Asymptotic power performance of test statistic based on Kendall's τ under generalized partially linear regression model

Sthitadhi Das*
Department of Mathematics, Brainware University, West Bengal , India

Abstract

Recently, Das et al. (2022) considered a simple partially linear model $Y = Z\beta + m(X) + \epsilon$ to construct consistent tests based on several nonparametric measures of association including Kendall's τ, for testing independence between X and ϵ. However, the assumptions on ϵ are generally made for realised values of the parametric and nonparametric covariates Z and X respectively. An extension of such model naturally involves more than one regressor, both parametric and nonparametric, and one obtains a generalized partially linear model $Y = \beta_1 X_1 + \ldots + \beta_p X_p + m(W_1, \ldots, W_q) + \epsilon$ where p parametric regressors $X_1, \ldots, X_p$ and q nonparametric regressors $W_1, \ldots, W_q$ jointly explain Y with ϵ being the error in explanation. The objective is to verify the validity of the assumptions on ϵ for realised values of the all $(p+q)$ regressors. Hence, testing independence between joint regressors $(X_1, \ldots, X_p, W_1, \ldots, W_q)$ and ϵ is a necessity in this situation. Before going to perform the test, it is necessary to estimate the parameters of the model by Robinson (1988)'s technique as well as the nonparametric regression function by a suitable kernel density estimation method. Afterwards, a V-statistic on the basis of Kendall's τ is formed by setting up appropriate null hypothesis as well as contiguous alternatives. The objective is to establish consistent test procedures involving the concerning test statistics. A study on such consistency is performed through deriving asymptotic power curve of the test statistic and the motive lies in improving the power further. In addition, data analysis is necessary to substantiate the consistency of the test statistic.

1 Introduction

Testing association(s) between two or more variables is a common topic of interest in the literature of statistics. Under the setup of regression, one or more variable(s) is (are) studied with the help of one or several independent variable(s) pertinent to the context. A generic idea could be studying a variable of interest through a number of concomitants, which is well-described by a generalized partially linear regression model. Such a model has the mathematical representation $Y = \beta_1 X_1 + \ldots + \beta_p X_p + m(W_1, \ldots, W_q) + \epsilon$, *i.e.* 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 $m(\cdot, \ldots, \cdot)$ of q independent variables $W_1, \ldots, W_q$ with unknown mathematical form where $p, q \geq 2$. The usual assumptions on the error ϵ are taken 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) $Var(\epsilon|X_1, \ldots, X_p, W_1, \ldots, W_q) = \sigma^2(X_1, \ldots, X_p, W_1, \ldots, W_q)$. Various analytical aspects involving a generalized partially linear model, including estimation of the parameters and the nonparametric regression function and the asymptotic properties of the estimators, were extensively carried out by **Robinson** (1988), **Andrews** (1995), **Qi Li** (1996), **Hamilton** (1997), **Liu et al.** (1997) etc. **Shiller** (1984) studied an earlier cost curve in the utility industry by considering a partially linear regression model. **Brown et al.** (2016) considered a partially linear model with intercept, unlike the setup without intercept considered here.

Utilisation of some measures of association namely **Kendall** (1961)'s τ, **Bergsma et al.** (2014)'s τ^* and **Szekely et al.** (2007)'s $dCov$ was done by **Das et al.** (2022) in a partially linear model with single parametric and semi-parametric regressors. Although they tested association between the sole nonparametric regressor and random error in that setup, the idea could be more generalized further. Considering all the regressors in the underlying model, the association of them with the error component could also be a topic of investigation. Let us divide the whole discussion into several sections as follows. **Ferraccioli et al.** (2023) proposed nonparametric test *viz.* eigen sign-flip test, studied the properties of the score statistic in penalized semiparametric regression etc.

Section 2 provides the methods for estimating $\beta_1, \ldots, \beta_p$ as well as $m(\cdot, \ldots, \cdot)$, whereas the hypotheses are developed on the basis of the estimated parametric and nonparametric parts of the model in Section 3. Proposition of test statistic T_n, the V-statistic corresponding to Kendall's τ, is discussed in Section 4. The asymptotic power of T_n

*sthdas999@gmail.com

is analyzed using the theory of asymptotic distribution of a non-degenerate V-statistic. Data analysis is carried out in Section 5 to check the applicability of Kendall's τ through T_n by deriving simulated p-values of T_n for different sample sizes.

2 Model estimation

We have to estimate the parameters β_1, ..., β_p by *Robinson's method* and the nonparametric regression function $g(\cdot, \ldots, \cdot)$ by suitable kernel smoothing method(s). The model under consideration is re-expressed as

$$Y_i = \underset{\sim}{X}_i^T \underset{\sim}{\beta} + m(\underset{\sim}{W}_i) + \epsilon_i, \;\; i = 1, \ldots, n. \tag{1}$$

where $\underset{\sim}{X}_i = (X_{i1}, \ldots, X_{ip})^T$, $\underset{\sim}{W}_i = (W_{i1}, \ldots, W_{iq})^T$, $i = 1, \ldots, n$. Taking $E(\cdot|\underset{\sim}{W}_i)$ to both sides of (1), we get

$$\begin{aligned}
E(Y_i|\underset{\sim}{W}_i) &= E(\underset{\sim}{X}_i|\underset{\sim}{W}_i)^T\underset{\sim}{\beta} + m(\underset{\sim}{W}_i) + E(\epsilon_i|\underset{\sim}{W}_i) \\
\implies E(Y_i|\underset{\sim}{W}_i) &= E(\underset{\sim}{X}_i|\underset{\sim}{W}_i)^T\underset{\sim}{\beta} + m(\underset{\sim}{W}_i) + E_{\underset{\sim}{X}_i}[E(\epsilon_i|\underset{\sim}{X}_i, \underset{\sim}{W}_i)] \\
\implies E(Y_i|\underset{\sim}{W}_i) &= E(\underset{\sim}{X}_i|\underset{\sim}{W}_i)^T\underset{\sim}{\beta} + m(\underset{\sim}{W}_i) \;\; (\because\; E(\epsilon_i|\underset{\sim}{X}_i, \underset{\sim}{W}_i) = 0) \\
\implies E(Y_i|\underset{\sim}{W}_i = \underset{\sim}{w}) &= E(\underset{\sim}{X}_i|\underset{\sim}{W}_i = \underset{\sim}{w})^T\underset{\sim}{\beta} + g(\underset{\sim}{w}) \\
i.e. \;\; g_Y(\underset{\sim}{w}) &= g_{\underset{\sim}{X}}(\underset{\sim}{w})^T\underset{\sim}{\beta} + g(\underset{\sim}{w}), \text{ say.}
\end{aligned} \tag{2}$$

Then, by subtracting (2) from (1), we obtain

$$\begin{aligned}
Y_i - g_Y(\underset{\sim}{w}) &= (\underset{\sim}{X}_i - g_{\underset{\sim}{X}}(\underset{\sim}{w}))^T\underset{\sim}{\beta} + \epsilon_i \\
\implies \epsilon_{Yi} &= \epsilon_{\underset{\sim}{X}i}^T\underset{\sim}{\beta} + \epsilon_i, \text{ where} \\
Y_i &= g_Y(\underset{\sim}{W}_i) + \epsilon_{Yi} \qquad (3)\\
\underset{\sim}{X}_i &= g_{\underset{\sim}{X}}(\underset{\sim}{W}_i) + \epsilon_{\underset{\sim}{X}i}, \;\; i = 1, \ldots, n \qquad (4)\\
\implies \tilde{\underset{\sim}{\beta}} &= \left(\sum_{i=1}^n \epsilon_{\underset{\sim}{X}i}\epsilon_{\underset{\sim}{X}i}^T\right)^{-1}\left(\sum_{i=1}^n \epsilon_{\underset{\sim}{X}i}\epsilon_{Yi}\right). \qquad (5)
\end{aligned}$$

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 $i = 1, \ldots, n$ so that a feasible estimator of β can be determined. From models (3) and (4), the estimators of the regression functions $g_Y(\cdot)$ and $g_{\underset{\sim}{X}}(\cdot)$ are to be derived by using Nadaraya-Watson estimation of kernel density (NW) method as follows. Note that,

$$g_Y(\underset{\sim}{w}) = E(Y|\underset{\sim}{W} = \underset{\sim}{w}) \text{ where } \underset{\sim}{w} = (w_1, \ldots, w_q)' = \int_{-\infty}^{\infty} y \cdot f_{Y|\underset{\sim}{W}}(y|\underset{\sim}{w})\, dy = \int_{-\infty}^{\infty} y \cdot \frac{h_{Y,\underset{\sim}{W}}(y, \underset{\sim}{w})}{\psi_{\underset{\sim}{W}}(\underset{\sim}{w})}\, dy \tag{6}$$

where $h_{Y,\underset{\sim}{W}}(\cdot,\cdot)$ is the joint *p.d.f.* of $(Y, \underset{\sim}{W})$ and $\psi_{\underset{\sim}{W}}(\cdot)$ is the *p.d.f.* of $\underset{\sim}{W}$, assuming they both exist. Now, the kernel density of $\underset{\sim}{W}$ is estimated as

$$\widehat{\psi}_{\underset{\sim}{W}}(\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_1 \ldots h_q}\; k\left(\frac{w_1 - W_{i1}}{h_1}, \ldots, \frac{w_q - W_{iq}}{h_q}\right) \tag{7}$$

at $\underset{\sim}{W} = \underset{\sim}{w}$. Here, $k(\cdot, \ldots, \cdot)$ is the q-dimensional kernel density function of $\underset{\sim}{W}$. By taking $k(\cdot, \ldots, \cdot)$ as a multiplicative kernel, i.e. $k(\theta_1, \ldots, \theta_q) = \prod_{j=1}^q k(\theta_j)\; \forall\; (\theta_1, \ldots, \theta_q)' \in \mathbb{R}^q$, we can further simplify the expression of $\widehat{\psi}_{\underset{\sim}{W}}(\cdot)$ as

$$\widehat{\psi}_{\underset{\sim}{W}}(\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_1 \ldots h_q}\; k_1\left(\frac{w_1 - W_{i1}}{h_1}\right) \ldots k_q\left(\frac{w_q - W_{iq}}{h_q}\right) = \frac{1}{n}\sum_{i=1}^n \left\{\prod_{j=1}^q \frac{1}{h_j}\; k_j\left(\frac{w_j - W_{ij}}{h_j}\right)\right\}$$

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

$$\widehat{h}_{Y,\underset{\sim}{W}}(y, \underset{\sim}{w}) = \frac{1}{nh_y h_1 \ldots h_q} \sum_{i=1}^{n} k_y\left(\frac{y - Y_i}{h_y}, \frac{w_1 - W_{i1}}{h_1}, \ldots, \frac{w_q - W_{iq}}{h_q}\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_y} k_y\left(\frac{y - Y_i}{h_y}\right) \left\{ \prod_{j=1}^{q} \frac{1}{h_j} k_j\left(\frac{w_j - W_{ij}}{h_j}\right) \right\}$$

where h_y is the bandwidth for estimating the $p.d.f.$ of Y. Then, at $\underset{\sim}{W} = \underset{\sim}{w}$, $\widehat{g}_Y(\underset{\sim}{w})$ is estimated as

$$\widehat{g}_Y(\underset{\sim}{w}) = \int_{-\infty}^{\infty} y \cdot \frac{\widehat{h}_{Y,\underset{\sim}{W}}(y, \underset{\sim}{w})}{\widehat{\psi}_{\underset{\sim}{W}}(\underset{\sim}{w})}\, dy = \frac{\frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{y}{h_y} k_y\left(\frac{y - Y_i}{h_y}\right) dy \cdot \left\{ \prod_{j=1}^{q} \frac{1}{h_j} k_j\left(\frac{w_j - W_{ij}}{h_j}\right) \right\}}{\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{q} \frac{1}{h_j} k_j\left(\frac{w_j - W_{ij}}{h_j}\right)}. \tag{8}$$

