

$$E(V_{(n+2-r)}) \sim \frac{\log n}{n} \longrightarrow 0 \tag{24}$$

where $V_{(s)}$ is the s^{th} order statistic among $\{V_{(1)}, \ldots, V_{(n)}\}$ based on the uniform spacings $V_i = U_{(i)} - U_{(i-1)}$'s $\forall\; i = 1, \ldots, n$. $V_{(s)}$ is also called the s^{th} ordered uniform spacing, $1 \leq s \leq n$. $U_{(i)}$ is the i^{th} order statistic based on $\{U_1, \ldots, U_n\}$ obtained from $Uniform(a, b)$ distribution, $a < b$, $1 \leq i \leq n$.

Along with Assumptions 1 and 2, let us further assume

Assumption 3: $X_1^*, \ldots, X_n^*$ (as defined earlier) are *i.i.d.* random variables with distribution function G_{X^*}.

Assumption 4: $Y_1^*, \ldots, Y_n^*$ (as defined earlier) are obtained from the model $Y_i^* = g(X_i^*)+\epsilon_i^*$, $i = 1, \ldots, n$, with $g(\cdot)$ having bounded derivative, ϵ^* having bounded probability density function and $E(\epsilon_i^*|X_i^*) = 0 \ \forall \ i = 1, \ldots, n$.

Based on Assumption 1-4, we develop the following theorems (Theorem 7, 8 and 9) regarding the limiting properties of $T_{n,i}$'s, $i = 1, 2, 3$. In each theorem, part (i) detects the order of degeneracy attached to each of $T_{n,i}$'s, $i = 1, 2, 3$. Part (ii) and part (iv) are directly followed from (i), describing the limiting distributions of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$. Part (ii) establishes the consistency of each of the test statistics. Suppose $\epsilon^{*(2)}$ has the CDF $H^*_{\epsilon^{*(2)}}(\cdot)$.

Theorem 7: (i) $T_{n,1}$ has kernel of order of degeneracy 0.

(ii) $T_{n,1} \xrightarrow{P} 0$ under H_0.

(iii) Under H_0, $\sqrt{n}(T_{n,1} - E(T_{n,1})) \xrightarrow{L} N(0, 4\xi_1)$.

(iv) Under H_n, $\sqrt{n}(T_{n,1} - E(T_{n,1})) \xrightarrow{L} N(\mu_1, 4\xi_1)$, where

$$\mu_1 = 2\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[2\int_{-\infty}^{x^*}\int_{-\infty}^{y^*} dG_{X^*}(u^*)dH^*_{\epsilon^{*(2)}}(v^*) + 2\int_{x^*}^{\infty}\int_{y^*}^{\infty} dG_{X^*}(u^*)dH^*_{\epsilon^{*(2)}}(v^*)]dK_{X^*,\epsilon^*}(x^*, y^*) \tag{25}$$

and,

$$\xi_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[2\int_{-\infty}^{x^*}\int_{-\infty}^{y^*} dG_{X^*}(u^*)dH^*_{\epsilon^{*(2)}}(v^*) + 2\int_{x^*}^{\infty}\int_{y^*}^{\infty} dG_{X^*}(u^*)dH^*_{\epsilon^{*(2)}}(v^*)]^2 dG_{X^*}(x^*)dH_{\epsilon^*}(y^*). \tag{26}$$

Theorem 8: (i) $T_{n,2}$ has kernel of order of degeneracy 1.

(ii) $T_{n,2} \xrightarrow{P} 0$ under H_0.

(iii) The asymptotic distribution for $T_{n,2}$ under H_0 is given by

$$n(T_{n,2} - E(T_{n,2})) \xrightarrow{L} \sum_{k=1}^{\infty} \lambda_k\{\Gamma_k^2 - 1\}$$

where Γ_1, Γ_2, ... are *iid* $N(0, 1)$ random variables, λ_k's are the eigenvalues associated with

$$\begin{aligned} l(x, y) &= E[sign\{|X^*_{(1)} - X^*_{(2)}| + |X^*_{(3)} - X^*_{(4)}| - |X^*_{(1)} - X^*_{(3)}| - |X^*_{(2)} - X^*_{(4)}|\} \\ &\quad \times sign\{|Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}| - |Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| - |Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}|\} \\ &\quad |X^*_{(1)} = x^*, Y^{*(2)}_{(1)} = y^*]. \end{aligned}$$

(iv) The asymptotic distribution for $T_{n,2}$ under H_n is given by

$$n(T_{n,2} - E(T_{n,2})) \xrightarrow{L} \sum_{k=1}^{\infty} \lambda_k \{(\Gamma_k + a_k)^2 - 1\} \tag{27}$$

where Γ_1, Γ_2, ... are *iid* $N(0,1)$ random variables, λ_k's are the eigenvalues associated with $l(x^*, y^*)$ given in (iii). The quantities a_k's are defined as

$$a_k = \int h\, f_k(x^*) f_k(y^*)\, dG_{X^*}(x^*) dH^*_{\epsilon^{*(2)}}(y^*). \tag{28}$$

where f_k's are the eigenfunctions corresponding to λ_k's, $k = 1, 2, \ldots$.

Theorem 9: (i) $T_{n,3}$ has kernel of order of degeneracy 1.

(ii) $T_{n,3} \xrightarrow{P} 0$ under H_0.

(iii) The asymptotic distribution for $T_{n,3}$ under H_0 is given by

$$n(T_{n,3} - E(T_{n,3})) \xrightarrow{L} \sum_{k=1}^{\infty} \lambda^*_k \{\Gamma^{*^2}_k - 1\}$$

where Γ^*_1, Γ^*_2, ... are *iid* $N(0,1)$ random variables, λ^*_k's are the eigenvalues associated with

$$\begin{aligned} l^*(x^*, y^*) \quad = \quad & E[\{|X^*_{(1)} - X^*_{(2)}| + |X^*_{(3)} - X^*_{(4)}| - |X^*_{(1)} - X^*_{(3)}| - |X^*_{(2)} - X^*_{(4)}|\} \\ & \times\{|Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}| - |Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| - |Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}|\} \\ & |X^*_{(1)} = x^*, Y^{*(2)}_{(1)} = y^*]. \end{aligned}$$

(iv) The asymptotic distribution for $T_{n,3}$ under H_n is given by

$$n(T_{n,3} - E(T_{n,3})) \xrightarrow{L} \sum_{k=1}^{\infty} \lambda^*_k \{(\Gamma^*_k + a^*_k)^2 - 1\} \tag{29}$$

where Γ^*_1, Γ^*_2, ... are *iid* $N(0,1)$ random variables, λ^*_k's are the eigenvalues associated with $l^*(x^*, y^*)$ given in (iii). The quantities a^*_k's are defined as

$$a^*_k = \int h\, f^*_k(x^*) f^*_k(y^*)\, dG_{X^*}(x^*) dH^*_{\epsilon^{*(2)}}(y^*). \tag{30}$$

where f^*_k's are the eigenfunctions corresponding to λ^*_k's, $k = 1, 2, \ldots$.

