On testing independence between regressors and error in generalized partially linear regression

Sthitadhi Das*, Saran Ishika Maiti† and Ayan Pal‡
Department of Mathematics, Brainware University, West Bengal , India
Department of Statistics, Visva Bharati, Santiniketan, West Bengal , India
Department of Statistics, University of Burdwan, West Bengal, India

Abstract

Under the set-up of generalized partially linear regression model, $Y = \underset{\sim}{X}'\underset{\sim}{\beta} + g(\underset{\sim}{W}) + \epsilon$ with parametric covariates $\underset{\sim}{X} = (X_1, \ldots, X_p)' \in \mathbb{R}^p$ and nonparametric covariates $\underset{\sim}{W} = (W_1, \ldots, W_q)' \in \mathbb{R}^q$, this article proposes a test whether $(\underset{\sim}{X}, \underset{\sim}{W})$ is independent of ϵ. By developing nondegenerate V-statistics based on Spearman's ρ_s and Kendall's τ, the test is executed against a sequence of contiguous alternatives. The test statistics are built on considering general order differences of the observed as well as estimated responses. Asymptotic properties of proposed test statistics are derived for checking the individual efficiencies under the assumed null hypothesis. Also, the potentiality of test statistics is investigated on some real datasets.

Keywords: Generalized partially linear regression model, Nonparametric regression model, Spearman's ρ_s, Kendall's τ, Measures of association, Contiguous alternatives, Asymptotic power, V-statistic
Mathematics Subject Classification: 62G05, 62G07, 62G09, 62G10, 62G20, 62G30, 62G35

1 Introduction

Testing association(s) (or independence) between two or more variables is a common topic of interest in the literature of statistics. In particular, under statistical regression setting, testing non-association between regressor and error is much sought-after issue of same regard. Keeping a steady beat, the same may be examined for a generalized partially linear regression model with mathematical representation

$$Y = \beta_1 X_1 + \ldots + \beta_p X_p + g(W_1, \ldots, W_q) + \epsilon. \tag{1}$$

Here Y is the variable of interest that is partly studied by p regressors $X_1, \ldots, X_p$ along with unknown constants (or, parameters) $\beta_1, \ldots, \beta_p$ as well as a Lipschitz continuous function $g(\cdot, \ldots, \cdot)$ of q independent variables $W_1, \ldots, W_q$ with unknown mathematical form where $p, q \geq 2$. The error ϵ is assumed as (i) $E(\epsilon | X_1, \ldots, X_p, W_1, \ldots, W_q) = 0$ for all $(X_1, \ldots, X_p, W_1, \ldots, W_q)$, (ii) $E(\epsilon^2 | X_1, \ldots, X_p, W_1, \ldots, W_q) = \sigma^2(X_1, \ldots, X_p, W_1, \ldots, W_q)$. Various analytical aspects on a generalized partially linear model, encompassing estimation of parameters, the nonparametric regression function and the asymptotic properties of the estimators, were extensively carried out by Robinson (1988)[12], Andrews (1995)[1], Qi Li (2000)[9], Hamilton (1997)[6], Liu et al. (1997)[10] etc. Later, Das et al. (2022)[4] proposed test of association between the sole nonparametric regressor variable and random error in the semiparametric regression model $Y = Z\beta + m(X) + \epsilon$ using popular nonparametric measures of association namely Kendall's τ, Bergsma (2006)[2]'s τ^* and Szekely et al. (2007)[13]'s *dcov* in the context of simple nonparametric regression and semiparametric model.

This article ventures out that erstwhile effort in more general set-up where more than one parametric and nonparametric regressors are involved as described in (1). If the regressors are jointly independent to ϵ then the homoskedasticity of error is inevitable; otherwise ϵ is heteroskedastic as its variance depends on the observed values of the $(p + q)$ regressors.

The entire article is organized as follows. Section 2 provides estimation of $\beta_1, \ldots, \beta_p$ using Robinson's method

*dsd.mt@brainwareuniversity.ac.in
†saranishika.maiti@Visvabharati.ac.in
‡apal@stat.buruniv.ac.in

as well as $g(\cdot,\dots,\cdot)$ by Nadaraya-Watson kernel density estimation method, followed by development of relevant hypotheses in Section 3. The nondegenerate V-statistics based on Kendall's τ and Spearman's ρ_s are constructed in Section 4 for further progress on test construction. Their asymptotic distributions are derived in Section 5, thereby leading for the computation of their asymptotic efficiencies and powers. Section 6 unfurls data analysis to check the utilities of the test statistics, by furnishing p-values, obtained through resampling technique for different sample sizes.

2 Estimation of model

The framework of generalized semiparametric regression model, in this article, has the following mathematical representation

$$
\begin{aligned}
& Y = \beta_1 X_1 + \ldots + \beta_p X_p + g(W_1,\dots,W_q) + \epsilon && (2)\\
\iff\quad & Y = \underset{\sim}{X}'\underset{\sim}{\beta} + g(\underset{\sim}{W}) + \epsilon, \text{ where } \underset{\sim}{X} = (X_1,\dots,X_p)',\ \underset{\sim}{W} = (W_1,\dots,W_q)' \text{and } \underset{\sim}{\beta} = (\beta_1,\dots,\beta_p)' && (3)\\
\iff\quad & Y_i = \underset{\sim}{X}_i'\underset{\sim}{\beta} + g(\underset{\sim}{W}_i) + \epsilon_i\,,\ i = 1,\dots,n && (4)
\end{aligned}
$$

where Y is scalar response and $\beta_1,\dots,\beta_p$ are p parameters corresponding to the covariates $X_1,\dots,X_p$. The nonparametric regression function $g(W_1,\dots,W_q)$, based on nonparametric regressors $W_1,\dots,W_q$, is a Lipschitz continuous function defined on $\mathbb{R}^q$. To estimate the parameters $\beta_1,\dots,\beta_p$ and $g(\cdot,\dots,\cdot)$, we apply Robinson (1988)[12]'s method. Taking $E(\cdot|\underset{\sim}{W}_i)$ to both sides of (4), we obtain $E(Y_i|\underset{\sim}{W}_i) = E(\underset{\sim}{X}_i|\underset{\sim}{W}_i)'\underset{\sim}{\beta} + g(\underset{\sim}{W}_i)$ and subtracting it from (4) we get

$$
\begin{aligned}
Y_i - g_Y(\underset{\sim}{W}_i) &= (\underset{\sim}{X}_i - g_{\underset{\sim}{X}}(\underset{\sim}{W}_i))'\underset{\sim}{\beta} + \epsilon_i \implies \epsilon_{Yi} = \epsilon_{\underset{\sim}{X}i}'\underset{\sim}{\beta} + \epsilon_i, \text{ where}\\
Y_i &= g_Y(\underset{\sim}{W}_i) + \epsilon_{Yi} && (5)\\
\underset{\sim}{X}_i &= g_{\underset{\sim}{X}}(\underset{\sim}{W}_i) + \epsilon_{\underset{\sim}{X}i},\ i = 1,\dots,n && (6)\\
\implies \tilde{\underset{\sim}{\beta}} &= \left(\sum_{i=1}^{n} \epsilon_{\underset{\sim}{X}i}\epsilon_{\underset{\sim}{X}i}'\right)^{-1}\left(\sum_{i=1}^{n} \epsilon_{\underset{\sim}{X}i}\epsilon_{Yi}\right). && (7)
\end{aligned}
$$

Here, $\tilde{\underset{\sim}{\beta}}$ is not a feasible estimator of β. We need to estimate the errors $\epsilon_{\underset{\sim}{X}i}$ and ϵ_{Yi} for $i = 1,\dots,n$ so that a feasible estimator of β can be determined. From models (5) and (6), the estimators of the regression functions $g_Y(\cdot)$ and $g_{\underset{\sim}{X}}(\cdot)$ are determined by applying the Nadaraya-Watson estimation of kernel density (NW) method. Note that,

$$
g_Y(\underset{\sim}{w}) \;=\; E(Y|\underset{\sim}{W} = \underset{\sim}{w}) \;=\; \int_{-\infty}^{\infty} y \cdot \frac{\mathcal{K}_{Y,\underset{\sim}{W}}(y,\underset{\sim}{w})}{\psi_{\underset{\sim}{W}}(\underset{\sim}{w})}\, dy \tag{8}
$$

where $f_{Y|\underset{\sim}{W}}(\cdot|\underset{\sim}{w})$ is the conditional $p.d.f.$ of $Y|\underset{\sim}{W}$ at $\underset{\sim}{W} = \underset{\sim}{w}$, $\mathcal{K}_{Y,\underset{\sim}{W}}(\cdot,\cdot)$ is the joint $p.d.f.$ of $(Y,\underset{\sim}{W})$ and $\psi_{\underset{\sim}{W}}(\cdot)$ is the $p.d.f.$ of $\underset{\sim}{W}$, all are assumed to exist. The kernel density of $\underset{\sim}{W}$ is estimated at $\underset{\sim}{w}$ as

$$
\widehat{\psi}_{\underset{\sim}{W}}(\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1\dots h_q}\; k\left(\frac{w_1 - W_{i1}}{h_1},\dots,\frac{w_q - W_{iq}}{h_q}\right) \;=\; \frac{1}{n}\sum_{i=1}^{n}\Big\{\prod_{j=1}^{q}\frac{1}{h_j}\; k_j\left(\frac{w_j - W_{ij}}{h_j}\right)\Big\} \tag{9}
$$

where $(W_{i1},\dots,W_{iq})' \equiv \underset{\sim}{W}_i$ for $i = 1,\dots,n$, $k_j(\cdot)$'s are the kernel density functions of W_j's, $j = 1,\dots,q$; $h_1,\dots,h_q$ are the bandwidths (> 0) for estimation of kernel density functions of $W_1,\dots,W_q$. In similar manner, the joint $p.d.f.$ of $(Y,\underset{\sim}{W})$ is estimated as

$$
\widehat{\mathcal{K}}_{Y,\underset{\sim}{W}}(y,\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_y}\,k_y\left(\frac{y - Y_i}{h_y}\right)\Big\{\prod_{j=1}^{q}\frac{1}{h_j}\,k_j\left(\frac{w_j - W_{ij}}{h_j}\right)\Big\} \tag{10}
$$

where h_y is the bandwidth for estimating the $p.d.f.$ of Y. Therefore, $\widehat{g}_Y(\cdot)$ is estimated as

$$\widehat{g}_Y(\underset{\sim}{w}) = \int_{-\infty}^{\infty} y \cdot \frac{\widehat{\mathcal{K}}_{Y,\underset{\sim}{W}}(y,\underset{\sim}{w})}{\widehat{\psi}_{\underset{\sim}{W}}(\underset{\sim}{w})}\, dy = \frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}\frac{1}{h_j}k_j\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}}{\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{q}\frac{1}{h_j}k_j\left(\frac{w_j - W_{ij}}{h_j}\right)}. \tag{11}$$