Let, $\dfrac{y - Y_i}{h_y} = v \implies y = Y_i + v\, h_y \implies dy = h_y\, dv$.

$$\begin{aligned} \therefore \int_{-\infty}^{\infty} \frac{y}{h_y} k_y\left(\frac{y - Y_i}{h_y}\right) dy &= \int_{-\infty}^{\infty} \frac{Y_i + v\, h_y}{h_y} k_V(v)\, h_y\, dv \\ &= \int_{-\infty}^{\infty} Y_i \int_{-\infty}^{\infty} k_V(v)\, dv + h_y \int_{-\infty}^{\infty} v\, k_V(v)\, dv = Y_i \cdot 1 + h_y \cdot 0 = Y_i. \end{aligned}$$

Using (8), we derive the estimator of $g_Y(\underset{\sim}{w})$ as

$$\widehat{g}_Y(\underset{\sim}{w}) = \frac{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{q} \frac{1}{h_j} k_j\left(\frac{w_j - W_{ij}}{h_j}\right) \right\} Y_i}{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{q} \frac{1}{h_j} k_j\left(\frac{w_j - W_{ij}}{h_j}\right) \right\}}. \tag{9}$$

In similar way, we proceed to estimate $g_{\underset{\sim}{X}}(\cdot)$ given by

$$g_{\underset{\sim}{X}}(\underset{\sim}{w}) = E(\underset{\sim}{X} | \underset{\sim}{W} = \underset{\sim}{w}) = \int_{\mathbb{R}^p} \underset{\sim}{x} \cdot g_{\underset{\sim}{X}|\underset{\sim}{W}}(\underset{\sim}{x}|\underset{\sim}{w})\, d\underset{\sim}{x} = \int_{\mathbb{R}^p} \underset{\sim}{x} \cdot \frac{h_{\underset{\sim}{X},\underset{\sim}{W}}(\underset{\sim}{x}, \underset{\sim}{w})}{\psi_{\underset{\sim}{W}}(\underset{\sim}{w})}\, d\underset{\sim}{x}. \tag{10}$$

Firstly, the joint probability density function of $(\underset{\sim}{X}, \underset{\sim}{W})$ is estimated as

$$\widehat{h}_{\underset{\sim}{X},\underset{\sim}{w}}(\underset{\sim}{x}, \underset{\sim}{w}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{m=1}^{p} \frac{1}{a_m} k_{m;X_i}\left(\frac{x_m - X_{im}}{a_m}\right) \right\} \times \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\}. \tag{11}$$

where $a_1, \ldots, a_p$ are the bandwidths for estimating the kernel densities of $X_1, \ldots, X_p$ and $b_1, \ldots, b_q$ are the bandwidths for estimating the kernel densities of $W_1, \ldots, W_q$. Now, the estimation of $g_{\underset{\sim}{X}}(\underset{\sim}{w})$ is elaborated as

$$\begin{aligned} \widehat{g}_{\underset{\sim}{X}}(\underset{\sim}{w}) &= \int_{\mathbb{R}^p} \underset{\sim}{x} \cdot \frac{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{m=1}^{p} \frac{1}{a_m} k_{m;X_i}\left(\frac{x_m - X_{im}}{a_m}\right) \right\} \times \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\}}{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\}}\, d\underset{\sim}{x} \\ &= \frac{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\} \times \int_{\mathbb{R}^p} \prod_{m=1}^{p} \left\{ \frac{x_m}{a_m} k_{m;X_i}\left(\frac{x_m - X_{im}}{a_m}\right) dx_m \right\}}{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\}} \\ &= \frac{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{m=1}^{p} X_{im} \right\} \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\}}{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{z=1}^{q} \frac{1}{b_z} k_{z;W_i}\left(\frac{w_z - W_{iz}}{b_z}\right) \right\}}. \end{aligned} \tag{12}$$

After estimating $\widehat{g}_Y(\cdot)$ and $\widehat{g}_{\underset{\sim}{X}}(\cdot)$ by all the four methods, the errors are further estimated as

$$\begin{aligned} \widehat{\epsilon}_{Yi} &= Y_i - \widehat{g}_Y(\underset{\sim}{W}_i) && (13)\\ \widehat{\epsilon}_{\underset{\sim}{X}i} &= \underset{\sim}{X}_i - \widehat{g}_{\underset{\sim}{X}}(\underset{\sim}{W}_i) && (14) \end{aligned}$$

followed by feasible estimation of $\underset{\sim}{\beta}$ as

$$\widehat{\underset{\sim}{\beta}} = \left(\sum_{i=1}^{n} \widehat{\epsilon}_{\underset{\sim}{X}i} \widehat{\epsilon}_{\underset{\sim}{X}i}^{T} \right)^{-1} \left(\sum_{i=1}^{n} \widehat{\epsilon}_{\underset{\sim}{X}i} \widehat{\epsilon}_{Yi} \right). \tag{15}$$

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

$$Y_i - \underset{\sim}{X}_i^T \underset{\sim}{\beta} = m(\underset{\sim}{W}_i) + \epsilon_i \implies Y_i^{'} = m(\underset{\sim}{W}_i) + \epsilon_i, \; i = 1, \ldots, n. \tag{16}$$

where $Y_i^{'}$ is the *transformed response* defined as $(Y_i - \underset{\sim}{X}_i^T \underset{\sim}{\beta})$, $i = 1, \ldots, n$. Now, based on the *i.i.d.* observations $(Y_i^{'}, \underset{\sim}{W}_i)$'s, $i = 1, \ldots, n$, the estimator of $m(\cdot)$ is obtained as

$$\begin{aligned} m(\underset{\sim}{w}) = E(Y^{'} | \underset{\sim}{W} = \underset{\sim}{w}) &= \int_{-\infty}^{\infty} y^{'} \phi_{Y^{'}|\underset{\sim}{W}}(y^{'}|\underset{\sim}{w}) \, dy^{'} \quad \left(\text{where } \phi_{Y^{'}|\underset{\sim}{W}}(\cdot|\underset{\sim}{w}) \text{ is the p.d.f. of } Y^{'}|\underset{\sim}{W} \right) \\ &= \int_{-\infty}^{\infty} y^{'} \cdot \frac{f_{Y^{'},\underset{\sim}{W}}(y^{'}, \underset{\sim}{w})}{\psi_{\underset{\sim}{W}}(\underset{\sim}{w})} \, dy^{'} \\ \implies \widehat{m}(\underset{\sim}{w}) = \widehat{E(Y^{'} | \underset{\sim}{W} = \underset{\sim}{w})} &= \int_{-\infty}^{\infty} y^{'} \cdot \frac{\frac{1}{n} \sum_{i=1}^{n} \{ k_{y^{'}} \left(\frac{y^{'} - Y_i^{'}}{h_{y^{'}}} \right) \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right) \}}{\frac{1}{n} \sum_{i=1}^{n} \{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_i - W_{ij}}{h_j} \right) \}} \, dy^{'} \\ &= \frac{\frac{1}{n} \sum_{i=1}^{n} \left(\int_{-\infty}^{\infty} y^{'} \, k_{y^{'}} \left(\frac{y^{'} - Y_i^{'}}{h_{y^{'}}} \right) dy^{'} \right) \times \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right)}{\frac{1}{n} \sum_{i=1}^{n} \{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right) \}} \\ &= \frac{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right) \right\} Y_i^{'}}{\frac{1}{n} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right) \right\}} \; \widehat{=} \; \sum_{i=1}^{n} \lambda_i(\underset{\sim}{w}) \, \widehat{Y}_i^{'} \end{aligned}$$

where $\lambda_i(\underset{\sim}{w}) = \dfrac{\left\{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right) \right\}}{\sum_{i=1}^{n} \left\{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{w_j - W_{ij}}{h_j} \right) \right\}}$ and $\widehat{Y}_i^{'} = (Y_i - \underset{\sim}{X}_i^T \widehat{\underset{\sim}{\beta}})$, $i = 1, \ldots, n$.

The predicted response $\widehat{Y}$ is obtained furthermore as

$$\widehat{Y} \;=\; \widehat{\beta}_1 X_1 + \ldots + \widehat{\beta}_p X_p + \widehat{m}\,(W_1, \ldots, W_q) \tag{17}$$

a function of Y, $X_1, \ldots, X_p$, $W_1, \ldots, W_q$ where

$$\widehat{m}\,(W_1, \ldots, W_q) \;=\; \sum_{i=1}^{n} \left[\frac{\left\{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{W_j - W_{ij}}{h_j} \right) \right\}}{\sum_{i=1}^{n} \left\{ \prod_{j=1}^{q} k_{j;W_i} \left(\frac{W_j - W_{ij}}{h_j} \right) \right\}} \right] \cdot (Y_i - \underset{\sim}{X}_i^T \widehat{\underset{\sim}{\beta}}).$$

3 Hypotheses of interest

The usual motivation in a nonparametric or semiparametric regression model lies in testing homoscedasticity of error (**Goldfeldt et al.** (1965)). If $\mathbf{U}$ be a L-dimensional vector of regressors ($L \geq 2$) and δ be the scalar error, then $E(\delta^2|\mathbf{U}) = c$ with c being a constant when δ is homoscedastic. Otherwise, $E(\delta^2|\mathbf{U}) = f(\mathbf{U})$ where $f(\cdot) > 0$ and depends on $\mathbf{U}$, then δ is termed as heteroscedastic error. In other words, independence of $\mathbf{U}$ and δ ensures the homoscedasticity of δ. Hence, the association between $\mathbf{U}$ and δ plays a pivotal role in a nonparametric or semiparametric regression setup. One may be interested to test if $\mathbf{U}$ and δ are independent or not.

Driven by similar idea, the independence between jointly distributed regressors $X_1, \ldots, X_p, W_1, \ldots, W_q$ and error ϵ could be a matter of investigation in our proposed partially linear regression setup. The natural hypotheses of interest in this regard are $H_0 : \underset{\sim}{Z} \perp\!\!\!\perp \epsilon$ and $H_1 : \underset{\sim}{Z} \not\perp\!\!\!\perp \epsilon$ where $\underset{\sim}{Z} = (\underset{\sim}{X}, \underset{\sim}{W})$ is the $(p+q)$-dimensional vector of regressors. But we have to work on the hypotheses further due to unobservability of ϵ by defining a suitable function of ϵ on $\mathbb{R}$, say ϵ, which is indeed tractable. Moreover, a real-valued function of the $(p+q)$ regressors is a convenient one to furnish the testing procedure, say $\tilde{F}(X_1, \ldots, X_p, W_1, \ldots, W_q)$. Then the null hypothesis H_0 further implies that $\tilde{F}(X_1, \ldots, X_p, W_1, \ldots, W_q) \perp\!\!\!\perp s(\epsilon)$ whereas the alternative hypothesis suggests dependence of them.