Proofs of all three theorems are furnished in Appendix 1.

5.3. Examples on asymptotic power calculation

To check on the performance of asymptotic power curves of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ with respect to different values of the mixing constant γ introduced in (13) we consider the values of γ from 0 to 10. We investigate on power against the H_0 in reference with these three statistics when the different percentage of missingness occurs in Y values under missing at random (MCAR) structure. All those missing values are refilled by NW estimation process as well as local linear smoothing (ILLS) as elaborately discussed in Section 3. Thereafter, the power functions for $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ are found for the imputed set of (X^*, Y^*) under $n = 100$. We generate such 500 sets of bootstap sample.

Let us pick up a couple of examples from Einmahl *et al.*(2008) where the conditional distributions of the error ϵ^* for given value of the covariate X^*, along with the joint proper distribution of (X^*, ϵ^*) are proposed. Epanechnikov kernel is used as the kernel function in the expression of the test statistics. Note that for each of the examples under consideration, the null model is taken as independent bivariate normal, *i.e.*, $f_{X^*,\epsilon^*}(.,.) = \frac{1}{2\pi}e^{-\frac{\epsilon^{*2}+x^{*2}}{2}}$. Since under H_0, $F_{X^*,\epsilon^*}(.,.) = G_{X^*}(.)H_{\epsilon^*}(.)$, μ_1 and ξ_1 in (25) and (26) are theoretically found out using the integral of standard normal variable. The rest of the results related to $T_{n,2}$ and $T_{n,3}$ are deduced by approximating infinite sum of weighted chi-square by finite one (taking upto the tenth term of (27) and (29)).

Example 1: $k_{X^*,\epsilon^*}(x^*, e^*)$ is such that $(\epsilon^*|X^* = x^*) \sim N(0, \frac{1+5x^*}{100})$ with $X^* \sim N(0, 1)$.

Example 2: $k_{X^*,\epsilon^*}(x^*, e^*)$ is such that $(\epsilon^*|X^* = x^*) \stackrel{\mathcal{D}}{=} Cauchy(0, x^{*^2})$ with $X^* \sim N(0, 1)$.

Percentages of missingness are chosen as 5%, 10% and 20% respectively. For each example, power curves of three statistics under complete data (without missing value) and other three missing proportion cases are drawn (a total of eight figures). The red line denotes the power curve of $T_{n,1}$, whereas the green and blue lines denote the power curves of $T_{n,2}$ and $T_{n,3}$ respectively. Due to space constraint, the power curves obtained only through LLS imputation technique in $n = 100$ are provided here. Appendix 2 contains the detailed and comparative tables of power calculation derived by both NW estimation and ILS technique taking sample size 100 with bootsrap size 500.