Similarly, $g_{\underset{\sim}{X}}(\underset{\sim}{w}) = E(\underset{\sim}{X}|\underset{\sim}{W} = \underset{\sim}{w})$ is estimated as $\widehat{g}_{\underset{\sim}{X}}(\underset{\sim}{w}) = \dfrac{\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}\frac{1}{h_j}k_j\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}\underset{\sim}{X}_i}{\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}\frac{1}{h_j}k_j\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}}$. Next, the errors are estimated as $\hat{\epsilon}_{Yi} = Y_i - \widehat{g}_Y(\underset{\sim}{W}_i)$ and $\hat{\epsilon}_{\underset{\sim}{X}i} = \underset{\sim}{X}_i - \widehat{g}_{\underset{\sim}{X}}(\underset{\sim}{W}_i)$. Then, a feasible estimator of $\underset{\sim}{\beta}$ is obtained as $\widehat{\underset{\sim}{\beta}} = \left(\sum_{i=1}^{n}\hat{\epsilon}_{\underset{\sim}{X}i}\hat{\epsilon}_{\underset{\sim}{X}i}^{'}\right)^{-1}\left(\sum_{i=1}^{n}\hat{\epsilon}_{\underset{\sim}{X}i}\hat{\epsilon}_{Yi}\right)$. Now, we have to estimate the nonparametric regression function $g(\cdot)$. Note that, the semiparametric model can be transformed to a nonparametric regression model as

$$Y_i - \underset{\sim}{X}_i^{'}\underset{\sim}{\beta} = g(\underset{\sim}{W}_i) + \epsilon_i \implies Y_i^{'} = g(\underset{\sim}{W}_i) + \epsilon_i,\ i = 1,\dots,n \tag{12}$$

where $Y_i^{'}$ is the *transformed response* as defined $(Y_i - \underset{\sim}{X}_i^{'}\underset{\sim}{\beta})$, $i = 1,\dots,n$. Now, based on the $i.i.d.$ observations $(Y_i^{'}, \underset{\sim}{W}_i)$'s, $i = 1,\dots,n$, the expression of $g(\underset{\sim}{w})$ is derived as

$$g(\underset{\sim}{w}) = E(Y^{'}|\underset{\sim}{W} = \underset{\sim}{w}) = \int_{-\infty}^{\infty} y^{'}\phi_{Y^{'}|\underset{\sim}{W}}(y^{'}|\underset{\sim}{w})\, dy^{'} = \int_{-\infty}^{\infty} y^{'} \cdot \frac{f_{Y^{'},\underset{\sim}{W}}(y^{'},\underset{\sim}{w})}{\psi_{\underset{\sim}{W}}(\underset{\sim}{w})}\, dy^{'} \tag{13}$$

where $\phi_{Y^{'}|\underset{\sim}{W}}(\cdot|\underset{\sim}{w})$ is the $p.d.f.$ of $(Y^{'}|\underset{\sim}{W})$ and $f_{Y^{'},\underset{\sim}{W}}(\cdot,\cdot)$ is the joint $p.d.f.$ of $(Y^{'},\underset{\sim}{W})$. Then,

$$\begin{aligned}\widehat{g}(\underset{\sim}{w}) &= \int_{-\infty}^{\infty} y^{'} \cdot \frac{\frac{1}{n}\sum_{i=1}^{n}\left\{k_{y^{'}}\left(\frac{y^{'} - Y_i^{'}}{h_{y^{'}}}\right)\prod_{j=1}^{q}k_{j;W_i}\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}}{\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}k_{j;W_i}\left(\frac{w_i - W_{ij}}{h_j}\right)\right\}}\, dy^{'} \\ &= \sum_{i=1}^{n} M_i Y_i^{'} \text{ where } M_i = \frac{\left\{\prod_{j=1}^{q}k_{j;W_i}\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}}{\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}k_{j;W_i}\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}},\ i = 1,\dots,n\end{aligned} \tag{14}$$

which is finally estimated as

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}k_{j;W_i}\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}\widehat{Y}_i^{'}}{\frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q}k_{j;W_i}\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}} = \sum_{i=1}^{n} M_i\widehat{Y}_i^{'} = \sum_{i=1}^{n} M_i(Y_i - \underset{\sim}{X}_i^{'}\widehat{\underset{\sim}{\beta}}).$$

3 Construction of Hypotheses and test statistics

Here, the hypotheses of interest to check independence between $(\underset{\sim}{X},\underset{\sim}{W})$ and ϵ are considered as follows.

$$H_0 : (\underset{\sim}{X},\underset{\sim}{W}) \perp\!\!\!\perp \epsilon \text{ against } H_1 : (\underset{\sim}{X},\underset{\sim}{W}) \not\perp\!\!\!\perp \epsilon \tag{15}$$

where '$\perp\!\!\!\perp$' stands for independence. However, development of any test procedure under (15) is quite inconvenient, as ϵ is an unobservable quantity. Therefore, a bonafide substitute of ϵ is required in (15) to re-frame the null hypothesis. Let us define r-th order difference of ϵ and Y as

$$\epsilon^*(r) = \sum_{j=1}^{r+1}(-1)^{j-1}\binom{r}{j-1}\epsilon_j \text{ and } Y^*(r) = \sum_{j=1}^{r+1}(-1)^{j-1}\binom{r}{j-1}Y_j.$$

where $\epsilon_1, \ldots, \epsilon_{r+1}$ and $Y_1, \ldots, Y_{r+1}$ are $(r+1)$ *i.i.d.* errors and responses respectively. Hence, by defining a general r-th order difference of ϵ^*, as denoted $\epsilon^*(r)$, we only verify if $\epsilon^*(r) \approx Y^*(r)$, where $Y^*(r)$ is the r-th order difference of Y^*.

Theorem 3.1. *$\epsilon^*(r)$, which has maximum k-th order absolute moment among all possible linear functions $\sum_{j=1}^{r+1} u_j\epsilon_j$ with real coefficients u_j's, is approximated as $Y^*(r)$.*

Under a simple partially linear regression setup $Y = Z\beta + m(X) + \epsilon$, **Das et al.** (2022)[4] considered third order difference of ϵ^*, to develop a test of independence between nonparametric covariate X and random error ϵ. The proof of Theorem 3.1 is elaborately discussed in Appendix-II. We are in quest of a general r-th order difference $\epsilon^*_{(r)}$. Therefore, we transform null hypothesis as $H_0 : (\underset{\sim}{X}, \underset{\sim}{W}) \perp\!\!\!\perp \epsilon^*(r)$ which approximately implies that $(\underset{\sim}{X}, \underset{\sim}{W}) \perp\!\!\!\perp Y^*(r)$. Furthermore, any function of $(\underset{\sim}{X}, \underset{\sim}{W})$ is independent to $Y^*(r)$.

Proposition 1. *$\widehat{Y}^*(r)$, the r-th order difference of $\widehat{Y}$, can be approximated as a function of $(\underset{\sim}{X}, \underset{\sim}{W})$.*

Therefore, H_0 further implies that $\widehat{Y}^*(r) \perp\!\!\!\perp Y^*(r)$ and the new hypotheses are formulated as

$$H_0 : \widehat{Y}^*(r) \perp\!\!\!\perp Y^*(r) \text{ against } H_1 : \widehat{Y}^*(r) \not\perp\!\!\!\perp Y^*(r). \tag{16}$$

Since the objective is to carry out a consistent test of 16, a conventional approach is to define a contiguous sequence of all possible alternative hypotheses with the aid of theory of contiguity due to Le Cam (1960)[3]. Such a sequence converges to the null hypothesis of interest as the sample size increases. Using Le Cam's first lemma, we construct the following sequence of contiguous alternatives (Das et al. (2022)[4])

$$H_n : \tilde{G}_{n;\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*, y^*) = \left(1 - \frac{\mu}{\sqrt{n}}\right) G_{0;\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*, y^*) + \frac{\mu}{\sqrt{n}} G_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*, y^*) \tag{17}$$

where $\tilde{G}_{n;\widehat{Y}^*(r),Y^*(r)}(\cdot,\cdot)$ is the joint cumulative distribution function of $(\widehat{Y}^*(r), Y^*(r))$ under H_n, $G_{0;\widehat{Y}^*(r),Y^*(r)}(\cdot,\cdot)$ and $G_{\widehat{Y}^*(r),Y^*(r)}(\cdot,\cdot)$ are the joint cumulative distribution functions of $(\widehat{Y}^*(r), Y^*(r))$ under H_0 and H_1 respectively. $\mu(> 0)$ is the mixing parameter. H_n is constructed as a sequence of contiguous alternatives provided that the corresponding joint densities of $(\widehat{Y}^*(r), Y^*(r))$ under H_0 and H_n and their marginal densities exist.