To obtain a well specified null hypothesis to fulfill our objective, let us proceed in the following manner. Let us provide the expressions for 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.

Theorem 3.1. $Y^*(r) \approx \epsilon^*(r)$*. Moreover, among all possible linear functions* $v(\epsilon_1, \ldots, \epsilon_{r+1}) = \sum_{j=1}^{r+1} \alpha_j \epsilon_j$ *with* $\alpha_j\ s \in \mathbb{R}$*,* $\epsilon^*(r)$ *has maximum k-th order absolute moment.*

The above theorem is quite meaningful in this context. It is evident that $\epsilon^*(1)$ can be well approximated by $Y^*(1)$, but one can obtain a test of homoscedasticity of ϵ based on $\epsilon^*(1)$. Hence, Einmahl et al. (2008a) as well as Das et al. (2021) considered $\epsilon^*(2)$ and approximated it by $Y^*(2)$ to construct testing scheme for independence between nonparametric covariate X and error ϵ. In addition, under a simple partially linear regression setup $Y = Z\beta + m(X) + \epsilon$, Das et al. (2022) considered $\epsilon^*(3)$ to develop a test of independence between sole nonparametric covariate X and random error ϵ. The proof of Theorem 3.1 is elaborately discussed in **Appendix-II**11 to understand the consideration of general order difference of ϵ in this regard.

Finally, we get transformed null hypothesis as $H_0 : (\underset{\sim}{X}, \underset{\sim}{W}) \perp\!\!\!\perp \epsilon^*(r)$ which further implies approximately that $(\underset{\sim}{X}, \underset{\sim}{W}) \perp\!\!\!\perp Y^*(r)$. Again, as another implication of H_0, any continuous function of $(\underset{\sim}{X}, \underset{\sim}{W})$ is independent to $Y^*(r)$.

Here, the predicted response $\widehat{Y}$ given by (17) is a function of $(Y, \underset{\sim}{X}, \underset{\sim}{W})$. The following proposition is helpful to proceed with $\widehat{Y}$ in this testing scheme. Proof is available in **Appendix-II**.

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

Therefore, the null hypothesis further implies that $H_0 : \widehat{Y}^*(r) \perp\!\!\!\perp Y^*(r)$ and subsequently the hypotheses of interest are derived as

$$H_0 : \widehat{Y}^*(r) \perp\!\!\!\perp Y^*(r) \text{ against } H_1 : \widehat{Y}^*(r) \not\perp\!\!\!\perp Y^*(r). \tag{18}$$

From (18) it is further explained that $H_0 : \mathcal{M}(\widehat{Y}^*(r), Y^*(r)) = 0$ against $H_1 : \mathcal{M}(\widehat{Y}^*(r), Y^*(r)) \neq 0$, where $\mathcal{M}$ is the concerning nonparametric measure of association between $\widehat{Y}^*(r)$ and $Y^*(r)$.

Now, in this testing of hypotheses problem, it is more consequential to set up a consistent test procedure, which is less likely in case of (18). A conventional wayout is to define a sequence of contiguous alternative hypotheses as described and studied by **Le Cam** (1990). Such a sequence converges to the null hypothesis with increasing sample size, hence it is possible to achieve a consistent testing methodology by taking contiguous alternatives against the null hypothesis.

Le Cam (1990) defined contiguity of two sequences of probability measures, say P_n and Q_n. Suppose these two measures are defined on $(\Omega_n, \mathcal{F}_n)$, where Ω_n is the sequence of sample spaces and $\mathcal{F}_n$ is the sequence of Borel σ-fields constituted by the classes of subsets of Ω_n. Then, Q_n to be contiguous with respect to P_n if

$P_n(A_n) \longrightarrow 0 \implies Q_n(A_n) \longrightarrow 0$ as $n \to \infty$, where $A_n \in \mathcal{F}_n$ is the sequence of measurable sets. This contiguous relationship of P_n and Q_n is denoted as $P_n \triangleleft Q_n$.

Le Cam's first lemma (discussed by Hajek et al. (1999)) is utilised to detect contiguity between any two arbitrary sequences of probability measures through its important corollary provided by Van der Vaart (2002) as follows.

Lemma 3.1. *If* $\log \dfrac{dQ_n}{dP_n} \xrightarrow{L} N(-\dfrac{\nu}{2}, \nu)$ *under* P_n $(\nu > 0)$*, then* $P_n \triangleleft Q_n$*.*

To construct a sequence of contiguous alternatives, Lemma 3.1 will be utilized. The proof is delineated by Van der Vaart. Let a sequence of alternative hypotheses is considered as follows.

$$H_n : \tilde{F}_{n;\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*, y^*) = \left(1 - \frac{\mu}{\sqrt{n}}\right) F_{0;\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*, y^*) + \frac{\mu}{\sqrt{n}} F_{\widehat{Y}^*(r),Y^*(r)}(\widehat{y}^*, y^*) \tag{19}$$

where $\tilde{F}_{n;\widehat{Y}^*(r),Y^*(r)}(\cdot,\cdot)$ is the joint CDF of $(\widehat{Y}^*(r), Y^*(r))$ under H_n, $F_{0;\widehat{Y}^*(r),Y^*(r)}(\cdot,\cdot)$ and $F_{\widehat{Y}^*(r),Y^*(r)}(\cdot,\cdot)$ are the joint CDFs of $(\widehat{Y}^*(r), Y^*(r))$ under H_0 and H_1 respectively, $\mu(>0)$ is the tuning parameter.

Theorem 3.2. *Under three following assumptions:*

(i) $\dfrac{\partial^2}{\partial\widehat{y}^*\partial y^*} F_0(\widehat{y}^*, y^*) < \infty$*, say* $f_0(\widehat{y}^*, y^*)$ $\forall$ $(\widehat{y}^*, y^*)$*,*

(ii) $\dfrac{\partial^2}{\partial\widehat{y}^*\partial y^*} F(\widehat{y}^*, y^*) < \infty$*, say* $f(\widehat{y}^*, y^*)$ $\forall$ $(\widehat{y}^*, y^*)$*,*

(iii) $E_{F_0}\left[\dfrac{f(\widehat{y}^*, y^*)}{f_0(\widehat{y}^*, y^*)} - 1\right]^2 < \infty$*,*

H_n *is a contiguous sequence of alternative hypotheses.*

Here $f_0(\cdot,\cdot)$ and $f(\cdot,\cdot)$ are the joint probability density function of $(\widehat{Y}^*(r), Y^*(r))$ under H_0 and H_1 respectively. To prove Theorem 3.2 the approach of **Das et al.** (2022) (p. 559) is evident. Next, we proceed to test H_0 against H_n, by constructing suitable test statistics based on the theory of non-degenerate V-statistic.

4 Test statistic and its distributions

To test the null hypothesis that $\widehat{Y}^*(r)$ is independent to $Y^*(r)$, a convenient approach is to develop robust test statistics. With the changing values of r, the asymptotic properties of the test statistics are expected to vary. For instance, one may study on the asymptotic power analysis of the test statistics for different choices of r. A frequently utilised measure of association, namely Kendall's τ, is considered in this backdrop to evaluate power performance of the test statistics made upon τ. A s discussed earlier, we proceed to test

$$H_0 : \tau(\widehat{Y}^*(r), Y^*(r)) = 0 \text{ against } H_1 : \tau(\widehat{Y}^*(r), Y^*(r)) \neq 0.$$

Quite familiar approach to construct nonparametric test statistics is to define V- statistics on the basis observations of the paired observations from a bivariate CDF. Here, let $(\widehat{y}_1^*, y_1^*), \ldots, (\widehat{y}_n^*, y_n^*)$ be a random sample of size n from the CDF of $(\widehat{Y}^*, Y^*)$. A V- statistic, to propose an unbiased estimator of the parameter of interest θ based on the obtained random sample, is defined as

$$V_n = n^{-m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \psi((\widehat{y}_{\alpha_1}^*(r), y_{\alpha_1}^*(r)), \ldots, (\widehat{y}_{\alpha_m}^*(r), y_{\alpha_m}^*(r)))$$

where ψ is the kernel of V_n involving m *i.i.d.* pairs $(\widehat{y}_{\alpha_1}^*(r), y_{\alpha_1}^*(r)), \ldots, (\widehat{y}_{\alpha_m}^*(r), y_{\alpha_m}^*(r))$ such that

$$E[\psi((\widehat{y}_{\alpha_1}^*(r), y_{\alpha_1}^*(r)), \ldots, (\widehat{y}_{\alpha_m}^*(r), y_{\alpha_m}^*(r)))] = \theta$$

, $\{\alpha_1, \ldots, \alpha_m\}$ being the set of m distinct integers $\in \{1, \ldots, n\}$.

Here, we consider the nonparametric measure of association Kendall's τ to perform the testing of hypothesis, hence $m = 2$ and the related kernel is defined as

$$\psi((\widehat{y}_{\alpha_1}^*(r), y_{\alpha_1}^*(r)), (\widehat{y}_{\alpha_2}^*(r), y_{\alpha_2}^*(r))) = sign\{(\widehat{y}_{\alpha_1}^* - \widehat{y}_{\alpha_2}^*)(y_{\alpha_1}^*(r) - y_{\alpha_2}^*(r))\}. \tag{20}$$

The corresponding V-statistic to test H_0 against H_n is regarded as the test statistic of interest in our discussion. The form of the test statistic is provided as

$$T_n^{<r>} \;=\; n^{-2}\sum_{\alpha_1=1}^{n}\sum_{\alpha_2=1}^{n} sign\{(\widehat{y}^*_{\alpha_1}(r)-\widehat{y}^*_{\alpha_2}(r))(y^*_{\alpha_1}(r)-y^*_{\alpha_2}(r))\}. \tag{21}$$

Since $\psi((\widehat{y}^*_{\alpha_1}(r), y^*_{\alpha_1}(r)), (\widehat{y}^*_{\alpha_2}(r), y^*_{\alpha_2}(r)))$ has order of degeneracy 0 (**Das et al.** (2022)), $T_n^{<r>}$ is a nondegenerate V-statistic indeed. In such situation, the asymptotic law of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ is a Gaussian one with mean 0 and constant variance as discussed by **Zhou et al. (2020)**. The limiting distributions of V_n, both under H_0 and H_n, are required for furnishing asymptotic power analysis of $T_n^{<r>}$ for different r eventually.

Theorem 4.1. *Under H_0,*
$\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) \xrightarrow{L} \mathcal{Z}^0 \sim N(0, 4\eta_{1,2}(r))$*, provided* $E[\psi^2((\widehat{Y}^*_1(r), Y^*_1(r)), (\widehat{Y}^*_2(r), Y^*_2(r)))] < \infty$*, where $\eta_{s;m}$ is defined as*

$$\eta_{s;m}(r) \;=\; Var_{H_0}[E(\psi((\widehat{Y}^*_1(r), Y^*_1(r)), \ldots, (\widehat{Y}^*_m(r), Y^*_m(r)))\Big|(\widehat{Y}^*_1(r), Y^*_1(r)), \ldots, (\widehat{Y}^*_s(r), Y^*_s(r)))]. \tag{22}$$

The proof is furnished by Zhou et al. (2020).