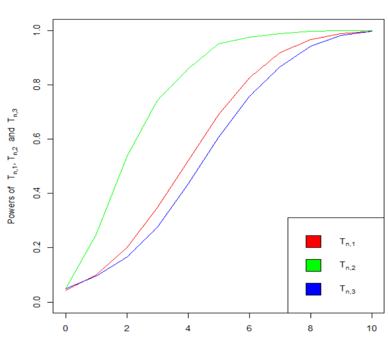

Figure 1: Power for Example 1 against γ in no missing setup

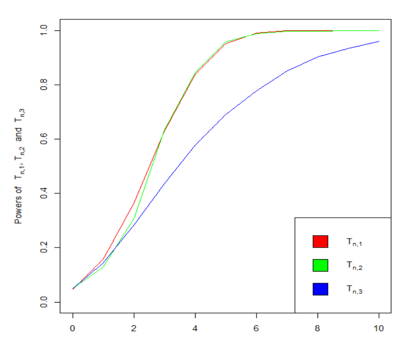

Figure 2: Power for Example 1 against γin 5% MCAR setup

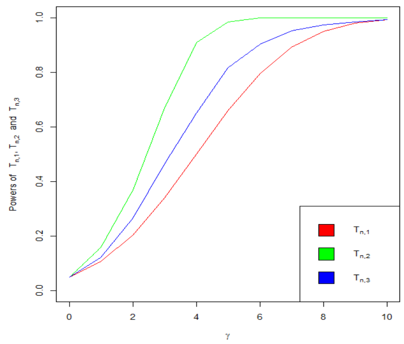

Figure 3: Power for Example 1 against γ in 10% MCAR setup

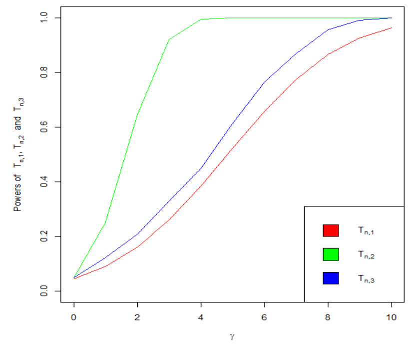

Figure 4: Power for Example 1 against γ in 20% MCAR setup

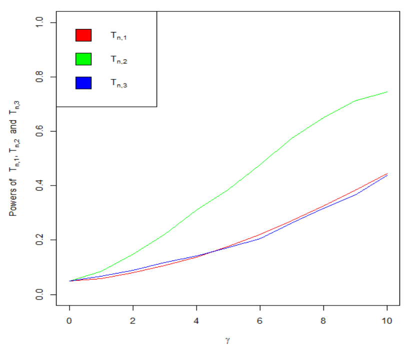

Figure 5: Power for Example 2 against γ in no missing setup

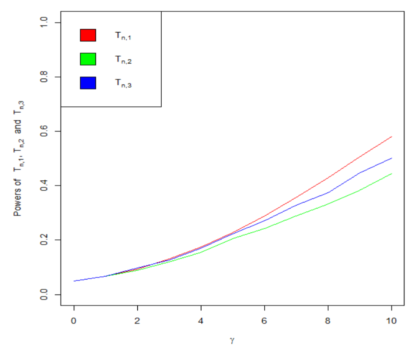

Figure 6: Power for Example 2 against γ in 5% MCAR setup

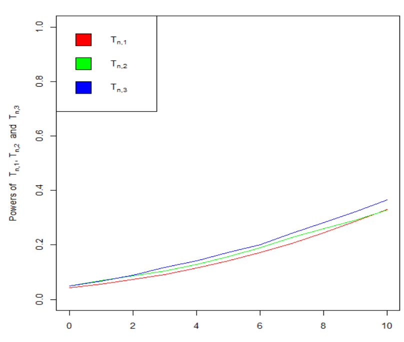

Figure 7: Power for Example 2 against γ in 10% MCAR setup

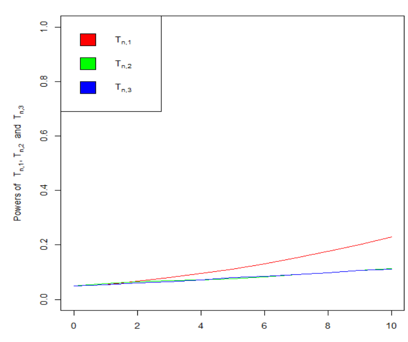

Figure 8: Power for Example 2 against γ in 20% MCAR setup

Although for no missing case power exerted by $T_{n,2}$ performs better across the mixing constant γ, in presence of missingness its power gets deteriorated as compared with the power by Kendall's tau, *i.e.* $T_{n,1}$. In contrast, power by distance based measure $T_{n,3}$ behaves not so well for all choices of missingness. Imputation done by local linear smoothing also does not change the scenario. In applying rank based test when observations on Y are missing does not guarantee the universal superiority in power. The more the counts in bivariate pairing in test statistic; lesser will be the power with the increase of missingness. Since in $T_{n,2}$ four bivariate pairs are in use, impact of missingness hits it more sharply than $T_{n,1}$. Plausible imputation can not improve the downfall as well.

Additionally, under normally distributed alternative the power exerted by all three statistics are quite reasonable and closer to 1. In contrast, Example 2 dealing with Cauchy alternatives experiences poorer power performance. Cauchy distribution being a heavy tailed distribution might be a good indicator of how sensitive the tests are to departures from normality, *i.e.* in presence of extreme observations. Although in no missing case the proposed $T_{n,2}$ holds its superiority, it fails to hold that in missing cases. In fact more the missingness worse the power comes out.
The entire simulation exercise is performed by R 4.0.5.

6. Real data analysis

In this segment of real data analysis, we choose out Abalone Data collected by the Department of Primary Industry and Fisheries, Tasmania. The data is available online in UCI Machine Learning Repository Data Set page (https://archive.ics.uci.edu/ml/datasets/Abalone).

The primary objective of this zoological data is to predict the age of abalone (a common species of marine gastropod molluscs, mainly inhabited in warm seas) from different physical measurements. This data consists of 4177 observations each having 10 qualitative and quantitative characters. Among those there are 9 independent characters, based on the physical measurements – viz, sex (nominal), length (in mm) for longest shell measurement, diameter (in mm) perpendicular to length, height (in mm) with meat in shell, whole weight (in grams) of abalone, shucked weight (in grams) *i.e.* weight of meat, viscera weight (in grams) *i.e.* gut weight (after bleeding), shell weight (in grams) after being dried, rings (integer) and one dependent variable — age (in years).

In our study, we pick up a single nonparametric regressor, *viz.*, shell weight after being dried (X in grams) and the regressand, *viz.* age (Y in years). For the sake of preciseness, we select first 100 observations instead of the whole. As a preliminary exploratory analysis, let us highlight the scatter plot on age against scaled shell weights below. The plot projects positive association with weakly linear tendency.

In order to incite readers' interest, the group of histograms (Figure 1) on underlying distributions of the response variable Y for complete case as well as for of several percentage of missingness is provided. In this figure, the missing observations are imputed by Nadaraya Watson estimator. Also the kernel density inlay is curved over each histogram. The underlying distribution is mildly right skewed which remains almost same not only in complete case but also in imputed distributions under 5%,10% and 20% missingness. Therefore imputation does not trigger any significant change in the underlying distribution.

To test the independence of X and ϵ we carry out bootstrap tests on 200 resamples having 100 sample observations in each set. At first, the observed values of the test statistics under the null hypothesis are obtained for the fixed sample size 100. Suppose the b^{th} resample of $T_{n,k}$ be $T_{n,k}^{\,b}$, $b = 1, \ldots, 200$, $k = 1, 2, 3$. The estimated p-value of $T_{n,k}$ is computed as $\frac{\#\{T_{n,k}^{\,b} > T_{n,k}^{*}\}}{200}$, $b = 1, \ldots, 200$, $k = 1, 2, 3$ where $T_{n,k}^{*}$ is the observed value of $T_{n,k}$ under H_0.

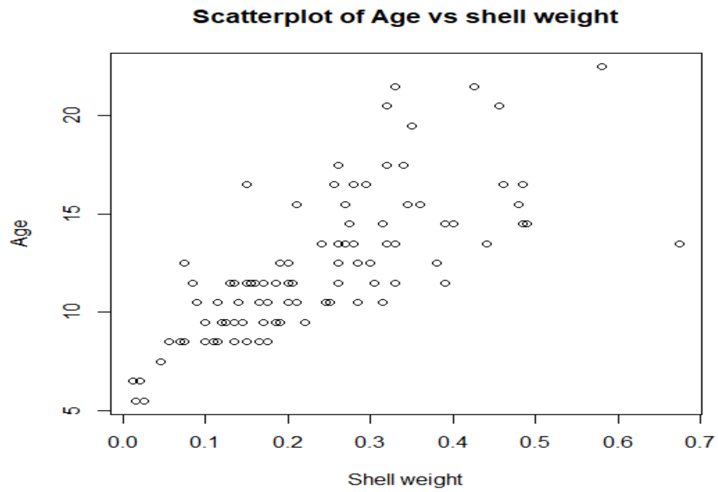

Figure 9: Scatter diagram

The same is repeated for (i) complete case (with 100 observations in each bootstrap set); (ii) 5% randomly missing observations, (iii) 10% randomly missing observations and (iv) 20% randomly missing observations. In each of the missing scenario, the missing observations are imputed by NW estimation as well as ILLS estimation and p-value is reported accordingly. Higher the p-value stronger is the evidence in favour of H_0. Tacitly speaking, for this data, under missingness, each p-value indicates preference towards H_0.