Next, we construct non-degenerate test statistics based on two nonparametric measures of association Kendall's τ and Spearman's ρ_s. A V- statistic, to propose an unbiased estimator of the parameter of interest θ based on the obtained random sample, is defined as

$$V_n = n^{-m}\sum_{u_1=1}^{n}\cdots\sum_{u_m=1}^{n}\psi((\widehat{y}^*_{u_1}(r), y^*_{u_1}(r)), \ldots, (\widehat{y}^*_{u_m}(r), y^*_{u_m}(r)))$$

where ψ is the kernel of V_n involving m *i.i.d.* pairs $(\widehat{y}^*_{u_1}(r), y^*_{u_1}(r)), \ldots, (\widehat{y}^*_{u_m}(r), y^*_{u_m}(r))$ such that

$$E[\psi((\widehat{y}^*_{u_1}(r), y^*_{u_1}(r)), \ldots, (\widehat{y}^*_{u_m}(r), y^*_{u_m}(r)))] = \theta$$

, $\{u_1, \ldots, u_m\}$ being the set of m distinct integers $\in \{1, \ldots, n\}$.
The concept of V-statistic is utilized to propose the test statistics, as provided below.

$$T_n^{<r>} = n^{-2}\sum_{u_1=1}^{n}\sum_{u_2=1}^{n} sign\{(\widehat{y}^*_{u_1}(r) - \widehat{y}^*_{u_2}(r))(y^*_{u_1}(r) - y^*_{u_2}(r))\} \tag{18}$$

$$S_n^{<r>} = 3n^{-3}\sum_{u_1=1}^{n}\sum_{u_2=1}^{n}\sum_{u_3=1}^{n} sign\{(\widehat{y}^*_{u_1}(r) - \widehat{y}^*_{u_2}(r))(y^*_{u_1}(r) - y^*_{u_3}(r))\}. \tag{19}$$

where $(\widehat{y}_1^*(r), y_1^*(r)), \ldots, (\widehat{y}_n^*(r), y_n^*(r))$ are *i.i.d.* samples on $(\widehat{Y}^*(r), Y^*(r))$. It is noteworthy that as an implication of H_0, $\tau(\widehat{Y}^*(r), Y^*(r)) = 0$ as well as $\rho_s(\widehat{Y}^*(r), Y^*(r)) = 0$. The measures take nonzero values under dependence of $(\widehat{Y}^*(r), Y^*(r))$. Note that, Kendall's τ has the kernel $sign\{(\widehat{y}_{u_1}^*(r) - \widehat{y}_{u_2}^*(r))(y_{u_1}^*(r) - y_{u_2}^*(r))\}$ and the kernel assumed by Spearman's ρ_s is $3sign\{(\widehat{y}_{u_1}^*(r) - \widehat{y}_{u_2}^*(r))(y_{u_1}^*(r) - y_{u_3}^*(r))\}$, $1 \leq u_1 \neq u_2 \neq u_3 \leq n$.

Proposition 2. *Under H_0, the kernel of $T_n^{<r>}$ has order of degeneracy 0.*

Proposition 3. *Under H_0, the kernel of $S_n^{<r>}$ has order of degeneracy 0.*

Hence, the asymptotic distributions of the test statistics 18 and 19 under both H_0 and H_n are asymptotically normal, as deduced in the next section.

4 Asymptotic distributions of test statistics

The asymptotic distribution of a nondegenerate V-statistic was determined by Zhou et al. (2021)[14]. In similar way, the asymptotic distributions of $S_n^{<r>}$ and $T_n^{<r>}$ under H_0 and H_n can be derived. Proofs are available in Das et al. (2022)[4].

Theorem 4.1. *Under H_0,*
$\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) \xrightarrow{L} N(0, 4\xi_1(r))$, *provided* $E[h^2((\widehat{Y}_1^*(r), Y_1^*(r)), (\widehat{Y}_2^*(r), Y_2^*(r)))] < \infty$,
where

$$\xi_1(r) \quad = \quad Var_{H_0}[E(h((\widehat{Y}_1^*(r), Y_1^*(r)), (\widehat{Y}_2^*(r), Y_2^*(r)))\Big|(\widehat{Y}_1^*(r), Y_1^*(r)))] \;=\; \frac{1}{9}.$$

Theorem 4.2. *Under H_n,*
$\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) \xrightarrow{L} N(\Upsilon^{(r)}, 4\xi_1(r))$, *where*

$$\Upsilon^{(r)} = \lim_{n\to\infty} Cov_{H_0}\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})), \log \frac{d\tilde{G}_n}{dG_0}\right).$$

In similar way the limiting distributions of $S_n^{<r>}$ are also derived, given by the following theorems.

Theorem 4.3. *Under H_0, provided that* $E[h^2((\widehat{Y}_1^*(r), Y_1^*(r)), (\widehat{Y}_2^*(r), Y_2^*(r)), (\widehat{Y}_3^*(r), Y_3^*(r)))] < \infty$,
$\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})) \xrightarrow{L} N(0, 4\nu_1(r))$, *where*

$$\nu_1(r) = Var_{H_0}[E\{h((\widehat{Y}_1^*(r), Y_1^*(r)), (\widehat{Y}_2^*(r), Y_2^*(r)), (\widehat{Y}_3^*(r), Y_3^*(r)))\Big|(\widehat{Y}_1^*(r), Y_1^*(r))\}] \;=\; 1.$$

Theorem 4.4. *Under H_n,*
$\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})) \xrightarrow{L} N(\Delta^{(r)}, 4\nu_1(r))$, *where* $\Delta^{(r)} = \lim_{n\to\infty} Cov_{H_0}\left(\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})), \log \frac{d\tilde{G}_n}{dG_0}\right)$.

4.1 Asymptotic efficiency

Before testing H_0 against H_n asymptotic efficiencies of $T_n^{<r>}$ and $S_n^{<r>}$ are investigated. Efficient test statistics harbor consistent tests. We determine the relative efficiencies (RE) of the test statistics due to Pitman, for increasing values μ under H_n. An m-dependent nondegenerate V-statistic V_n has mean square error (MSE) under H_n is

$$MSE_{H_n}(V_n) = Var_{H_n}(V_n) + (Bias_{H_n}(V_n))^2 = \frac{m^2\kappa_{1,m}}{n} + E^2_{H_n}(V_n - E(V_n)) = \frac{m^2\kappa_{1,m} + \theta^2}{n} = \frac{m^2\kappa_{1,m} + \mu^2\,\lambda^2}{n}. \tag{20}$$

where $\kappa_{1,m} = Var_{H_0}[\psi((a_1, b_1), \ldots, (a_m, b_m))|a_1, b_1]$; $(a_1, b_1), \ldots, (a_m, b_m)$ being m *i.i.d.* paired observations on the jointly distributed variables of interest (A, B). Also, $\theta = E_{H_n}(\sqrt{n}(V_n - E(V_n))) = \mu\lambda$ where

$$\lambda = \lim_{n\to\infty} Cov_{H_0}\left(\sqrt{n}(V_n - E_{H_0}(V_n)), \log \frac{d\tilde{G}_n}{dG_0}\right).$$

If $\mu \uparrow (\mu + D)$ for $D > 0$, then the relative efficiency of V_n for mixing parameter $(\mu + D)$ with respect to V_n for mixing parameter μ is computed as

$$RE_{\mu}^{(r)}(V_n) = \frac{MSE_{H_n}(V_n; \mu, r)}{MSE_{H_n}(V_n; \mu + D, r)} = \frac{\dfrac{m^2\kappa_{1,m}}{n} + \dfrac{\mu^2\lambda^2}{n}}{\dfrac{m^2\kappa_{1,m}}{n} + \dfrac{(\mu+D)^2\lambda^2}{n}}. \tag{21}$$

For large n, $\dfrac{m^2\kappa_{1,m}}{n} \simeq 0$. Therefore $MSE_{H_n}(V_n) \simeq \dfrac{\mu^2\lambda^2}{n}$, which further implies $eff_{H_n}(V_n; \mu, r) \simeq F^* \cdot \dfrac{n}{\mu^2\,\lambda^2}$ where F^* is the inverse of *Fisher's information* for the joint distribution of (A, B). Then,

$$AARE_{\mu;D}^{(r)}(V_n) \simeq \frac{\dfrac{\mu^2\lambda^2}{n}}{\dfrac{(\mu+D)^2\lambda^2}{n}} = \left(\frac{\mu}{\mu + D}\right)^2. \tag{22}$$

which is *asymptotic approximated relative efficiency (AARE)* of $V_{n;\mu+D}$ compared to $V_n;\mu$. Clearly, $AARE_{\mu}^{(r)}(V_n) \uparrow$ as $\mu \uparrow$, for $n \to \infty$. Also, it is independent of order of difference r and the order of V_n *i.e.* m. From (17), it is evident that $\left(1 - \dfrac{\mu}{\sqrt{n}}\right) > 0$ as well as $\dfrac{\mu}{\sqrt{n}} > 0$, which implies that $0 \leq \mu \leq \sqrt{n}$. Figure 1 captures AARE of V_n behaves as the values of μ differ.