Theorem 4.2. *Under H_n,* $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) \xrightarrow{L} \mathcal{Z}^1 \sim N(\Upsilon^{(r)}, 4\eta_{1,2}(r))$*, where*

$$\Upsilon^{(r)} = \lim_{n\to\infty} Cov_{H_0}\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})),\, \log\frac{d\tilde{F}_n}{dF_0}\right).$$

Moreover, $\Upsilon^{(r)}$ can be derived as

$$\begin{aligned}
\Upsilon^{(r)} &= \lim_{n\to\infty} Cov_{H_0}\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})),\, \left\{\frac{\mu}{\sqrt{n}}\left(\frac{f}{f_0}-1\right) - \frac{\mu^2}{2n}\left(\frac{f}{f_0}-1\right)^2 + o(n^{-1})\right\}\right)\\
&= \lim_{n\to\infty} E_{H_0}\left[\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))\right)\cdot\left\{\frac{\mu}{\sqrt{n}}\left(\frac{f}{f_0}-1\right) - \frac{\mu^2}{2n}\left(\frac{f}{f_0}-1\right)^2 + o(n^{-1})\right\}\right]\\
&= \mu\, E_{H_0}\left[\left\{T_n^{<r>} - E_{H_0}(T_n^{<r>})\right\}\left(\frac{f}{f_0}-1\right)\right].
\end{aligned}$$

Le Cam's third lemma is useful to complete the proof of Theorem 4.2. Elaborated approach is available in **Das et al.** (2022). Next, the asymptotic power of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ needs to be derived so that whether $T_n^{<r>}$ is a consistent test statistic or not can be examined.

5 Asymptotic power of test statistic

The statistical power of a test statistic determines how it behaves when the null of hypothesis lacks its rationality. A reasonably powerful test statistic is always desired to perform a given test of hypothesis. However, consistency plays a key role to delineate the utility of the test statistic. In this circumstance, the setup of contiguous alternatives is considered only to develop a consistent test statistic $T_n^{<r>}$.

The asymptotic power of $T_n^{<r>}$ is determined as

$$\begin{aligned}
P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\kappa) &= 1 - P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) \le \pi_\kappa)\\
&= 1 - \Phi\left(\frac{\pi_\kappa - \Upsilon^{(r)}}{\sqrt{4\eta_{1,2}(r)}}\right) = \Phi\left(\frac{\Upsilon^{(r)} - \pi_\kappa}{\sqrt{4\eta_{1,2}(r)}}\right)
\end{aligned} \tag{23}$$

where π_κ is a point such that $P_{H_0}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\kappa) = \kappa$ with $0 < \kappa < 1$ and $\Phi(\cdot)$ is the CDF of a standard normal law. For $\mu = 0$, $\Upsilon^{(r)} = 0$ for which asymptotic power and size of $T_n^{<r>}$ are equal. Otherwise, as $\mu\uparrow$, the limiting power of $T_n^{<r>}$ increases, provided the probability of concordance exceeds 0.5.

Proposition 2. $\Upsilon^{(r)} = \mu(2P_{c;H_1} - 1)$ *where $P_{c;H_1}$ is the probability of concordance of n i.i.d. samples from* $(\widehat{Y}^*(r), Y^*(r))$.

Next, the consistency of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ as well as its bigger power for increasing sample size need to be checked to perform the relevant power study eventually.

Proposition 3. *For $n^* > n$, $P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\kappa) < P_{H_n}(\sqrt{n^*}(T_{n^*}^{<r>} - E(T_{n^*}^{<r>})) > \pi_\kappa)$ and $P_{H_n}(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\kappa) \uparrow 1$ as $\mu \uparrow$ and $n \to \infty$.*

Now we shall consider various examples to study the power performance of $T_n^{<r>}$ against the tuning parameter μ by taking sample size $n = 1000$ and the orders of difference $r = 2, 3, 4, 5, 10$.

5.1 Examples

I. We consider a generalized partially linear model $Y = \beta_1 X_1 + \beta_2 X_2 + m(W_1, W_2) + \epsilon$ with usual assumptions on error ϵ, *viz.*, (i) $E(\epsilon | X_1 = x_1, X_2 = x_2, W_1 = w_1, W_2 = w_2) = 0$ and (ii) $E(\epsilon^2 | X_1 = x_1, X_2 = x_2, W_1 = w_1, W_2 = w_2) = \sigma^2(x_1, x_2, w_1, w_2)$ for all (x_1, x_2, w_1, w_2). The joint distribution of $(X_1, X_2, W_1, W_2)^T$ is $\mathbb{N}_4\left(\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0.18 & -0.06 & 0.22 & -0.13\\ -0.06 & 0.14 & -0.28 & 0.19\\ 0.22 & -0.28 & 0.2 & 0.17\\ -0.13 & 0.19 & 0.17 & 0.25\end{pmatrix}\right)$ which is independent to $\epsilon \sim N(0, 0.015)$ under H_0. Also, Y has a t-distribution with 2 degrees of freedom. The nonparametric regression function is considered as $m(W_1, W_2) = 0.45 W_1 W_2 - 0.25 W_1^2 W_2 + W_2^3$. The conditional error distributions for $\underset{\sim 1}{c} \in \mathbb{R}^4$ under H_1 are

Example 1. *Under H_1, $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, W_1, W_2)^T \sim N(0,\, 0.015\, |1 + \underset{\sim 1}{c}^T \underset{\sim}{Z}|)$*

Example 2. *Under H_1, $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, W_1, W_2)^T \overset{\mathcal{D}}{=\!=} \dfrac{\chi^2_{\underset{\sim}{Z}} - d_{\underset{\sim}{Z}}}{\sqrt{2\, d_{\underset{\sim}{Z}}}}$ where $d_{\underset{\sim}{Z}} = |1 + \underset{\sim 1}{c}^T \underset{\sim}{Z}|^{-1}$ and $\chi^2_{\underset{\sim}{Z}} \sim \chi^2_{\lceil d_{\underset{\sim}{Z}} \rceil}$,*

The asymptotic power curves of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ for $r = 2, 3, 4, 5, 10$ against μ are displayed next.

(a) Asymptotic powers of $T_n^{<r>}$ in Example 1 (b) Asymptotic powers of $T_n^{<r>}$ in Example 2

Figure 1: Asymptotic powers of $T_n^{<r>}$'s in Example 1 and Example 2 where $\underset{\sim 1}{c} = (-1.5, -1.7, 1.2, 1.3)'$

(a) Asymptotic powers of $T_n^{<r>}$ in Example 1 (b) Asymptotic powers of $T_n^{<r>}$ in Example 2

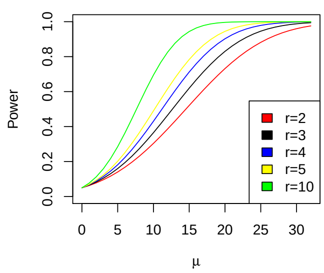
Figure 2: Asymptotic powers of $T_n^{<r>}$'s in Example 1 and Example 2 where $\underset{\sim 1}{c} = (5, -9, 4, 11)'$

II. A generalized partially linear model $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + m(W_1, W_2) + \epsilon$ is considered along with assumptions on ϵ as (i) $E(\epsilon | X_1 = x_1, X_2 = x_2, X_3 = x_3, W_1 = w_1, W_2 = w_2) = 0$ and (ii) $E(\epsilon^2 | X_1 = x_1, X_2 = x_2, X_3 = x_3, W_1 = w_1, W_2 = w_2) = \sigma^2(x_1, x_2, x_3, w_1, w_2)$ for all $(x_1, x_2, x_3, w_1, w_2)$. The joint distribution of $(X_1, X_2, X_3, W_1, W_2)^T$ is $\mathbb{N}_5 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.18 & -0.06 & 0.09 & 0.32 & -0.18 \\ -0.06 & 0.14 & -0.17 & 0.25 & 0.15 \\ 0.09 & -0.17 & 0.25 & -0.3 & -0.23 \\ 0.32 & 0.25 & -0.3 & 0.2 & 0.17 \\ -0.18 & 0.15 & -0.23 & 0.17 & 0.25 \end{pmatrix} \right)$ which is independently distributed to $\epsilon \sim N(0, 0.015)$ under H_0. In addition, $Y \sim t_2$. The nonparametric regression function is considered as $m(W_1, W_2) = 0.45 W_1 W_2 - 0.25 W_1^2 W_2 + W_2^3$. The conditional error distributions for $\underset{\sim 2}{c} \in \mathbb{R}^5$ under H_1 are

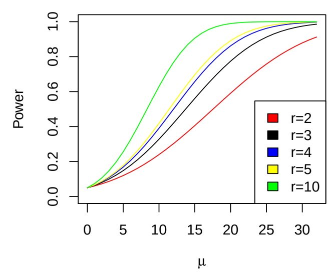
Example 3. *Under H_1,* $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, X_3, W_1, W_2)^T \sim N(0,\, 0.015\, |1 + \underset{\sim 2}{c}^T \underset{\sim}{Z}|)$

Example 4. *Under H_1,* $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, X_3, W_1, W_2)^T \stackrel{\mathcal{D}}{==} \dfrac{\chi^2_{\underset{\sim}{Z}} - d'_{\underset{\sim}{Z}}}{\sqrt{2\, d'_{\underset{\sim}{Z}}}}$ *where* $d'_{\underset{\sim}{Z}} = |1 + \underset{\sim 2}{c}^T \underset{\sim}{Z}|^{-1}$ *and* $\chi^2_{\underset{\sim}{Z}} \sim \chi^2_{\lceil d'_{\underset{\sim}{Z}} \rceil}$.

The asymptotic power curves of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ for $r = 2, 3, 4, 5, 10$ against μ values are presented below.

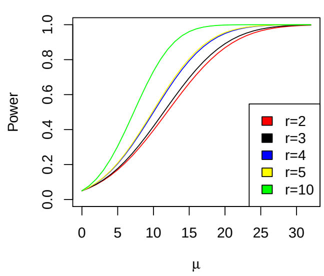


(a) Asymptotic powers of $T_n^{<r>}$ in Example 3

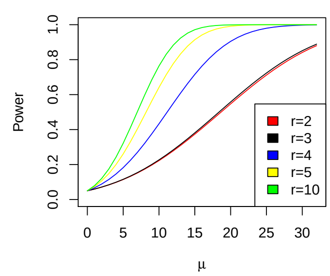


(b) Asymptotic powers of $T_n^{<r>}$ in Example 4

Figure 3: Asymptotic powers of $T_n^{<r>}$'s in Example 3 and Example 4 where $\underset{\sim 2}{c} = (-1.5, -1.7, 1.2, 1.3, -3.5)'$



(a) Asymptotic powers of $T_n^{<r>}$ in Example 3



(b) Asymptotic powers of $T_n^{<r>}$ in Example 4

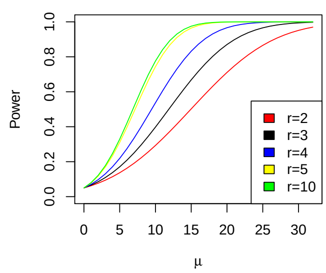
Figure 4: Asymptotic powers of $T_n^{<r>}$'s in Example 3 and Example 4 where $\underset{\sim 2}{c} = (-0.13, -2.1, 5.6, -0.95, 1.4)'$