Table 1: Table showing p-values of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ under missingness estimated by N-W & ILLS imputation respectively

Statistic	Complete case	p-values					
		N-W			ILLS		
		5%	10%	20%	5%	10%	20%
$T_{n,1}$	0.575	0.375	0.480	0.415	0.490	0.515	0.635
$T_{n,2}$	0.680	0.940	0.930	0.900	0.980	0.989	0.890
$T_{n,3}$	0.660	0.900	0.920	0.880	0.980	0.999	0.905

7. Conclusion

In this article we have investigated the performance of three statistics– two rank based and one distance based, in the presence of MCAR missingness of observations. These tests are consistent. Powers are calculated under contiguous alternatives. For complete case situation $T_{n,2}$ shows best staging over $T_{n,1}$ and $T_{n,3}$ in both Gaussian and the heavy tailed distribution Cauchy but $T_{n,2}$ is not robust enough in presence of constant proportion of missingness. Specifically for non Gaussian alternative, missingness yields poor power exerted by $T_{n,2}$ and $T_{n,3}$ as compared to that by $T_{n,1}$. On the other hand, estimation of missing responses by imputated local linear smoothing (ILLS) method may yield a better power over that deduced by Nadaraya Watson (N-W) method, still those results are not convincing enough for non Gaussian distribution. Therefore, applying a rank based test statistic in testing of independence under nonparametric regression set up in presence of missingness would not

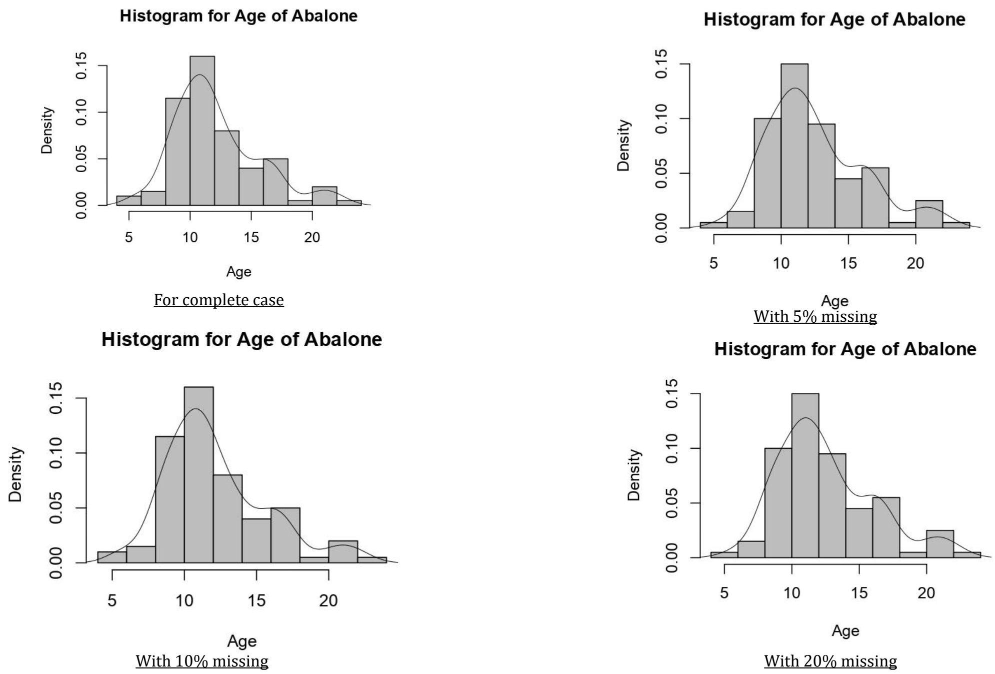

Figure 10: Histograms and regression curve in-lays for complete and missing cases

add substantial amount of power. In order to deal with such a situation few other distance based measure on distribution functions, *e.g.* Kolmogorov-Smirnov or Cramer-Von-Mises might be given a thought. It is to be noted that Alvo *et al.* (1995) proposed a new class of measures of rank correlation which are formed on a notion of distance between incomplete rankings. This approach utilizes the information on the positions of the actual observations relative to the string of incomplete observations. This mechanism would compensate for missing values and may be used as consistent test statistic in same context too.

In missing situation the strongest assumption that is commonly made is that the data are missing completely at random (MCAR) as probability that any variable is missing can not depend on any other variable in the model of interests. But for most data sets, the MCAR assumption is unlikely to be precisely specified, specially in design data. In those cases, a much weaker assumption, missing at random (MAR) is more common in practice. In MAR, the missingness of response depends on another observed variable. Therefore, effectivity of $T_{n,2}$ may be more worth investigating subject under MAR situation as compared with the performance by $T_{n,1}$ and $T_{n,3}$, considering a certain probability distribution of missingness.

Acknowledgments

Both authors are thankful to the anonymous referee for the valuable suggestions. The corresponding author is indebted to Dr. Ujjwal Das, Associate Professor, Indian Institute of Management, Udaipur, India for sharing his initial idea of this problem. Also, the corresponding author like to thank the Chair Editor for his meticulous proof reading which led to substantial improvement of this paper.

References

Afifi, A. A. and Elashoff (1966). R. M. Missing observations in multivariate statistics:I.review of literature. *Journal of the American Statistical Association*, **61**, 595-604.

Alvo, M.and Cabilio, P. (1995). Rank correlation methods for missing data. *The Canadian Journal of Statistics*, **13**, 345-358.