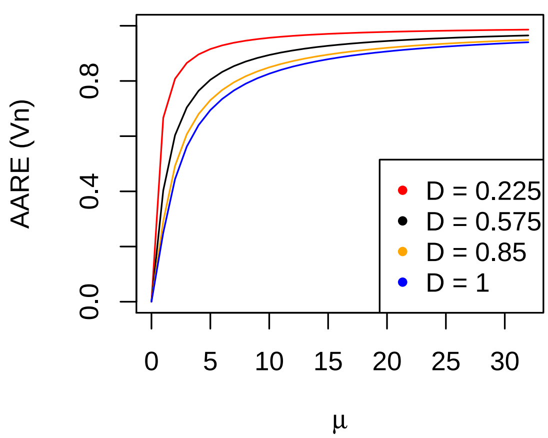


Figure 1: AAREs of V_n against μ for $n = 1000$

Smaller the values of D, steeper is $AARE$s curve. The AAREs of $T_n^{<r>}$ and $S_n^{<r>}$ are indeed the same and the calculations are presented in Appendix-I.

4.2 Asymptotic power

Another necessary characteristic of an appropriate test statistic is its power, specially when the sample size is large. The asymptotic power of an efficient V-statistic establishes its worth in testing independence between jointly distributed random variables. The limiting distributions of $T_n^{<r>}$ and $S_n^{<r>}$, as presented beforehand, are required to formulate the expressions for their asymptotic powers. $T_n^{<r>}$'s asymptotic power is computed as

$$P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\alpha) \quad = \quad 1 - \Phi\left(\frac{\pi_\alpha - \Upsilon^{(r)}}{\sqrt{4\xi_1(r)}}\right) \;=\; \Phi\left(\frac{\Upsilon^{(r)} - \pi_\alpha}{\sqrt{4\xi_1(r)}}\right)$$

where π_α is a point that satisfies $P_{H_0}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\alpha) = \alpha,\ 0 < \alpha < 1$. $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. If $\mu = 0$, then $\Upsilon^{(r)} = 0$, implying the asymptotic power and size are equal. Similarly, the asymptotic power of $S_n^{<r>}$ is calculated as

$$P_{H_n}(\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})) > \pi_\beta) \quad = \quad 1 - \Phi\left(\frac{\pi_\beta - \Delta^{(r)}}{\sqrt{4\nu_1(r)}}\right) \quad = \quad \Phi\left(\frac{\Delta^{(r)} - \pi_\beta}{\sqrt{4\nu_1(r)}}\right)$$

where π_β satisfies $P_{H_0}(\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})) > \pi_\beta) = \beta$ for $0 < \beta < 1$. $\mu = 0$ implies $\Delta^{(r)} = 0$, yielding asymptotic power as the size of the test.

In general, $\Upsilon^{(r)}$ is determined as $(2P_c - 1)$ where P_c denotes the probability of concordance of **two** arbitrary paired observations among n *i.i.d.* samples from $(\widehat{Y}^*(r), Y^*(r))$ for computation of τ. P_c is expressed as

$$P_c = P(\widehat{Y}^*_1(r) < \widehat{Y}^*_2(r), Y_1^*(r) < Y_2^*(r)) + P(\widehat{Y}^*_1(r) > \widehat{Y}^*_2(r), Y_1^*(r) > Y_2^*(r)).$$

The explicit expression of P_c was provided by Dhar et al. (2018)[5]. On the other hand, $\Delta^{(r)} = 3P_c^{'} - 3P_d^{'}$ where $P_c^{'}$ is the probability of concordance of **three** arbitrary paired observations among n *i.i.d.* samples from $(\widehat{Y}^*(r), Y^*(r))$ and $P_d^{'}$ denotes the probability of discordance of those three paired observations for computation of ρ_s, expressed as

$$P_c^{'} = 3P(\widehat{Y}^*_1(r) < \widehat{Y}^*_2(r), Y_1^*(r) < Y_3^*(r)) + 3P(\widehat{Y}^*_1(r) > \widehat{Y}^*_2(r), Y_1^*(r) > Y_3^*(r))$$

$$\text{and } P_d^{'} = 3P(\widehat{Y}^*_1(r) > \widehat{Y}^*_2(r), Y_1^*(r) < Y_3^*(r)) + 3P(\widehat{Y}^*_1(r) < \widehat{Y}^*_2(r), Y_1^*(r) > Y_3^*(r)).$$

The next proposition is presented in order to establish consistency of the two test statistics $T_n^{<r>}$ and $S_n^{<r>}$.

Proposition 4. *(i) For $n^* > n$, $P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\alpha) < P_{H_n}(\sqrt{n^*}(T_{n^*}^{<r>} - E(T_{n^*}^{<r>})) > \pi_\alpha)$ and $P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\alpha) \uparrow 1$ as $\mu \uparrow$ and $n \to \infty$.*

(ii) For $n^* > n$, $P_{H_n}(\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})) > \pi_\beta) < P_{H_n}(\sqrt{n^*}(S_{n^*}^{<r>} - E(S_{n^*}^{<r>})) > \pi_\beta)$ and $P_{H_n}(\sqrt{n}(S_n^{<r>} - E_{H_0}(S_n^{<r>})) > \pi_\beta) \uparrow 1$ as $\mu \uparrow$ and $n \to \infty$.

We consider a pair of examples to assess the power performances of the test statistics against μ by taking sample size $n = 1000$ and $r = 2, 3, 4, 5, 10$ as the orders of difference. Also, $\epsilon \sim N(0, 0.015)$ and $Y \sim t_2$ (t-distribution with 2 d.f.) under H_0.

Example 1. *Let us consider a generalized partially linear model $Y = \beta_1 X_1 + \beta_2 X_2 + g(W_1, W_2) + \epsilon$ with error assumptions*
(i) $E(\epsilon | X_1 = x_1, X_2 = x_2, W_1 = w_1, W_2 = w_2) = 0$ for all (x_1, x_2, w_1, w_2),
(ii) $E(\epsilon^2 | X_1 = x_1, X_2 = x_2, W_1 = w_1, W_2 = w_2) = \sigma^2(x_1, x_2, w_1, w_2)$. Moreover, $(X_1, X_2, W_1, W_2) \sim \mathbb{N}_4 \left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \left(\begin{matrix}0.18 & -0.06\\ -0.06 & 0.14\\ 0.22 & -0.28\\ -0.13 & 0.19\end{matrix}\right.\right.$
The nonparametric regression function $g(W_1, W_2) = 0.45W_1W_2 - 0.25W_1^2W_2 + W_2^3$. The conditional error is distributed as $\left(\epsilon \,\middle|\, X_1, X_2, W_1, W_2\right) \sim N\left(0,\ 0.015\left|1 - 0.13X_1 - 2.1X_2 + 5.6W_1 - 0.95W_2\right|\right)$ under H_1.

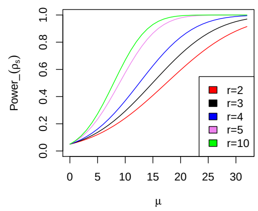
Example 2. *Another generalized partially linear model is considered as $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + g(W_1, W_2, W_3) + \epsilon$ with assumptions on ϵ as*
(i) $E(\epsilon | X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5, W_1 = w_1, W_2 = w_2, W_3 = w_3) = 0$ for all $(x_1, x_2, x_3, x_4, x_5, w_1, w_2, w_3)$,
(ii) $E(\epsilon^2 | X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5, W_1 = w_1, W_2 = w_2, W_3 = w_3) = \sigma^2(x_1, x_2, x_3, x_4, x_5, w_1, w_2, w_3)$ for all $(x_1, x_2, x_3, x_4, x_5, w_1, w_2, w_3)$. The joint distribution of $(X_1, X_2, X_3, X_4, X_5, W_1, W_2, W_3)^T$ is

$$\mathbb{N}_8\left(\begin{pmatrix}0\\0\\0\\0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix} 0.18 & -0.06 & 0.09 & -0.13 & 0.16 & -0.17 & 0.31 & -0.22\\ -0.06 & 0.14 & -0.17 & 0.26 & -0.14 & 0.18 & 0.25 & -0.33\\ 0.09 & -0.17 & 0.25 & 0.33 & -0.18 & -0.24 & -0.15 & 0.15\\ -0.13 & 0.26 & 0.33 & 0.32 & 0.05 & -0.14 & -0.07 & 0.12\\ 0.16 & -0.14 & -0.18 & 0.05 & 0.24 & 0.26 & -0.18 & -0.03\\ -0.17 & 0.18 & -0.24 & -0.14 & 0.26 & 0.11 & -0.21 & -0.19\\ 0.31 & 0.25 & -0.15 & -0.07 & -0.18 & -0.21 & 0.27 & 0.14\\ -0.22 & -0.33 & 0.15 & 0.12 & -0.03 & -0.19 & 0.14 & 0.40 \end{pmatrix}\right).$$

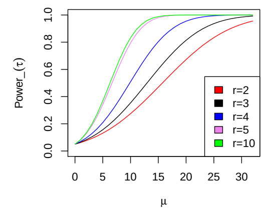
We take $g(W_1, W_2, W_3) = 0.36W_1^3 - 0.25W_2^2W_3 - 0.11W_3^2W_1 + 0.08W_1W_2W_3$. $(\epsilon | X_1, X_2, X_3, X_4, X_5, W_1, W_2, W_3) \sim N(0, 0.015\,|1 + 8X_1 + 5X_2 - 7X_3 - 4X_4 + 3.1X_5 - 6W_1 + 3W_2 + 6W_3|)$ under H_1.