III. The generalized partially linear model $Y = \beta_1 X_1 + \beta_2 X_2 + m(W_1, W_2, W_3) + \epsilon$ is considered with assumptions on ϵ as (i) $E(\epsilon | X_1 = x_1, X_2 = x_2, W_1 = w_1, W_2 = w_2, W_3 = w_3) = 0$ and (ii) $E(\epsilon^2 | X_1 = x_1, X_2 = x_2, W_1 = w_1, W_2 = w_2, W_3 = w_3) = \sigma^2(x_1, x_2, w_1, w_2, w_3)$ for all $(x_1, x_2, w_1, w_2, w_3)$. The joint distribution of $(X_1, X_2, W_1, W_2, W_3)^T$ is $\mathbb{N}_5\left(\begin{pmatrix}0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0.18 & -0.06 & 0.32 & -0.18 & -0.24\\ -0.06 & 0.14 & 0.25 & 0.15 & -0.18\\ 0.32 & 0.25 & 0.2 & 0.17 & -0.22\\ -0.18 & 0.15 & 0.17 & 0.25 & 0.11\\ -0.24 & -0.18 & -0.22 & 0.11 & 0.27\end{pmatrix}\right)$ which is independently distributed to $\epsilon \sim N(0, 0.015)$ under H_0 and $Y \sim t_2$. The nonparametric regression function is considered as $m(W_1, W_2, W_3) = 0.45W_1W_2 - 0.25W_1^2W_3 + W_3^3$. The conditional error distributions for $\underset{\sim 3}{c} \in \mathbb{R}^5$ under H_1 are

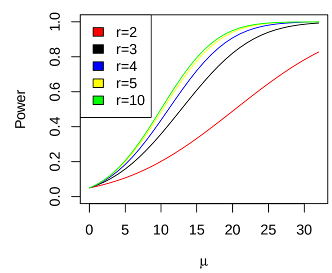
Example 5. *Under* H_1, $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, W_1, W_2, W_3)^T \sim N(0,\, 0.015\,|1 + \underset{\sim 3}{c}^T \underset{\sim}{Z}|)$

Example 6. *Under* H_1, $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, W_1, W_2, W_3)^T \overset{\mathcal{D}}{=\!=} \dfrac{\chi^2_{\underset{\sim}{Z}} - d^*_{\underset{\sim}{Z}}}{\sqrt{2\, d^*_{\underset{\sim}{Z}}}}$ *where* $d^*_{\underset{\sim}{Z}} = |1 + \underset{\sim 3}{c}^T \underset{\sim}{Z}|^{-1}$ *and* $\chi^2_{\underset{\sim}{Z}} \sim \chi^2_{\lceil d^*_{\underset{\sim}{Z}} \rceil}$.

The asymptotic power curves of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ for $r = 2, 3, 4, 5, 10$ against μ values are presented below.

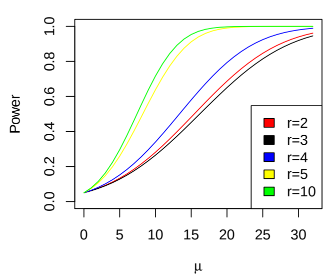


(a) Asymptotic powers of $T_n^{<r>}$ in Example 5

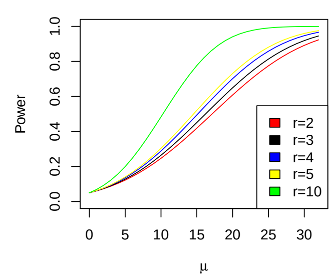


(b) Asymptotic powers of $T_n^{<r>}$ in Example 6

Figure 5: Asymptotic powers of $T_n^{<r>}$'s in Example 5 and Example 6 where $\underset{\sim 3}{c} = (2.5, 4, -6.2, -5.5, 3.1)'$



(a) Asymptotic powers of $T_n^{<r>}$ in Example 5



(b) Asymptotic powers of $T_n^{<r>}$ in Example 6

Figure 6: Asymptotic powers of $T_n^{<r>}$'s in Example 5 and Example 6 where $\underset{\sim 3}{c} = (-0.77, 0.34, 0.72, -0.56, 0.65)'$

IV. The generalized partially linear model $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + m(W_1, W_2, W_3) + \epsilon$ is considered with assumptions on ϵ as (i) $E(\epsilon|X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5, W_1 = w_1, W_2 = w_2, W_3 = w_3) = 0$ and (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 $\mathbb{N}_5\left(\begin{pmatrix} \underset{\sim 5}{0} \\ \underset{\sim 3}{0} \end{pmatrix}, \begin{pmatrix} \Sigma_5 & \Sigma_{53} \\ \Sigma_{35} & \Sigma_3 \end{pmatrix}\right)$ where

$$\Sigma_5 = \begin{pmatrix} 0.18 & -0.06 & 0.09 & -0.13 & 0.16 \\ -0.06 & 0.14 & -0.17 & 0.26 & -0.14 \\ 0.09 & -0.17 & 0.25 & 0.33 & -0.18 \\ -0.13 & 0.26 & 0.33 & 0.32 & 0.05 \\ 0.16 & -0.14 & -0.18 & 0.05 & 0.24 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 0.11 & -0.21 & -0.19 \\ -0.21 & 0.27 & 0.14 \\ -0.19 & 0.14 & 0.4 \end{pmatrix},$$

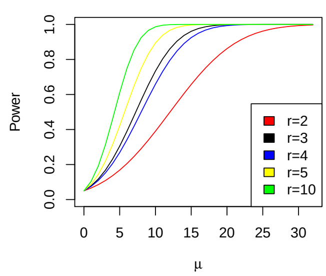
$$\Sigma_{53} = \begin{pmatrix} -0.17 & 0.31 & -0.22 \\ 0.18 & 0.25 & -0.33 \\ -0.24 & -0.15 & 0.15 \\ -0.14 & -0.07 & 0.12 \\ 0.26 & -0.18 & -0.03 \end{pmatrix}, \quad \Sigma_{35} = \Sigma_{53}^T = \begin{pmatrix} -0.17 & 0.18 & -0.24 & -0.14 & 0.26 \\ 0.31 & 0.25 & -0.15 & -0.07 & -0.18 \\ -0.22 & -0.33 & 0.15 & 0.12 & -0.03 \end{pmatrix}.$$

Moreover, $(X_1, X_2, X_3, X_4, X_5, W_1, W_2, W_3)^T$ is independently distributed to $\epsilon \sim N(0, 0.015)$ under H_0 and $Y \sim t_2$. The nonparametric regression function is considered as $m(W_1, W_2, W_3) = 0.36W_1^3 - 0.25W_2^2W_3 - 0.11W_3^2W_1 + 0.08W_1W_2W_3$. The conditional error distributions for $\underset{\sim 4}{c} \in \mathbb{R}^8$ under H_1 are

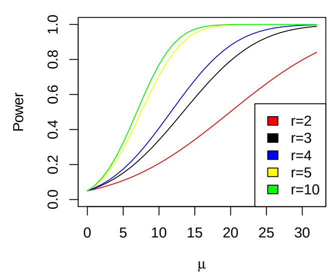
Example 7. *Under* H_1, $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, X_3, X_4, X_5, W_1, W_2, W_3)^T \sim N(0,\ 0.015\ |1 + \underset{\sim 4}{c}^T \underset{\sim}{Z}|)$

Example 8. *Under* H_1, $\epsilon \,\Big|\, \underset{\sim}{Z} = (X_1, X_2, X_3, X_4, X_5, W_1, W_2, W_3)^T \overset{\mathcal{D}}{==} \dfrac{\chi^2_{\underset{\sim}{Z}} - d^*_{\underset{\sim}{Z}}}{\sqrt{2\, d^*_{\underset{\sim}{Z}}}}$ *where* $d^*_{\underset{\sim}{Z}} = |1 + \underset{\sim 4}{c}^T \underset{\sim}{Z}|^{-1}$ *and* $\chi^2_{\underset{\sim}{Z}} \sim \chi^2_{\lceil d^*_{\underset{\sim}{Z}} \rceil}$.

The asymptotic power curves of $\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))$ for $r = 2, 3, 4, 5, 10$ against μ values are presented below.



(a) Asymptotic powers of $T_n^{<r>}$ in Example 7



(b) Asymptotic powers of $T_n^{<r>}$ in Example 8

Figure 7: Asymptotic powers of $T_n^{<r>}$'s in Example 7 and Example 8 where $\underset{\sim 4}{c} = (5.92, -3.78, -10.66, 8.89, -5.45, 9.65, 8.35, -7.89)'$

Various examples on partially linear models with specified conditional error structures reveal on enhancement of the asymptotic powers of $T_n^{<r>}$ due to improved values of r. All the power curves are consistent also, as they close to 1 for $\mu \uparrow$. So, the association between jointly distributed covariates and random error is quite sensitive; falsity of the null hypothesis is well captured by $T_n^{<r>}$ as r increases. Hence, the robustness of a nonparametric statistic is important to detect the presence dependence in a partially linear model.

It is quite noteworthy that the power curves of the test statistics $T_n^{<10>}$'s (indicated by green graphs) exhibit significantly high power curves in almost all the instances above. Moreover, in case of conditional error distribution being Gaussian, the performances of $T_n^{<10>}$'s are better compared to the situations where conditional errors are standardized chi-squared random variables. In few cases, the power curves shown by test statistics $T_n^{<4>}$ and $T_n^{<5>}$ of observed and predicted responses are too close to the power curves of $T_n^{<10>}$. As usual, $T_n^{<2>}$ emerges as lowest performer compared to the other ones, although having power curves being almost same with the power curves of

$T_n^{<3>}$ in some cases. Such increment of power performances of $T_n^{<r>}$'s with $r = 2, 3, 4, 5, 10$ further motivates us to consider a test statistic $T_n^{<r>}$ with higher r to conduct the testing of hypotheses with greater consistency. All the values of asymptotic powers of $T_n^{<r>}$'s are provided in **Appendix-I**. To substantiate the purpose of testing of association between regressors and error further, the relevance of $T_n^{<r>}$'s must be studied for real datasets also.

6 Real data analysis

7 Conclusion

8 Acknowledgement

9 Reference

10 Appendix-I

11 Appendix-II

11.1 Proof of Theorem 3.1

Note that,

$$\begin{aligned}
Y_{i+1} - Y_i &= \underset{\sim}{\beta}^T(\underset{\sim}{X}_{i+1} - \underset{\sim}{X}_i) + \{m(\underset{\sim}{W}_{i+1}) - m(\underset{\sim}{W}_i)\} + (\epsilon_{i+1} - \epsilon_i), \;\; i = 1, \ldots, n \\
\implies P\left(|(Y_{i+1} - Y_i) - (\epsilon_{i+1} - \epsilon_i)| > \delta\right) &= P\left(|\underset{\sim}{\beta}^T(\underset{\sim}{X}_{i+1} - \underset{\sim}{X}_i) + \{m(\underset{\sim}{W}_{i+1}) - m(\underset{\sim}{W}_i)\}| > \delta\right) \\
&\leq P\left(|\underset{\sim}{\beta}^T(\underset{\sim}{X}_{i+1} - \underset{\sim}{X}_i)| + |m(\underset{\sim}{W}_{i+1}) - m(\underset{\sim}{W}_i)| > \delta\right) \\
&= P\left(|\sum_{k=1}^{p} \beta_k (X_{i+1\,k} - X_{i\,k})| + |m(\underset{\sim}{W}_{i+1}) - m(\underset{\sim}{W}_i)| > \delta\right) \;;\; \delta > 0. \quad (24)
\end{aligned}$$

Since $m(\cdot, \ldots, \cdot)$ is Lipschitz continuous on $\mathbb{R}^q$, therefore $|m(\underset{\sim}{W}_{i+1}) - m(\underset{\sim}{W}_i)| \leq K \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q$ where $K(> 0)$ is a constant and $||\cdot||_q$ is the q-th norm of a vector in $\mathbb{R}^q$.