Bairamov, I., Berred, A., and Stepanov, A. (2010). Limit results for ordered uniform spacings. *Statistical Papers*, **51**, 227-240.

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

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

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, **89**, 81-87.

Chu, C. K. and Cheng, P. E. (1995). Nonparametric regression estimation with missing data. *Journal of Statistical Planning and Inference*, **48**, 85-99.

Das, S., Halder, S., and Maiti, S.I. (2022). An extended approach to test of independence between error and covariates under nonparametric regression model. *Thailand Statistician*, **21**, 19-36.

Dhar, S. S., Bergsma, W., and Dassios, A (2018). Testing Independence of Covariates and Errors in Non-parametric Regression. *Scandinavian Journal of Statistics*, **45**, 421-443.

Einmahl, J. H. and Van Keilegom, I.(2008). Tests for independence in nonparametric regression. *Statistica Sinica*, **18**, 601-615.

Einmahl, J. H. and Van Keilegom, I.(2008). Specification tests in nonparametric regression. *Journal of Econometrics*, **143**, 88-102.

Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *The Annals of Statistics*, **5**, 110-123.

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

Hartley, H. O. and Hocking R. R. (1971). The analysis of Incomplete Data. *Biometrics*, **27**, 783-823.

Hlavka Z., Huskova M., and Meintanis S. G. (2011). Tests for independence in nonparametric heteroscedastic regression models. *Journal of Multivariate Analysis*, **102**, 816-27.

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

Lehmann E. L. and Romano, J. P. (2005). *Testing of Statistical Hypotheses, 3rd Ed.*. Springerlink.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*. 3rd Edition. John Wiley & Sons Inc.

Neumeyer, N. (2009). Testing independence in nonparametric regression. *Journal of Multivariate Analysis*, **100**, 1551-1566.

Serfling, R. J. (1981). *Approximation Theorems of Mathematical Statistics*. John-Wiley & Sons, Inc.

Van Der Vaart A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

ANNEXURE

Appendix 1

Proof of Theorem 1

The expansion of $\log L_n$ takes the form as follows

$$\begin{aligned}\log L_n &= \log \prod_{i=1}^{n} \frac{f_{n;X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} \\ &= \log \prod_{i=1}^{n} \Big\{ \frac{(1-\frac{\gamma}{\sqrt{n}}) f_{X^*,\epsilon^*}(x_i^*, e_i^*) + \frac{\gamma}{\sqrt{n}} k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} \Big\} \\ &= \sum_{i=1}^{n} \log \Big\{ \frac{(1-\frac{\gamma}{\sqrt{n}}) f_{X^*,\epsilon^*}(x_i^*, e_i^*) + \frac{\gamma}{\sqrt{n}} k_{X^*,\epsilon}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} \Big\}.\end{aligned}$$

With the aid of Taylor's expansion of $\log(1+r)$, $r > -1$ as well as the weak law of large numbers, $\log L_n$ is further expanded as

$$\sum_{i=1}^{n} \frac{\gamma}{\sqrt{n}} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right) - \frac{\gamma^2}{2n} \sum_{i=1}^{n} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right)^2 + O_P(n^{-1/2}). \tag{31}$$

Then,

$$\log L_n - \textstyle\sum_{i=1}^{n} \frac{\gamma}{\sqrt{n}} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right) + \frac{\gamma^2}{2n} \sum_{i=1}^{n} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right)^2 = O_P(n^{-1/2}) \longrightarrow 0 \ \ as \ \ n \to \infty.$$

Define a sequence of random variables W_n as $\sum_{i=1}^{n} \frac{\gamma}{\sqrt{n}} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right)$. With the help of Lindeberg's condition, the asymptotic distribution of W_n is developed as $\frac{W_n - E(W_n)}{\sqrt{Var(W_n)}} \xrightarrow{L} N(0,1)$ under H_0, where

$$E_{H_0}(W_n) = \sum_{i=1}^{n} \frac{\gamma}{\sqrt{n}} E_{H_0} \left(\frac{k_{X^*,\epsilon}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right) = 0$$

and $Var_{H_0}(W_n) = \frac{\gamma^2}{n} \sum_{i=1}^{n} E_{H_0} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right)^2 = \gamma^2 E_{H_0} \left(\frac{k_{X^*,\epsilon^*}}{f_{X^*,\epsilon^*}} - 1 \right)^2$. Hence under H_0,

$$W_n \xrightarrow{L} N \left(0, \gamma^2 E_{H_0} \left(\frac{k_{X^*,\epsilon^*}}{f_{X^*,\epsilon^*}} - 1 \right)^2 \right).$$

Another sequence of random variables $V_n = \frac{\gamma^2}{2n} \sum_{i=1}^{n} \left(\frac{k_{X^*,\epsilon^*}(x_i^*, e_i^*)}{f_{X^*,\epsilon^*}(x_i^*, e_i^*)} - 1 \right)^2$ weakly converges to $\frac{\gamma^2}{2} E_{H_0} \left(\frac{k_{X^*,\epsilon^*}}{f_{X^*,\epsilon^*}} - 1 \right)^2$. So, $\log L_n - W_n + V_n = o_p(1)$. Slutsky's theorem further ensures that the limiting distribution of the sequence of random variables $M_n = W_n - V_n$ converges to a random variable M such that

$$M \sim N\left(-\frac{1}{2}\gamma^2 E_{H_0}\left(\frac{k}{f}-1\right)^2, \gamma^2 E_{H_0}\left(\frac{k}{f}-1\right)^2\right). \tag{32}$$

Summing up all, one can conclude that $\log L_n - M_n = o_p(1)$, *i.e.* $\log L_n$ has the limiting distribution which is identical with that of limiting distribution of M_n, *i.e.* $N(-\frac{1}{2}\sigma, \sigma)$ where $\sigma = \gamma^2 E_{H_0}\left(\frac{k}{f}-1\right)^2$. Thereafter, the Corollary 5.1 of lemma 5.1 is sufficient enough in establishing the fact that H_n is a contiguous sequence of alternatives due to asymptotic normality of $\log L_n$. Notationally, contiguity can be expressed as $F_{X^*,\epsilon^*} \triangleleft F_{n;X^*,\epsilon^*}$.