The asymptotic powers curves, as furnished below, project that

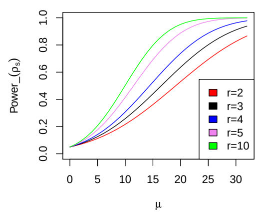
1. The asymptotic power increases as we increase the the order of difference (r).
2. Moreover under lower order Gaussian set-up (Example 1), $S_n^{<r>}$ portrays comparatively abrupt sloping while power curve in $T_n^{<r>}$ exhibits similar pattern in higher and lower order of Gaussian distributions.



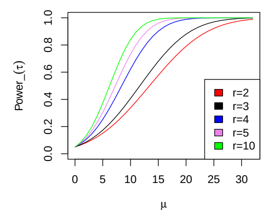
Asymptotic powers of $S_n^{<r>}$ in Example 1



Asymptotic powers of $T_n^{<r>}$ in Example 1



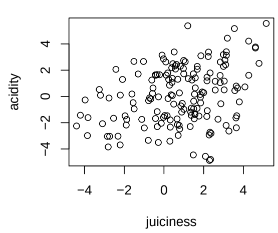
Asymptotic powers of $S_n^{<r>}$ in Example 2



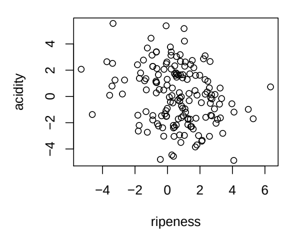
Asymptotic powers of $T_n^{<r>}$ in Example 2

5 Real data analysis

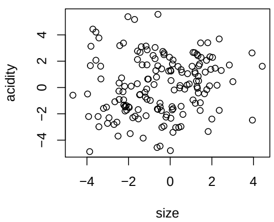
We consider *Apple quality dataset* (url: `https://www.kaggle.com/datasets/nelgiriyewithana/apple-quality`) in order to investigate the efficiency of $S_n^{<r>}$'s and $T_n^{<r>}$'. This data has 9 variables, *viz.*, 'A-id' (unique identifier for each apple), 'size' (size of apple'), 'weight' (weight of apple), 'sweetness' (degree of sweetness of apple), 'crunchiness' (texture indicating the crunchiness of the apple), 'juiciness' (level of juiciness of apple), 'ripeness' (stage of ripeness of apple), 'quality' (overall quality of apple) and 'acidity' (acidity level of apple). We study the regressand'acidity' with the help of 6 independent variables 'size', 'weight', 'sweetness', 'crunchiness', 'juiciness' and 'ripeness'. Primarily, volume of the dataset is trimmed to 150 from actual size 4001 to avoid computational complications, followed by scatterplots showing the association of 'acidity' with each of the independent variables.



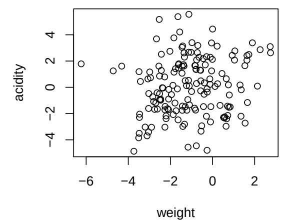
Scatterplot of acidity vs juiciness



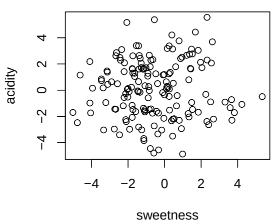
Scatterplot of acidity vs ripeness



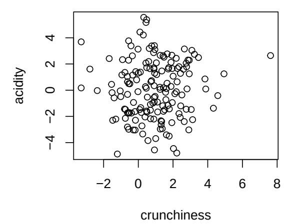
Scatterplot of acidity vs size



Scatterplot of acidity vs weight



Scatterplot of acidity vs sweetness



Scatterplot of acidity vs crunchiness

From the scatterplots, presented above, one can observe that both ‘juiciness’ and ‘ripeness’ share almost linear relationships with ‘acidity’ while first one showing positive linear association and the second one having negative linear relationship with ‘acidity’. The association of ‘acidity’ with rest of the four variables is not explicit at all, hence nonparametric relationships may be assumed between those four regressors and the response variable ‘acidity’.

The proposed partially linear regression model is a $Y = \beta_1 X_1 + \beta_2 X_2 + g(W_1, W_2, W_3, W_4) + \epsilon$ where X_1 and X_2 denote ‘ripeness’ and ‘juiciness’ and W_1, W_2, W_3, W_4 symbolize ‘size’, ‘crunchiness’, ‘weight’ and ‘sweetness’ respectively with Y indicating ‘acidity’. By Robinson’s method as described in Section 2, estimated values of β_1 and β_2 are obtained as $\widehat{\beta}_1 \approx 0.0377$ and $\widehat{\beta}_2 \approx -0.0842$. The r-th order differences of Y and $\widehat{Y}$ are calculated, followed by computations of test statistics. Furthermore, bootstrap samples are generated from the set of bivariate observations on the r-th order differences of Y and $\widehat{Y}$ and the test statistics are evaluated at all stages of resampling. The p-values of $S_n^{<r>}$ and $T_n^{<r>}$ are estimated at chosen level of significance $\alpha = 0.05$ for $r = 2, 3, 4, 5, 10$, presented below. It is to be noted that the more the order of difference, p-value suggests less evidence in favour of H_0.

Resampling size (B)	**p-values of $S_n^{<r>}$**					**p-values of $T_n^{<r>}$**				
	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$
500	0.082	0.063	0.055	0.049	0.044	0.064	0.056	0.054	0.048	0.04
1000	0.064	0.06	0.054	0.048	0.039	0.058	0.055	0.051	0.046	0.036

Table 1: Simulated p-values of $S_n^{<r>}$ and $T_n^{<r>}$ for $r = 2, 3, 4, 5, 10$

6 Concluding remarks

In this article, we have constructed some consistent tests, pertinent for generalized partially linear regression setup. The test statistics are based on two rank based measures Spearman’s ρ_s and Kendall’s τ and they are nondegenerate V-statistics. Such tests are evident to check whether the assumptions on error ϵ are relevant or not. If independence between the regressors and error holds further we proceed on the validity of homoscedastic error. Both the tests based on $S_n^{<r>}$ and $T_n^{<r>}$ are efficient as well as consistent with improving order of difference r. Also $T_n^{<r>}$’s performance is better compared to $S_n^{<r>}$ as usual. One may perform this test by considering a degenerate test statistic. The order of degeneracy indeed affects the power of a nonparametric test statistic. It is noteworthy that incorporation of more regressors in underlying model enhances the asymptotic powers of the concerning test statistics as shown by Example 1 and Example 2. Apart from ρ_s and τ, there can be numerous measures of association providing nondegenerate test statistics.

Since the limiting local powers of $S_n^{<r>}$ and $T_n^{<r>}$ increase as $r \uparrow$, the p-values of the test statistics would go down in such instance(s). The real data study reveals that the probability of rejection of true null hypothesis gets lowered when the test statistic of interest involves higher order difference of observed and estimated acidity level.

As a meaningful conclusion, the association between higher order difference of estimated response and observed response causes monotonicity in power performances of test statistics in this regard. Furthermore, as a prospective future introspection, r can be improved from 10 to achieve more powerful tests.

7 Acknowledgement

The authors are thankful to the anonymous referee as well as the editor of the journal for considering the article.

References

[1] Andrews, D. W. (1995). Nonparametric kernel estimation for semiparametric models. Econometric Theory, 11(3), 560-586.

[2] Bergsma, W. P. (2006). A new correlation coefficient, its orthogonal decomposition and associated tests of independence. *https://doi.org/10.48550/arXiv.math/0604627.*

[3] Cam Le, L. (1960). Locally asymptotically normal families of distributions. University of California Publications in Statistics, 3, 37-98.

[4] Das, S. & Maiti, S. I. (2022). On the Test of Association Between Nonparametric Covariate and Error in Semiparametric Regression Model. *Journal of the Indian Society for Probability and Statistics*, 23(2), 541-564.

[5] Dhar, S. S., Dassios, A., and Bergsma, W. (2018). Testing Independence of Covariates and Errors in Nonparametric regression. Scandinavian Journal of Statistics, 45, 421-443.

[6] Hamilton, S. A., & Truong, Y. K. (1997). Local linear estimation in partly linear models. Journal of Multivariate Analysis, 60(1), 1-19.

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

[8] Lévy, P. (1939). Sur la division d'un segment par des points choisis au hasard. CR Acad. Sci. Paris, 208, 147-149.

[9] Li, Q. (2000). Efficient estimation of additive partially linear models. International Economic Review, 41(4), 1073-1092.

[10] Liu, Z., Liu, Z., Lu, X., & Lu, X. (1997). Root-n-consistent semiparametric estimation of partially linear models based on k-nn method. Econometric Reviews, 16(4), 411-420.

[11] Pyke, R. (1965). Spacings. Journal of the Royal Statistical Society: Series B (Methodological), 27(3), 395-436.

[12] Robinson, P. M. (1988). Root-N-consistent semiparametric regression. Econometrica, 56(4), 931-954.

[13] Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. The Annals of Statistics, 35(6), 2769-2794.

[14] Zhou, Z., Mentch, L., & Hooker, G. (2021). V-statistics and variance estimation. Journal of Machine Learning Research, 22(287), 1-48.