$\therefore$ (24) $\leq P\left(|\sum_{k=1}^{p} \beta_k (X_{i+1\,k} - X_{i\,k})| + K \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta\right)$.

Define, $X_{i+1\,k} - X_{i\,k} = D_{ik}$ for $k = 1, \ldots, p$. Then,

$$\begin{aligned}
& P\left(|\sum_{k=1}^{p} \beta_k (X_{i+1\,k} - X_{i\,k})| + K \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta\right) \\
= \;& P\left(|\sum_{k=1}^{p} \beta_k D_{ik}| + K \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta\right) \\
= \;& \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} P\left(|\sum_{k=1}^{p} \beta_k D_{ik}| + K \cdot ||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta | D_{ik} = d_{ik}\right) \times \prod_{k=1}^{p} dH_{D_{ik}}(d_{ik}) \\
= \;& \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} P\left(||\underset{\sim}{W}_{i+1} - \underset{\sim}{W}_i||_q > \delta'\right) \times \prod_{k=1}^{p} dH_{D_{ik}}(d_{ik}), \text{ where } \delta' = \frac{\left(\delta - |\sum_{k=1}^{p} \beta_k d_{ik}|\right)}{K} \\
= \;& \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} P\left[\left(\sum_{m=1}^{n} |W_{i+1\,m} - W_{i\,m}|^q\right)^{\frac{1}{q}} > \delta'\right] \times \prod_{k=1}^{p} dH_{D_{ik}}(d_{ik}),
\end{aligned}$$

where

$$\begin{aligned}
P\left[\left(\sum_{m=1}^{q} |W_{i+1\,m} - W_{i\,m}|^q\right)^{\frac{1}{q}} > \delta'\right] &= P\left[|W_{i+1\,1} - W_{i\,1}| > \left(\delta'^{\,q} - \left(\sum_{m=2}^{q} |W_{i+1\,m} - W_{i\,m}|^q\right)\right)^{\frac{1}{q}}\right] \\
&= P\left[|W_{i+1\,1} - W_{i\,1}| > \left(\delta'^{\,q} - \left(\sum_{m=2}^{q} |T_{im}|^q\right)\right)^{\frac{1}{q}}\right], \quad (25)
\end{aligned}$$

where $T_{im} = (W_{i+1\,m} - W_{i\,m})$ for $m = 2, \ldots, q$; $i = 1, \ldots, n-1$.

Now, without loss of generality, the model is re-expressed by ordering the W_1-observations as

$$\begin{aligned}
Y_1^* &= \beta_1 X_{11}^* + \ldots + \beta_p X_{1p}^* + g(W_{11}^*, \ldots, W_{1q}^*) + \epsilon_1^* \\
Y_2^* &= \beta_1 X_{21}^* + \ldots + \beta_p X_{2p}^* + g(W_{21}^*, \ldots, W_{2q}^*) + \epsilon_2^* \\
&\vdots \\
Y_n^* &= \beta_1 X_{n1}^* + \ldots + \beta_p X_{np}^* + g(W_{n1}^*, \ldots, W_{nq}^*) + \epsilon_n^*
\end{aligned}$$

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, (25) can be further deduced as

$$
\begin{aligned}
& P\left[|W_{i+1\,1}^* - W_{i\,1}^*| > \left(\delta'^{\,q} - \left(\sum_{m=2}^{q} |T_{im}^*|^q\right)\right)^{\frac{1}{q}}\right], \text{ where } T_{im}^* = W_{i+1\,m}^* - W_{i\,m}^* \\
= \;& \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} P\left[|W_{i+1\,1}^* - W_{i\,1}^*| > \left(\delta'^{\,q} - \left(\sum_{m=2}^{q} |T_{im}^*|^q\right)\right)^{\frac{1}{q}} |T_{im}^* = t_{im}^*\right] \times \prod_{m=2}^{q} dF_{T_{im}^*}(t_{im}^*) \\
= \;& \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} P\left[|W_{i+1\,1}^* - W_{i\,1}^*| > \delta''\right] \times \prod_{m=2}^{q} dF_{T_{im}^*}(t_{im}^*), \text{ where } \delta'' = \left(\delta'^{\,q} - \left(\sum_{m=2}^{q} |t_{im}^*|^q\right)\right)^{\frac{1}{q}}. \qquad (26)
\end{aligned}
$$

Here, $(W_{i+1\,1}^* - W_{i\,1}^*)$ denotes the spacing between i-th and $(i+1)$-th order statistics on W_1, $i = 1, \ldots, n-1$. Suppose $F_{W_1}(\cdot)$ is the CDF of W_1. For any $\delta^* > 0$,

$$
\begin{aligned}
P\left[|W_{i+1\,1}^* - W_{i\,1}^*| > \delta^*\right] &= P\left[|F_{W_1}^{-1}(F_{W_1}(W_{i+1\,1}^*)) - F_{W_1}^{-1}(F(W_{i\,1}^*))| > \delta^*\right], \\
&\quad \text{by assuming } F_{W_1}^{-1}(\cdot) \text{ exists} \\
&= P\left[|F_{W_1}^{-1}(U_{i+1\,1}^*) - F_{W_1}^{-1}(U_{i\,1}^*)| > \delta^*\right], \text{ where } U_{p1}^* \; (1 \leq p \leq n-1) \text{ is the} \\
&\quad \text{p-th ordered uniform random variable on U(0,1) distribution} \\
&= P\left[|F_{W_1}^{-1'}(\xi_{i;i+1})| \cdot |U_{i+1\,1}^* - U_{i\,1}^*| > \delta^*\right], \text{ where } F_{W_1}^{-1'}(\xi_{i;i+1}) \text{ is the} \\
&\quad \text{first order derivative of } F_{W_1}^{-1}(\cdot) \text{ at } \xi_{i;i+1} \in (U_{i1}^*, U_{i+1\,1}^*) \\
&= P\left[|U_{i+1\,1}^* - U_{i\,1}^*| > \delta_i^{**}\right], \text{ where } \delta_i^{**} = (\delta^* / |F_{W_1}^{-1'}(\xi_{i;i+1})|) \\
&\leq P\left[\sup_{1 \leq i \leq n-1} |U_{i+1\,1}^* - U_{i\,1}^*| > \inf_{1 \leq i \leq n-1} \delta_i^{**}\right] \\
&= P\left[\Delta_{n-1;n} > R\right], \text{ where } \Delta_{n-1;n} \text{ is the maximal spacing based on} \\
&\quad (n-1) \text{ uniform spacings } (U_{i+1\,1}^* - U_{i1}^*)s \text{ and } R = \inf_{1 \leq i \leq n-1} \delta_i^{**} \\
&= P\left[\frac{n\Delta_{n-1;n}}{\log n} > \frac{nR}{\log n}\right] \longrightarrow 1 - \exp\left(-\exp\left(-\frac{nR}{\log n}\right)\right) \quad (\textbf{Levy}\ (1939)) \qquad (27)
\end{aligned}
$$

Then, it is clear that (26) $\leq 1 - \exp\left(-\exp\left(-\dfrac{nR}{\log n}\right)\right)$.
$\because \dfrac{nR}{\log n} \longrightarrow \infty$ as $n \longrightarrow \infty$. $\therefore \exp\left(-\dfrac{nR}{\log n}\right) \longrightarrow 0$ as $n \longrightarrow \infty$, which further implies the R.H.S. of (27) tends to $1 - \exp(0) = 0$. Then, $|W_{i+1\,1}^* - W_{i\,1}^*| = o_p(1)$, which further implies that $\sup_{i \in \{1,\ldots,n-1\}} |W_{i+1\,1}^* - W_{i\,1}^*| = o_p(1)$.

Therefore, (24) to (27) implies that $P\left(|(Y_{i+1}^* - Y_i^*) - (\epsilon_{i+1}^* - \epsilon_i^*)| > \delta\right) \longrightarrow 0$, i.e. $Y_{i+1}^* - Y_i^* = \epsilon_{i+1}^* - \epsilon_i^* + o_p(1)$; $i = 1, \ldots, n-1$, which is the first order difference of Y^*, expressed in terms of the first order difference of ϵ^*.

We would like to propose a more general setup of hypotheses by defining a general r-th order difference of ϵ, denoted as $\epsilon^*(r)$. We need to verify if $\epsilon^*(r) \approx Y^*(r)$, where $Y^*(r)$ is the r-th order difference of Y.

Let us define a function of (r+1) observations on ϵ as $\psi(\epsilon_1, \ldots, \epsilon_{r+1}) = \alpha_1\epsilon_1 + \ldots + \alpha_{r+1}\epsilon_{r+1}$ such that α_i's are real coefficients of ϵ_i's, $i = 1, \ldots, (r+1)$. Moreover, $\sum_{i=1}^{r+1} \alpha_i = 0$ with $\alpha_i \neq 0 \; \forall \; i$.
The k-th order absolute raw moment of $\psi(\epsilon_1, \ldots, \epsilon_{r+1})$ is

$$
\begin{aligned}
E|\psi(\epsilon_1, \ldots, \epsilon_{r+1})|^k = E|\sum_{i=1}^{r+1} \alpha_i \epsilon_i|^k \; &\leq \; E\left[\sqrt{\sum_{i=1}^{r+1} \alpha_i^2} \cdot \sqrt{\sum_{i=1}^{r+1} \epsilon_i^2}\right]^k \quad \text{(using Cauchy-Schwartz inequality)} \\
&= \left(\sum_{i=1}^{r+1} \alpha_i^2\right)^{k/2} E\left[\sqrt{\sum_{i=1}^{r+1} \epsilon_i^2}\right]^k = \eta(\alpha_1, \ldots, \alpha_{r+1}) \cdot E\left[\sqrt{\sum_{i=1}^{r+1} \epsilon_i^2}\right]^k \qquad (28)
\end{aligned}
$$

where $\eta(\alpha_1,\ldots,\alpha_{r+1}) = \left(\sum_{i=1}^{r+1}\alpha_i^2\right)^{k/2}$. Now, observe that,

$$\eta(\alpha_1,\ldots,\alpha_{r+1}) \le \max_{\alpha_1,\ldots,\alpha_{r+1}} \eta(\alpha_1,\ldots,\alpha_{r+1}) \ \forall\ \alpha_1,\ldots,\alpha_{r+1}\in\mathbb{R} \text{ with } \sum_{i=1}^{r+1}\alpha_i = 0,\ \alpha_i \ne 0.$$

Note that, for $r = 2$,

$$\begin{aligned}\eta(\alpha_1,\alpha_2,\alpha_3) = \alpha_1^2+\alpha_2^2+\alpha_3^2 &= (\alpha_1+\alpha_2)^2 - 2\alpha_1\alpha_2 + \alpha_3^2\\ &= (-\alpha_3)^2 - 2\alpha_1(-\alpha_1-\alpha_3)+\alpha_3^2\\ &= 2(\alpha_3^2+\alpha_1^2+\alpha_3\alpha_1)\\ &= 2\left(\frac{\alpha_3^3-\alpha_1^3}{\alpha_3-\alpha_1}\right) = \zeta(\alpha_1,\alpha_3), \text{ say.}\end{aligned}$$