Proof of Theorem 7

(i) Suppose the kernel of $T_{n,1}$ is denoted by $\psi((X^*_{(1)}, Y^{*(2)}_{(1)}), (X^*_{(2)}, Y^{*(2)}_{(2)}))$. One can simplify its form as

$$\begin{aligned}\psi_1(x^*, y^*) &= E[\psi((X^*_{(1)}, Y^{*(2)}_{(1)}), (X^*_{(2)}, Y^{*(2)}_{(2)}))|X^*_{(1)} = x^*, Y^{*(2)}_{(1)} = y^*]\\ &= E[sign\{(X^*_{(1)} - X^*_{(2)})(Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)})\}|X^*_{(1)} = x^*, Y^{*(2)}_{(1)} = y^*]\\ &= 2P[(X^*_{(1)} - X^*_{(2)})(Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}) > 0|X^*_{(1)} = x^*, Y^{*(2)}_{(1)} = y^*] - 1.\end{aligned}$$

Now under H_0 one can determine that

$$E_{(X^*_{(1)}, Y^{*(2)}_{(1)})}[\psi_1(X^*_{(1)}, Y^{*(2)}_{(1)})] = E_{(X^*_{(1)}, Y^{*(2)}_{(1)}),(X^*_{(2)}, Y^{*(2)}_{(2)})}[\psi((X^*_{(1)}, Y^{*(2)}_{(1)}), (X^*_{(2)}, Y^{*(2)}_{(2)}))] = 0.$$

Then, $\xi_1 = Var[\psi_1(X^*_{(1)}, Y^{*(2)}_{(1)})] = E[\psi_1^2(X^*_{(1)}, Y^{*(2)}_{(1)})] > 0$, where $Y^{*(2)}$ is approximately identically distributed with $\epsilon^{*(2)}$. Therefore, $\xi_0 = 0$ and $\xi_1 > 0$ is enough to conclude that ψ has order of degeneracy 0.

(ii) From Theorem 4 it is clear that the variance of $T_{n,1}$ gets approximated as $\frac{4\xi_1}{n}$ for large n, and $E[sign\{(X^*_{(i)} - X^*_{(j)})(Y^{*(2)}_{(i)} - Y^{*(2)}_{(j)})\}] = 0 \; \forall\; 1 \leq i < j \leq n$ as $P[(X^*_{(i)} - X^*_{(j)})(Y^{*(2)}_{(i)} - Y^{*(2)}_{(j)}) > 0] = P[(X^*_{(i)} - X^*_{(j)})(Y^{*(2)}_{(i)} - Y^{*(2)}_{(j)}) < 0]$ under H_0. One may conclude that $T_{n,1} \xrightarrow{P} 0$ as $E(T_{n,1}) = 0$ and $var(T_{n,1}) \to 0$ for $n \to \infty$ under H_0.

(iii) Deducing the asymptotic variance in Theorem 4 when $n \to \infty$, we derive the asymptotic distribution of $\sqrt{n}(T_{n,1} - E(T_{n,1}))$ under H_0. To prove this part of the theorem, any standard textbook on nonparametric inference would suffice.

(iv) Directed from the Le Cam's third lemma (Dhar *et al.* (2018)) the asymptotic distribution of $(\sqrt{n}(T_{n,1} - E(T_{n,1})), \log L_n)$ converges to $N_2\left(\begin{pmatrix}0\\ -\frac{\theta}{2}\end{pmatrix}, \begin{pmatrix}4\xi_1 & \tau\\ \tau & \theta\end{pmatrix}\right)$, $\theta > 0$ under H_0. Then it is easy to determine the limiting distribution of $\sqrt{n}(T_{n,1} - E(T_{n,1}))$ under H_n as $N(0 + \tau, 4\xi_1)$ *i.e.* $N(\tau, 4\xi_1)$. Hence $\tau = \lim_{n\to\infty} cov_{H_0}(\sqrt{n}(T_{n,1} - E(T_{n,1})), \log L_n)$ which can be finally derived as

$$2\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[2\int_{-\infty}^{x^*}\int_{-\infty}^{y^*} dG_{X^*}(u^*)dH_{\epsilon^*}(v^*) + 2\int_{x^*}^{\infty}\int_{y^*}^{\infty} dG_{X^*}(u^*)dH_{\epsilon^*}(v^*) - 1]dK_{X^*,\epsilon^*}(x^*, y^*).$$

Proof of Theorem 8

(i) The simplification of the kernel of $T_{n,2}$ is done as

$$\begin{aligned} & a(X^*_{(1)}, X^*_{(2)}, X^*_{(3)}, X^*_{(4)})a(Y^*_{(1)}, Y^*_{(2)}, Y^*_{(3)}, Y^*_{(4)}) \\ = \; & 2I(|X^*_{(1)} - X^*_{(2)}| + |X^*_{(3)} - X^*_{(4)}| - |X^*_{(1)} - X^*_{(3)}| - |X^*_{(2)} - X^*_{(4)}| > 0, \\ & |Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}| - |Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| - |Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}| > 0) + \\ & 2I(|X^*_{(1)} - X^*_{(2)}| + |X^*_{(3)} - X^*_{(4)}| - |X^*_{(1)} - X^*_{(3)}| - |X^*_{(2)} - X^*_{(4)}| < 0, \\ & |Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}| - |Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| - |Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}| < 0) - 1 \\ = \; & 2P(|Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}| - |Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| - |Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}| < 0) - 1 \\ = \; & \tilde{a}((X^*_{(1)}, Y^{*(2)}_{(1)}), (X^*_{(2)}, Y^{*(2)}_{(2)}), (X^*_{(3)}, Y^{*(2)}_{(3)}), (X^*_{(4)}, Y^{*(2)}_{(4)})) \end{aligned} \tag{33}$$

where $I(\cdot)$ is an indicator function. Now define, for $c = 0, \ldots, 4$,

$$\begin{aligned} & \tilde{a}_c((x^*_{(1)}, y^{*(2)}_{(1)}), \ldots, (x^*_{(c)}, y^{*(2)}_{(c)})) \\ = \; & E[\tilde{a}((x^*_{(1)}, y^{*(2)}_{(1)}), \ldots, (x^*_{(c)}, y^{*(2)}_{(c)}), (X^*_{(c+1)}, Y^{*(2)}_{(c+1)}), \ldots, (X^*_{(4)}, Y^{*(2)}_{(4)}))] \end{aligned}$$