8 Appendix-I

μ	Powers of $S_n^{<r>}$					Powers of $T_n^{<r>}$				
	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
1	0.0605	0.0624	0.0651	0.072	0.0754	0.0617	0.0645	0.0697	0.081	0.0832
5	0.1203	0.1369	0.1614	0.2335	0.2728	0.1309	0.1561	0.2086	0.3406	0.3669
10	0.2413	0.2934	0.3696	0.5753	0.6686	0.2744	0.3534	0.509	0.7947	0.8326
15	0.4088	0.5029	0.6267	0.8659	0.9301	0.4693	0.6019	0.8041	0.9802	0.9884
20	0.5952	0.7115	0.8362	0.9786	0.9941	0.672	0.8139	0.9545	0.9995	0.9998
25	0.7619	0.8662	0.949	0.9984	0.9998	0.8335	0.9366	0.9942	1	1
30	0.8817	0.9515	0.989	0.9999	1	0.932	0.9847	0.9996	1	1

Table 2: Asymptotic powers of $S_n^{<r>}$ and $T_n^{<r>}$ for $r = 2, 3, 4, 5, 10$ in Example 1

μ	Powers of $S_n^{<r>}$					Powers of $T_n^{<r>}$				
	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$	$r=2$	$r=3$	$r=4$	$r=5$	$r=10$
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
1	0.0595	0.0612	0.0629	0.0666	0.0695	0.064	0.0663	0.0736	0.078	0.0837
5	0.1123	0.1258	0.1415	0.1763	0.2066	0.1511	0.1735	0.2526	0.3042	0.3727
10	0.2165	0.2585	0.3076	0.415	0.5032	0.3377	0.4066	0.6225	0.7325	0.8402
15	0.3617	0.4406	0.5272	0.6916	0.798	0.577	0.6801	0.9015	0.9602	0.9897
20	0.5305	0.6364	0.7387	0.8879	0.9516	0.7903	0.8794	0.9884	0.998	0.9999
25	0.6939	0.8016	0.8869	0.9732	0.9936	0.9222	0.9697	0.9994	1	1
30	0.8257	0.9108	0.9626	0.9959	0.9995	0.979	0.9951	1	1	1

Table 3: Asymptotic powers of $S_n^{<r>}$ and $T_n^{<r>}$ for $r = 2, 3, 4, 5, 10$ in Example 2

9 Appendix-II

9.1 Proof of Theorem 3.1

Note that, for $i = 1, \ldots, n$ and $\delta > 0$,

$$\begin{aligned} P\left(|(Y_{i+1} - Y_i) - (\epsilon_{i+1} - \epsilon_i)| > \delta\right) &= P\left(|\underset{\sim}{\beta}^T(\underset{\sim}{X}_{i+1} - \underset{\sim}{X}_i) + \{g(\underset{\sim}{W}_{i+1}) - g(\underset{\sim}{W}_i)\}| > \delta\right) \\ &\leq P\left(|\sum_{s=1}^{p} \beta_s (X_{i+1\,s} - X_{i\,s})| + |g(\underset{\sim}{W}_{i+1}) - g(\underset{\sim}{W}_i)| > \delta\right). \end{aligned} \tag{23}$$

Due to Lipschitz continuity of $g(\cdot)$, therefore $|g(\underset{\sim}{W}_{i+1}) - g(\underset{\sim}{W}_i)| \leq C \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q$, $C > 0$.

$\therefore (23) \leq P\left(|\sum_{s=1}^{p} \beta_s (X_{i+1\,s} - X_{i\,s})| + C \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta\right) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} P\left(||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta'\right) \times \prod_{s=1}^{p} dH_{D_{is}}(d_{is})$,

where $D_{is} = X_{i+1\,s} - X_{i\,s}$ for $s = 1, \ldots, p$ and $\delta' = C^{-1}\left(\delta - |\sum_{s=1}^{p} \beta_s d_{is}|\right)$. Moreover,

$$P\left[\left(\sum_{m=1}^{q} |W_{i+1\,m} - W_{i\,m}|^q\right)^{\frac{1}{q}} > \delta'\right] = P\left[|W_{i+1\,1} - W_{i\,1}| > \left(\delta'^{\,q} - \left(\sum_{m=2}^{q} |W_{i+1\,m} - W_{i\,m}|^q\right)\right)^{\frac{1}{q}}\right] \tag{24}$$

for $m = 2, \ldots, q$; $i = 1, \ldots, n-1$. Next, the model is re-expressed by ordering the W_1-observations as

$$Y_K^* \quad = \quad \beta_1 X_{K1}^* + \ldots + \beta_p X_{Kp}^* + m(W_{K1}^*, \ldots, W_{Kq}^*) + \epsilon_K^*; \;\; K = 1, \ldots, n$$

where $\{W_{11}^*, \ldots, W_{n1}^*\}$ are the n observations on W_1 such that $W_{11}^* \leq \ldots \leq W_{n1}^*$. Corresponding to the ordered observations of W_1, the observations on $W_2, \ldots, W_q$ as well as $X_1, \ldots, X_p$ are the induced ordered observations. The responses $\{Y_1^*, \ldots, Y_n^*\}$ are called induced ordered responses corresponding to $\{W_{11}^*, \ldots, W_{n1}^*\}$.

Then, (24) is further deduced as $\int_{\mathbb{R}} \ldots \int_{\mathbb{R}} P\left[|W_{i+1\,1}^* - W_{i\,1}^*| > \delta^{''}\right] \times \prod_{m=2}^{q} dF_{T_{im}^*}(t_{im}^*)$, where $T_{im}^* = W_{i+1\,m}^* - W_{i\,m}^*$ and $\delta^{''} = \left({\delta'}^{q} - \left(\sum_{m=2}^{q} |t_{im}^*|^q\right)\right)^{\frac{1}{q}}$. Observe that, $(W_{i+1\,1}^* - W_{i1}^*)$ is the i-th spacing (Pyke (1965))[11] on W_1, $i = 1, \ldots, n-1$. Suppose $F_{W_1}(\cdot)$ is the CDF of W_1. For any $\delta^* > 0$, it is possible to deduce $P\left[|W_{i+1\,1}^* - W_{i1}^*| > \delta^*\right]$ as $P\left[\frac{n\Delta_{n-1;n}}{\log\, n} > \frac{nR}{\log\, n}\right]$, where $\Delta_{n-1;n}$ is the maximal spacing based on $(n-1)$ uniform spacings $(U_{i+1\,1}^* - U_{i1}^*)$s and $R = \inf_{1 \leq i \leq n-1} \delta^*/|F_{W_1}^{-1'}(\xi_{i;i+1})|$. Due to to **Lévy** (1939)[8], one can verify that

$$P\left[\frac{n\Delta_{n-1;n}}{\log\, n} > \frac{nR}{\log\, n}\right] \;\longrightarrow\; 1 - \exp\left(-\exp\left(-\frac{nR}{\log\, n}\right)\right). \tag{25}$$

Therefore, $\int_{\mathbb{R}} \ldots \int_{\mathbb{R}} P\left[|W_{i+1\,1}^* - W_{i\,1}^*| > \delta^{''}\right] \times \prod_{m=2}^{q} dF_{T_{im}^*}(t_{im}^*) \leq 1 - \exp\left(-\exp\left(-\frac{nR}{\log\, n}\right)\right) \longrightarrow 0$, which further implies the R.H.S. of (25) tends to $1 - e^0 = 0$. Then, $|W_{i+1\,1}^* - W_{i\,1}^*| = o_p(1) \implies \sup_{i \in \{1,\ldots,n-1\}} |W_{i+1\,1}^* - W_{i\,1}^*| = o_p(1)$.

Hence, $P\left(|(Y_{i+1}^* - Y_i^*) - (\epsilon_{i+1}^* - \epsilon_i^*)| > \delta\right) \longrightarrow 0$, $i = 1, \ldots, n-1$, where $(Y_{i+1}^* - Y_i^*)$ is the first order difference of Y^* and $(\epsilon_{i+1}^* - \epsilon_i^*)$ the first order difference of ϵ^*. Next, we need to verify if $\epsilon^*(r) \approx Y^*(r)$, where $Y^*(r)$ is the r-th order difference of Y and $\epsilon^*(r)$ is the r-th order difference of ϵ. Define $L(\epsilon_1, \ldots, \epsilon_{r+1}) = \alpha_1\epsilon_1 + \ldots + \alpha_{r+1}\epsilon_{r+1}$ with $\alpha_1, \ldots, \alpha_{r+1} \in \mathbb{Z}$ and $\sum_{i=1}^{r+1} \alpha_i = 0$, $\epsilon_1, \ldots, \epsilon_{r+1}$ are $(r+1)$ *i.i.d.* errors. The k-th order absolute moment of $L(\epsilon_1, \ldots, \epsilon_{r+1})$ is

$$\begin{aligned} E\Big|L(\epsilon_1, \ldots, \epsilon_{r+1})\Big|^k = E\Big|\sum_{i=1}^{r+1} \alpha_i\epsilon_i\Big|^k \;\; &\leq \;\; E\left[\sqrt{\sum_{i=1}^{r+1} \alpha_i^2} \cdot \sqrt{\sum_{i=1}^{r+1} \epsilon_i^2}\right]^k \;\; \text{(using Cauchy-Schwartz inequality)} \\ &= \;\; \left(\sum_{i=1}^{r+1} \alpha_i^2\right)^{k/2} E\left[\sqrt{\sum_{i=1}^{r+1} \epsilon_i^2}\right]^k \;=\; \eta(\alpha_1, \ldots, \alpha_{r+1}) \cdot E\left[\sqrt{\sum_{i=1}^{r+1} \epsilon_i^2}\right]^k \end{aligned} \tag{26}$$