Then, we get $\log\zeta = \log 2 + \log(\alpha_3^3-\alpha_1^3) - \log(\alpha_3-\alpha_1)$ and maximize with respect to α_1, α_3 as

$$\begin{aligned}\frac{\partial\log\zeta}{\partial\alpha_1} = 0 &\implies \frac{-3\alpha_1^2}{\alpha_3^3-\alpha_1^3} + \frac{1}{\alpha_3-\alpha_1} = 0\\ &\implies -3\alpha_1^2 + (\alpha_3^2+\alpha_1^2+\alpha_3\alpha_1) = 0\\ &\implies -2\alpha_1^2+\alpha_1\alpha_3+\alpha_3^2 = 0 \end{aligned}\tag{29}$$

$$\begin{aligned}\text{and } \frac{\partial\log\zeta}{\partial\alpha_3} = 0 &\implies \frac{3\alpha_3^2}{\alpha_3^3-\alpha_1^3} - \frac{1}{\alpha_3-\alpha_1} = 0\\ &\implies 3\alpha_3^2 - (\alpha_3^2+\alpha_1^2+\alpha_3\alpha_1) = 0\\ &\implies 2\alpha_3^2-\alpha_1^2-\alpha_1\alpha_3 = 0 \end{aligned}\tag{30}$$

$$\begin{aligned}\therefore\ (29)+(30) &\implies -3\alpha_1^2+3\alpha_3^2 = 0\\ &\implies \alpha_1^2 = \alpha_3^2 \implies \alpha_1 = \pm\alpha_3.\end{aligned}$$

If $\alpha_1 = -\alpha_3$, then $\alpha_1+\alpha_3 = 0 \implies \alpha_2 = 0$. But α_i's are nonzero with $\sum_{i=1}^{3}\alpha_i = 0$.
$\therefore \alpha_1 = \alpha_3 \implies \alpha_2 = -2\alpha_1$; $\alpha_1 \ne 0$. Then, $\zeta(\alpha_1,\alpha_3)$ is maximized at (α_1,α_1) and hence $\eta(\alpha_1,\alpha_2,\alpha_3)$ has maximum value at $(\alpha_1,-2\alpha_1,\alpha_1)$, $\alpha_1\ne 0$.
Now, from (28), the maximum value of k-th order absolute moment of $(\alpha_1\epsilon_1 - 2\alpha_1\epsilon_2+\alpha_1\epsilon_3)$ satisfies the following inequality

$$E|\alpha_1\epsilon_1-2\alpha_1\epsilon_2+\alpha_1\epsilon_3|^k \le |\alpha_1|^k\{1^2+(-2)^2+1^2\}^{k/2}\ E\left(\sqrt{\sum_{i=1}^{3}\epsilon_i^2}\right)^k$$

for any $\alpha_1\ne 0$. Without loss of generality, take $\alpha_1 = 1$. Then, for $\alpha_1 = 1$, among all possible linear contrasts of ϵ_1, ϵ_2 and ϵ_3, the contrast $(\epsilon_1-2\epsilon_2+\epsilon_3)$ has maximum absolute raw moment of order $k(\in\mathbb{N})$. One can also write

$$E|\alpha_1\epsilon_1-2\alpha_1\epsilon_2+\alpha_1\epsilon_3|^k \ge E|\delta_1\epsilon_1+\delta_2\epsilon_2+\delta_3\epsilon_3|^k$$

for all possible linear functions $(\delta\epsilon_1+\delta_2\epsilon_2+\delta_3\epsilon_3)$.

Let us define $g(\epsilon_1,\epsilon_2,\epsilon_3) = \epsilon_1-2\epsilon_2+\epsilon_3$, termed as the *second order difference of* ϵ based on $\epsilon_1,\epsilon_2,\epsilon_3$, as $g(\epsilon_1,\epsilon_2,\epsilon_3)$ has minimum variance. It is to be noted that $(\epsilon_1-2\epsilon_2+\epsilon_3)$ can be interpreted as the first order difference of two first order differences of ϵ, i.e. $(\epsilon_1-\epsilon_2)-(\epsilon_2-\epsilon_3)$. Then, $(\epsilon_1-\epsilon_2)-(\epsilon_2-\epsilon_3) = \epsilon_1^*(1)-\epsilon_2^*(1) \approx Y_1^*(1)-Y_2^*(1)$, *i.e.* $\epsilon^*(2)\approx Y^*(2)$.
For the case $r = 3$, note that

$$\begin{aligned}\eta(\alpha_1,\alpha_2,\alpha_3,\alpha_4) &= \alpha_1^2+\alpha_2^2+\alpha_3^2+\alpha_4^2 \text{ subject to } \alpha_1+\alpha_2+\alpha_3+\alpha_4 = 0\\ &= (\alpha_2+\alpha_3+\alpha_4)^2+\alpha_2^2+\alpha_3^2+\alpha_4^2\\ &= \zeta(\alpha_2,\alpha_3,\alpha_4), \text{ say.}\end{aligned}$$

Now, ζ is maximized by solving the following equations.

$$\frac{\partial\zeta}{\partial\alpha_2} = 0 \implies 2\alpha_2 + \alpha_3 + \alpha_4 = 0 \tag{31}$$
$$\frac{\partial\zeta}{\partial\alpha_3} = 0 \implies \alpha_2 + 2\alpha_3 + \alpha_4 = 0 \tag{32}$$
$$\frac{\partial\zeta}{\partial\alpha_4} = 0 \implies \alpha_2 + \alpha_3 + 2\alpha_4 = 0 \tag{33}$$

$$\therefore (31) + (33) - (32) \implies 2(\alpha_2 + \alpha_4) = 0 \implies \alpha_2 = -\alpha_4. \tag{34}$$

$$\text{and } \ (33) - 2 \times (31) \implies \alpha_3 = -3\alpha_2 \ \textit{ i.e. } \ \alpha_3 = 3\alpha_4. \tag{35}$$

Then, $\zeta(\alpha_2, \alpha_3, \alpha_4)$ is maximized at $(-\alpha_4, 3\alpha_4, \alpha_4)$, $\alpha_4 \neq 0$. Note that, $\alpha_1 = -(-\alpha_4 + 3\alpha_4 + \alpha_4) = 3\alpha_4$. Therefore, $\eta(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is maximized at $(-3\alpha_4, -\alpha_4, 3\alpha_4, \alpha_4)$, $\alpha_4 \neq 0$, i.e.

$$E| - 3\alpha_4\epsilon_1 - \alpha_4\epsilon_2 + 3\alpha_4\epsilon_3 + \alpha_4\epsilon_4|^k \quad \geq \quad E|\beta\epsilon_1 + \beta_2\epsilon_2 + \beta_3\epsilon_3 + \beta_4\epsilon_4|^k, \ \text{ for any } \beta_1, \beta_2, \beta_3, \beta_4 \in \mathbb{R}$$

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 involving ϵ_i's, $i = 1, 2, 3, 4$. Taking $\alpha_4 = 1$, the linear contrast $(\epsilon_4 - 3\epsilon_1 + 3\epsilon_4 - \epsilon_2)$ [or equivalently, $(\epsilon_1 - 3\epsilon_2 + 3\epsilon_3 - \epsilon_4)$ 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 of ϵ_1, ϵ_2, ϵ_3 and ϵ_4. Here, the function $(\epsilon_1 - 3\epsilon_2 + 3\epsilon_3 - \epsilon_4)$ is known as the *third order difference of* ϵ based on four i.i.d. observations ϵ_1, ϵ_2, ϵ_3 and ϵ_4. In fact, it can be defined as 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 $\{(\epsilon_1 - \epsilon_2) - 2(\epsilon_2 - \epsilon_3) + (\epsilon_3 - \epsilon_4)\}$. By similar logic, one can conclude that $\epsilon^*(3) \approx Y^*(3)$.

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 is easy to generalize the whole context for a general order difference r of ϵ which finally yields $\epsilon^*(r) \approx Y^*(r)$.

Furthermore, to show that $\epsilon^*(r)$ has the maximal k-th order absolute moment among all possible linear functions of $\epsilon_1, \ldots, \epsilon_{r+1}$, let us denote this assertion by $P(r)$, $r \geq 2$. Previously, we showed that $P(r)$ holds true for $r = 2, 3$. Now, suppose for $r = m - 1$, $P(m-1)$ is true, i.e.

$$v(\epsilon_1^*, \ldots, \epsilon_m^*) = \sum_{j=1}^{m} \alpha_j \epsilon_j^* \,; \ \alpha_1, \ldots, \alpha_m \in \mathbb{R} \text{ with } \alpha_j = (-1)^{j-1}\binom{m-1}{j-1}.$$

Then, we need to verify if $P(m)$ is true. Since, $P(m-1)$ holds, therefore

$$E|\sum_{j=1}^{m} \{(-1)^{j-1}\binom{m-1}{j-1}\}\epsilon_j^*|^k \geq E|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^k \ \ \forall \ k \in \mathbb{N}, \ \text{ for any } v(\epsilon_1^*, \ldots, \epsilon_m^*).$$

Note that,

$$\begin{aligned}
\sum_{j=1}^{m+1}(-1)^{j-1}\binom{m}{j-1}\epsilon_j^* &= \sum_{j=1}^{m}(-1)^{j-1}\binom{m}{j-1}\epsilon_j^* + (-1)^m\epsilon_{m+1}^* \\
&= \sum_{j=1}^{m}(-1)^{j-1}\left[\binom{m-1}{j-1} + \binom{m-1}{j-2}\right]\epsilon_j^* + (-1)^m\,\epsilon_{m+1}^* \\
&= \sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^* + \sum_{j=2}^{m}(-1)^{j-1}\binom{m-1}{j-2}\epsilon_j^* + (-1)^m\,\epsilon_{m+1}^* \\
&= \sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^* + \sum_{j=2}^{m+1}(-1)^{j-1}\binom{m-1}{j-2}\epsilon_j^* \\
&= \sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^* + \sum_{l=1}^{m}(-1)^{l}\binom{m-1}{l-1}\epsilon_{l+1}^*.
\end{aligned}$$

Now, $v(\epsilon_1^*, \ldots, \epsilon_{m+1}^*)$ can be further simplified as

$$v(\epsilon_1^*, \ldots, \epsilon_{m+1}^*) = \sum_{j=1}^{m+1} \alpha_j \epsilon_j^* = \sum_{j=1}^{m} \alpha_j \epsilon_j^* + \alpha_{m+1}\epsilon_{m+1}^* = v(\epsilon_1^*, \ldots, \epsilon_m^*) + \alpha_{m+1}\epsilon_{m+1}^*.$$

$$\begin{aligned}
\therefore \quad & |\sum_{j=1}^{m+1}(-1)^{j-1}\binom{m}{j-1}\epsilon_j^*|^k - |v(\epsilon_1^*, \ldots, \epsilon_{m+1}^*)|^k \\
= \; & |\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^* - \sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^k - |v(\epsilon_1^*, \ldots, \epsilon_m^*) + \alpha_{m+1}\epsilon_{m+1}^*|^k \\
\geq \; & \left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*| - |\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|\right)^k - \left(|v(\epsilon_1^*, \ldots, \epsilon_m^*)| + |\alpha_{m+1}\epsilon_{m+1}^*|\right)^k \\
& (\because |a-b| \geq |a| - |b| \text{ and } |a+b| \leq |a| + |b|) \\
= \; & |\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^k + \sum_{t=1}^{k}\binom{k}{t}|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t \times |\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t} \\
& -|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^k + \sum_{t=1}^{k}\binom{k}{t}|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^t|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t} \text{ for any } \alpha_1, \ldots, \alpha_m \in \mathbb{R}.
\end{aligned}$$

Taking expectations both sides, we get

$$\begin{aligned}
& E|\sum_{j=1}^{m+1}(-1)^{j-1}\binom{m}{j-1}\epsilon_j^*|^k - E|v(\epsilon_1^*, \ldots, \epsilon_{m+1}^*)|^k \\
\geq \; & \left[E|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^k - E|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^k\right] \\
+ \; & \sum_{t=1}^{k}\binom{k}{t}\left[E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - E\left(|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^t|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right)\right]
\end{aligned} \tag{36}$$

$\because E|\sum_{j=1}^{m}(-1)^j\binom{m-1}{j-1}\epsilon_j^*|^k \geq E|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^k$, for any k.