and, $\xi_c = Var[\tilde{a}_c((X_{(1)}, Y^{*(2)}_{(1)}), \ldots, (X_{(c)}, Y^{*(2)}_{(c)}))]$.
In equation (33), $|Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}|$ and $|Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}|$ can be written into following two inequalities as $|Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| \le |Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(2)} - Y^{*(2)}_{(3)}|$ and
$|Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}| \le |Y^{*(2)}_{(2)} - Y^{*(2)}_{(3)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}|$. Then,
$P(Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}, Y^{*(2)}_{(1)} > Y^{*(2)}_{(4)})$
$= P(Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}, Y^{*(2)}_{(1)} > Y^{*(2)}_{(4)}, Y^{*(2)}_{(3)} > Y^{*(2)}_{(1)}) + P(Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}, Y^{*(2)}_{(1)} > Y^{*(2)}_{(4)}, Y^{*(2)}_{(3)} \le Y^{*(2)}_{(1)}) = \frac{1}{4!} \times 6 = \frac{1}{4}$. Similarly, $P(Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}, Y^{*(2)}_{(1)} \le Y^{*(2)}_{(4)})$ is calculated as $\frac{1}{4}$.
Then, $P(Y^{*(2)}_{(2)} < Y^{*(2)}_{(3)}) = \frac{1}{2} = P(Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)})$.
Finally we obtain $2P(|Y^{*(2)}_{(1)} - Y^{*(2)}_{(2)}| + |Y^{*(2)}_{(3)} - Y^{*(2)}_{(4)}| - |Y^{*(2)}_{(1)} - Y^{*(2)}_{(3)}| - |Y^{*(2)}_{(2)} - Y^{*(2)}_{(4)}| < 0) = 2\min\left(\frac{1}{2}, \frac{1}{2}\right) = 1$. Therefore,
$E[\tilde{a}((X^*_{(1)}, Y^{*(2)}_{(1)}), (X^*_{(2)}, Y^{*(2)}_{(2)}), (X^*_{(3)}, Y^{*(2)}_{(3)}), (X^*_{(4)}, Y^{*(2)}_{(4)}))] = 0$.
On the other hand, the expression of ξ_1 is same as
$cov[\tilde{a}((X^*_{(1)}, Y^{*(2)}_{(1)}), (X^*_{(2)}, Y^{*(2)}_{(2)}), (X^*_{(3)}, Y^{*(2)}_{(3)}), (X^*_{(4)}, Y^{*(2)}_{(4)}))]$ which equals
$\{1 + 4P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}, Y^{*(2)}_{(5)} > Y^{*(2)}_{(6)}] - 2P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}] - 2P[Y^{*(2)}_{(5)} > Y^{*(2)}_{(6)}]\}$.
For four distinct numbers (i_1, i_2, i_3, i_4) with $1 \le i_1 \ne i_2 \ne i_3 \ne i_4 \ne 7$ it is easy to verify that
$P[Y^{*(2)}_{(i_1)} > Y^{*(2)}_{(i_2)} > Y^{*(2)}_{(i_3)} > Y^{*(2)}_{(i_4)}] = \frac{6}{4!} = \frac{1}{4}$ and furthermore $P[Y^{*(2)}_{(i_1)} > Y^{*(2)}_{(i_2)}] = \frac{1}{2}$.
Then $\xi_1 = 1 + 4 \cdot \frac{1}{4} - 2 \cdot \frac{1}{2} - 2 \cdot \frac{1}{2} = 0$.
Consequently, the computation of ξ_2 becomes necessary to verify whether it is equal to 0 or not. ξ_2 is evaluated further as $\{1 + 4P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}, Y^{*(2)}_{(2)} > Y^{*(2)}_{(5)}] - 2P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)}] - 2P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(5)}]\}$ which equals $4 \times 5 \times 6 \times P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(3)} > Y^{*(2)}_{(5)} > Y^{*(2)}_{(6)} > Y^{*(2)}_{(4)} > Y^{*(2)}_{(1)}] + 4 \times 5 \times 6 \times P[Y^{*(2)}_{(2)} > Y^{*(2)}_{(5)} > Y^{*(2)}_{(3)} > Y^{*(2)}_{(6)} > Y^{*(2)}_{(4)} > Y^{*(2)}_{(1)}] =$

$2 \times \frac{4\times5\times6}{6!} = \frac{1}{3} > 0$. So $\xi_2 > 0$, which naturally implies that the order of degeneracy of $T_{n,2}$ is 1.

(ii) It is to be noted that $(|X^*_{(i)} - X^*_{(j)}| + |X^*_{(k)} - X^*_{(l)}| - |X^*_{(i)} - X^*_{(k)}| - |X^*_{(j)} - X^*_{(l)}|)(|Y^{*(2)}_{(i)} - Y^{*(2)}_{(j)}| + |Y^{*(2)}_{(k)} - Y^{*(2)}_{(l)}| - |Y^{*(2)}_{(i)} - Y^{*(2)}_{(k)}| - |Y^{*(2)}_{(j)} - Y^{*(2)}_{(l)}|) = O_p(\frac{\log n}{n}), 1 \leq i < j < k < l \leq n$ by Result 5.1 originally introduced by Bairamov *et al.* (2010).
The distribution function of $(|\epsilon^{*(2)}_{(i)} - \epsilon^{*(2)}_{(j)}| + |\epsilon^{*(2)}_{(k)} - \epsilon^{*(2)}_{(l)}| - |\epsilon^{*(2)}_{(i)} - \epsilon^{*(2)}_{(k)}| - |\epsilon^{*(2)}_{(j)} - \epsilon^{*(2)}_{(l)}|)$ is $\int_{-\infty}^{\infty} \left\{ H_{\epsilon^*}\left(y^* + \frac{t}{2}\right) - H_{\epsilon^*}\left(y^* - \frac{t}{2}\right) \right\} dH_{\epsilon^*}(y^*)$, denoted by $H^*_{\epsilon^{*(2)}}(t)$. Also the distribution function of $\epsilon^{*(2)}$ is approximately equal to the distribution function of $Y^{*(2)}$. One can derive that $a(X^*_{(i)}, X^*_{(j)}, X^*_{(k)}, X^*_{(l)})a(Y^{*(2)}_{(i)}, Y^{*(2)}_{(j)}, Y^{*(2)}_{(k)}, Y^{*(2)}_{(l)}) \longrightarrow 0$ in probability for $1 \leq i < j < k < l \leq n$ under H_0. Consequently a final conclusion becomes inevitable that $T_{n,2} \xrightarrow{P} 0$ as $n \to \infty$.