where $\eta(\alpha_1, \ldots, \alpha_{r+1}) = \left(\sum_{i=1}^{r+1} \alpha_i^2\right)^{k/2}$. Observe that, $\eta(\alpha_1, \ldots, \alpha_{r+1}) \leq \max_{\alpha_1, \ldots, \alpha_{r+1} \neq 0} \eta(\alpha_1, \ldots, \alpha_{r+1})$ with $\sum_{i=1}^{r+1} \alpha_i = 0$. Taking $r = 2$, one can deduce $\eta(\alpha_1, \alpha_2, \alpha_3) = 2\left(\frac{\alpha_3^3 - \alpha_1^3}{\alpha_3 - \alpha_1}\right) = S(\alpha_1, \alpha_3)$, say. Then, $\log S = \log\, 2 + \log\,(\alpha_3^3 - \alpha_1^3) - \log\,(\alpha_3 - \alpha_1)$. Maximizing $\log S$ with respect to α_1, α_3 *i.e.* solving the following equations

$$\frac{\partial \log S}{\partial \alpha_1} = 0 \quad \implies \quad -2\alpha_1^2 + \alpha_1\alpha_3 + \alpha_3^2 = 0 \tag{27}$$

$$\text{and} \quad \frac{\partial \log S}{\partial \alpha_3} = 0 \quad \implies \quad 2\alpha_3^2 - \alpha_1^2 - \alpha_1\alpha_3 = 0 \tag{28}$$

one gets $\alpha_1 = \pm\alpha_3$. But α_i's $\neq 0$, hence $\alpha_1 = \alpha_3 \implies \alpha_2 = -2\alpha_1$. So, $\eta(\alpha_1, \alpha_2, \alpha_3)$ has maximum value at $(\alpha_1, -2\alpha_1, \alpha_1)$. Also, from (26), the maximum value of k-th order absolute moment of $(\alpha_1\epsilon_1 - 2\alpha_1\epsilon_2 + \alpha_1\epsilon_3)$ satisfies $E|\alpha_1\epsilon_1 - 2\alpha_1\epsilon_2 + \alpha_1\epsilon_3|^k \leq |\alpha_1|^k\{1^2 + (-2)^2 + 1^2\}^{k/2}\, E\left(\sqrt{\sum_{i=1}^{3} \epsilon_i^2}\right)^k$, for any nonzero integer α_1. If $\alpha_1 = \pm 1$, the linear contrast $\pm(\epsilon_1 - 2\epsilon_2 + \epsilon_3)$ has minimum variance as well as maximum k-th absolute raw moment among all possible linear contrasts $(\delta_1\epsilon_1 + \delta_2\epsilon_2 + \delta_3\epsilon_3)$. Here $L(\epsilon_1, \epsilon_2, \epsilon_3) = \epsilon_1 - 2\epsilon_2 + \epsilon_3$ is termed as the *second order difference of ϵ* based on $\epsilon_1, \epsilon_2, \epsilon_3$, which is indeed a first order difference of two first order differences of ϵ, i.e. $(\epsilon_1 - \epsilon_2) - (\epsilon_2 - \epsilon_3) = \epsilon_1^*(1) - \epsilon_2^*(1) \approx Y_1^*(1) - Y_2^*(1), \implies \epsilon^*(2) \approx Y^*(2)$.

Similarly, one can verify that for $r = 3$, $\eta(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2$ subjected to $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$ has maximum value at $(-\alpha_4, 3\alpha_4, \alpha_4)$, $\alpha_4 \neq 0$ and $E| - 3\alpha_4\epsilon_1 - \alpha_4\epsilon_2 + 3\alpha_4\epsilon_3 + \alpha_4\epsilon_4|^k \geq E|t_1\epsilon_1 + t_2\epsilon_2 + t_3\epsilon_3 + t_4\epsilon_4|^k$; $t_1, t_2, t_3, t_4 \in \mathbb{R}$, i.e. $|\alpha_4|^k E|\epsilon_4 - 3\epsilon_1 + 3\epsilon_3 - \epsilon_2|^k$ is maximum among the k-th order absolute moment of all possible linear functions $(t_1\epsilon_1 + t_2\epsilon_2 + t_3\epsilon_3 + t_4\epsilon_4)$. For $\alpha_4 = \pm 1$, the linear contrast $\pm(\epsilon_4 - 3\epsilon_1 + 3\epsilon_4 - \epsilon_2)$ [or equivalently, $\pm(\epsilon_1 - 3\epsilon_2 + 3\epsilon_3 - \epsilon_4)$ as the errors are i.i.d.] has the minimum variance as well as maximum k-th order moment among all the k-th order absolute moments of linear functions $(t_1\epsilon_1 + t_2\epsilon_2 + t_3\epsilon_3 + t_4\epsilon_4)$. The function $(\epsilon_1 - 3\epsilon_2 + 3\epsilon_3 - \epsilon_4)$ is denoted as the *third order difference of ϵ* based on four i.i.d. observations ϵ_1, ϵ_2, ϵ_3 and ϵ_4, which is the first order difference of two second order differences of ϵ as $\{(\epsilon_1 - 2\epsilon_2 + \epsilon_3) - (\epsilon_2 - 2\epsilon_3 + \epsilon_4)\}$, or the second order difference of three first order differences of ϵ's as $\{(\epsilon_1 - \epsilon_2) - 2(\epsilon_2 - \epsilon_3) + (\epsilon_3 - \epsilon_4)\}$. Similarly, $\epsilon^*(3) \approx Y^*(3)$.

Therefore, the second and third order differences of i.i.d. errors constitute best possible linear functions having highest second and third order absolute moments among all possible linear functions of errors respectively. It can be concluded finally that for a general order difference r of ϵ, $\epsilon^*(r) \approx Y^*(r)$. Also, $\epsilon^*(r)$ has the maximal k-th order absolute moment among all possible linear functions of $\epsilon_1, \ldots, \epsilon_{r+1}$.

9.2 Proof of Proposition 1

Taking $p = 1 = q$, one can obtain $\widehat{\beta} = \left(\sum_{i=1}^{n} \hat{\epsilon}_{Xi}\hat{\epsilon}_{Xi}^T\right)^{-1} \left(\sum_{i=1}^{n} \hat{\epsilon}_{Xi}\hat{\epsilon}_{Yi}\right)$ where $\hat{\epsilon}_{Yi} = Y_i - \widehat{g}_Y(W_i)$ and $\hat{\epsilon}_{Xi} = X_i - \widehat{g}_X(W_i)$ based on random sample of size n (Y_i, X_i, W_i), $i = 1, \ldots, n$, from (Y, X, W). Then,

$$\widehat{Y} = X\widehat{\beta} + \widehat{g}(W) = S(Y, X, W) = S(X\beta + g(W) + \epsilon, X, W) = S(Z + \epsilon, X, W), \text{ say, where } Z = X\beta + g(W).$$

Using Taylor's theorem, the expansion of S upto first order approximation is given by

$$S(X\beta + g(W) + \epsilon, X, W) \simeq S(X\beta + g(W), X, W) + \begin{pmatrix}\epsilon\\0\\0\end{pmatrix}^T \begin{pmatrix}\frac{\partial S}{\partial Z}\\ \frac{\partial S}{\partial X}\\ \frac{\partial S}{\partial W}\end{pmatrix}\Bigg|_{\substack{Z=\epsilon\\X=0\\W=0}} = S(X\beta + g(W), X, W) + \epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right).$$

Assume that $\sup_{X,W\in\mathbb{R}} \left|\frac{\partial}{\partial Z} S(Z, X, W)\Big|_{Z=\epsilon}\right| < \infty$, say $\mathbf{A}$. It is to be noted that

$$S^{(r)}(X\beta + g(W) + \epsilon, X, W) \simeq S^{(r)}(X\beta + g(W), X, W) + \left[\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right](r)$$

where $S^{(r)}$ denotes the r-th order difference of S and $\left[\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right](r)$ is the r-th order difference of $\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)$.
Furthermore, $\sup_{X,W,\epsilon\in\mathbb{R}} \left|\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right| = \mathbf{A}\epsilon$ that implies $\left|\left[\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right](r)\right| = \mathbf{A}\epsilon^*(r)$.
Then,

$$S^{(r)}(X\beta + g(W) + \epsilon, X, W) - S^{(r)}(X\beta + g(W), X, W) \simeq \left[\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right](r)$$

$$\implies P\left(|S^{(r)}(X\beta + g(W) + \epsilon, X, W) - S^{(r)}(X\beta + g(W), X, W)| > \delta\right) \simeq P\left(\left|\left[\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right](r)\right| > \delta\right).$$

Now, it is worth to realize that

$$P\left(\left|\left[\epsilon \cdot \left(\frac{\partial S}{\partial Z}\Big|_{Z=\epsilon}\right)\right](r)\right| > \delta\right) \leq P\left(\left|\mathbf{A}\epsilon^*(r)\right| > \delta\right) = P\left(\left|\epsilon^*(r)\right| > \delta/|\mathbf{A}|\right) = P\left(\epsilon^*(r) > \delta/|\mathbf{A}|\right) + P\left(\epsilon^*(r) < -\delta/|\mathbf{A}|\right).$$

Using Markov's inequality, we get $P\left(\epsilon^*(r) > \delta/|\mathbf{A}|\right) \leq \dfrac{E[\epsilon^*(r)]}{\delta/|\mathbf{A}|} = \dfrac{|\mathbf{A}|}{\delta} \displaystyle\sum_{j=1}^{r+1} (-1)^{j-1} \binom{r}{j-1} E(\epsilon_j)$,
where $E(\epsilon_j) = E_{X,W} E(\epsilon_j | X, W) = 0$. Then, $P\left(\epsilon^*(r) > \delta/|\mathbf{A}|\right) = 0$. In similar manner, $P\left(\epsilon^*(r) < -\delta/|\mathbf{A}|\right) = 0$.
$\therefore P\left(|S^{(r)}(X\beta + g(W) + \epsilon, X, W) - S^{(r)}(X\beta + g(W), X, W)| > \delta\right) \simeq 0 \implies S^{(r)}(X\beta + g(W) + \epsilon, X, W) \approx S^{(r)}(X\beta + g(W), X, W)$. The proposition can be proved in similar way for $p, q > 1$.