$\therefore$ RHS of (36) $\geq \sum_{t=1}^{k}\binom{k}{t}\left[E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - E\left(|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^t|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right)\right].$

Now,

$$\begin{aligned}
& E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - E\left(|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^t|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right) \\
= \; & cov\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t, |\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - cov\left(|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^t, |\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right) \\
& +E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t\right)E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) \\
& -E\left(|v(\epsilon_1^*, \ldots, \epsilon_m^*)|^t\right)E\left(|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right); \text{ for } t = 1, \ldots, k.
\end{aligned} \tag{37}$$

Note that, for any $v(\cdot,\ldots,\cdot)$,

$$\begin{aligned}
E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t\right) &\geq E\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t\right) \\
\implies E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t\right)E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) &\geq E\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t\right) \\
&\quad \times E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right)
\end{aligned}$$

$$\begin{aligned}
\implies E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t\right)E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - E\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t\right)E\left(|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right) \\
\geq E\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t\right)\{E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - E\left(|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right)\}. \quad (38)
\end{aligned}$$

To prove if **RHS** of (38) ≥ 0 or not. Define, $\tilde{\epsilon^*}_l = \epsilon_{l+1}^*$, $l = 1,\ldots,m$. Then,

$$\begin{aligned}
E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) &= E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\tilde{\epsilon^*}_l|^{k-t}\right) \\
&\geq E\left(|v(\tilde{\epsilon^*}_1,\ldots,\tilde{\epsilon^*}_m)|^t\right) \text{ for any } v(\cdot,\ldots,\cdot).
\end{aligned}$$

Take $v(\tilde{\epsilon^*}_1,\ldots,\tilde{\epsilon^*}_m) = \alpha_{m+1}\tilde{\epsilon^*}_m = \alpha_{m+1}\epsilon_{m+1}^*$; $\alpha_{m+1} \neq 0$. Therefore,

$$E\left(|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) \geq E\left(|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right).$$

Also, $E\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t\right) \geq 0 \ \forall \ \epsilon_1^*,\ldots,\epsilon_m^* \implies$ RHS of (38) ≥ 0.
Again, $cov\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t, |\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right) = 0$ as $\epsilon_{m+1}^* \perp\!\!\!\perp \epsilon_1^*,\ldots,\epsilon_m^*$. Therefore, (37) can be expressed as

$$\begin{aligned}
& E\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t|\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) - E\left(|v(\epsilon_1^*,\ldots,\epsilon_m^*)|^t|\alpha_{m+1}\epsilon_{m+1}^*|^{k-t}\right) \\
\geq \quad & cov\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t, |\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\epsilon_{l+1}^*|^{k-t}\right) \\
= \quad & cov\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t, |\sum_{l=1}^{m}(-1)^{l-1}\binom{m-1}{l-1}\tilde{\epsilon^*}_l|^{k-t}\right) \\
= \quad & cov\left(|\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^t, |\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|^{k-t}\right), \text{ due to distributional equality} \\
i.e. \quad & cov\left(Z^t, Z^{k-t}\right) \geq 0, \text{ where } Z = |\sum_{j=1}^{m}(-1)^{j-1}\binom{m-1}{j-1}\epsilon_j^*|(\geq 0) \text{ with } k \geq t.
\end{aligned}$$

Therefore, one can further imply that RHS of (36) ≥ 0, which finally leads us to conclude that

$$E|\sum_{j=1}^{m+1}(-1)^{j-1}\binom{m}{j-1}\epsilon_j^*|^k \geq E|v(\epsilon_1^*,\ldots,\epsilon_{m+1}^*)|^k$$

i.e. $P(m)$ is true. Hence, by mathematical induction, we are able to show that the r-th order difference of ϵ^* based on $\epsilon_1^*,\ldots,\epsilon_{r+1}^*$ has the maximum k-th moment among all possible choices of $g(\epsilon_1^*,\ldots,\epsilon_{r+1}^*)$.

11.2 Proof of Proposition 1

It is to be noted that for $p = 1 = q$, $\widehat{\beta} = \left(\sum_{i=1}^{n} \hat{\epsilon}_{Xi}\hat{\epsilon}_{Xi}^{T}\right)^{-1}\left(\sum_{i=1}^{n} \hat{\epsilon}_{Xi}\hat{\epsilon}_{Yi}\right)$ 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,

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

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

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

We further assume that $\sup\limits_{X,W\in\mathbb{R}}\left|\frac{\partial}{\partial Z}\zeta(Z, X, W)\Big|_{Z=\epsilon}\right| < \infty$, say $\mathbf{A}$. It is to be noted that

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

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

$$\begin{aligned}\left|\zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| &\leq \left|\zeta^{(r)}(X\beta + m(W), X, W)\right| + \left|\epsilon\cdot\left(\frac{\partial\zeta}{\partial Z}\Big|_{Z=\epsilon}\right)\right|(r)\\&\leq \left|\zeta^{(r)}(X\beta + m(W), X, W)\right| + \mathbf{A}\epsilon(r).\end{aligned}$$

By using the inequality $|s_1 - s_2| \geq |s_1| - |s_2|$ for $s_1, s_2 \in \mathbb{R}$, one can verify that

$$\left|\zeta^{(r)}(X\beta + m(W), X, W)\right| - \left|\zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| \geq -\mathbf{A}\epsilon(r)$$

$$\implies \left|\zeta^{(r)}(X\beta + m(W), X, W) - \zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| \geq -\mathbf{A}\epsilon(r)$$

$$\text{i.e. } -\left|\zeta^{(r)}(X\beta + m(W), X, W) - \zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| \leq \mathbf{A}\epsilon(r) \leq \mathbf{A}|\epsilon(r)|$$

For any positive quantity ϕ,

$$P\left(-\left|\zeta^{(r)}(X\beta + m(W), X, W) - \zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| \geq \phi\right) \leq P(|\mathbf{A}\epsilon(r)| \geq \phi) \longrightarrow 0 \text{ as } n \to \infty.$$

Therefore,

$$-\left|\zeta^{(r)}(X\beta + m(W), X, W) - \zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| = o_P(1)$$

$$\implies \left|\zeta^{(r)}(X\beta + m(W), X, W) - \zeta^{(r)}(X\beta + m(W) + \epsilon, X, W)\right| = o_P(1)$$

from which it is easy to conclude that

$$\zeta^{(r)}(X\beta + m(W) + \epsilon, X, W) = \widehat{Y}^{*}(r) \approx \zeta^{(r)}(X\beta + m(W), X, W), \text{ a function of (X,W).}$$

A higher order generalization of the above validates our main proposition in similar manner when $p, q > 1$.

11.3 Proof of Proposition 2

Note that,

$$\Upsilon^{(r)} = \lim_{n\to\infty} cov_{H_0}\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})), \log\frac{d\tilde{F}_n}{dF_0}\right)$$

where

$$\begin{aligned}
\frac{d\tilde{F}_n}{dF_0}(\widehat{y}^*, y^*) &= \frac{d}{dF_0}\left[\left(1-\frac{\mu}{\sqrt{n}}\right)F_0(\widehat{y}^*, y^*) + \frac{\mu}{\sqrt{n}}F(\widehat{y}^*, y^*)\right] \\
&= \left(1-\frac{\mu}{\sqrt{n}}\right) + \frac{\mu}{\sqrt{n}}\cdot\frac{dF}{dF_0}(\widehat{y}^*, y^*) \;=\; 1+\frac{\mu}{\sqrt{n}}\left(\frac{dF}{dF_0}(\widehat{y}^*, y^*)-1\right) \\
\implies \log\frac{d\tilde{F}_n}{dF_0}(\widehat{y}^*, y^*) &= \log\left[1+\frac{\mu}{\sqrt{n}}\left(\frac{dF}{dF_0}(\widehat{y}^*, y^*)-1\right)\right] \\
&= \frac{\mu}{\sqrt{n}}\left(\frac{dF}{dF_0}(\widehat{y}^*, y^*)-1\right) - \frac{1}{2}\left\{\frac{\mu}{\sqrt{n}}\left(\frac{dF}{dF_0}(\widehat{y}^*, y^*)-1\right)\right\}^2 + \frac{1}{3}\left\{\frac{\mu}{\sqrt{n}}\left(\frac{dF}{dF_0}(\widehat{y}^*, y^*)-1\right)\right\}^3 \\
&\quad -\ldots\infty.
\end{aligned}$$

Then,

$$\begin{aligned}
cov_{H_0}\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})), \log\frac{d\tilde{F}_n}{dF_0}\right) &= \mu\cdot E_{H_0}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\cdot\left(\frac{dF}{dF_0}-1\right)\right) \\
&\quad -\frac{\mu^2}{2\sqrt{n}}\cdot E_{H_0}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\cdot\left(\frac{dF}{dF_0}-1\right)^2\right) \\
&\quad +\frac{\mu^2}{2\sqrt{n}}\cdot E_{H_0}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\cdot\left(\frac{dF}{dF_0}-1\right)^3\right) - \ldots\infty \\
\implies cov_{H_0}\left(\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})), \log\frac{d\tilde{F}_n}{dF_0}\right) &= \mu\cdot E_{H_0}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\cdot\left(\frac{dF}{dF_0}-1\right)\right) \\
&= \mu\cdot E_{H_0}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\cdot\frac{dF}{dF_0}\right) \\
&\quad -\mu\cdot E_{H_0}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\right) \\
&= \mu\cdot\int_{\mathbb{R}^2}(T_n^{<r>} - E_{H_0}(T_n^{<r>}))\cdot\frac{dF}{dF_0}\cdot dF_0 \\
&= \mu\cdot E_{H_1}\left((T_n^{<r>} - E_{H_0}(T_n^{<r>}))\right) \\
&= \mu\, E_{H_1}(T_n^{<r>}) = \mu(2P_{c;H_1}-1).
\end{aligned}$$

11.4 Proof of Proposition 3

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

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

i.e. for increasing sample size n, the power of $T_n^{<r>}$ soars and approaches to 1.
Moreover, as $\mu\uparrow$,

$$P_{H_n}\left[\sqrt{n}(T_n^{<r>} - E_{H_0}(T_n^{<r>})) > \pi_\kappa\right] \;=\; \Phi\left(\frac{\mu\, E_{H_1}(T_n^{<r>}) - \pi_\kappa}{\sqrt{4\eta_{1,2}(r)}}\right) \longrightarrow 1.$$