(iii) Due to Serfling (1981)'s theorem on the asymptotic distribution of a degenerate U-statistic presented by Theorem 5, it is quite straightforward to derive the limiting distributional form of $n(T_{n,2} - E(T_{n,2}))$ under H_0.

(iv) To furnish the elaborate proof regarding the asymptotic distribution of $n(T_{n,2} - E(T_{n,2}))$ under H_n, Theorem 6 by Gregory (1977) is required.

Proof of Theorem 9

In similar way to the proof of Theorem 8, Theorem 9 can also be proved.

Appendix 2

Table 2: Powers of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ for Example 1 for complete and missing cases using N-W and ILLS imputation

γ	Powers of test statistics in MCAR setup using NW estimation												Powers of test statistics in MCAR setup using ILLS								
	No missing			5% missing			10% missing			20% missing			5% missing			10% missing			20% missing		
	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$
0	0.042	0.05	0.05	0.053	0.05	0.05	0.05	0.05	0.05	0.054	0.05	0.05	0.047	0.05	0.05	0.05	0.05	0.05	0.046	0.05	0.05
1	0.1	0.25	0.095	0.103	0.06	0.188	0.156	0.064	0.12	0.157	0.087	0.087	0.157	0.129	0.143	0.108	0.16	0.123	0.091	0.252	0.124
2	0.201	0.537	0.166	0.18	0.085	0.474	0.355	0.1	0.249	0.343	0.143	0.148	0.366	0.309	0.284	0.205	0.371	0.268	0.161	0.647	0.209
3	0.347	0.742	0.276	0.286	0.139	0.796	0.604	0.148	0.468	0.579	0.206	0.241	0.628	0.632	0.436	0.341	0.668	0.465	0.261	0.92	0.33
4	0.521	0.859	0.436	0.415	0.23	0.969	0.816	0.238	0.695	0.789	0.273	0.344	0.84	0.845	0.579	0.501	0.909	0.652	0.386	0.995	0.45
5	0.691	0.95	0.606	0.554	0.356	0.997	0.938	0.381	0.877	0.92	0.325	0.501	0.951	0.957	0.691	0.661	0.984	0.818	0.523	1	0.613
6	0.827	0.975	0.759	0.686	0.483	1	0.985	0.5	0.968	0.978	0.359	0.66	0.99	0.989	0.777	0.796	1	0.904	0.658	1	0.766
7	0.918	0.988	0.866	0.798	0.628	1	0.998	0.63	0.996	0.995	0.391	0.758	0.999	0.998	0.85	0.892	1	0.953	0.775	1	0.869
8	0.967	0.996	0.943	0.882	0.758	1	1	0.753	1	0.999	0.411	0.84	1	0.998	0.902	0.951	1	0.974	0.866	1	0.957
9	0.989	0.999	0.982	0.938	0.852	1	1	0.851	1	1	0.421	0.898	1	1	0.934	0.981	1	0.986	0.927	1	0.991
10	0.997	0.999	0.996	0.97	0.924	1	1	0.913	1	1	0.42	0.948	1	1	0.96	0.993	1	0.994	0.964	1	1

Table 3: Powers of $T_{n,1}$, $T_{n,2}$ and $T_{n,3}$ for Example 2 for complete and missing cases using N-W and ILLS imputation

γ	Powers of test statistics in MCAR setup using N-W estimation												Powers of test statistics in MCAR setup using ILLS								
	No missing			5% missing			10% missing			20% missing			5% missing			10% missing			20% missing		
	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$	Power of $T_{n,1}$	Power of $T_{n,2}$	Power of $T_{n,3}$
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.044	0.05	0.05	0.05	0.05	0.05
1	0.059	0.084	0.066	0.063	0.083	0.067	0.085	0.065	0.061	0.065	0.058	0.051	0.066	0.067	0.068	0.057	0.069	0.067	0.055	0.058	0.054
2	0.08	0.149	0.09	0.085	0.148	0.088	0.146	0.073	0.077	0.087	0.06	0.051	0.094	0.09	0.097	0.073	0.086	0.089	0.066	0.064	0.06
3	0.106	0.221	0.117	0.112	0.206	0.104	0.231	0.08	0.093	0.115	0.067	0.055	0.13	0.119	0.126	0.092	0.104	0.118	0.08	0.069	0.064
4	0.138	0.31	0.142	0.145	0.29	0.124	0.338	0.103	0.113	0.149	0.069	0.056	0.175	0.154	0.17	0.115	0.128	0.141	0.095	0.072	0.072
5	0.176	0.386	0.173	0.184	0.396	0.151	0.46	0.121	0.132	0.189	0.079	0.056	0.228	0.205	0.222	0.141	0.157	0.173	0.112	0.075	0.08
6	0.22	0.478	0.206	0.229	0.488	0.182	0.586	0.146	0.15	0.235	0.09	0.057	0.289	0.243	0.271	0.172	0.189	0.201	0.131	0.083	0.085
7	0.27	0.573	0.263	0.28	0.57	0.211	0.704	0.181	0.18	0.287	0.106	0.063	0.357	0.289	0.327	0.206	0.227	0.242	0.152	0.091	0.091
8	0.325	0.651	0.316	0.336	0.64	0.244	0.803	0.194	0.207	0.343	0.122	0.064	0.43	0.332	0.374	0.244	0.259	0.283	0.176	0.098	0.098
9	0.383	0.712	0.365	0.395	0.692	0.287	0.879	0.213	0.245	0.403	0.136	0.069	0.506	0.382	0.446	0.286	0.29	0.321	0.201	0.106	0.107
10	0.445	0.745	0.437	0.458	0.72	0.325	0.932	0.238	0.288	0.465	0.146	0.072	0.581	0.444	0.501	0.331	0.327	0.365	0.229	0.113	0.112