9.3 Proof of Proposition 2

The variance of kernel of $T_n^{<r>}$ is computed under H_0 as

$$\begin{aligned}
\xi_1(r) &= Var\left[E(h((\widehat{Y}^*_1(r),Y^*_1(r)),(\widehat{Y}^*_2(r),Y^*_2(r)))\Big|(\widehat{Y}^*_1(r),Y^*_1(r)))\right] = Var\left[\psi(\widehat{Y}^*_1(r),Y^*_1(r))\right],\ \text{ say, where}\\
\psi(\widehat{Y}^*_1(r),Y^*_1(r)) &= E\left[h((\widehat{Y}^*_1(r),Y^*_1(r)),(\widehat{Y}^*_2(r),Y^*_2(r)))\Big|(\widehat{Y}^*_1(r),Y^*_1(r))\right]\\
&= P\left[\widehat{Y}^*_1(r)>Y^*_1(r)>\widehat{Y}^*_2(r),Y^*_2(r)|\widehat{Y}^*_1(r),Y^*_1(r)\right]+P\left[\widehat{Y}^*_1(r)<Y^*_1(r)<\widehat{Y}^*_2(r),Y^*_2(r)|\widehat{Y}^*_1(r),Y^*_1(r)\right]\\
&\quad -P\left[\widehat{Y}^*_1(r)>Y^*_1(r)<\widehat{Y}^*_2(r),Y^*_2(r)|\widehat{Y}^*_1(r),Y^*_1(r)\right]-P\left[\widehat{Y}^*_1(r)<Y^*_1(r)>\widehat{Y}^*_2(r),Y^*_2(r)|\widehat{Y}^*_1(r),Y^*_1(r)\right]
\end{aligned}$$

which is finally deduced as $\left(2G_{0;\widehat{Y}^*(r)}(\widehat{Y}^*_1(r))-1\right)\left(2G_{0;Y^*(r)}(Y^*_1(r))-1\right)$. $G_{0;\widehat{Y}^*(r)}(\cdot)$ and $G_{0;Y^*(r)}(\cdot)$ are the marginal CDFs of $\widehat{Y}^*(r)$ and $Y^*(r)$ respectively under H_0. Then,

$$\begin{aligned}
Var\left[\psi(\widehat{Y}^*_1(r),Y^*_1(r))\right] &= Var\left[\left(2G_{0;\widehat{Y}^*(r)}(\widehat{Y}^*_1(r))-1\right)\left(2G_{0;Y^*(r)}(Y^*_1(r))-1\right)\right]\\
&= E\left(2G_{0;\widehat{Y}^*(r)}(\widehat{Y}^*_1(r))-1\right)^2 E\left(2G_{0;Y^*(r)}(Y^*_1(r))-1\right)^2\\
&\quad -E^2\left(2G_{0;\widehat{Y}^*(r)}(\widehat{Y}^*_1(r))-1\right)E^2\left(2G_{0;Y^*(r)}(Y^*_1(r))-1\right)
\end{aligned}$$

Since $G_{0;\widehat{Y}^*(r)}(\cdot), G_{0;Y^*(r)}(\cdot) \overset{indep.}{\sim} U(0,1)$, therefore $E\left(2G_{0;\widehat{Y}^*(r)}(\widehat{Y}^*_1(r))-1\right)^2 = 4\cdot\frac{1}{12}=\frac{1}{3}=E\left(2G_{0;Y^*(r)}(Y^*_1(r))-1\right)^2$. Also, $E\left(2G_{0;\widehat{Y}^*(r)}(\widehat{Y}^*_1(r))-1\right)=E\left(2G_{0;Y^*(r)}(Y^*_1(r))-1\right)=0$. Therefore, $\xi_1(r)=\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{9}$.

9.4 Proof of Proposition 3

Here, the kernel is $\psi((a_1,b_1),(a_2,b_2),(a_3,b_3)) = 3sign\{(a_1-a_2)(b_1-b_3)\}$ for three arbitrary bivariate points $(a_1,b_1),(a_2,b_2),(a_3,b_3)$. Define,

$$\begin{aligned}
f_1(\widehat{Y}^*_1(r),Y^*_1(r)) &= E[3\,sign\{(\widehat{Y}^*_1(r)-\widehat{Y}^*_2(r))(Y^*_1(r)-Y^*_3(r))\}|(\widehat{Y}^*_1(r),Y^*_1(r))]\\
&= 3\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} sign(\widehat{Y}^*_1(r)-\widehat{y}^*_2(r))sign(Y^*_1(r)-y^*_3(r))\times\prod_{i=2,3} dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_i(r),y^*_i(r))\\
&= 3\int_{\mathbb{R}^2} sign(\widehat{Y}^*_1(r)-\widehat{y}^*_2(r))\,I(Y^*_1(r))\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_2(r),y^*_2(r)).
\end{aligned}$$

where

$$\begin{aligned}
I(Y^*_1(r)) &= \int_{\mathbb{R}^2} sign(Y^*_1(r)-y^*_3(r))\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r))\\
&= \int_{\mathbb{R}}\int_{\mathbb{A}_1} sign(Y^*_1(r)-y^*_3(r))\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r))\\
&\quad +\int_{\mathbb{R}}\int_{\mathbb{A}_1^c} sign(Y^*_1(r)-y^*_3(r))\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r)),\ \text{ where } \mathbb{A}_1=\{y^*_3(r):y^*_3(r)<Y^*_1(r)\}\\
&= \int_{\mathbb{R}}\int_{-\infty}^{Y^*_1(r)} (+1)\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r))+\int_{\mathbb{R}}\int_{Y^*_1(r)}^{\infty}(-1)\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r))\\
&= \int_{\mathbb{R}}\int_{-\infty}^{Y^*_1(r)} dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r))-\int_{\mathbb{R}}\int_{Y^*_1(r)}^{\infty} dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_3(r),y^*_3(r))\\
&= H_{Y^*(r)}(Y^*_1(r))-\left[1-H_{Y^*(r)}(Y^*_1(r))\right]=2H_{Y^*(r)}(Y^*_1(r))-1.
\end{aligned}$$

$$\therefore\ f_1(\widehat{Y}^*_1(r),Y^*_1(r)) = 3\left[2H_{Y^*(r)}(Y^*_1(r))-1\right]\int_{\mathbb{R}^2} sign(\widehat{Y}^*_1(r)-\widehat{y}^*_2(r))\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_2(r),y^*_2(r)).$$

In similar way of deduction of $I(Y^*_1(r))$, one can compute $\int_{\mathbb{R}^2} sign(\widehat{Y}^*_1(r)-\widehat{y}^*_2(r))\,dG_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*_2(r),y^*_2(r))$ as

$\left[2G_{\widehat{Y}^*(r)}(\widehat{Y}^*_1(r)) - 1\right]$. Here, $G_{\widehat{Y}^*(r)}(\cdot)$ and $H_{Y^*(r)}(\cdot)$ are the marginal *c.d.f.*s of $\widehat{Y}^*(r)$ and $Y^*(r)$ respectively and they are uniformly distributed on [0,1]. Therefore, $f_1(\widehat{Y}^*_1(r), Y^*_1(r)) = 3\left[2G_{\widehat{Y}^*(r)}(\widehat{Y}^*_1(r)) - 1\right]\left[2H_{Y^*(r)}(Y^*_1(r)) - 1\right]$, which further implies $\nu_1 = Var(f_1(\widehat{Y}^*_1(r), Y^*_1(r))) = 9 \cdot \frac{1}{3} \cdot \frac{1}{3} = 1 > 0$ under H_0.

9.5 Proof of Proposition 4

For sample size n^* such that $n^* > n$,

$$\begin{aligned}
P_{H_n}\left[\sqrt{n}(T_n^{(r)} - E_{H_0}(T_n^{(r)})) > t_\kappa\right] &= P_{H_n}\left[\sqrt{\frac{n^*}{n}} \cdot \sqrt{n}(T_n^{(r)} - E_{H_0}(T_n^{(r)})) > \sqrt{\frac{n^*}{n}} \cdot t_\kappa\right] \\
&= P_{H_n}\left[\sqrt{n^*}(T_n^{(r)} - E_{H_0}(T_n^{(r)})) > \sqrt{\frac{n^*}{n}} \cdot t_\kappa\right] \\
&< P_{H_n}\left[\sqrt{n^*}(T_n^{(r)} - E_{H_0}(T_n^{(r)})) > t_\kappa\right] \overset{asy.}{\equiv} P_{H_n}\left[\sqrt{n^*}(T_n^{(r)} - E_{H_0}(T_n^{(r)})) > t_\kappa\right],
\end{aligned}$$

i.e. for increasing sample size n, the power of $T_n^{(r)} \uparrow$ and tends to 1. Moreover, as $\mu \uparrow$,

$$P_{H_n}\left[\sqrt{n}(T_n^{(r)} - E_{H_0}(T_n^{(r)})) > t_\kappa\right] = \Phi\left(\frac{\mu\, E_{H_1}(T_n^{(r)}) - t_\kappa}{\sqrt{4\xi_1(r)}}\right) \longrightarrow 1.$$